Habilitation à diriger des recherches

Université de Strasbourg Spécialité
SCIENCES ÉCONOMIQUES

Fabrice LE LEC

Essais en économie expérimentale et comportementale

Soutenue le 28 novembre 2019

PUBLICATION 1 : Michal Krawczyk et Fabrice Le Lec (2010), “ ‘Give me a chance!’, An experiment in social decision under risk, ***Experimental Economics***, 13: 500-511.

Exp Econ (2010) 13: 500–511
DOI 10.1007/s10683-010-9253-5

'Give me a chance!' An experiment in social decision under risk

Michal Krawczyk · Fabrice Le Lec

Received: 13 May 2008 / Accepted: 23 August 2010 / Published online: 16 September 2010

Abstract This paper reports the results of a 'probabilistic dictator game' experiment in which subjects were given an option to share chances to win a prize with a dummy player. Using a within-subject design we manipulated two aspects of the decision, the relative cost of sharing and the nature of the lottery: the draws were either independent for the two players ('noncompetitive' condition) or one's success meant other's failure ('competitive' condition). We also asked for decisions in a standard, non-probabilistic, setting. The main results can be summarized as follows: first, a substantial fraction of subjects do share chances to win, also in the competitive treatments, thus showing concern for the other player that cannot be explained by outcome-based models. Second, subjects share less in the competitive treatment than in other treatments, indicating that procedural fairness alone cannot explain the data. Overall, these results suggest that models aiming at generalizing social concerns to risky environments will have to rely on a mix of distributive and procedural fairness.

Keywords Social preferences · Other-regarding preferences · Inequity aversion · Social concern · Social preferences under risk · Procedural fairness

JEL Classification A13 · C65 · C72 · D63 · D03

M. Krawczyk
Faculty of Economic Sciences, University of Warsaw, 44/50 Dluga, 00-241 Warsaw, Poland
e-mail: mkrawczyk@wne.uw.edu.pl

F. Le Lec (✉)
Lille Economics and Management UMR CNRS 8179, Catholic University of Lille, 41 rue du Port, 59000 Lille, France
e-mail: fabrice.lelec@icl-lille.fr

1 Introduction

Popular newspapers are full of stories of individuals taking life-threatening risks to help others, such as jumping in icy waters to save a drowning fellow citizen or intervening when others are assaulted in the street. Historical accounts may be even more dramatic, as evidenced by the case of thousands of families across Europe hiding persecuted Jews during the second World War. To name further examples from everyday life, a student may help a less-able friend before an exam (or during an exam by cheating), an employee may play fair against a competitor applying for the same position in the firm, firms may choose not to take unfair advantage when competing for a contract, competitors may freely exchange strategic information in markets, living donors share organs with anonymous donees, etc. In all of these examples, individuals facing a risky environment voluntarily give up some chance of own success (or even survival) to enhance the perspectives faced by the other. Of course, the idea that individuals are partly motivated by concern for their neighbors has been widely supported by numerous experimental studies in the last decades. As a result, several economic models aiming at capturing other-regarding concerns have emerged, mostly based on inequity aversion (Fehr and Schmidt 1999; Bolton and Ockenfels 2000) or on maximization of both total payoff (social utilitarianism) and of the payoff of the worst-off agent (maximin) as in Charness and Rabin (2002). Yet, these models only address social concerns on the basis of consequences (outcomes) and are not meant to deal with risky situations as in the examples given above. Even if supplemented with further assumptions, they cannot, as we will show, account for sharing on chances, especially in the case when there is ultimately only one winner.

This theoretical situation contrasts with the fact that most daily situations have a risky component and many prosocial behaviors take place under risk: in principle any resource that is voluntarily passed from one agent (time, information, objective chances) and that can increase the probability to obtain certain reward makes a relevant case. Some path-breaking studies have addressed this issue, suggesting indeed that socially-oriented preferences play a role in risky environment. Bolton et al. (2005) provide evidence of other-regarding preference in the presence of risk, as subjects played the Ultimatum Game with chances to win some prize: second players tended to reject proposals when offered overly low chances to win. A three-person dictator experiment by Karni et al. (2008) studies whether an agent may want to give up (a bit of) her (expected) payoff to achieve greater procedural fairness between two other agents. The studies by Brennan et al. (2008) and Güth et al. (2008) seek to establish whether subjects care about the amount of risk assumed by another individual. Their main finding is that although some subjects care about the allocation of expected payoffs, they do not seem to treat risk for themselves and others in the same way: they are only willing to pay to reduce the variance for their own payoff but not for others'. Overall, these studies indicate that risk appears to be taken into account by socially-oriented subjects, even though much is still left to have a precise view on how risk and outcomes interact in fairness consideration.

Here, we investigate experimentally this issue, using two natural generalizations of social concern models to risky environments: on the one hand a *consequentialist* model where social concerns only take into account consequences and on the other hand a *procedural* one for which the initial distribution of probability matters primarily, *e.g.* fairness is based on expected payoffs (Trautmann 2009). The consequentialist hypothesis implies that the social utility of given consequences is simply weighted by the probability of the corresponding event as suggested in Charness and Rabin (2002) whereas for the procedural view, *ex ante* probability distributions are part of the evaluation of fairness. The two models lead to sharp differences when applied to some of the situations sketched above. In particular, in the case of two competitors for the same reward, the final outcome is necessarily totally asymmetrical: the consequentialist model predicts selfish behavior but the purely procedural model does not. Yet, when the chances of obtaining the reward are independent for the agents, the consequentialist model may now allow some prosocial behavior but for the purely procedural one, the situation is similar to the previous case. Using a variation of the Dictator Game, we study this question by comparing three different procedures to allocate money, for which the two models predict differently: a deterministic one playing the role of a control treatment; a 'competitive' one, where the dictator allocates mutually-exclusive chances (such that only one subject can earn the total stake); and a 'non-competitive' one where the dictator allocates independent chances (so outcomes may be fair *ex post*). If social concerns can be entirely captured by procedural considerations, then the two probabilistic treatments, *i.e.* competitive and non-competitive, should generate identical behaviors. In contrast, if subjects only care about consequences, no socially oriented behavior should appear in the competitive treatment. The results show that social concerns are indeed affected by risk, but they cannot be accounted for by a purely consequentialist model, nor by a purely procedural model. More specifically, a substantial fraction of subjects choose non-selfishly when given an option to share chances to win mutually-exclusive rewards, suggesting that the purely consequentialist view falls short. Yet, since choices in this condition are less generous than when probabilities are independent, the purely procedural view is not totally accurate either. The data hence suggests to extend current models by relying on a mix of procedural and consequentialist considerations.

The remainder of the paper is organized as follows. The next section describes the design and procedures. Section 3 exposes the most important findings of our experiment and contrast them with the theoretical models and eventually the last section concludes.

2 Design and theoretical background

2.1 Design and experimental issues

At the beginning of the experiment, each subject was assigned the role of player A ('dictator') or player B ('dummy'). This was fixed throughout the experiment, to prevent subjects from justifying selfish behavior on the grounds that the 'victim' will have a chance to be the dictator herself in the following round. For the same reason we did not to use the strategy method, that is to ask all subjects what their choice would be as a dictator and then to randomize roles, a method favored in many related

Table 1 The nine treatments

	(10; 30)	(20; 20)	(30; 10)
Certain outcome	Deterministic-10	Deterministic-20	Deterministic-30
Non-competitive lottery	NonCompetitive-10	NonCompetitive-20	NonCompetitive-30
Competitive lottery	Competitive-10	Competitive-20	Competitive-30

experiments (Andreoni and Miller 2002; Güth et al. 2008): it may promote apparently selfish behaviors since role allocation is procedurally fair. It was also important that subjects would not perceive the allocation of roles as just random, and procedurally fair, so role assignment was based on arrival time but not in a monotonic way not to reduce possible "entitlement effect" (Gächter and Riedl 2005).[1]

In all treatments, each dictator had to distribute 10 tokens between herself and a dummy. What each token represented varied during the experiment on two dimensions. The prizes at stake, *i.e.* the maximum amount that each player could earn in a given round, and the lottery used to allocate these prizes, were systematically manipulated. The prizes, in euro, were (10; 30), (20; 20) or (30; 10) for the dictator and dummy respectively. The second, more fundamental, dimension is the type of the game (hereafter game type). In the Deterministic condition, each token simply corresponds to one tenth of the prize, so that keeping all the tokens means winning the full prize. For example, if prizes are (30; 10), keeping 7 tokens and passing 3 to the other player results in earnings of 21 and 3 euro for the dictator and dummy respectively. In the other two conditions, tokens represent a probabilistic analogue of a fraction of the prize: in the Competitive condition each token stands for a 10% chance of winning the prize, with the events of winning for both players being mutually exclusive. For example, with prizes (30; 10) and a distribution of tokens (7; 3) as above, a 10-sided die is rolled: if the outcome is in $\{1, 2, \ldots, 7\}$, A wins 30 euro and B nothing; otherwise B wins 10 euro and A nothing. In the NonCompetitve condition each token still represents a 10% chance of winning own prize, yet chances are independent for A and B. With the numbers given above, two dice are rolled, one giving A a 70% chance of winning his prize and another giving B hers with a 30% chance. The combination of both dimensions generates nine treatments, which are summarized in Table 1. In order to correlate behavior in different situations and increase statistical power for treatment comparisons, we implemented a within-subject design in the spirit of Andreoni and Miller (2002): our design can be seen as an adaptation of theirs with two additional probabilistic conditions–Competitive and NonCompetitive. Every subject played all nine games, grouped in three blocks: Deterministic, Competitive and NonCompetitive. Relevant instructions were distributed at the beginning of each block and followed by control questions. This was implemented to ensure that subjects were not confused about specific payment schemes. Every erroneous answer

[1] Subjects were explicitly told: "Role assignment has been decided by the order of arrival: some participants have picked yellow cards corresponding to the role of A-participant and the others some corresponding to B-participant. Note that in general it does not mean that for instance early arrivals correspond to any specific role".

to control questions was recorded and had to be corrected by the participant in order to proceed. Block order was different across sessions and the order of particular rounds, *i.e.* with different prizes within each block, was randomly manipulated at the individual level. Participants did not get any feedback between rounds, and each dictator was matched with a new dummy at every round, and had no information about the specific features of the decisions to come. This was so in order to restrict attempt to behave fairly across rounds by taking into account the different values of tokens. For instance, it was very difficult if not impossible to try to both equalize and maximize overall expected payoffs of both subjects. At the end, it was announced which of the nine rounds would be picked for payment and the relevant decision of the dictator was revealed to the dummy player she was matched with. Finally, if needed, risk was resolved with one or two 10-sided dice being rolled publicly.

After that, a series of three individual decisions under risk was run, in order to establish individual attitudes towards risk. We used the Interactive Multiple Price List design (Andersen et al. 2006), to elicit certainty equivalent for a 10, 50 and 90 percent chance to get 20 euro. One third of the subjects were randomly picked for payment in this second part of the experiment, and if chosen, subjects were only paid for one of the tasks, once again randomly determined. The experiment was run in October and December 2007 in the laboratory of the Center for Research in Experimental Economics and Political Decision Making (CREED) in Amsterdam. It was programmed using z-Tree (Fischbacher 2007). In total 128 subjects, mostly undergraduate students at the University of Amsterdam, participated. Average earnings for an experiment lasting from 70 to 100 minutes equaled about 22 euro, including a 7.50 euro show-up fee.

2.2 Theoretical predictions

Independently of game types, rational selfish players will obviously keep all the tokens. In contrast, socially oriented players will pass some tokens at least in some of the treatments. In the Deterministic treatments, the implications of existing models are very well known: inequity aversion models predict that tokens passed will increase with own prize whereas Charness and Rabin's model is more indeterminate. As in Andreoni and Miller's (2002) study, the different definitions of social concerns, especially inequality aversion and maximin vs. efficiency, generate different predictions with varying prizes. This aspect has been thoroughly investigated in previous studies so we will only focus on the predictions concerning game types.

Indeed, the real specificity of our design is to allow to discriminate between two natural generalizations of social preferences under risk. First we consider a purely consequentialist model, based on the natural assumption that preferences are represented by the product of the social utility on outcomes and the corresponding probabilities.[2] In the Competitive conditions, this implies that if there is some social distribution that is preferred to the other, then the decision-maker should choose it with probability 1. For instance, in Competitive-20, only two distributions are possible $(20, 0)$ and $(0, 20)$, all models predict that the former is preferred to the latter, so the dictator should choose to keep all tokens.

[2] See Charness and Rabin, p. 853 and the former working paper version, footnote 46.

Prediction 1 *A purely consequentialist other-regarding agent will keep all the tokens*[3] *in the competitive treatments while she will not in the deterministic treatment.*

Note that this is independent of the agent's individual attitude towards risk, be it the shape of the utility for money or a probability transformation function.

In contrast, for models based on procedural fairness, individuals are assumed to care about *ex ante* fairness. This is true in particular for the "process Fehr-Schmidt model" (Trautmann 2009). In this case, whether the game type is Competitive or Non-Competitive should have no impact: the self-interest term and the fairness function take the same values in both cases. This qualitative prediction easily generalizes to different definitions of social concerns such as quasi-maximin or efficiency. More generally, any model of the following form would yield identical choices in the two probabilistic game type treatments:

$$v(x) = \sum_x \phi(p_k).u(x_k^1) + F\left(\sum_x \psi(p_k^1).v(x_k^1), \sum_x \psi(p_k^2).v(x_k^2)\right) \qquad (1)$$

with u and v being utility functions for money, ϕ and ψ probability weighting functions, F the fairness term, p_k the probability that the kth outcome in the social lottery x is drawn randomly, and x_k^1 and x_k^2 being players 1 and 2's payoffs. Such a generic procedural model does not depend on whether probabilities and payoff are valued in the same way in the self-interest term and the fairness function, nor does it rely on a specific definition of social concern. In particular, it can be that own payoffs and probabilities are valued through a usual generalized expected utility function, whereas in the fairness function they might be considered through expected payoffs.[4] This general form leads to the following prediction:

Prediction 2 *A purely procedural other-regarding agent will give the same number of tokens in Competitive and Non-Competitive.*

These two general models therefore provide testable predictions for social preferences under risk.

3 Results

Figure 1 gives an overview of the distribution of the number of tokens given in the nine treatments for decision-makers. First, a large fraction of choices in each treatment corresponds to keeping all the tokens. Giving nothing is always the median

[3] Except in the case of Competitive-30 where (0, 30) is preferred to (10, 0).

This case is rather unusual, and, should it occur, is easy to identify since the dictator should give *all* the tokens to the dummy. So on a general note, whatever consequentialist model is considered, *extreme* choice in the Competitive treatments are predicted.

[4] It is noteworthy that this particular instance of procedural social preferences is compatible with the results exhibited in Brennan et al. (2008) and Güth et al. (2008), and especially the model suggested in the second study in Eq. (2).

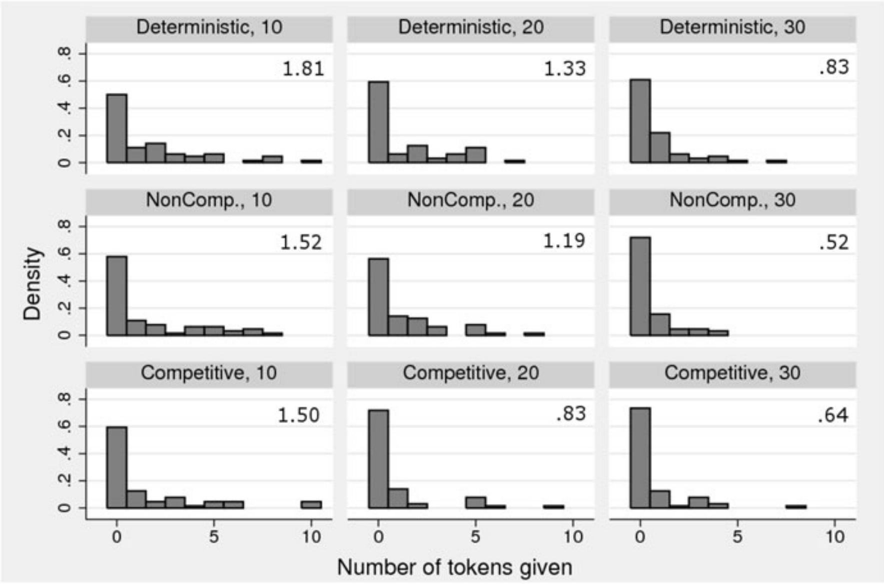


Fig. 1 Frequencies of dictators' choices and mean numbers of given tokens in all treatments

choice and may be chosen by as much as three-quarters of individuals, as in treatments Competitive-30 and Noncompetitive-30. Overall, the mean number of tokens given is 1.13. Yet non-selfish choices cannot be regarded simply as 'mistakes': they seem too numerous and not correlated with the number of erroneous answers to the control questions. On a general note, the non-selfish choices are distributed quite evenly between 1 and 5. Interestingly, the classic DG case (Deterministic-20) seems to be the only one in which sharing tokens equally (5-5) appears to be somewhat prominent as in previous studies, though it was actually chosen by only 11% of subjects. Moreover, it is hard to see a strong tendency to equalize earnings or expected earnings in asymmetric treatments: apart from the standard DG with 20 euro, there is hardly any peak at the equalizing choice in the Deterministic treatments. Regarding the effect of prize, our results hence mimic Andreoni and Miller's findings, with a skew of the distribution towards more selfish behaviors.

3.1 The effect of game type

Comparing results across game types, it appears that decisions are less selfish in the Deterministic case than in the Competitive one, and less selfish in the NonCompetitive than and the Competitive one. For the standard prize of (20; 20), dictators' behaviors seem different between Deterministic and Competitive on the one hand, and between Competitive and NonCompetitive on the other hand. A Wilcoxon signed-rank test, a non-parametric test based on individual differences in two matched-pair conditions, turns out to be significant, with $p = 0.02$ and $p = 0.05$ respectively. Since only prosocial subjects can be plausibly affected by game treatment, we also performed these tests on this subgroup only. Prosocial subjects were defined, in accordance with

Table 2 Regression with individual fixed effects

'Given tokens'	Coefficient	[95% Conf. Interval]
Prize/10	−0.48***	[−0.60; −0.30]
Classic	0.38***	[0.11; 0.66]
Noncompetitive	0.21	[−0.05; 0.47]
Roundwithinblock	−0.04	[−0.10; 0.18]
Block	−0.24***	[−0.39; −0.09]
Errors	−0.02	[−0.18; 0.15]

*Significant at the 0.1 level

**Significant at the 0.05 level

***Significant at the 0.01 level

$R^2 = 0.60$, $Adj\ R^2 = 0.54$

all social preferences models, as subjects that gave at least one token over the three deterministic treatments. That composes a group of 35 subjects. The significance of individual differences between Deterministic and Competitive and between Non-Competitive and Competitive is confirmed when we focus on the prosocial dictators only with $p < 0.001$ and $p = 0.05$ respectively with a signed-rank Wilcoxon test. The average number of tokens given ranges from 1.34 in Competitive-20 to 1.98 for Non-Competitve-20 to finally reach 2.43 in Deterministic-20. It is also worth noting that choices made by prosocial dictators in the Competitive treatment are quite far from the extreme predictions of the outcome-based models, *i.e.* that they should keep all the tokens. The picture is less clear for other prizes, only some of the pair-wise comparisons between game types for prosocial dictators appear to be significant at conventional level. Yet, an overall comparison of all treatments for prosocial dictators exhibits a monotonic effect of prize and risk, NonCompetitive-30 being the only exception. Overall, there seems to be the following pattern in the data: choices are more generous in Deterministic than in Non-Competitive and more generous in this latter condition than in Competitive. This is especially true and salient for prosocial dictators as well as for the symmetric prize, as can be expected. It is also interesting to note that, while being generally much more generous the dummies' hypothetical choices exhibit essentially identical patterns of treatment effects with some additional noise.

A multivariate regression[5] tends to confirm these findings (Table 2). Taking into account individual fixed effects, the analysis show that the difference between Competitive and Deterministic is strongly significant, and the difference with NonCompetitive is weakly significant ($p = 0.12$). Likewise, own prize has a strong negative impact on the number of tokens given. New insights concern the impact of other variables: subjects turn out to be more generous at the beginning than later on (the

[5] Another way to look at the data is to use a Tobit analysis (left-censored at 0 and clustered by subjects). The limit of this type of analysis though is that using Tobit with fixed effects does not provide an unbiased estimator. We hence performed a regular Tobit analysis of the number of tokens given clustered by subjects on the same covariates, which yields similar results as the ones described above.

variable “block” is significant). Since learning effects are ruled out because no feedback was provided, it seems likely that subjects who shared in early rounds came to feel entitled to act more selfishly later on. Similar findings have been exhibited in the past, and a possible psychological explanation is the so-called ‘moral credential bias’ (Monin and Miller 2001). Yet, there seems to be no reason to think that this time effect is in one way or another linked to the specific ordering of treatments in different sessions, given that the general findings of treatment effects seem robust enough to still hold when controlling for time. It is also reassuring to see that other design-based variables have no significant influence on dictators’ behaviors: the order within block (“round in block”) as well as errors made in the control questions in the corresponding block (“errors”) do not seem to matter. Results are similar when regressing the number of given tokens with dummy variables for prize, instead of the prize value as above. We have also performed some analysis of demographic variables, finding that relatively unexperienced, female, non-economist subjects tended to give more.

Finally, we have included in the analysis our measures of risk aversion from the second experimental task. In particular, it is interesting to check whether strongly risk-averse subjects showed stronger treatment effects (sharing is relatively more attractive in the deterministic case for these subjects). We performed numerous tests on different measures of risk attitude, and the overall results were mixed, only sporadically weakly significant.

3.2 Results w.r.t. theoretical models

As in many other studies, the hypothesis that all subjects maximize their own payoffs falls short in all situations. Still, it is important to put forth that about half the subjects did behave completely selfishly. Thus, the effects of treatments that we observe are necessarily driven by the behavior of a minority of subjects.

Although models based on the idea of inequality aversion predict correctly in some treatments, *e.g.* the standard DG, there is overall little support for it in our data: choices actually increase in own prize, thus accentuating differences between final outcomes. The failure to explain aggregate data in terms of inequality aversion could be due to heterogeneity of players. For instance a few efficiency-oriented subjects could strongly influence the distribution of tokens kept across prizes. Yet, when focusing on individual data, even if we apply a very loose selection criterion and consider as inequality-averse those subjects with choices weakly decreasing in own prize, we find only 2 subjects conforming to it. On the contrary, models based on the assumption that at least some subjects tend to maximize *ceteris paribus* the total surplus as proposed by Charness and Rabin do find some support. These results are roughly consistent with other studies showing some concern for efficiency (Andreoni and Miller 2002; Engelmann and Strobel 2004; Charness and Rabin 2002).

More to the point, our results cast doubts on the ability of the two benchmark models, the purely consequentialist and the purely procedural, to account for the data. Focusing on the standard prize $(20, 20)$, we clearly observe that in the Competitive treatments, prosocial dictators’ offers are far from being null and that they make different choices in Competitive and NonCompetitive: the average number of tokens varies from 1.34 to 1.98, the difference being significant at $p = 0.05$ in a Wilcoxon signed-rank test. To be on the safe side, we also ran some regression analysis limited to this prize value but with the presence of the 'block' variable, in order to check for possible time effects in this difference. A Tobit analysis and a regression with fixed effects give similar results: Deterministic as well as NonCompetitive in comparison with Competitive give coefficients that are negative and highly significant (for the Tobit, $p = 0.009$ and $p = 0.009$, and for the regression $p = 0.018$ and $p = 0.024$). These results are overall hard to reconcile with the models presented in section 2 and the corresponding predictions.

Indeed, a consequentialist model cannot account for the fact that in the Competitive treatment offers are different from 0. Prediction 1 seems thus at odds with our experimental results. Yet, purely procedural models do not do much better, since they cannot account for part of the data either. They cannot explain the difference between Competitive and NonCompetitive. This is even more striking for models that assume that the self-interest term as well as the fairness one are determined by expected payoffs, since they predict that Deterministic and Competitive would yield the same results while they generate even more difference than NonCompetitive vs. Competitive. A possible conclusion is that both classes of models capture part of the evidence.

Such results could then be the effect of some heterogeneity among subjects, some of them being consequentialist while others may be procedural. To investigate this possibility, we looked at individual data to see whether prosocial subjects could be easily classified as purely procedural or purely consequentialist. We hence checked how many subjects would be in line with prediction 1 or 2. A subject was categorized as a *procedural subject* if the sum of tokens kept in Competitive and NonCompetitive did not differ by more than 10 percent: for instance if a subject kept 20 tokens in the three NonCompetitive treatments, then she would be classified as procedural if she kept between 18 and 22 tokens in Competitive. Given the relatively high numbers of token kept, this classification is rather in favor of the procedural hypothesis. In total, 57% of the prosocial subjects turned out to satisfy it. For the consequentialist model, we adopted a similar feature: all prosocial subjects that kept all tokens or gave less than 10 percent of the tokens in the Competitive treatments were considered as consequentialists. This is again a classification rule rather in favor of the consequentialist hypothesis since some subjects who gave little in the Deterministic treatments can end up giving more in the Competitive one and still be classified as consequentialist. This corresponds to 40% of the prosocial subjects. Yet, the numbers of consequentialist and procedural subjects do not add up since there is a common area for these two classes, namely in the neighborhood of all tokens kept by the subject. We find indeed that 25% of the prosocial subjects are both. Overall this gives the following distribution: 32% are unambiguously procedural, 15% are unambiguously consequentialist, and 25% are indetermined consequentialist or procedural, and 28% do no belong to any of the two. As a consequence, it seems that some heterogeneity of procedural/consequentialist prosocial subjects may explain a substantial share of the effects of game type in our experiment, yet a non-negligible part of prosocial subjects appears to rely on either a mix of procedural/consequentialist concerns or on different ones. Our data hence indicate that an accurate model of social preference under risk should integrate a mix of procedural and consequentialist social concern, both to potentially account for all subjects' behaviors and to have a general model for whose consequentialist as well as procedural agents would be particular cases.

Another, possibly complementary, explanation may lie in the idea of impression management.[6] Subjects may be willing to give something, regardless of the meaning of tokens, just to make a good impression on the dummy or on themselves. This idea of impression management has recently received quite some support in empirical studies (Koch and Normann 2008; Murnighan et al. 2001; Dana et al. 2007). It would explain why in the Competitive treatments, subjects still give, for instance, one token to the other participant. Combined with some consequentialist view of fairness, it may also explain why choices in Competitive and in NonCompetitive treatments are different: the possibility of both the dummy and the dictator winning the prize may compensate the possibility that both lose and add some incentives to give. The fact that a significant share (25%) of prosocial subjects are classified as both consequentialist and procedural seems to also point in this direction.

4 Conclusion

We take a few main lessons from the results of the experiment. First, consistent with previous literature studying other contexts, we find on top of inequality-aversive behavior evidence for efficiency-oriented choices, as well as evidence of socially oriented choices in the presence of risk, more specifically when subjects share chances rather than money.

Second, using our simple design we find clear evidence that a purely consequentialist model of social preference is not a promising way to model social behavior in risky situations in the sense that it can only account for a small share of subjects's behaviors, and cannot explain the aggregated data. But similarly, purely procedural models, even though they seem to account for a larger share of individual choice, seems to fall short too, since they would predict identical behaviors in both the Competitive and NonCompetitive cases. More generally, a substantial share of subjects do not correspond to any of those two models. These results suggest that any adequately descriptive model of social preferences under risk need to rely on a combination of procedural and consequentialist motives, in order to be able to account for the aggregate data as well as some specific individual behaviors but also to allow consequentialist types and procedural types as special cases.

Acknowledgements We thank Gary Charness, Martin Kocher, Serge Macé, Frans van Winden, the editor, Jordi Brandts and two anonymous referees, as well as participants in seminars and conferences in Warsaw, Zurich, Munich, Amsterdam and Rennes. The financial support from the European Commission via the Marie Curie Research and Training Network ENABLE and CREED of the University of Amsterdam is gratefully acknowledged.

[6] The credit for this alternative explanation must be given to an anonymous referee.

References

Andersen, S., Harrsion, G., Lau, M., & Rutstrom, E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, *9*, 383–405.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: an experimental test of the consistency of preferences for altruism. *Econometrica*, *70*, 737–754.

Bolton, G. E., & Ockenfels, A. (2000). ERC: a theory of equity, reciprocity and competition. *American Economic Review*, *90*, 166–193.

Bolton, G. E., Brandts, J., & Ockenfels, A. (2005). Fair procedures: evidence from games involving lotteries. *Economic Journal*, *115*, 1054–1076.

Brennan, G., Güth, W., Gonzalez, L., & Levati, M. V. (2008). Attitudes toward private and collective risks in individual and strategic choice situations. *Journal of Economic Behavior and Organization*.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quartely Journal of Economics*, *117*, 817–869.

Dana, J., Weber, R., & Kuang, J. (2007). Exploiting moral wriggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*, 67–80.

Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review*, *94*, 857–869.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*, 817–868.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*, 171–178.

Gächter, S., & Riedl, A. (2005). Moral property rights in bargaining with infeasible claims. *Management Science*, *51*, 249–263.

Güth, W., Levati, V., & Ploner, M. (2008). On the social dimension of time and risk preferences: an experimental study. *Economic Inquiry*, *46*, 261–272.

Karni, E., Salmon, T., & Sopher, B. (2008). Individual sense of fairness: an experimental study. *Experimental Economics*, *11*, 174–179.

Koch, A. K., & Normann, H.-T. (2008). Giving in dictator games: regard for others or regard by others? *Southern Economic Journal*, *75*, 223–231.

Monin, B., & Miller, D. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, *81*, 33–43.

Murnighan, J. K., Oeschb, J. M., & Pillutlac, M. (2001). Player types and self-impression management in dictatorship games: two experiments. *Games and Economic Behavior*, *37*, 388–414.

Trautmann, S. (2009). A tractable model of process fairness under risk. *Journal of Economic Psychology*, *30*, 803–813.

PUBLICATION 2 : Michal Krawczyk et Fabrice Le Lec (2016), "Dictating the Risk: Comment", ***American Economic Review***, 106(3): 836-839.

American Economic Review 2016, 106(3): 836–839
http://dx.doi.org/10.1257/aer.20130779

Dictating the Risk: Experimental Evidence on Giving in Risky Environments: Comment†

By Michal Krawczyk and Fabrice Le Lec*

Based on experimental dictator games with probabilistic prospects, Brock, Lange, and Ozbay (2013) conclude that neither ex post nor ex ante comparisons can fully account for observed behavior. We argue that their conclusion that ex ante comparisons cannot explain the data is at best weakly supported by their results, and do so on three grounds: (i) the absence of significant differences between the most relevant treatments, (ii) the implicit assumption of subjects' risk neutrality, and (iii) the asymmetry of treatments regarding the disclosure of dictators' choice. (*JEL* C72, D63, D64, D81)

In "Dictating the Risk: Experimental Evidence on Giving in Risky Environments," Brock, Lange, and Ozbay (2013)—henceforth, BLO—report the results of a series of experimental dictator games with probabilistic prospects. Their main conclusion is that a mix of ex post (outcome-based, "consequentialist") comparisons and ex ante ("procedural") comparisons is necessary to account for observed behavior. Although we feel that both considerations are indeed likely to be important, the experimental evidence BLO report may be analyzed in a different light. In particular, we believe the claim that ex ante comparisons cannot explain the data is at best weakly supported by their results. We will argue this point on three grounds. First, the most comparable treatments (*T1*, *T4*, and *T5*) to discriminate between ex ante and ex post social considerations do not yield any significant difference. Second, interpreting the difference between treatments *T2* or *T3* and *T1* as evidence for the inadequacy of ex ante social concerns presupposes subjects' risk neutrality. Third, the nondisclosure of dictators' choices may generate a systematic bias toward lower generosity (the "moral wiggle room") in all treatments but *T1*, a tendency which can be mistaken for ex post concerns.

I. No Significant Treatment Effects (*T1* versus *T4*, *T5*)

The essence of BLO is to examine variations of the standard dictator game by manipulating the meaning of the 100 "tokens," any number of which the first player can voluntarily pass to the second one. In the baseline treatment (*T1*), one token is

* Krawczyk: University of Warsaw, 44/50 Dluga Street, 00-241 Warsaw, Poland (e-mail: mkrawczyk@wne.uw.edu.pl); Le Lec: University Paris-1 Pantheon-Sorbonne, Centre d'Economie de la Sorbonne, 96-112 boulevard de L'Hopital, 75013 Paris, France (e-mail: fabrice.le-lec@univ-paris1.fr). The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.
† Go to http://dx.doi.org/10.1257/aer.20130779 to visit the article page for additional materials and author disclosure statement(s).

worth \$0.10. The two treatments which most directly test the influence of ex post/ex ante comparisons are *T4* and *T5*. In both, each token represents 1 percent probability to earn a prize of \$10, the difference being that each player's chances are mutually exclusive in *T4* and independent in *T5*. As stated by the authors, these three treatments should yield similar behaviors if only ex ante comparison is at play and different ones if a mix of both ex ante/ex post concerns is present. BLO find no significant difference between *T1* and *T4* nor between *T1* and *T5* (see Table 5 of BLO): the average number of tokens seems to be lower in *T4* and *T5*, but not significantly so. This lack of significance holds for the entire sample as well as for a subsample of subjects who at some point gave at least one token. If ex ante concerns fall short of explaining behaviors, one would expect some differences in these particular treatments.

II. The Assumption of Risk Neutrality (*T1* versus *T2*, *T3*)

Hence, BLO mostly base their claim of the insufficiency of ex ante concerns on *T2* and *T3*. In these treatments, each token kept by the dictator is worth \$0.10 but each token passed to the recipient corresponds to a 1 percent chance to get \$10 in *T2* and 2 percent chance to get \$5 in *T3*. At first glance, BLO's finding that subjects give less in *T2* and *T3* than in *T1* points at the insufficiency of ex ante comparisons: all three are fully equivalent in terms of expected values and should not trigger different behaviors. But this only holds if one assumes that ex ante comparisons are made on *expected payoff*. Given that there is massive evidence that most individuals are risk averse, that is perhaps a bold assumption.

Indeed, the issue when comparing *T1* with *T2* and *T3* is that there is, in *T2* and *T3*, an asymmetry between the (subjective) value of the tokens for the dictator and for the receiver, while there is none for *T1*. If subjects feel that a 1 percent probability to earn \$10 is of less value than \$0.10 (to the recipient who starts with nothing), then there are good reasons, even if only motivated by ex ante comparisons, to give less in *T2*. Andreoni and Miller (2002) among others showed that less is given when giving is less efficient. An ex ante concerned agent, driven at least partly by efficiency as in Charness and Rabin (2002), would give less in *T2*/*T3* than in *T1*, simply because keeping the tokens is more efficient, when considering the total ex ante welfare. In other words, based on risk preference only, one should expect at least some subjects to give less in *T2* and *T3* than in *T1*. As no measure of individual risk aversion is available, results in *T2* and *T3* could very well be the result of non-expected, payoff-based ex ante comparison.

III. The Asymmetric "Moral Wiggle Room" (*T1* versus *T2–T5*)

Studies (such as Dana, Cain, and Dawes 2006) showed that when the receiver does not know exactly the dictator's action, the amount sent tends to be lower, and in many cases almost null. BLO acknowledge explicitly the existence of such a phenomenon, but may underestimate its systematic consequences. In their design and for all treatments, the receiver only knows the final outcome, and not the treatment selected for payment nor the dictator's choice. This is desirable to measure intrinsic social consideration, as opposed to "social impression" management. But

Table 1—Stylized Predictions and Results on Tokens Passed, by Treatment

Models/hypotheses	Stylized predictions/findings
Risk neutrality, no moral wiggle room	
pure ex post comparison	$x_1 > x_3 > x_2 > x_5 > x_4 = 0$
pure ex ante comparison	$x_1 = x_3 = x_2 = x_5 = x_4$
ex post/ex ante mix	$x_1 > x_3 > x_2 > x_5 > x_4 > 0$
Pure ex ante comp. with risk aversion	$x_1 > x_3 > x_2 > x_5 = x_4$
Pure ex ante comp. with moral wiggle room	$x_1 > x_3 = x_2 = x_5 = x_4$
BLO findings	$x_1 > x_3 \sim x_2 \sim x_5 \sim x_4 > 0$[a]
Krawczyk and Le Lec's (2010) findings	$x_1 >$ $x_5 > x_4 > 0$

[a] Gifts are higher in *T1* than in any other treatment, but only significantly so for *T2* and *T3*.

it implies a systematic difference between *T1* and *T2–T5*. In *T1*, there is little room for hiding: if this treatment is implemented, the receiver will know how much the dictator gave,[1] while in the other treatments, the receiver will only know his final earnings (0 or 100, and occasionally 50) and not the number of tokens passed. The dictator is in two different situations: in *T1*, she has to choose an amount knowing that (in most cases) the receiver will know her decision while in the other treatments she does so knowing that the receiver will not. One additional observation about BLO's data may give support to this explanation: the 50/50 split is frequently chosen in *T1*, but it is not as prominent in other treatments. In particular, it is only half as frequent in *T3* as in *T1*, although in either case it would implement perfect ex post equality with certainty. The natural interpretation is that subjects do not care only about achieving ex post equality but also about avoiding an appearance of being selfish: choosing 40 in *T1* would reveal the receiver a departure from the "50-50 norm" (Andreoni and Bernheim 2009). When they can act selfishly at no cost to their image (treatments 2–5), subjects often seize the opportunity. The "moral wiggle room" may explain more generosity in *T1* than in other conditions and be confounded with the insufficiency of ex ante comparisons. A stronger case that ex post considerations are also needed can be made based on our earlier experiment (Krawczyk and Le Lec 2010) where differences between equivalents of *T1* versus *T4* and *T5* were observed. The important difference was that dictators' decisions for the treatment selected for payment were disclosed, thus avoiding the moral wiggle room confound.

In sum, the insufficiency of ex ante concerns may not be fully supported by the overall results. Table 1 provides a systematic comparison between results and predictions: no single explanation seems to rationalize the whole set of results. Alternative interpretations to BLO's, such as risk aversion or the moral wiggle room, seem to fare at least as well. In this light, the need for a mix of ex ante and ex post concerns may still be an open question.

[1] The only exceptions being 0, 50, and 100: outcomes which are also possible with other treatments.

REFERENCES

Andreoni, James, and B. Douglas Bernheim. 2009. "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica* 77 (5): 1607–36.

Andreoni, James, and John Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70 (2): 737–54.

Brock, J. Michelle, Andreas Lange, and Erkut Y. Ozbay. 2013. "Dictating the Risk: Experimental Evidence on Giving in Risky Environments." *American Economic Review* 103 (1): 415–37.

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117 (3): 817–69.

Dana, Jason, Daylian M. Cain, and Robyn M. Dawes. 2006. "What You Don't Know Won't Hurt Me: Costly (but Quiet) Exit in Dictator Games." *Organizational Behavior and Human Decision Processes* 100 (2): 193–201.

Krawczyk, Michal, and Fabrice Le Lec. 2010. "'Give Me a Chance!' An Experiment in Social Decision under Risk." *Experimental Economics* 13 (4): 500–11.

TRAVAUX EN COURS MENTIONÉ DANS LA SYNTHÈSE ET DANS l'OEUVRE SCIENTIFIQUE CAR EN L'ETAT D'ÊTRE PUBLIÉ 1 : Michal Krawczyk et Fabrice Le Lec (2018), "Eliciting distributional preferences". révision soumise pour le **Journal of Economic Behavior and Organization**. [Liste CNRS: Rang 2]

NB: La version présentée est celle ayant été invitée à resoumettre, initialement soumise au *Journal of Economic Behavior and Organization* en 2018.

Eliciting distributional preferences

Michal Krawczyk[*], Fabrice Le Lec[†]

January 13, 2018

Abstract:
We develop an experimental method of eliciting distributional social preferences. The method relies on the simple idea that a series of binary choices allows to gather estimates of indifference points, which ultimately leads to the observation of individual indifference curves between one's own payoff and another individual's earnings. This model-free, non-parametric approach uncovers patterns of distributive social preferences at the individual level and relies on very few minimal assumptions (consistency and continuity of preferences). It permits to test the accuracy of the distributional preference models proposed in the literature, but also various shared assumptions (linearity, the relevance of equality and the distinction between favourable and unfavourable domains, monotonicity in own payoff). Overall, our findings give support to the basic assumptions used in the modeling of distributional preferences, as well as previous conclusions on the strong heterogeneity of social concerns across individuals. Our results also shed light on a number of open issues in the experimental literature, especially the persistence of somewhat divergent results concerning distributional preferences: we observe that the most *common* types of preferences are quasi-maximin, but the most *intense* are those based on some anti-social component (inequity aversion, envy and spitefulness). This double heterogeneity, qualitative and quantitative, implies an aggregate (averaged) preference of the type of inequity aversion while the median preference is of a quasi-maximin type. Eventually, we use allocations elicited to be indifferent to test and quantify the strength of ex ante concerns, i.e. distributive social preferences including risk; Subjects are offered the choice between lotteries whose consequences have been previously elicited as indifferent. We find that ex ante concerns do play a role, but that this role is rather weak, approximately twice as low as ex post concerns.

Keywords: social preference, fairness models, distributive justice, procedural justice, ex ante comparison, indifference curves, altruism, concern for efficiency, in-

[*]University of Warsaw, 44/50 Dluga St, 00-241 Warsaw, Poland, mkrawczyk@wne.uw.edu.pl

[†]Université Paris-1 Panthéon Sorbonne, Centre d'Economie de la Sorbonne UMR CNRS 8174, Maison des Sciences Economiques - 106-112 Boulevard de l'Hôpital - 75647 Paris cedex 13 - France, fabrice.le-lec@univ-paris1.fr

equity aversion
JEL classification: C9, D03

1 Introduction

A large body of literature, both empirical and theoretical, has questioned the relevance of the conventional assumption of individuals' self-interest. Many laboratory experiments have shown that in some circumstances people were willing to sacrifice some of their resources to improve another individual's situation (such as in the well-known Dictator Game), while in other cases individuals were eager to sacrifice some of their earnings to reduce the other's welfare (as in the Ultimatum Game). Several models of social preferences have been developed to try to account for these empirical regularities, either by specifying utility functionals (Levine, 1998; Fehr and Schmidt, 1999; Charness and Rabin, 2002) or by the use of standard consumer theory approach (Andreoni and Miller, 2002; Cox, Friedman, and Gjerstad, 2007). In the social preference literature, two main – often complimentary– approaches have been distinguished: static preferences over social allocations, i.e. *distributional preferences* and reciprocity-based preferences. We focus here only on the first type. Of course the question how individuals allocate resources in the absence of a strategic concerns is of interest by itself. Moreover, distributional preference also generally represent a central building block of more complex models, including those based on reciprocity, as they indicate what individuals are expected to consider fair vs. unfair or kind vs. unkind, as in Charness and Rabin (2002) and Falk and Fischbacher (2006)). Last but not least, no experimental consensus seems to have been reached on the issue of the most relevant models or principles, nor on their distributions in various populations (Engelmann and Strobel, 2004; Fehr, Naef, and Schmidt, 2006; Binmore and Shaked, 2010; Cooper and Kagel, 2009; Cox and Sadiraj, 2010).

This relative lack of consensus may stem from the use of different experimental methods to test for distributional preferences. Heterogeneous methods may indeed trigger or reveal different concerns or different aspects of social preferences. Two main approaches can be found in the literature. Some studies (Engelmann and Strobel, 2004; Charness and Rabin, 2002) mostly involve testing a given model (or a set of competing models) in situations that by nature are quite specific (because they allow to test models' predictions). In these studies, subjects face discrete choice, typically with as few as two or thee allocations, each allocation corresponding to a type of concerns. In general, these methods do not allow to get a full picture of the individuals' preferences, since they only generate a single data point, or a few of them, per subject. Similarly, they do not (unless such typical situations are repeated with varying values to test the intensity of each concerns) allow measuring intensity of social concerns very precisely – i.e. how strong is a given type of social concern with respect to self-interested monetary concern). These studies mostly estimate the distribution in the subject sample of social concerns. In other experimental investigations, subjects have to choose a single point from a continuum (Andreoni and Miller, 2002; Fisman, Kariv, and Markovits, 2007; Engel, 2011; Bolton, Katok, and Zwick, 1998) or a large choice set. The researchers then observe which allocation is chosen, and compare this choice with what is compatible with the predictions of the different models. In some of these studies (Andreoni and Miller, 2002), the choice set is varied to infer, under the assumptions of revealed preference theory, the type of subjects' preferences. In these studies, the specific opportunity sets used (or equivalently: the price and the budget) may have a major effect on the findings, as highlighted by Bardsley (2008a) and List (2007). The observed decisions may also suffer from some well-known behavioral biases when choosing among many alternatives, for instance the "compromise effect". Moreover, the inference of the underlying preferences (or equivalently utility functions) relies heavily on assumptions on their shape,[1] or even the optimality of individuals' choice in one-shot sometimes complicated situations. All these methods have specific advantages, are well-suited for some specific experimental purposes and have been fruitfully applied to specific research questions such as the extent to which individuals depart from the self-interest benchmark as well as the distribution of social concerns in typical populations.

To complement these studies, we propose to approach the issue experimentally in the most basic and fundamental way, by measuring systematically preferences over a (relatively) large area of the orthant made of two individuals' allocations (the decision-maker's and a third party's). To do so, subjects face binary choices between a reference allocation and numerous alternative allocations. These alternative allocations are varied systematically in order to establish the set of allocations preferred to the reference one (the upper contour), the set of allocations dominated by the reference one (the lower contour) and the indifference curve to the reference allocation. This provides a picture of a subject's preferences as (locally) exhaustive as possible. Under the minimal assumptions that preferences are consistent and continuous, it gives a model-free method of inference of distributional preferences. This method not only provides a way to test the models proposed in the literature, but also has several other advantages. First, it permits to test the underlying (and often shared) assumptions of social preference models such as linearity of utility and the distinction between favourable and unfavourable domains. Second, it provides a picture of social preferences at the individual level. It hence allows to establish finely the distributions of social concerns in the subject sample, and to observe correlations between conceptually different aspects of social preferences, for instance the relation between the intensity of social motivations and the type of concerns.

[1]Although most assumptions may be innocuous, the one of concavity is in effect a potential problem: most models are linear, at least piece-wise, which may result in many cases in corner solutions and discontinuities. This may imply a lack of robustness, due to extreme sensitivity to minor changes in the budget constraint for instance.

The method is in spirit close to Kerschbamer's (2015) work, who estimates two points indifferent to an equal allocation, from which he infer the nature of social concerns in the favourable and unfavourable domains. More specifically, he elicits the decision-maker's payoff $\bar{x}$ that is indifferent to a baseline situation of equality, (x, x), and a situation in which the other's payoff is increased: $(x, x) \sim (\bar{x}, x+\delta)$ for a fixed positive δ. Likewise he elicits $\underline{x}$, such that $(x, x) \sim (\underline{x}, x-\delta)$. Our approach can be seen as a generalization of this "geometric" approach to distributional preferences. There are several ways in which we expand upon this approach, to some extent corresponding to Kerschbamer's own discussion of the important ramifications of his work. A key assumption of Kerschbamer' method is that it relies on the assumption that qualitative change (technically: the change in the sign of the partial derivative of the utility function wrt the other person's payoff) is only allowed at the equal allocation (which is what most prominent models predict). That is the reason why equal allocation has to be the reference allocation for the indifference curve. This have three possibly limiting consequences: first, quantitative change in the strength of social concerns is not measured; second, qualitative change except at the equal situation may not be observed[2]; and third, the focus put experimentally on the equal allocation, which appears as the reference allocation in all the binary decisions, may trigger salience effects (Güth, Huck, and Müller, 2001) or experimental demands effects (Bardsley, 2008a; List, 2007), and may reveal more about sharing norms and image concerns than "genuine" social preferences (Andreoni and Bernheim, 2009). In our experiment, we start with an unequal reference allocation, reducing the above-mentioned unwanted effects. We also elicit six, rather than just three points on each indifference curve, which allows us to detect changes in the slope (including the sign) even within the favourably unequal and within the unfavourably unequal allocations. We also add to the precision of our elicitation, by using the iterative multiple price list (IMPL) of Andersen, Harrsion, Lau, and Rutstrom (2007), which allows to estimate more precisely the indifference points. Finally, we elicit two indifference curves per individual.

Our results are quite consistent with the general findings of the literature but also shed some new light on individual distributional preferences. Concerning the

[2]Although the central role played by equality is not only intuitive but broadly–yet indirectly most of the time–empirically supported, some pattern of social preferences that do not assign a special role to equality are perfectly plausible. For example, an individual may be altruistic towards poorer individuals up to a certain threshold–perhaps depending on her income/payoff–and indifferent once this threshold is passed; likewise, an individual may be competitive up to a certain threshold, that is up to a point where she is sufficiently ahead of the other not to care anymore, or even to consider the other party's fate unfortunate and reverse to altruism: giving to the very poor but competing with the intermediate and rich. Similarly, an individual may be envious of others in her relative neighbourhood, in terms of payoff, but not to those far away, either because much poorer or because much richer ("We envy those who are near us in time, place, age, or reputation. [...] So too we compete with those who follow the same ends as ourselves [...] Hence the saying, Potter against potter." (Aristotle, 4th c. BC/2004)).

specific shapes of the indifference curves, our results are broadly consistent with previous experimental studies on distributional preferences: as found previously by Engelmann and Strobel (2004), the majority of our subjects exhibit only prosocial concerns, and the concerns of this majority is closely convergent with Charness and Rabin's quasi-maximin model. Moreover, we observe a lot of heterogeneity at the individual level, in line with Fisman, Jakiela, and Kariv (2014). In addition, we find that almost all subjects can be accounted for by an existing model (or a plausible extension) as suggested by Kerschbamer's results. Yet, and quite counter-intuitively, aggregate (averaged) preferences are best described by inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). This discrepancy between the aggregate and individual level stems from two main features revealed by the data. First, even though inequity averse individuals are fewer, they are more strongly socially oriented than the other types, and second, some composition effects arise with some infrequent types: envious, spiteful and competitive individuals, although not numerous, combine with purely maximin subjects to create an average picture of inequity aversion. These phenomena, and the quantitative measurement we provide, may partly explain why the consensus has been hard and slow to reach in the experimental literature. For instance, in a situation of strong unfavorable inequity, if the task proposed to subjects is to choose between two allocations, then the quasi-maximin and altruistic will dominate because they are more numerous. However, if the task seeks to establish whether subjects will sacrifice some of their resource to decrease other's payoff, strongly inequity averse individuals, reinforced by a few envious and spiteful ones, will lead to a rather high average willingness to pay for the other's misfortune. Put differently, we observe two types of heterogeneity: a qualitative heterogeneity–individuals have different types of concerns– that is not independent of a quantitative heterogeneity–the strengths of these concerns. Finally, we also find that equality plays, as hypothesized in most models, a central role even though our experimental protocol does not make equal allocations salient. We also confirm that the indifference curves are roughly piece-wise linear at the monetary levels involved in the experiment, monotony holds for the vast majority of subjects and finally, social concerns are consistent for varying stakes.

We complement this analysis by studying the relative importance of ex ante social preferences (based on expected payoff for instance), a topic raised and identified in several recent studies (Krawczyk and Le Lec, 2010). Indeed, having elicited indifference curve, we are able to construct lotteries to which individuals should be completely indifferent if they are only concerned with final allocations, but not if they care about the ex-ante distribution of expected payoffs. We find some evidence of ex ante concern but estimate that they play a rather weak role, and are strongly dominated by plain monetary allocation concerns.

The remainder of the paper is organized as follows: the next section provides the details of our experimental protocol and the corresponding theoretical predictions, the third section analyzes the aggregate results, while the fourth focuses on individual data, the fifth part exposes the results of the ex-ante preferences part of the experiment, while the last section provides some discussion and concluding remarks.

2 Theoretical motivation

Subjects participating in our experiment were anonymously matched in pairs consisting of Person 1 and Person 2. They were told that Person 1 would make a number of choices, one of which would be randomly selected to affect the earnings of both. The experiment was divided into two stages with different objectives. In the first stage, we elicited indifference curves of distributive preferences, testing at the individual level the accuracy of several prominent models of social preferences. In the second stage, we investigated whether such social motives play a role in risky environments, and in particular if subjects tend to consider that fairness should take into account *ex ante* comparisons (for instance expected payoffs of both participants) and not only outcomes.

2.1 Stage 1: Measuring distributional preferences

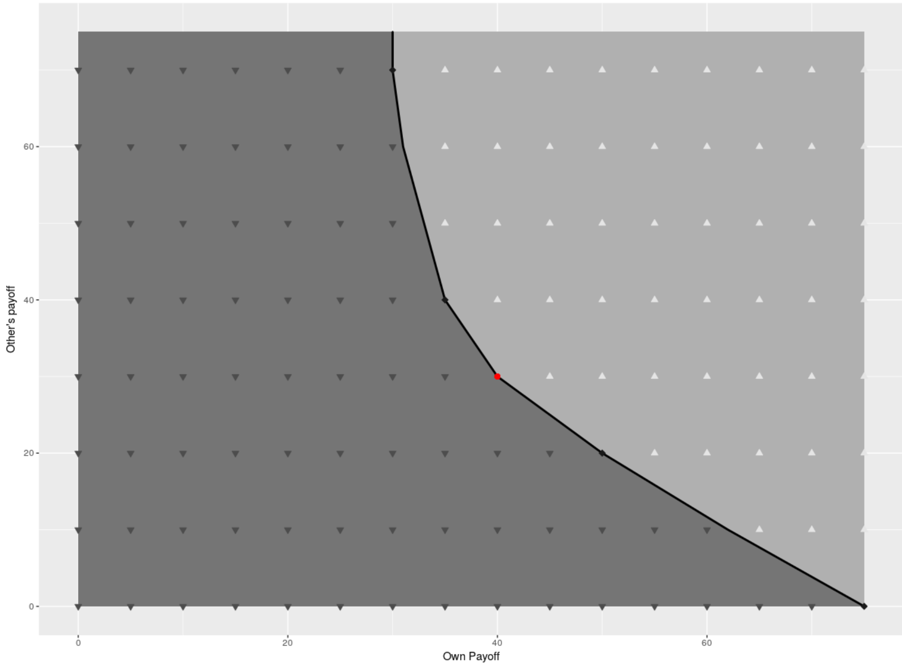
2.1.1 Distributional preferences for two individuals

Denote the set of monetary allocations for two individuals by X^2, with $X = [0, M]$ and M the maximal payoff for both. It is assumed that Participant 1 (the decision-maker) has preferences over X^2 with the usual properties. To each allocation A in X^2, corresponds three sets: the indifference curves (composed of composed of allocations $B \sim A$), the better-set or upper contour (composed of allocations $B \succ A$) and the dominated set or lower contour (composed of allocations B with $A \succ B$). We propose to use these basic framework to experimentally study distributional preferences at the individual level. The idea of eliciting experimentally indifference curves or maps has been carried out in the early days in experimental economics (MacCrimmon and Toda, 1969), and even before in psychology (Thurstone, 1931), but seems not to have been recently used except perhaps in the risk literature (see Hey and Strazzera 1989 for instance).[3] We believe that because of their non-parametric and (almost) hypothesis-free properties, such methods are particularly well suited for distributional preferences.

In practice, we fix A as a reference allocation and subject participant to a series of binary questions involving varying alternative allocations B. With the variations being systematic, we are able to infer the common contours of the better-set and

[3] A recent exception is provided by Decancq and Nys 2017, who apply fruitfully similar methods in the domain of wealth and health.

the dominated set, that is the indifference curve. An example of such a representation is given in Figure 1: the preference of an hypothetical individual displayed with respect to the red dot allocation is represented by the colored area (dark grey for the worse set, light grey for the better set) and the indifference curve (in black). Each triangle and diamond represents the theoretical answer of the hypothetical individual to the binary choice involving the reference allocation (red dot) and the corresponding triangle/diamond.



Upward triangle: measured preference for the alternative option; downward triangle: measured preference for the reference allocation; Red Dot: reference allocation; Diamonds: measured indifference

Figure 1: *A representation of hypothetical distributional preference and its corresponding binary measures.*

It is clear that based on a series of binary choices, the distributional preference map can be relatively straightforwardly estimated. To do so, we only need the minimal assumption of continuity.[4] The example displayed in Fig. 1 is a very standard preference relation (i.e., monotonic, convex, and in effect corresponding to Cox, Friedman, and Gjerstad's model, 2007), but the general measurement technique extends straightforwardly to less standard cases. Practical questions of course arise (number of binary choices, noise and errors, precision of indifference

[4]Strictly speaking, in any two points elicited to be preferred to the reference allocation, any shape for the indifference curve is possible, in principle. But major local changes are extremely implausible.

estimates, etc.), but these questions can easily be addressed by the use of standard experimental techniques. One of the serious issues is to obtain a relatively precise estimate of the indifference curve, which, provided that individual preferences are relatively well-behaved, conveys almost all important information about the individual's preference.

2.1.2 Eliciting indifference curves

Eliciting indifference curves with precision is the critical objective of our experiment. It is likely (yet to be confirmed empirically) that a large share of subjects will have monotonic preferences in their own payoff. Under this assumption, eliciting the indifference curve automatically provides the upper contour and the lower contour of the reference allocation. Estimating precisely indifference points can be achieved by using a method similar to the iterative Multiple Pricing List (Andersen, Harrsion, Lau, and Rutstrom, 2008). Each allocation decision (as represented in Fig. 1) belongs to a list of binary questions involving the reference allocation and some fixed payoff for the other party (y), with a varying payoff (x) for the decision-maker. For instance, in the first row of Fig. 1, the list of alternative options in the series of binary decisions is $\{(0,70),(5,70),(10,70),...,(70,70)\}$. For some x large enough, allocation $(x,70)$ will be chosen, while for all the options before, i.e. where the decision maker's payoff was less than x, the decision-maker will choose the reference allocation. Suppose for example that x is equal to 20. To find out exactly which value between 15 and 20 makes the subject indifferent, we add a sub-list of allocations ranging from $(15,70)$ to $(20,70)$, again eliciting binary choices. Incentive compatibility is easily achieved as instead of drawing a binary decision among all, we select a row and then any number in the entire range (between 0 and 70 in our example). By repeating the procedure for $y \in \{0,10,20,...,70\}$, we obtain a rich representation of the individual's distributional preferences.

2.1.3 Relation to existing models of distributional preferences

In the simple case of two agents, the predictions from the various models in the literature are easily derived. Under the assumption that the sign of the partial derivative of utility wrt the other's payoff only change at the equal allocation, Kerschbamer (2015) shows that there are only nine possible types. In terms of utility, the social motives (such as inequity aversion, quasi-maximin, altruism, etc.) can be expressed in terms of signs of the partial derivatives of the utility function $u(x,y)$ with x being the payoff of the decision-maker and y being the passive subject's payoff. Depending on the specific model, these partial derivatives may vary depending on the relation between x and y. Most notably, for some models, e.g. inequity aversion and quasi-maximin, the attitudes towards the other's payoff (and possibly one's own earning) depend on the ordering thereof. Most important classes of models of social preferences can thus be characterized by the sign of the partial derivative $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$ for the two domains: the decision-maker favorable domain, *i.e.*, $x > y$ and the decision-maker unfavorable domain, *i.e.*, $y > x$.

For purely self-interested individuals, we straightforwardly have $\frac{\partial u}{\partial x} > 0$ and $\frac{\partial u}{\partial y} = 0$ since they are by definition indifferent to others' fate but motivated by their own payoff. For altruistic and efficiency-driven ones, $\frac{\partial u}{\partial x} > 0$ and $\frac{\partial u}{\partial y} > 0$: the utility increases in both the decision-maker's and the passive subject's payoffs. Altruism here includes Andreoni and Miller's types, such as Constant Elasticity of Substitution, including perfect substitution and Cobb-Douglas, but not Leontieff preferences (see below). Indeed, altruism and efficiency lead to the same generic pattern, since increasing anyone's payoff leads to an increase in efficiency (understood as the sum of payoffs). For spiteful individuals (as well as competitive individuals defined as having their utility increasing in x and $x - y$), we have $\frac{\partial u}{\partial x} > 0$ and $\frac{\partial u}{\partial y} < 0$: they tend to prefer higher payoff for themselves but experience disutility from the other's. For inequity averse individuals, two cases need to be distinguished: in the case of unfavorable inequity aversion, i.e. $x < y$, we have $\frac{\partial u}{\partial x} > 0$ and $\frac{\partial u}{\partial y} < 0$, since increasing the decision-maker's payoff reduces inequity and favors self-interest, while increasing the other's payoff widens inequity. In the case of favorable inequity aversion, i.e. $x > y$, one gets $\frac{\partial u}{\partial y} > 0$ because increasing the other's payoff reduces, ceteris paribus, inequity, and $\frac{\partial u}{\partial x} > 0$ because the benefit of higher payoff typically outweighs the cost of greater inequality. Admittedly, in the case of extreme inequity aversion, we would have $\frac{\partial u}{\partial x} < 0$. We focus on the case where $0 < \frac{\partial u}{\partial x} < \frac{\partial u}{\partial y}$, which corresponds to the assumption made in prominent inequity aversion models (Fehr and Schmidt, 1999). Regarding Charness and Rabin's (2002) model, for maximin individuals, whose utility depends positively on one's individual payoff and the payoff of the worst-off agent, we have $\frac{\partial u}{\partial x} > 0$ and $\frac{\partial u}{\partial y} > 0$ in the favorable domain, and $\frac{\partial u}{\partial x} > 0$ and $\frac{\partial u}{\partial y} = 0$ in the unfavorable one. For the quasi-maximin model (that is the maximin model plus efficiency concerns), we have in both domains, $\frac{\partial u}{\partial x} > 0$ and $\frac{\partial u}{\partial y} > 0$.

In some specific cases, the preferences cannot be represented by a differentiable utility function. In particular, Leontieff preferences (Andreoni, 1988) fall into this category. As is well known, the corresponding indifference curves exhibit a typical "rectangular" pattern that can be directly be inferred from our experimental representation of preferences.

2.1.4 Theoretical Predictions for Indifference Curves

Given that the shape of the indifference curve is given by the marginal rate of substitution between the decision maker's payoff and the other's, i.e. $\frac{dy}{dx} = -\frac{\partial u}{\partial x} / \frac{\partial u}{\partial y}$, generic predictions can be straightforwardly derived and are displayed in the first two columns of Table 1. For convenience, we use $\frac{dx}{dy}$ rather than $\frac{dy}{dx}$ in order to get rid of infinite values in the case of vertical indifference curves, i.e. pure self-interest. In addition, we relate these generic models to their specific instances found in the literature. First, regarding inequity aversion, the Fehr and Schmidt (1999) model is linear, while Bolton and Ockenfels (2000)'s is more general in allowing non-constant partial derivatives. One important feature of the Fehr and Schmidt (1999) model is also that unfavorable inequity aversion is stronger than favorable inequity aversion: although it is not theoretically crucial, it plays an essential part in the ability of the model to organize experimental stylized facts. Charness and Rabin's model is also linear, which once again is more a question of tractability and ease-of-use than a fundamental requirement. Yet, it yields an interesting property for the indifference curve, which is that the slope of the indifference curve in the favorable domain (both maximin and efficiency tend to increase y) is steeper than the one in the unfavorable domain (only efficiency pushes towards an increase of y, while maximin favors x). That would result in a kink at the equal allocation, that is $\lim_{x^+ \to x} s_f(x^+, x) > \lim_{x^- \to x} s_u(x^-, x)$, even in the non-linear version of the model. This kink is the only difference in terms of qualitative predictions between quasi-maximin and constant elasticity altruism: while the slope of the indifference curves becomes gradually less steep in x in the latter, they should exhibit a kink at the equal allocation for the former.

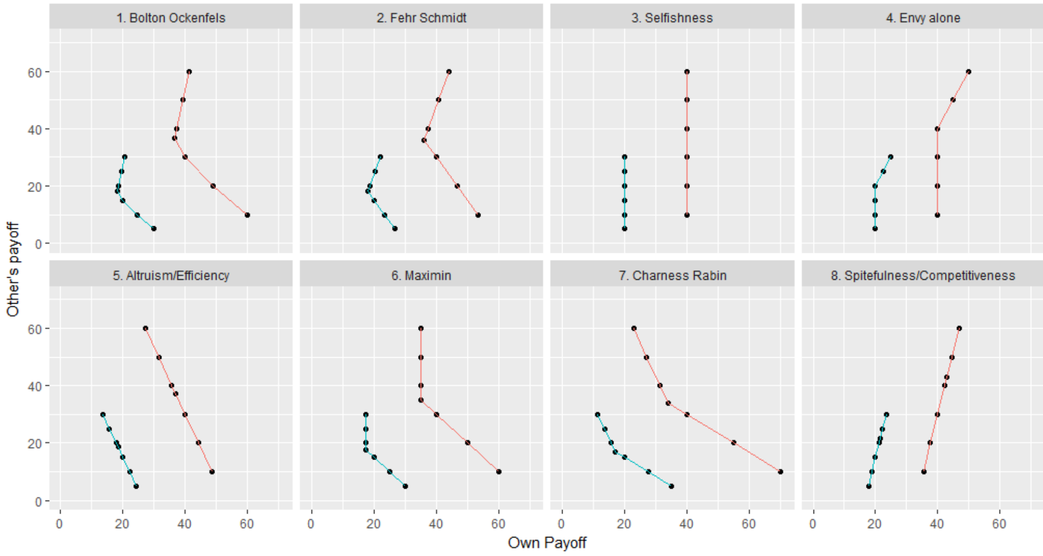
	$s_f = \frac{dx}{dy}$ favorable	$s_u = \frac{dx}{dy}$ unfavorable	s_f vs. s_u	remarks
Self-interest	0	0		
Quasi-Maximin	Neg	Neg	$s_f < s_u < 0$	s_f, s_u constant
Maximin	Neg	0		
Efficiency	Neg	Neg		
Altruistic	Neg	Neg	$s_f = s_u$ for Levine	$\frac{d^2y}{dx^2} > 0$ for Cox et al. 2007, Andreoni and Miller (2002)
Inequity Averse	Neg	Pos	$\mid s_f \mid < \mid s_u \mid$	s_f, s_u const. for F-S
Envious	0	Pos		
Spiteful	Pos	Pos	$s_f = s_u$ for Levine	
Competitive	Pos	Pos		

Table 1: *Predictions for slopes of indifference curves*

The specific predictions are provided in the rightmost columns of Table 1. Several typical indifference curves derived from different models and corresponding to the specific starting points that were used in our experiment are also represented graphically in Figure 2.

2.2 Stage 2: Measuring ex ante and ex post concerns

Several studies have put forth that individuals not only have regards for ex post allocations of monetary payoffs, but also take into consideration the distribution of chances of obtaining some payoff (Krawczyk and Le Lec, 2010; Bolton, Brandts, and Ockenfels, 2005; Brock, Lange, and Ozbay, 2013). The intuition is quite straightforward: an inequity-averse or maximin individual may (relatively) dislike both allocations $(50, 0)$ and $(0, 50)$ with certainty, and find a lottery that gives



Linear versions of altruism/efficiency, envy and spitefulness/competitiveness models depicted.

Figure 2: *Examples of typical indifference curves based on models*

equal chances to both allocations preferable to any of the two sure allocations. The lottery provides an equal expected payoff to both individuals, or maximizes the minimal expected payoff. Such a rather intuitive preference cannot be accounted for ex post concerns alone. The issue of the importance of ex ante concerns (and more generally the interaction between risk and outcomes in a social setting) is still an open question, with studies finding a substantial role of risk and ex ante concerns (Krawczyk and Le Lec, 2010; Brock, Lange, and Ozbay, 2013) without being necessarily conclusive about their relative importance (Krawczyk and Le Lec, 2016), and others finding little effect of the risk dimension (Rohde and Rohde, 2011; Brennan, González, Güth, and Levati, 2008).

Having indifference points estimated at the individual level indeed provides an opportunity to test the role of risk in other-regarding preferences, and more specifically the role of ex ante concerns (that include the distribution of chances) and ex post ones (that rely only on the final consequences). By definition, individuals are indifferent between any two certain allocations A and B on their indifference curves, and if they have only outcome-based distributional preferences, any lottery whose support is $\{A, B\}$ should be equivalent either to A or B or any other lottery with the same support. On the contrary, if the distribution of risk or of expected payoff matters, individuals may deviate from indifference between such lotteries and certain allocations.

By constructing lotteries based on previously elicited indifferent ex post allocations and testing whether or not individuals prefer such lotteries or any sure outcome on the indifference point, we are able to test the robustness of results obtained mostly in dictator-like experimental configurations (Krawczyk and Le Lec, 2010; Brock, Lange, and Ozbay, 2013), that may trigger various experimental demand effects (Bardsley, 2008a), image concerns, etc. Moreover, under the assumption that distributional concerns have similar motivations (inequity aversion, maximin, total welfare, etc.) ex ante and ex post, we can *quantify* the relative importance of these two considerations based on both the choice of individuals between social lotteries and the former elicitation of individual distributional preferences. Indeed, from the existing literature, some evidence suggests that ex ante concerns may be enough to account for observed behavior (Krawczyk and Le Lec, 2016).

To both test ex ante concerns and estimate their strength relative to ex post ones, we develop in Appendix A a formal framework, in the spirit of Saito (2013), to determine the conditions under which individuals depart from indifference between lotteries with support in the indifference curve and certain allocations on this indifference curves.

3 Experimental design

3.1 Stage 1

As already discussed, we elicit indifference curves using a modified iterative multiple price list (Andersen, Harrsion, Lau, and Rutstrom, 2006, 2007) where subjects face a series of binary choices. We keep one alternative, labeled option A, unchanged, with payoffs x_A and y_A for Person 1 (the decision maker) and Person 2 (the passive participant) respectively. (x_A, y_A) is hence the *reference allocation.* For each additional point on the indifference curve we then set some amount for Person 2 to receive in case option B is chosen, y_B. We then let the subject choose between options A and B for different values of x_B, the amount that Person 1 receives in the case option B is chosen. The switching point determines the value of x_B such that (x_A, y_A) is (approximately) indifferent to (x_B, y_B). One such series of questions is shown in Table 2, with $x_A = 40$, $y_A = 30$ and $y_B = 10$ (all amounts in Polish zloty; 1PLN is equivalent to ca. .23EUR).

In each row subjects had to indicate whether they preferred Option A, Option B or were indifferent (I). For a typical pattern of starting with A in early rows and then eventually switching to B (possibly with one or more Is in between), subjects would then see a second iteration, in which the values of x_B would vary between the x_B from the last row before the first non-A and the first row after the last non-B. For instance, if a subject chose AAABBBBB, the new iteration of the

Option A	Option B
40 for you, 30 for the other	0 for you, 10 for the other
40 for you, 30 for the other	8 for you, 10 for the other
40 for you, 30 for the other	16 for you, 10 for the other
40 for you, 30 for the other	24 for you, 10 for the other
40 for you, 30 for the other	32 for you, 10 for the other
40 for you, 30 for the other	40 for you, 10 for the other
40 for you, 30 for the other	48 for you, 10 for the other
40 for you, 30 for the other	56 for you, 10 for the other
40 for you, 30 for the other	64 for you, 10 for the other
40 for you, 30 for the other	72 for you, 10 for the other
40 for you, 30 for the other	80 for you, 10 for the other
40 for you, 30 for the other	88 for you, 10 for the other
40 for you, 30 for the other	96 for you, 10 for the other
40 for you, 30 for the other	104 for you, 10 for the other

Table 2: *Typical sequence of binary tasks in iMPL*

table would start from the value of row 3 and end at the value of row 4. In the rare cases of inconsistent patterns of choice the procedure still applies: for instance AAAIBIBIBBBB, the second iteration table would start from the value of row 3 and end at the value of row 7.[5] With this method, we were able to determine one point of indifference with (x_A, y_A). An alternative way to elicit indifference curves would have been to fix the value of x_B and to vary y_B. However, the obvious weakness of this approach is that neither fully selfish subjects nor subjects averse to disadvantageous inequity would be indifferent between, say $(40, 30)$ and $(30, y)$, no matter how high the value of y could go.

As a reference allocation (the A option that does not vary in the iMPL), we preferred some unequal allocation in order to reduce the salience of the equal split[6] as well as to mitigate possible experimenter demand effect (Bardsley, 2008b). Since perfect equality plays a special role in several models, e.g. inequity aversion and quasi-maximin models, the use of a reference allocation not involving equal payoffs provides a tougher test for these models and more generally for the idea that equal allocations have a special status. For practical purpose, a mildly unequal reference allocation provides enough room in the advantageous (decision-maker's payoff is greater than the other participant's) as well as disadvantageous area (the other participant's payoff is greater than the decision-maker's) to investigate both

[5]These "extreme" values of y_B under second iteration were sometimes modified slightly in order to make sure that the number of rows was not greater than 14, that the increment of x_B was constant (typically equaling 1 PLN) and no decimals had to be used.

[6]The presence of a perfectly equal allocation seem to lead to exaggerated aversion to inequity (Ubel, Baron, and Asch, 2001; Güth, Huck, and Müller, 2001).

domains in a (roughly) balanced way. To check the robustness of our results, we elicited two indifference curves corresponding to two reference allocations: (40,30) in what we call a high-payoff task (H) and (20,15) in low-payoff task (L). For H, we vary y_B among $\{10, 20, 40, 50, 60\}$ and for L among $\{5, 10, 20, 25, 30\}$. For each condition, that ensures that we had two tables in which choosing B would make the payoff of Person 2 higher than Person 1's payoff under option A and two tables in which it would make it lower, and guarantees a balanced measurement of the favorable and unfavorable domains. The fact that both conditions are perfectly homothetic also allows to test for the consistency and robustness of the findings at the individual level.

Using iMPLs we had to decide on $x_B^{\max}$, the maximal amount offered in option B on a list (e.g., 104 in Table 2). We had it equal to $x_A + y_B - y_A = y_B + 10$ when y_B was greater than y_A, so that, for a given difference between Person 1 and Person 2's payoffs, assuming nearly anyone will choose the option under which these payoffs are the highest, *e.g.* would prefer $B = (50, 40)$ to $A = (40, 30)$). When y_B was lower than y_A, $x_B^{\max}$ was set to $x_A + 3(y_A - y_B)$, allowing subjects to express even very strong aversion to lowering other person's payoff when she is behind. For instance, in the H condition x in option B ($(x, 10)$) would run from zero to 100.[7]

The iMPL procedure typically allowed us to estimate indifference to the nearest PLN. We would take the mean value of x_B in the last row, in which subject preferred A and the value of x_B in the first row in which she preferred B, both in the last iteration of the procedure, as the indifference point. The value of x, for which a subject was indifferent between $(40, 30)$ and $(x, 10)$ will be denoted $x^H(10)$ and similarly, the value of x for which the subject is indifferent between $(20, 15)$ and $(x, 5)$ will be called $x^L(5)$. For each value of y_B, we obtain an indifference value for x_B. Repeating the procedure with five y_B and the two reference allocations, this allows to obtain five indifferent points on each of the two indifference curves, in addition to the reference allocation, based on 10 iterative multiple price lists, corresponding to approximately 280 binary choices. The whole set of tasks is summarized in Table 3.

3.2 Stage 2 experimental design

Once indifference points are estimated in Stage 1, we are able to construct lotteries whose outcomes are supposed to be indifferent to subjects. We use the generic theoretical predictions of Appendix A to assess experimentally the extent to which social preferences apply to *ex ante* and *ex post* social comparisons. The first set of

[7]This number was adjusted slightly if necessary to keep the increment constant as in Table 2 where the $x_B^{\max} = 104$ instead of 100.

Reference alloc. (x_A, y_A)	vs	Varying alloc. (x_B, y_B)
$(40, 30)$	vs	$(x_B, 10)$
$(40, 30)$	vs	$(x_B, 20)$
$(40, 30)$	vs	$(x_B, 40)$
$(40, 30)$	vs	$(x_B, 50)$
$(40, 30)$	vs	$(x_B, 60)$
$(20, 15)$	vs	$(x_B, 5)$
$(20, 15)$	vs	$(x_B, 10)$
$(20, 15)$	vs	$(x_B, 20)$
$(20, 15)$	vs	$(x_B, 25)$
$(20, 15)$	vs	$(x_B, 30)$

Table 3: *List of all iMPL, elicitation of* x_B^* *such that* $(x_A, y_A) \sim (x_B^*, y_B)$

choice tasks in Stage 2 is composed as follows: the program randomly picks three out of these six allocations on the indifference curve, denoted P, Q, and R[8] and lets the subject choose between R for sure and a lottery involving P or Q with probability one half each. Such a task was run three times with allocations P, Q, and R being reselected. As laid out in Proposition 1, the assumption of maximization of outcome-based utility leads to the prediction of perfect indifference in any of such cases because P, Q, and R were revealed equally good. However, if preferences are not based on outcome comparisons but on *ex ante* ones, *i.e.* based on the distribution of chances, choices should depart from pure indifference whenever P and Q are on different sides of the $x = y$ line (see Appendix A, Proposition 1). By measuring social motives on expected payoffs such as expected equity, expected efficiency, etc., as estimated at the individual level based on Stage 1's results, we are then able to see if departures from indifference are linked to ex ante considerations.

Additionally, in order to assess the strength of this preference, each question was supplemented with an additional one, in which the alternative non-preferred in the first step was made relatively better: in case the decision maker choose R, 3PLN was subtracted from the decision maker's payoff under R. If she chose the lottery, 3 PLN would be added to it. Suppose for example that someone preferred $[(42, 21), .5; (38, 51), .5]$ to $(41, 31)$, e.g. because the former would result in lower inequality in expected terms[9], we would then ask whether she still preferred it when the sure thing was slightly improved, by adding 3PLN to her sure payoff,

[8] More precisely, 1PLN was always added to all payoffs or subtracted from all payoffs to make it less transparent to subjects that these were their previously elicited indifference points.

[9] Note that $(40, 30)$ is the reference allocation while $(41, 20)$ and $(37, 50)$ are assumed here to have been judged by this subject as just as good in Stage 1. 1PLN was subsequently added to all payoffs to obscure the chained nature of the design, as explained before.

namely we would let her compare $[(42, 21), .5; (38, 51), .5]$ against $(44, 31)$ etc.[10] The initial Sure thing vs. Lottery choices will be denoted as SL and the choices with the additional 3PLN as SL'. Note that the number of tasks may vary for each subject given that if a participant chose Indifference in the initial task SL, then no SL' task is added.

Finally, in the last set of tasks (LL) we had subjects face lottery versus lottery choices. These were, again, constructed using indifferences elicited in the first part. This time we would use any three allocations P, Q, R identified as equally good, round them and modify slightly as before and ask the subject to choose between two risky prospects $(P, .5; Q, .5)$ vs. $(P, .5; R, .5)$. Again, as shown in Proposition 2 (Appendix A), ex ante social concerns are expected to push subjects towards the lottery whose outcomes are more spread. Five different choice tasks were asked in this set. The tasks faced by subject in the second stage are summarized in table 4. The more procedurally oriented an individual is, the stronger preference towards mixing (distant allocations) we expect.

3.3 Experimental procedures

The experimental sessions were run at the Laboratory of Experimental Economics, University of Warsaw, Poland. Participants were recruited from the local subject pool using ORSEE (Greiner, 2015). 13 experimental sessions were run with a total of 202 subjects making all the choices.[11] For obvious reasons, sessions were perfectly balanced in terms of allocation to roles: in all sessions half the subjects would become Person 1 and make consequential decisions and the other half would make analogous, yet hypothetical, choices as Person 2. Once subjects had read the instructions and been through control questions, everyone was told which role she or he was to play. Revealing it only after decisions could have distorted answers, as there are concerns in the literature about the role that such *ex post* role assignment could play. Such a procedure could substantially influence choices if subjects had "procedural" or ex-ante preferences (Fehr, Naef, and Schmidt, 2006; Krawczyk and Le Lec, 2010; Trautmann, 2009): a post-experiment random draw to determine roles could in expected terms make it fairer to choose selfishly in the stage task.

The experiment was computerized using LABsee software, developed by Robert Borowski.[12] Experimental instructions are available in Appendix D. Elicitation of

[10]This was 5 PLN in the pilot session but we noted that it would almost always reverse the preference.

[11]Additionally, an incentivized pilot with 12 subjects was run, with only high-stake choices. We refer to this session as Session 1, but do not include it in the analysis of the results, its effects on our conclusion being negligible.

[12]See www.labsee.pl

Stage 1's choices	Choice Tasks SL (3)
$P = (x_P, y_P)$ $Q = (x_Q, y_Q)$ $R = (x_R, y_R)$ $P \sim Q \sim R$ P, Q, R chosen randomly on the indif. curve	Option A: R^ϵ Option B: $\frac{1}{2}Q^\epsilon + \frac{1}{2}P^\epsilon$ $P^\epsilon = (x_P + \epsilon, y_P + \epsilon)$ $Q^\epsilon = (x_Q + \epsilon, y_Q + \epsilon)$ $R^\epsilon = (x_R + \epsilon, y_R + \epsilon)$ $\epsilon = \pm 1$
	Choice Tasks SL' (0 to 3)
$P = (x_P, y_P)$ $Q = (x_Q, y_Q)$ $R = (x_R, y_R)$ $P \sim Q \sim R$ same P, Q, R as in choice tasks SL	Option A: $R^\epsilon = (x_R + \epsilon \pm 3, y_R + \epsilon)$ Option B: $(P^\epsilon, .5; Q^\epsilon, .5)$ $P^\epsilon = (x_P + \epsilon, y_P + \epsilon)$ $Q^\epsilon = (x_Q + \epsilon, y_Q + \epsilon)$ $\epsilon = \pm 1$
	Choice Tasks LL (5)
$P = (x_P, y_P)$ $Q = (x_Q, y_Q)$ $R = (x_R, y_R)$ $P \sim Q \sim R$ P,Q,R chosen randomly on the indif. curve	Option A: $\frac{1}{2}P^\epsilon + \frac{1}{2}Q^\epsilon$ Option B: $\frac{1}{2}P^\epsilon + \frac{1}{2}R^\epsilon$ $P^\epsilon = (x_P + \epsilon, y_P + \epsilon)$ $Q^\epsilon = (x_Q + \epsilon, y_Q + \epsilon)$ $R^\epsilon = (x_R + \epsilon, y_R + \epsilon)$ $\epsilon = \pm 1$

Table 4: *Stage 2 tasks*

the high-stakes indifference curve was followed by low stakes in sessions 2, 3, 5, 10, 12, and 13, while the reverse was true for sessions 4, 6, 7, 8, 9, 11 and 14. Within a task (high-stake or low-stake), the order of subtasks, differentiated by their specific value of x_B, *e.g.*, 10, 20, 30, 50 or 60 in the high-stake case, was randomly determined at the individual level. At the end, one task was picked randomly and the decision of Person 1 was implemented to determine actual payments. Sessions lasted for approximated 70 minutes and subjects earned approx. 35 PLN (8 EUR) on average.

On a methodological note, since parameters of Stage 2 questions depended on Stage 1 answers, "switching" from A to B in Stage 1 "later" (i.e. in a lower row) than truly preferred could be beneficial to the subject, because she could later receive better deals in Stage 2.[13] This is a common problem in experimental literature using dynamic elicitation techniques (see e.g. van de Kuilen and Wakker 2009). Our subjects were, truthfully, told that "choosing the preferred option turns out to be optimal if this part is selected for real payment". The responses to the open-ended questions at the end of the experiment do not let us identify a single case in which a subject would realize the link between questions, so we do not believe that it distorted the answers.

Finally, a number of demographic variables were collected, including gender, age, number of siblings, academic major and year of study (if any), net household income per capita and experience with economic experiments in the past.

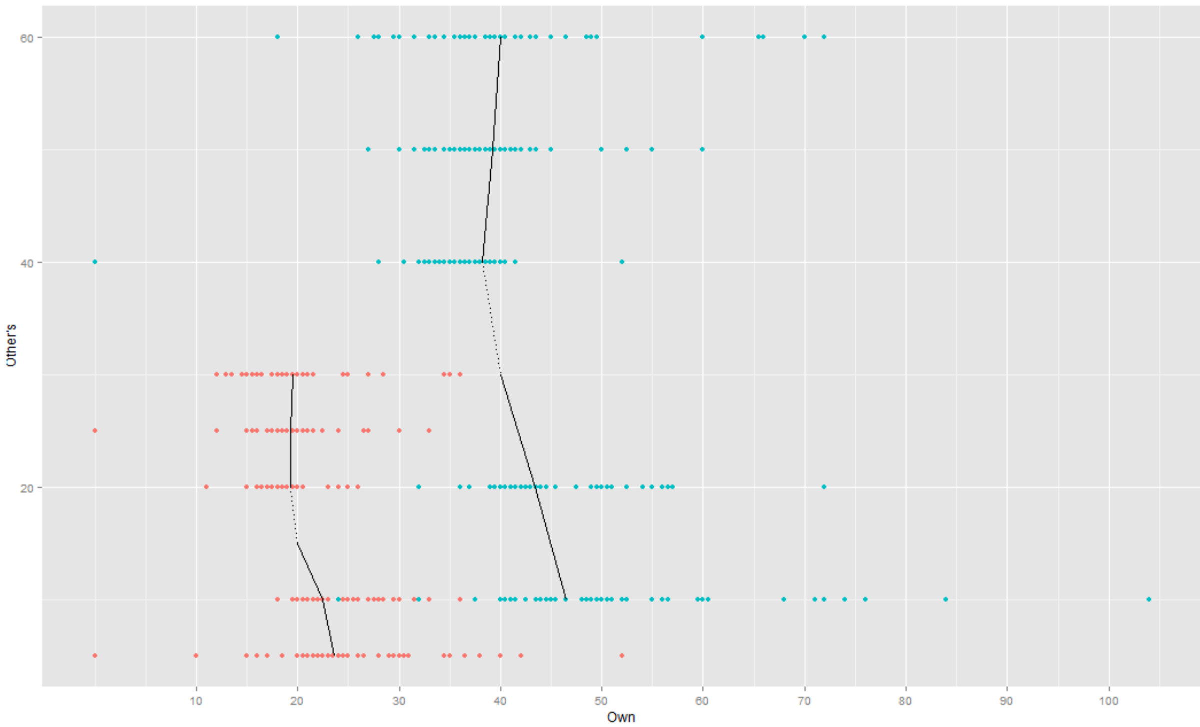
4 Results

Regarding the first stage of the experiment, the elicitation of distributional preferences, we distinguish the aggregate pattern and the individual analysis. Indeed, one of the main findings is that they diverge to a certain extent, a phenomenon explained by the heterogeneity of subjects in this regard. In each case, we analyze the data along three lines: descriptive statistics for the indifferent points, a piecewise linear estimation of indifference curves and a structural estimation based on a general linear specification of utility. The results of the second stage are presented in a similar fashion. All the analysis is based on the sample of subjects facing real incentives ($n = 101$).

4.1 Aggregate results of Stage 1: a mild inequity aversion

Table 5 displays mean and median choices in Stage 1; the choices are also graphically represented in Fig. 3. A substantial share of subjects depart from pure self-interest: they choose on average $(46.9, 10)$ as equivalent to $(40, 30)$, hence sacrificing almost 7PLN for the sake of the other's (relatively) low payoff. Something similar, in a scaled-down form, is observed for the case of low stakes, where on average the decision maker is indifferent between $(20, 15)$ and $(23.7; 5)$. Yet when the other's payoff reaches relatively high levels, we observe no direct departure from self-interest, for instance on average $(40; 60)$ is approx. equivalent to $(40, 30)$. The median choice shows a less pronounced pattern, probably because of a substantial share of purely self-interested individuals. Again, results for low stakes show the same tendency; it is worth noting how the two indifference curves seem to yield consistent numbers: the mean values for x in the case of low stakes are almost exactly half of their H counterparts, suggesting weak effects of stakes and that the elicitation procedure is robust to change in monetary stakes.

[13] Although, of course, this is hardly a viable strategy because subjects do not know how the chaining works.



The solid lines represent linear interpolation of average indifference points. The section above the reference allocation is dotted to emphasize that a kink at perfect equality (rather than straight line) is to be expected.

Figure 3: *Average indifference curves.*

	$x^L(5)$	$x^L(10)$	$x^L(20)$	$x^L(25)$	$x^L(30)$
Average	23.68	22.56	19.23	19.30	19.50
Median	21.50	21.50	19.50	19.50	19.50

	$x^H(10)$	$x^H(20)$	$x^H(40)$	$x^H(50)$	$x^H(60)$
Average	46.92	43.54	38.18	39.15	39.90
Median	42.50	41.50	39.50	39.50	39.50

Table 5: *Average and median x^L and x^H.*

Figure 3, which displays the "average" indifference curve, shows a general pattern quite similar to inequity aversion models: a pronounced kink near the equal situation, a slightly upward-sloping curve in the unfavorable domain, and a downward one in the favorable one. This is partly supported statistically by a series of paired t-tests (Table 6), where choices in situation close to equity and more extreme allocations (in the favorable and unfavorable domain) are tested against each other.

Comparisons	Differences	t (p)
$x^L(5)$ vs $x^L(10)$	1.13*	2.16 (.03)
$x^L(20)$ vs $x^L(25)$	-0.07	0.24 (.81)
$x^L(25)$ vs $x^L(30)$	-0.20	0.64 (.52)
$x^L(20)$ vs $x^L(30)$	-0.27	0.89 (.38)
$x^H(20)$ vs $x^H(30)$	3.60***	3.25 (.001)
$x^H(40)$ vs $x^H(50)$	-0.96	1.41 (.161)
$x^H(50)$ vs $x^H(60)$	-0.75	1.12 (.27)
$x^H(40)$ vs $x^H(60)$	-1.71*	1.68 (.09)

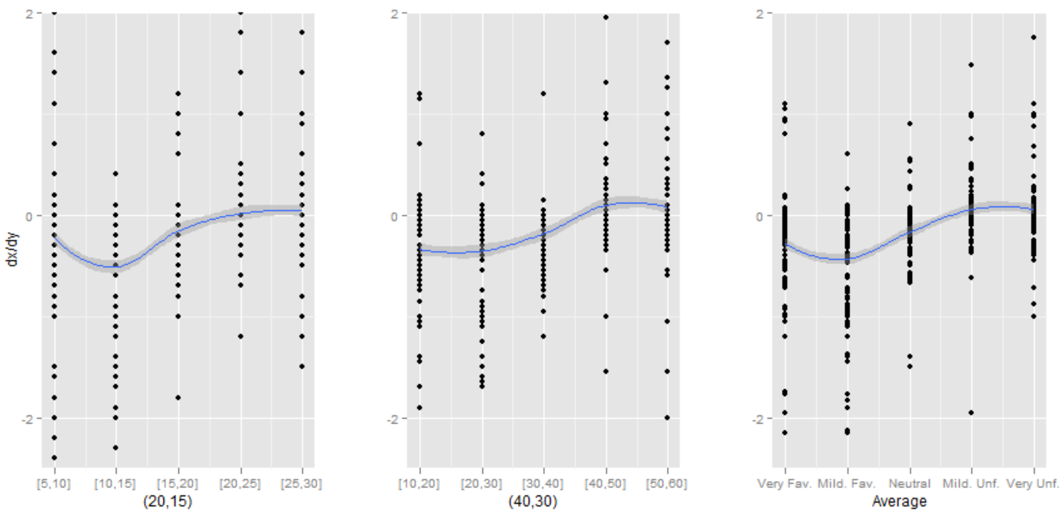
*=significant at the .10 level, **= significant at the .05 level, ***= significant at the .01 level

Table 6: *Paired Student-t Tests*

The picture drawn by these aggregate results can be summarized in three points. First, there is strong evidence in the favorable domain that people are willing to sacrifice some of their payoff to improve the other's situation, and this is the most striking deviation from the selfish benchmark in our data. Second, the average indifference curve seems to exhibit a kink between $(30, 40)$ and $(40, 30)$, or equivalently between $(15, 20)$ and $(20, 15)$, giving support to models that distinguish favorable and unfavorable domains. Third, it seems that on average the decision maker tends to mildly dislike situations where the other's payoff is higher than hers.

4.1.1 Piecewise Linear Indifference Curve Estimation

To complement this aggregate statistical description (based only on indifference points), we rely on a simple piecewise linear estimation of indifference curves. Thus, we estimate the slopes of the indifference curves elicited experimentally in between two points. Given the mild differences in monetary stakes in between two contiguous points, for instance $(10, x)$ and $(20, x)$, it seems appropriate to assume that such a linear piece-wise approximation should be sufficient. In practice, we compute the reverse of the slope in between two points since in the case of purely self-interested individuals the curve is vertical. The results are displayed in Fig. 4: we calculate the average of the relevant slope in the two tasks.[14] The results are



The blue line represents locally weighted smoothing (LOESS), with the grey ribbon corresponding to standard errors.

Figure 4: *Indifference curves slopes*

in line with the general statistical description of the previous subsection. Slopes seem to be on average negative in the favorable domain, and slightly positive in the unfavorable one.

We find very little evidence of curvature: both for high and low stakes, we find no statistically significant difference in pairwise comparisons of slopes when comparing two contiguous ranges in the same domain (favorable or unfavorable), e.g. [10, 20] and [20, 30] for high stakes. The only exception is that in the case of low stakes, the slope for [5, 10] seems to be greater than that of [10, 15] ($p < 0.001$, $t = 2.86$). If anything, this suggests an effect opposite to that of a concave shape of the underlying utility function (as in Cox et al. 2007 for instance).

[14]This seems justified empirically, as the H and L curves are similar and also theoretically in view of most of the models presented in section 2.

Because we find little curvature and little difference between shapes of H and L, the following general linear two-parameter utility model appears appropriate:[15]

$$\begin{aligned} u(x,y) &= x + \delta y \quad \text{if} \quad x < y \\ u(x,y) &= (1+\delta-\gamma)x + \gamma y \quad \text{if} \quad x \geq y \end{aligned} \tag{1}$$

Given that this simple specification determines uniquely the slopes of the indifference curves in the favorable and unfavorable domains, we estimate the slopes of the indifference curve by domain by domains by taking the average of the linear interpolation used previously. The aggregate results are given in Table 7 and tends to support our previous conclusions that individuals on average exhibit clear prosocial concerns in the favourable domain, and weak anti-social concerns in the unfavourable one.

	task		
	H	L	Both
Slope $\frac{dx}{dy}$ under fav. ineq. s_f	$-0.35^{ttt,www}$	$-0.37^{ttt,www}$	$-0.36^{ttt,www}$
Slope $\frac{dx}{dy}$ under unfav. ineq. $s_u = -\delta$	0.09^{t}	0.03	0.06^{t}
γ	$0.12^{tt,www}$	$0.19^{ttt,www}$	$0.20^{ttt,www}$

t,tt,ttt=significant at the .10 level, .05 and .01 level for a one-sample Student t-test against 0, w,ww,www=significant at the .10 level, .05 and .01 level for a one-sample signed-rank Wilcoxon test against 0

Table 7: *Average estimated slopes and parameters for the general linear model*

4.1.2 Structural Estimation

To complete this aggregate analysis, we estimate by maximum likelihood the parameters of the general linear model presented in Eq. (1). One advantage of doing so is that we can run this estimation procedure on the whole set of binary tasks (and not just estimated indifference points): this allows to take into account possibly inconsistent series of binary choices. Given that our goal in this section is to account for the aggregate patterns, the estimation is run on the pooled set of individuals, so the error term in the specification covers the individual randomness but more importantly the differences between individuals. The estimation assumption here is that all individuals have the same structural preferences, plus or minus individual specifics accounted by the stochastic error term.

[15] Of course alternative, equivalent parameterizations are possible, in particular, for riskless choices this formulation corresponds to equation (2) in Appendix A with $\delta = -\alpha/(1+\alpha)$ and $\gamma/(1+\delta-\gamma) = \beta/(1-\beta)$.

The estimates are presented in Table 8. All standard error estimates are clustered by individuals or by individuals and tasks. These estimates are consistent with the simple linear estimates of the previous subsection and even tend to give additional support to the inequity aversion models, which are the only ones that are consistent with both a positive γ and a negative δ.

	Logit	Probit
δ	-.108**†	-.114*
	(.036/.051)	(.046/.078)
γ	.238***†††	.255***†††
	(.045/.042)	(.053/.059)
μ	5.239***†††	11.244***†††
	(.521/.824)	(.921/1.719)
log(LL)	-7611	-7963
AIC	15230	15932
n	20,093	20,093
Clusters	101/202	101/202

Coefficients
(standard-errors with individual cluster/ standard-errors with individual and task cluster)
*= significant at the .05 level, **= significant at the .01 level, ***= significant at the .001 level.
Stars are used for individual clusters, daggers for individual and task clusters. p-values are obtained using the Z statistic.

Table 8: *Structural estimation results*

Overall, these two methods tend to support, at the aggregate level, the inequity aversion models. Indeed, based on the two types of estimations obtained, we would have, once the utility is rewritten (and renormalized), in the spirit of inequity aversion models:

	Based on PW Linear Est.	Based on Structural Est.
for $x < y$	$u(x,y) = x - 0.06(y-x)$	$u(x,y) = x - 0.12(y-x)$
for $x \geq y$	$u(x,y) = x - 0.14(x-y)$	$u(x,y) = x - 0.27(x-y)$

These two set of estimates are consistent with the prediction of inequity aversion models: unfavorable as well as favorable inequity aversion have a negative role in the utility function. Yet, in contrast to inequity aversion models, aversion to favorable inequity seems stronger than aversion to unfavorable inequity. But overall, the average preference, as elicited by our experimental procedure, seems to correspond to a mild version of inequity aversion models.

4.2 Individual data of Stage 1: Quasi-maximin dominates

Quite interestingly, individual types in generally do not correspond to the aggregate pattern. Our analyses of individual data reveals that the performance of the inequity aversion models at the aggregate level is largely an artefact of data composition: quasi-maximin and strictly prosocial considerations best describe the vast majority of individuals.

Individual indifference curves are displayed in Fig. 6. Perhaps the most striking feature is the high level of internal consistency. In most cases, the shapes of both curves are similar and only one subject exhibits straightforward inconsistency (#0902_4_19 whose curves are crossing).[16]

But, across individuals, the heterogeneity of indifference curves is apparent: several different patterns are necessary to account for the pool of individual curves. The second striking feature is that although it seemed to perform better than alternative models at the aggregate level, very few subjects (e.g. #1002_1_16) seem to exhibit an indifference curve shape typical of inequity aversion.

Overall, subjects can be divided into a few clear categories. First, a substantial share seems to be mostly or completely self-interested because their curves are perfectly (or close to perfectly) vertical. About one in four subjects exhibit this pattern (e.g. 0902_1_12). Second, a comparable group, including 0902_2_2, seems to show concern for others in favorable situations and pure self-interest in unfavorable situations (which contrast with the idea that individuals are, generally speaking, more sensitive to disadvantageous than to advantageous unfairness). This indifference curve pattern is typical of maximin concerns.[17] Around a fourth of subjects seem to exhibit this trait. The remaining subjects show other, rare patterns: self-interest in the favorable domain and inequity aversion in the unfavorable one, suggesting envy (e.g. the high-stakes line of 0902_4_9); upward trend in the entire domain, as consistent with spitefulness (0902_4_7), etc.

[16] Note that we do not observe any substantial correlation between inconsistencies at the level of iMPL tasks and apparently "surprising" shapes for indifference curves. For instance, subject #0902_4_19 exhibit perfect consistency at the level of iMPL tasks. When we focus on binary tasks, we observe very few inconsistencies. They occur in around a dozen of tasks where some subjects switch back and forth. These concern six subjects: in most cases, inconsistencies concern some isolated choice, e.g. AAABBBABB, and seem to be trembling errors. This is very little given that in total we have 1010 such tasks; thus only about 1% of choices are inconsistent, once again most of them being isolated cases.

[17] It can also be approximated by strongly curved Cox et al.'s 2007 specification, or by releasing constraints on the parameter in inequity aversion models, where unfavorable inequity aversion is assumed to be stronger than favorable one. Yet, we do not consider it as in the spirit of these models.

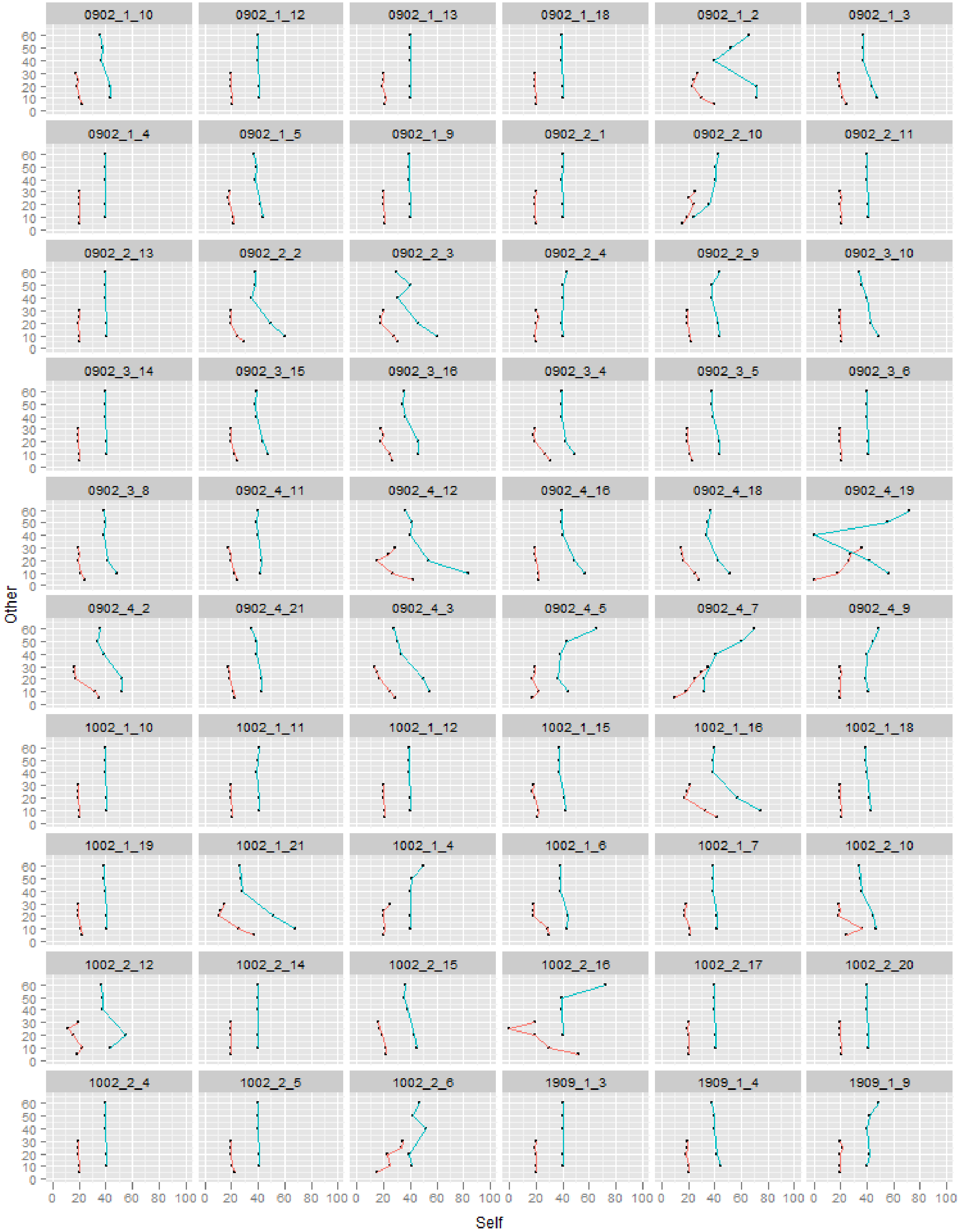


Figure 5: *Individual indifference curves*

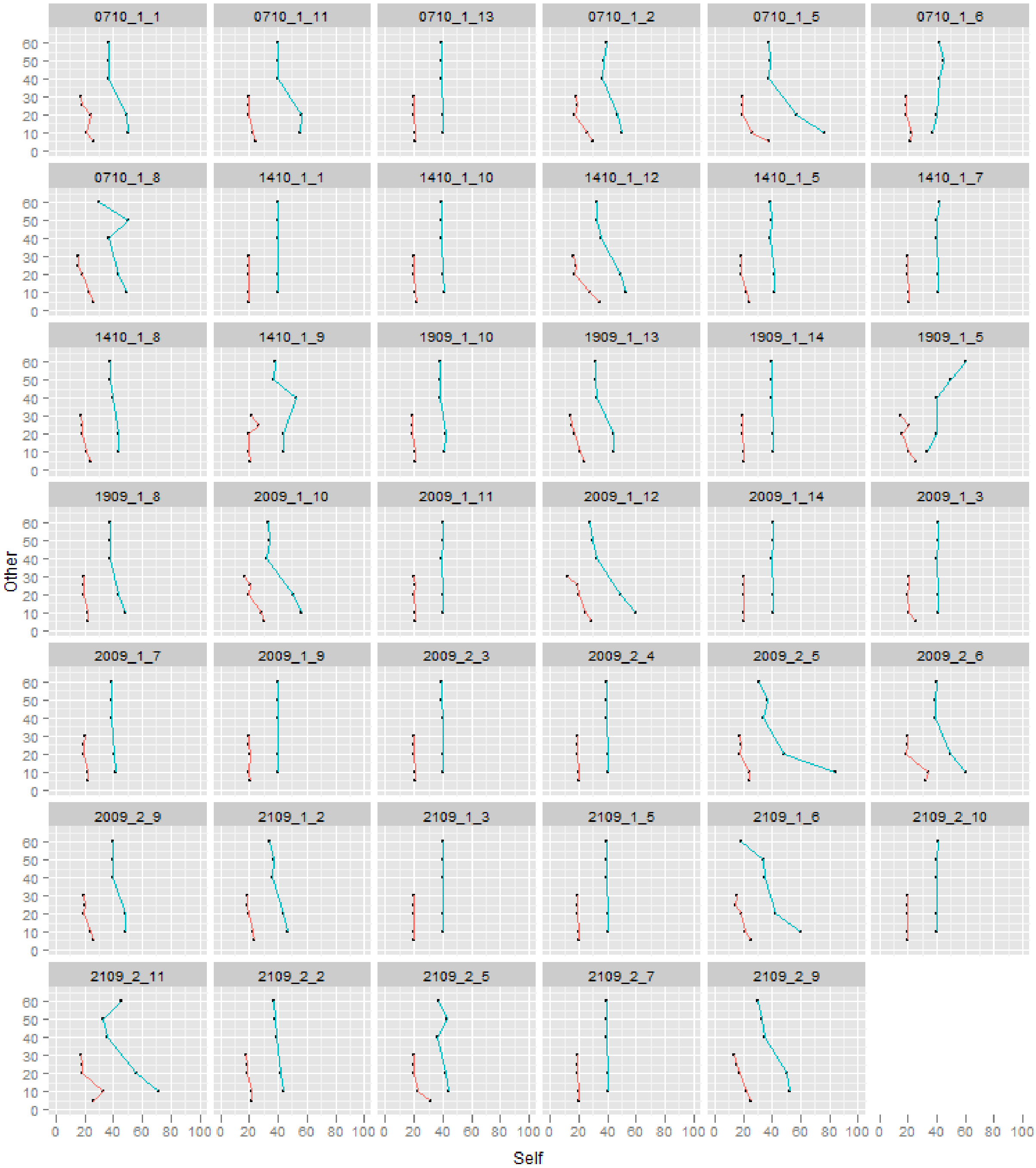


Figure 6: *Individual indifference curves (continued)*

To systematically categorize behavioral patterns we use the two estimation methods previously applied at the aggregate level to individual data. For each individual, we estimate maximum likelihood structural parameters (as in section 4.1.2) and slopes/piecewise linear parameters (as in section 4.1.1). Based on these two sets of estimates, we then classify individuals into categories of models (Table 1). An estimate was considered positive (resp. negative) if it is greater than or equal to .05 (or less than or equal to -.05). That corresponds, for the H indifference curve to a deviation of at least 1 PLN for each 20 PLN difference with the counterpart.[18] The results are displayed in Table 9.

	Self-interest	Maximin	Quasi-Maximin or Altruism	Inequity Aversion	Spitefulness	Envy	Unclassified
No. and fraction of subjects, based on PW Linear estimates	29 28.71%	29 28.71%	20 19.80%	12 11.88%	2 1.98%	9 8.91%	0 0%
No. and fraction of subjects, based on MLE (Logit or Probit)	21 20.79%	15 14.85%	27 26.73%	19 18.81%	7 6.93%	3 2.97%	9 8.91%

Table 9: *Individual classification: number and percentage*

The overall impression is that Charness and Rabin's (2002) model performs rather well at rationalizing individual choices: It can account for 62%-78% of individual (when including the self-interested individuals). In contrast with the aggregate analysis, the inequity aversion model seems to perform relatively poorly, as only 42%-50% of subjects adhere to it (including self-interested and envious individuals[19]). The general altruism/spitefulness model (whose Levine's and Cox et al.'s are instances) can account for 50%-55% of individual patterns (including self interested and spiteful types). If we only consider non-selfish individuals, the quasi-maximin model accounts for 52%-68% of them, while inequity aversion 29%-23% altruism/spitefulness 30%-42%. The conclusion is hence two-fold: first, it seems fair to conclude that quasi-maximin performs well (and in any case better than any alternative in the literature) in stark contrast with the results of the aggregate analysis. Second, the overall pattern is one of strong qualitative heterogeneity: all types are represented, and even quasi-maximin, the best performing model, accounts for only around two-thirds of all individuals.

The apparent discrepancy between the aggregate picture and the individual level stems from the fact that in addition to the qualitative heterogeneity, we observe that minority types, i.e., inequity averse, are more intensely sensitive to others'

[18] We considered also a threshold of .025 and even though it reshuffles the classification slightly, it does not fundamentally change the general picture: what it implies is that we observe less selfish types, with the efficiency and envy categories being the most affected.

[19] We do not include strictly maximin types in inequity aversion given that the fact that aversion to unfavorable inequity is assumed greater than aversion to favorable inequity is an essential feature of the model, allowing it to account for stylized experimental facts.

situations than are majority types, i.e., quasi-maximin individuals. Table 10 shows the average and median estimated parameters by category.[20] For the parameters that are directly comparable, inequity aversion is stronger: the β parameter can be twice as high there on average than in the quasi-maximin and maximin cases. Although the inequity averse individuals are much fewer than the broad category of maximin/quasi-maximin, the relative strength of this aversion can more than counter-balance, at the aggregate level, the mass of Charness and Rabin's types. In addition to that, spiteful and envious types, although not numerous, shape the aggregate pattern for the unfavorable domain in a way similar to that of inequity-averse subjects.

	Self-interest	Maximin	Quasi-Maximin or Altruism	Inequity Aversion	Spitefulness	Envy
PW linear estimates						
δ	-0.004 (0.000)	0.002 (0.000)	0.208 (0.131)	-0.560 (-0.206)	-0.275 (-0.275)	-0.276 (-0.225)
γ	0.028 (0.036)	0.263 (0.290)	0.390 (0.372)	0.471 (0.404)	-0.747 (-0.747)	0.017 (0.019)
n	29	29	20	12	2	9
MLE Logit estimates						
δ	0.001 (0.001)	-0.002 (-0.003)	0.250 (0.149)	-0.979 (-0.343)	-0.266 (-0.176)	-0.214 (-0.61)
γ	0.005 (0.003)	0.283 (0.146)	0.443 (0.401)	0.884 (0.654)	-0.527 (-0.451)	0.023 (0.032)
n	21	15	27	19	7	3
MLE Probit estimates						
δ	0.006 (0.003)	0.008 (0.003)	0.235 (0.143)	-0.774 (-0.320)	-0.537 (-0.420)	-0.434 (-0.445)
γ	0.016 (0.001)	0.255 (0.148)	0.545 (0.463)	0.859 (0.564)	- 0.319 (-0.221)	-0.020 (-0.022)
n	21	15	27	19	7	3

Table 10: *Average (median) estimated parameters by types*

These data can also be seen as additional evidence of a kink in indifference curves around equity. It is not only true that most individuals belong to a category implying a difference of sign between γ and δ, but even for the categories implying only a difference in magnitude (e.g., altruism), the difference between γ and δ seems to be substantial. That suggests indeed that individuals consider others' situations in different light in the favorable and unfavorable domains, supporting models that emphasize such a difference as well as the working hypothesis of Kerschbamer (2015).

To summarize, four interesting patterns emerge: first, inequity aversion models seem to perform better at the aggregate level; second, quasi-maximin (including strict maximin) seems to be the best model at accounting for individual patterns; third, the heterogeneity of motives and their intensity plays a very important role in shaping this discrepancy; fourth, most subjects seem to treat favorable and unfavorable domains quite differently. In a way, the three prominent types of

[20]In the case of the structural estimation, two outliers were removed for the computation of means due to the aberrant values obtained, with parameters being greater than 10, probably due to noise or inconsistent behavior, the corresponding error (μ) term being extremely high also.

models in social preference capture some aspect of our results: individuals are mostly Charness and Rabin types, a representative agent would look more like a Fehr and Schmidt or Bolton and Ockenfels one, whereas Levine's model based on the distribution of heterogeneous types capture an essential aspect of the data.

Two general conclusions are possible. Based on the current literature, it is perhaps illusory to pursue a single model of social preferences. *A contrario*, and this is a second possible interpretation, it is quite striking to see that very few individuals do not belong to one or the other of theoretical types, and the apparent qualitative heterogeneity could be accounted for in some future model based on some (perhaps even simple) combination of the various principles at work in the current models.

4.3 Socio-demographic analysis

Given that we have a fine measure of subjects' distributional preferences (not only slopes but specific parameters), it is of interest to study whether socio-demographic variables have an impact on both the type of distributional preferences and their intensity. We do so in two different yet complementary ways: first, we test how standard socio-demographic variables impact on the probability to be of a given type through a multinomial logistic regression, and second we test how these socio-demographic variables influence the intensity of these concerns. For both analyses, we rely on the piece-wise linear estimation, which presents less noise.

The estimated role of socio-demographic variables on distributional preferences are presented in Table 11. Male participants appear to be less socially oriented overall, and especially less prone to have anti-social concerns (inequity aversion, spitefulness and envy). This is somehow in line with the former studies (Almås, Cappelen, Sørensen, and Tungodden, 2010) showing that women tend to be more often inequity averse while men are more often motivated by efficiency. Moreover, we find that students in economics are, as is often observed (for instance, Frank, Gilovich, and Regan 1993), more often self-interested, and, possibly as a consequence, show less tendency to antisocial preferences. Inequity aversion seems also more prevalent in participant without an economic background (students of another major or non-students) in line with Fehr, Naef, and Schmidt's (2006) results.

The intensity of social concerns can be analyzed with an OLS regression of the individually estimated slopes in the favourable and unfavourable domains with the

	Maximin	Quasi-Maximin or Altruism	Inequity Aversion	Spitefulness	Envy
Gender (Male)	-0.84 (0.58)	0.67 (0.66)	-1.44* (0.78)	-18.60*** (<0.01)	-1.93** (0.93)
Major≠ Economics	1.30** (0.60)	-0.03 (0.64)	3.04***(1.17)	18.85*** (0.44)	0.13 (0.84)
Not a student	0.81 (0.94)	-0.93 (1.23)	3.32** (1.38)	1.84 (9.78)	-12.79*** (0.10)
	n = 101, reference type=Selfish, AIC:324.5				

*=significant at the .10 level, **= significant at the .05 level, ***= significant at the .01 level

Table 11: *Multinomial logisitic regression of types*

same demographic variables. The estimates are displayed in Table 12.[21]

Dep. variable	s_u	s_f
Gender (Male)	0.088	.103*
	(.062)	(.060)
Major ≠ Economics	-0.085	0.037
	(0.064)	(.063)
Not a student	-0.478***	0.207*
	(0.108)	(.106)
Constant	-0.016	0.111**
	(0.055)	(0.054)
$\bar{R}^2$	.147	.045
F(df)	6.739(3/97)	2.559 (3/97)
p	<0.001	0.059
n	101	101

*= significant at the .10 level, **= significant at the .05 level, ***= significant at the .01 level.

Table 12: *OLS estimates of slopes of indifference curves in the favourable and unfavourable domains*

Males tend to be more prosocial in the favorable domain, and less antisocial in the unfavourable one. The academic major does not seem to have a systematic impact, and non-students tend to show concern for equity: they are more prosocial in the favourable domain, and more antisocial in the unfavourable one.

Although some of these results should be taken with caution (due to the relatively small sample of purely antisocial types in the sample as well as non-students),

[21] We also ran the same estimations with additional socio-demographic variables such as income category (no effect), experience in the lab (number of experiments they take part in, which as no effect), age (with some effect, probably linked to the fact of not being a student) and the number of siblings (which increases the level of prosociality). The general results reported here are overall robust to these controls.

it still appears that economics student may be relatively different from the rest of the population, when compared to non-students in the analysis of intensity of preference and when compared to other students and non-students in the analysis of types. We also find that in contrast with the usual gender stereotypes, females are less prosocially oriented than males. That may, but only in part, be explained by more attention put towards equality as previously found in the experimental literature.

4.4 Hypothetical versus incentivized choice

A question of methodological interest is also to know how hypothetical decisions mimic (or depart from) properly incentivized ones. Indeed, in our experimental protocol (to prevent an effect of ex post role assignations), half the sample was assigned to the role of recipient and was merely hypothetically asked what they would do if they were in a position of deciding. While such non-incentivized decisions are usually not considered very highly in experimental economics (Hertwig and Ortmann, 2001), some authors have argued that incentives mostly reduce the variance, but may leave the central tendency unchanged (Camerer, Hogarth, Budescu, and Eckel, 1999), and even that not paying at all may yield more reliable results than paying little (Gneezy and Rustichini, 2000). To the best of our knowledge, there is very little systematic evidence on the type of hypothetical bias found in social preference elicitation tasks, even though there is compelling evidence that in contingent valuation (for public goods/service or environmental improvements), the bias is large (Murphy, Allen, Stevens, and Weatherhead, 2005). For plain distributional preference elicitation tasks in the lab, the folk wisdom among experimentalists (and economists in general) is that in an absence of any real cost, participants probably overstate their departure from self-interest. Given our data, we can contrast at a relatively precise level whether or not, this is the case. To do so we conduct the same analyses as for the incentivized subsample and compare the results.

Table 13 presents the same data for subjects in the hypothetical conditions as Table 5, as well as the difference with the incentivized conditions. What these raw figures suggest is that participants in the hypothetical condition tend to exhibit more prosocial concerns in the favourable domain (the average departure from the selfish benchmark is around twice as large in the hypothetical condition in the most unfavourable cases), and they tend to be less antisocial in the unfavourable domain. We observe in both the tasks and for all levels of payoffs that the variance is significantly larger for the hypothetical conditions, in line with previous research suggesting such an effect in other fields of experimental economics (Camerer, Hogarth, Budescu, and Eckel, 1999). A more carefull look at the results (see Appendix C) reveals a few features: first and overall, roughly the same proportions of individuals are in the same prosocial and antisocial categories, so that the qualitative categorization of prosocial/antisocial are not different between

	$x^L(5)$	$x^L(10)$	$x^L(20)$	$x^L(25)$	$x^L(30)$
Average	27.42	23.19	18.70	18.80	19.27
Difference	3.74	0.63	-0.53	-0.50	-0.23
Median	24.00	21.50	19.50	19.50	19.50
Difference	3.50	0	0	0	0

	$x^H(10)$	$x^H(20)$	$x^H(40)$	$x^H(50)$	$x^H(60)$
Average	52.65	45.23	38.06	38.62	37.96
Difference	5.73	1.69	-0.12	-0.53	-1.94
Median	47.00	43.00	39.50	39.50	39.50
Difference	4.50	1.50	0	0	0

Table 13: *Average and median x^L and x^H in the hypothetical condition and differences to the incentivized condition*

the hypothetical and invencentivized conditions. Second, however, conditional on the prosocial/antisocial types, there is a substantial difference in the intensity of the motivation, as highlighted in Fig. 7, Appendix C. In both the favourable and unfavourable domain, prosocial individuals are much more "generous", but in the unfavourable domain, antisocial individuals are also less antisocial. Overall, the hypothetical condition does not always push people to be further away from the selfish benchmark: antisocial participants in the unfavourable domain are actually closer to the selfish benchmark. A possible interpretation is that when individuals do not bear the costs of their decision, they tend to be attracted to a more prosocial choice: prosocial individuals exaggerate their (genuine) tendency to prosociality while antisocial individuals (in the unfavourable domain) attenuate their antisocial tendencies. This interpretation is roughly supported by the categorisation of individuals in the hypothetical conditions (see Table 15, Appendix C).

The socio-demographic analyses on hypothetical choices reveal roughly the same pattern as that in the incentivized condition: although we observe less effect of gender (no significant effect for any category), we still observe that students from an economic major are less likely to be inequity averse than those of a different economic major (see Tables 16 and 17, Appendix C). For the rest, the comparison are difficult to draw: more prosociality may mean that maximin turn into quasi-maximin for instance and spiteful and envious subsample also shrinks in the hypothetical condition.

Overall, we find that hypothetical answers in our experiment generate qualitatively similar results to that of the incentivized condition, but with a substantial shift towards a more prosocial (or less anti-social) behavior. This shift affects the categorization of individuals (for instance envious "become" inequity averse, or maximin "become" quasi-maximin) but leads to a distribution of type that is not qualitatively different from the one obtained in the incentivized condition (heterogeneity and roughly the same ordinal importance of types). The variance seems to increase in hypothetical tasks, as found in other area of experimental economics. But it seems fair to conclude that, although biased towards more prosocial choices, hypothetical data may still reveal some aspects of individuals' distributional preferences.

4.5 Stage 2 results on *ex ante* social concerns

As highlighted in the previous sections, we observe for most individuals a kink at the equal allocation. This kink means that, as argued in Appendix A, Proposition 1, that we can observe ex ante concerns through lotteries constructed on the basis of indifferent allocations as hypothesized in Sections 2 and 3. We start our analysis of behavior in the sure option vs. lottery (SL) tasks of Stage 2, in which ex ante concerns theoretically make subjects prefer the lottery, i.e. when P and Q are on different social domains (favourable or unfavourable). At the aggregate level, we observe a very balanced distribution of choices, with 43.56% of decision for option A (the sure option) and 45.87% for option B (the lottery), when aggregating the three choices of individuals. At this level, we do not observe much of a difference to the case where P and Q are on the same social domain (42.95 %). Yet, as stated in Proposition 1, Appendix A, individuals are expected to exhibit a preference for the lottery when two conditions are met: first, the two outcomes of the lottery are on different sides of the equal allocation diagonal; second, individual indifference curves are angled. To take this last condition into account, we median-split the sample of individuals on the basis of the difference in estimated parameters[22] in section 4.2. Restricting again the sample to the relevant lotteries, we obtain the following proportions: the sure option is chosen 37.8% of time, indifference 5.6%, and the lottery 56.7%. A pooled one-sample proportion test (restricted to non-indifference choices) gives that this distribution is weakly significantly different from the even split ($p = 0.083$, $n = 85$). Overall, in this particular task, we do observe at best a weak effect of *ex ante* comparison.

Focusing on the second set of tasks (lottery vs. lottery or LL tasks) may be a more relevant test since some inherent preference for or against uncertainty could taint choices in SL tasks, dwarfing any ex ante concerns. For these tasks, observe than when P, Q and R satisfy either condition 2. or 3. of Proposition 2 (Appendix A), we observe that 48.45% of decisions are in favour of the predicted lottery (and 11.60% for indifference and 39.95% for the other option). A pooled proportion test (restriced to non-indifference cases) gives a χ^2 statistic of 2.99 ($p = 0.08$). When restricting the sample to the half with the most pronunced $\alpha + beta$, we obtain $\chi^2 = 4.10$, $p = 0.04$. Once again, this can be taken, at best, as mild or partial evidence of the role of *ex ante* comparisons.

[22] That corresponds to the value $\alpha + \beta$ in Proposition 1, Appendix A.

To assess the strength of *ex ante* considerations as compared to *ex post* consideration, we refer to the simple linear utility model of eq. 2. We use the estimates of γ and δ to calculate α and β and then estimate the remaining ζ parameter by means of maximum likelihood estimation based on a logit specification, once again with fechnerian error. We get an estimate for ζ of around .635, with an individual clusterized standard-error of 0.117 ($Z = 3.11$, $p = 0.00018$, LL:-7010 and AIC:14023, $n = 1079$, 101 individual clusters).[23] In other words, *ex ante* social considerations seem to play a role in lottery choices, yet a much less important role than *ex post* concern. In sum, we find that a model mixing ex ante and ex post concerns seems to be relevant to account for the data (Krawczyk and Le Lec, 2010; Brock, Lange, and Ozbay, 2013), with an average weight for ex post concern of two-third and ex ante concerns of one third. This tends to be in line with previous findings that tend to show that ex ante concerns play a role, but a weaker one than outcome-based ones. The relative weakness of the ex ante concerns also cast doubts on some possibilities of accounting for social preferences on the basis of pure ex ante concerns (as suggested by Krawczyk and Le Lec 2016, Rohde and Rohde 2015 or Trautmann 2009 among others).

5 Conclusion

To summarize, one can point to several patterns. On the positive side, existing models, when taken together, do a rather good job at accounting for heterogeneous individuals' distributional patterns. On a more negative side, no model seems to be capable of explaining all or even a large majority of observations. Moreover, the heterogeneity of social concerns and their intensity lead to an apparently counter-intuitive finding that the aggregate preference corresponds to a mildly inequity averse type while the majority of subjects are characterized as quasi-maximin. This apparent discrepancy between the average and the median picture stems from the two findings: firstly, antisocial tendencies in the unfavourable domain, although infrequent, are often strong, and secondly composition effects play an important role in the sense that few envious and competitive types shift the aggregate tendency of indifference to the other's payoffs in the unfavourable domain towards a mild antisocial tendency. In sum, the most empirically convergent model depends on whether the focus is on the "average preferences" or the median ones.

This mixed picture probably explains the contradictory evidence one can find in the literature. For instance Engelman and Strobel (2004) find that in binary allocation tasks, inequity aversion models fail to describe choices for the majority of individuals, but the models still perform well to describe aggregate patterns

[23]We obtain similar results with a probit specification, with an estimated γ of 0.590, extremely significant $p < .001$. The results when taking the values of γ and δ from the structural estimation are similar: in between 0.44 and 0.71 with resp. probit and logit specifications, once removing the aberrant estimated parameters, as above.

such as experimental game behaviors, or average behavior in many allocation tasks (such as the DG). This may also explain the relative lack of consensus in the experimental literature about the most relevant models of distributional preferences. Even though there seems to be an emerging consensus on the superiority of the quasi-maximin model in the sense that it predicts most subjects' behavior (Sutter, Feri, Kocher, Martinsson, Nordblom, and Rützler, 2010), yet aggregation effects may in some circumstances render the inequity aversion relevant.

We also find that there seems to exist different social concerns in the favourable and unfavourable domains, and the specific role assigned to equal allocations find support in most of our individuals' measured preferences. In addition to the observed linearity of individuals' concerns, this seems to favour the simple and tractable linear models rather than concave specifications such as the constant elasticity models (Andreoni and Miller, 2002; Cox, Friedman, and Gjerstad, 2007).

Using the indifferent points elicited in the first stage of the experiment, we were also in capacity to test the strength of ex ante concerns (distributional concerns that take into account the risk distribution). Any departure from indifference between lotteries whose support is only composed of indifferent final allocations is a departure from purely ex post concerns, and may be used to test the strength of both considerations. We find that although not absent, ex ante concerns have a rather minor role, which is roughly half that of ex post considerations.

On a more methodological note, our method of eliciting distributional concerns could be expanded in several directions. Eliciting indifference curves with three individuals would be of obvious interest to distinguish efficiency concerns and pure altruism. Doing the same in response to another individual's choice may also be a relevant way to investigate reciprocity-based preferences. And finally, studying more indifference curves on perhaps a richer domain, with varying subject pools would allow to have a finer understanding of the effect of stakes, the generalizability of our findings and the demographics underlying distributional preferences.

References

Almås, I., A. W. Cappelen, E. Ø. Sørensen, and B. Tungodden (2010): "Fairness and the development of inequality acceptance," *Science*, 328(5982), 1176–1178.

Andersen, S., G. Harrsion, M. Lau, and E. Rutstrom (2006): "Elicitation using multiple price list formats," *Experimental Economics*, 9, 383–405.

——— (2007): "Valuation using multiple price list formats," *Applied Economics*, 39, 675–682.

——— (2008): "Eliciting Risk and Time Preferences," *Econometrica*, 76, 583–618.

Andreoni, J. (1988): "Why free ride?: strategies and learning in public goods experiments," *Journal of Public Economics*, 37(3), 291–304.

Andreoni, J., and B. D. Bernheim (2009): "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects," *Econometrica*, 77(5), 1607–1636.

Andreoni, J., and J. Miller (2002): "Giving according to GARP: an experimental test of the consistency of preferences for altruism," *Econometrica*, 70, 737754.

Aristotle, u. (4th c. BC/2004): *Rhetoric*. Kessinger Publishing.

Bardsley, N. (2008a): "Altruism or artifact? A note on dictator game giving," *Experimental Economics*, 11(2), 122–133.

——— (2008b): "Dictator game giving: altruism or artefact?," *Experimental Economics*, 11(2), 122–133.

Binmore, K., and A. Shaked (2010): "Experimental economics: Where next?," *Journal of Economic Behavior & Organization*, 73(1), 87–100.

Bolton, G. E., J. Brandts, and A. Ockenfels (2005): "Fair procedures: evidence from games involving lotteries," *Economic Journal*, 115, 1054–1076.

Bolton, G. E., E. Katok, and R. Zwick (1998): "Dictator game giving: Rules of fairness versus acts of kindness," *International Journal of Game Theory*, 27(2), 269–299.

Bolton, G. E., and A. Ockenfels (2000): "ERC: a theory of equity, reciprocity and competition," *American Economic Review*, 90, 166–193.

Brennan, G., L. G. González, W. Güth, and M. V. Levati (2008): "Attitudes toward private and collective risk in individual and strategic choice situations," *Journal of Economic Behavior & Organization*, 67(1), 253–262.

Brock, J. M., A. Lange, and E. Y. Ozbay (2013): "Dictating the risk: Experimental evidence on giving in risky environments," *The American Economic Review*, 103(1), 415–437.

Camerer, C. F., R. M. Hogarth, D. V. Budescu, and C. Eckel (1999): "The effects of financial incentives in experiments: A review and capital-labor-production framework," in *Elicitation of Preferences*, ed. by B. Fischhoff, and C. F. Manski, pp. 7–48. Springer.

Charness, G., and M. Rabin (2002): "Understanding social preferences with simple tests," *Quartely Journal of Economics*, 117, 817869.

Cooper, D., and J. H. Kagel (2009): "Other regarding preferences: a selective survey of experimental results," *Handbook of experimental economics*, 2.

Cox, J. C., D. Friedman, and S. Gjerstad (2007): "A tractable model of reciprocity and fairness," *Games and Economic Behavior*, 59(1), 17–45.

Cox, J. C., and V. Sadiraj (2010): "A tractable model of reciprocity and fairness," *Economic Inquiry*, 50(4), 920–931.

Decancq, K., and A. Nys (2017): "Non-parametric well-being comparisons," .

Engel, C. (2011): "Dictator games: A meta study," *Experimental Economics*, 14(4), 583–610.

Engelmann, D., and M. Strobel (2004): "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments," *American Economic Review*, 94, 857–869.

Falk, A., and U. Fischbacher (2006): "A theory of reciprocity," *Games and Economic Behavior*, 54, 293–315.

Fehr, E., M. Naef, and K. M. Schmidt (2006): "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: comment.," *American Economic Review*, 96, 1912–1917.

Fehr, E., and K. M. Schmidt (1999): "A theory of fairness, competition, and cooperation," *Quarterly Journal of Economics*, 114, 817–868.

Fisman, R., P. Jakiela, and S. Kariv (2014): "The distributional preferences of Americans," Discussion paper, National Bureau of Economic Research.

Fisman, R., S. Kariv, and D. Markovits (2007): "Individual preferences for giving," *The American Economic Review*, 97(5), 1858–1876.

Frank, R. H., T. Gilovich, and D. T. Regan (1993): "Does studying economics inhibit cooperation?," *The Journal of Economic Perspectives*, 7(2), 159–171.

Gneezy, U., and A. Rustichini (2000): "Pay enough or don't pay at all," *The Quarterly Journal of Economics*, 115(3), 791–810.

Greiner, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1(1), 114–125.

Güth, W., S. Huck, and W. Müller (2001): "The relevance of equal splits in ultimatum games," *Games and Economic Behavior*, 37(1), 161–169.

Hertwig, R., and A. Ortmann (2001): "Experimental practices in economics: A methodological challenge for psychologists?," *Behavioral and Brain Sciences*, 24(3), 383–403.

Hey, J. D., and E. Strazzera (1989): "Estimation of indifference curves in the Marschak-Machina triangle a direct test of the fanning out hypothesis," *Journal of Behavioral Decision Making*, 2(4), 239–260.

Holt, C. A., and S. K. Laury (2002): "Risk aversion and incentive effects," *American economic review*, 92(5), 1644–1655.

Kerschbamer, R. (2015): "The geometry of distributional preferences and a non-parametric identification approach: The Equality Equivalence Test," *European economic review*, 76, 85–103.

Krawczyk, M., and F. Le Lec (2010): "Give me a chance!An experiment in social decision under risk," *Experimental Economics*, 13(4), 500–511.

——— (2016): "Dictating the Risk: Experimental Evidence on Giving in Risky Environments: Comment," *American Economic Review*, 106(3), 836–39.

Levine, D. K. (1998): "Modeling altruism and spitefulness in experiments," *Review of Economic Dynamics*, 1(3), 593–622.

List, J. A. (2007): "On the interpretation of giving in Dictator Games," *Journal of Political Economy*, 115(3).

Loomes, G., and R. Sugden (1995): "Incorporating a stochastic element into decision theories," *European Economic Review*, 39(3), 641–648.

MacCrimmon, K. R., and M. Toda (1969): "The experimental determination of indifference curves," *The Review of Economic Studies*, 36(4), 433–451.

Moffatt, P. G., and S. A. Peters (2001): "Testing for the presence of a tremble in economic experiments," *Experimental Economics*, 4(3), 221–228.

Murphy, J. J., P. G. Allen, T. H. Stevens, and D. Weatherhead (2005): "A meta-analysis of hypothetical bias in stated preference valuation," *Environmental and Resource Economics*, 30(3), 313–325.

Rohde, I. M., and K. I. Rohde (2011): "Risk attitudes in a social context," *Journal of Risk and Uncertainty*, 43(3), 205–225.

——— (2015): "Managing social risks–tradeoffs between risks and inequalities," *Journal of Risk and Uncertainty*, 51(2), 103–124.

Saito, K. (2013): "Social preferences under risk: equality of opportunity versus equality of outcome," *The American Economic Review*, 103(7), 3084–3101.

Sutter, M., F. Feri, M. G. Kocher, P. Martinsson, K. Nordblom, and D. Rützler (2010): "Social preferences in childhood and adolescence-A large-scale experiment," .

Thurstone, L. L. (1931): "The indifference function," *The Journal of Social Psychology*, 2(2), 139–167.

Trautmann, S. T. (2009): "A tractable model of process fairness under risk," *Journal of Economic Psychology*, 30(5), 803–813.

Ubel, P., J. Baron, and D. Asch (2001): "Preference for equity as a framing effect," *Medical Decision Making*, 21(3), 180–189.

van de Kuilen, G., and P. Wakker (2009): "The midweight method to measure attitudes towards risk and ambiguity," *Management Science*, pp. 1–32.

Appendix A: formal framework for ex ante and ex post concerns

Generic Consequences of Ex Ante Social Concerns

Consider an agent whose social preferences are based on a mix of outcome-based and ex ante comparisons, as suggested by empirical evidence (Krawczyk and Le Lec, 2010; Brock, Lange, and Ozbay, 2013). To keep it tractable, we work with a linear model (Saito, 2013):

$$\begin{aligned} EU = E(X) - \zeta[\alpha E((max(Y-X,0)) + \beta E((max(X-Y,0))] \\ - (1-\zeta)[\alpha(max(EY-EX,0)) + \beta(max(EX-EY,0))]. \end{aligned} \quad (2)$$

$\zeta \in [0,1]$ is the weight attached to ex post comparisons: when it is equal to 1, we have outcome-based models: selfish when $\alpha = \beta = 0$, Fehr-Schmidt when $\alpha > \beta > 0$, maximin when $\alpha = 0, \beta > 0$, envy when $\alpha > 0, \beta = 0$ etc. By contrast, $\zeta = 0, \alpha > \beta > 0$ corresponds to Trautmann (2009). When $0 < \zeta < 1$, a mix of ex ante and ex post concerns results. Of course, an unconstrained, four-paramater model could be considered, but it seems plausible that someone showing inequity aversion in actual payoffs will display similar inequity aversion in expected payoffs (provided she cares about ex ante comparisons at all), etc.

We now show the conditions for which mixing among equally attractive allocations is desirable.

Proposition 1 *Consider three allocations P, Q, R that are perceived as equally good and a lottery L that gives P with probability $p \in (0,1)$ and Q otherwise.*

(i) If $x_P \leq y_P$ and $x_Q \leq y_Q$, then L and R are equally good

(ii) If $x_P \geq y_P$ and $x_Q \geq y_Q$, then L and R are equally good

(iii) If $x_P > y_P$ and $x_Q < y_Q$ or vice versa, then L is better than R if and only if $\zeta < 1$ and $-\alpha < \beta$

Points (i) and (ii) mean that mixing makes no difference if one of the parties is eventually better off no matter what, whereas (iii) specifies when mixing is desirable.

Proof. Points (i) and (ii) are straightforward. Intuitively, the two allocations Q and P are on the same side of the diagonal where the indifference curve is linear. For (iii), we will assume that $x_P > y_P$ and $x_Q < y_Q$, the other situation being fully analogous. We consider two cases:

1. Advantageous expected inequality, $px_P + (1-p)x_Q > py_Q + (1-p)y_Q$. The utility of the lottery is given by:

$$\begin{aligned} EU(L) &= p[x_P - \zeta\beta(x_P - y_P)] + (1-p)[x_Q - \zeta\alpha(y_Q - x_Q)] \\ &\quad -(1-\zeta)\beta[p(x_P - y_P) + (1-p)(x_Q - y_Q)] \end{aligned} \tag{3}$$

The utility of R is equal to the utility of P or Q, so can be written as follows:

$$\begin{aligned} EU(R) &= pEU(P) + (1-p)EU(Q) \\ &= p[x_P - \zeta\beta(x_P - y_P) - (1-\zeta)\beta(x_P - y_P)] \\ &\quad + (1-p)[x_Q - \zeta\alpha(y_Q - x_Q) - (1-\zeta)\alpha(y_Q - x_Q)] \end{aligned} \tag{4}$$

The difference between the two then simplifies to

$$EU(L) - EU(R) = (1-p)(1-\zeta)[\beta(y_Q - x_Q) + \alpha(y_Q - x_Q)] \tag{5}$$

We conclude that L is indeed preferred to R if an only if $\zeta < 1$ and $-\alpha < \beta$

2. Disadvantageous expected inequality, $px_P + (1-p)x_Q < py_Q + (1-p)y_Q$. The utility of the lottery is given by:

$$\begin{aligned} EU(L) &= p[x_P - \zeta\beta(x_P - y_P)] + (1-p)[x_Q - \zeta\alpha(y_Q - x_Q)] \\ &\quad -(1-\zeta)\alpha[p(y_P - x_P) + (1-p)(y_Q - x_Q)] \end{aligned} \tag{6}$$

The difference in utility becomes:

$$EU(L) - EU(R) = p(1-\zeta)[\alpha(x_P - y_P) + \beta(x_P - y_P)] \tag{7}$$

Again, this difference is positive if an only if $\zeta < 1$ and $\alpha > -\beta$, QED

Note that $\alpha > -\beta$ means that increasing other's payoff is at least as desirable when the other is behind as when the other is ahead. This is a highly plausible assumption and it is both predicted by most models and confirmed empirically, also in our sample, as we shall see. Thus, if mixing helps equalize expected payoffs and if the individual cares about comparison of expected payoff at all, then mixing is desirable, except for some very peculiar preferences. We now turn to comparing two different non-degenerate lotteries.

Proposition 2 *For any two lotteries L and L' whose supports are on the same indifference curve, with $L = pP + (1-p)Q$ and $L' = pP + (1-p)R$, we have*

1. *If $sign(x_P - y_P) = sign(x_Q - y_Q) = sign(x_R - y_R)$, then L and L' are equally good.*

2. *If* $sign(x_P - y_P) = sign(x_Q - y_Q) \neq sign(x_R - y_R)$*, then* L' *is preferred to* L *if and only if* $\zeta < 1$ *and* $-\alpha < \beta$*.*

3. *If* $sign(x_P - y_P) \neq sign(x_Q - y_Q) = sign(x_R - y_R)$ *and* $|x_R - y_R| > |x_Q - y_Q|$*, then* L' *is preferred to* L *if and only if* $\zeta < 1$*,* $-\alpha < \beta$*.*

If all the indifferent allocations are on the same domain (advantageous or disadvantageous) as in (1), then mixing does not bring any difference (because of the linearity of the basic model). Case (2) means that it is better to mix a given allocation with an allocation that is in the other domain rather than on the same domain. And for case (3), if the common allocation is mixed with two other allocations on the opposite domain, then it is better to mix with the one that is the further away from equality.

Proof We will sketch the proof for the case of advantageous expected inequality in both lotteries, i.e. $px_P + (1-p)x_Q > py_P + (1-p)y_Q$ and $px_P + (1-p)x_R > py_P + (1-p)y_R$, the remaining three being analogous. The (expected) utility of P is given by:

$$\begin{aligned} EU(P) &= x_P - \zeta[\alpha \max(y_p - x_p, 0) + \beta \max(x_P - y_P, 0)] \\ &\quad -(1-\zeta)[\alpha \max(y_p - x_p, 0) + \beta \max(x_P - y_P, 0)] \end{aligned}$$

Setting $A(x_P, y_P) = \alpha \max(y_P - x_P, 0) + \beta \max(x_P - y_P, 0)$:

$$EU(P) = x_P - \zeta A(x_P, y_P) - (1-\zeta)A(x_P, y_P)$$

Written differently, the ex post concern in the utility of P is given by:

$$x_P - \zeta A(x_P, y_P) = EU(P) + (1-\zeta)A(x_P, y_P)$$

The same straightforwardly holds for Q:

$$x_Q - \zeta A(x_Q, y_Q) = EU(Q) + (1-\zeta)A(x_Q, y_Q)$$

Now we can decompose $EU(L)$ as its expected ex post part (first line) and its ex ante component (second line):

$$\begin{aligned} EU(L) &= p\big(EU(P) + (1-\zeta)A(x_P, y_P)\big) + (1-p)\big(EU(Q) + (1-\zeta)A(x_Q, y_Q)\big) \\ &\quad -(1-\zeta)\beta(px_P + (1-p)x_Q - py_P - (1-p)y_Q) \end{aligned}$$

The ex ante component takes this value because of the assumption that $px_P + (1-p)x_Q > py_P + (1-p)y_Q$.

Likewise, since by assumption $px_P + (1-p)x_R > py_P + (1-p)y_R$:

$$\begin{aligned} EU(L') &= p\big(EU(P) + (1-\zeta)A(x_P, y_P)\big) + (1-p)\big(EU(R) + (1-\zeta)A(x_R, y_R)\big) \\ &\quad -(1-\zeta)\beta(px_P + (1-p)x_R - py_P - (1-p)y_R) \end{aligned}$$

We hence have:

$$\begin{aligned}EU(L)-EU(L') &= p\big(EU(P)+(1-\zeta)A(x_P,y_P)\big)+(1-p)\big(EU(Q)+(1-\zeta)A(x_Q,y_Q)\big)\\ &\quad -p\big(EU(P)+(1-\zeta)A(x_P,y_P)\big)-(1-p)\big(EU(R)+(1-\zeta)A(x_R,y_R)\big)\\ &\quad -(1-\zeta)\beta(px_P+(1-p)x_Q-py_P-(1-p)y_Q)\\ &\quad +(1-\zeta)\beta(px_P+(1-p)x_R-py_P-(1-p)y_R)\end{aligned}$$

Since $EU(P)=EU(Q)=EU(R)$ by assumption, we have:

$$\begin{aligned}EU(L)-EU(L') &= p(1-\zeta)A(x_P,y_P)+(1-p)(1-\zeta)A(x_Q,y_Q)\\ &\quad -p(1-\zeta)A(x_P,y_P)-(1-p)(1-\zeta)A(x_R,y_R)\\ &\quad -(1-\zeta)p\beta(x_P-y_P)-(1-\zeta)(1-p)\beta(x_Q-y_Q)\\ &\quad +(1-\zeta)p\beta(x_P-y_P)+(1-\zeta)(1-p)\beta(x_R-y_R)\end{aligned}$$

Rearranging:

$$\begin{aligned}EU(L)-EU(L') &= p(1-\zeta)[A(x_P,y_P)-A(x_P,y_P)-\beta(x_P-y_P)+\beta(x_P-y_P)]\\ &\quad +(1-p)(1-\zeta)[A(x_Q,y_Q)-A(x_R,y_R)-\beta(x_Q-y_Q)+\beta(x_R-y_R)]\end{aligned}$$

In the end, we have:

$$\begin{aligned}EU(L)-EU(L') &= (1-p)(1-\zeta)\\ &\quad \times[A(x_Q,y_Q)-A(x_R,y_R)-\beta(x_Q-y_Q)+\beta(x_R-y_R)] \qquad (8)\end{aligned}$$

Suppose as in case 1. that $x_Q > y_Q$ and $x_R > y_R$[24], then $A(x_Q,y_Q)=\beta(x_Q-y_Q)$ and $A(x_R,y_R)=\beta(x_R-y_R)$. Then $EU(L)-EU(L')=0$ and the two lotteries are equally good.

Now suppose as in 2. that $x_P > y_P$, $x_Q > y_Q$ and $x_R < y_R$,[25] then we have $A(x_Q,y_Q)=\beta(x_Q-y_Q)$ and $A(x_R,y_R)=-\alpha(x_R-y_R)$. After simplification, we have:

$$EU(L)-EU(L')=(1-p)(1-\zeta)(-\alpha-\beta)[y_R-x_R]$$

And $EU(L') > EU(L)$ iff $-\alpha < \beta$ and $\zeta < 1$.

Now suppose as in 3. that $x_P > y_P$, $x_Q < y_Q$ and $x_R < y_R$,[26] then:

$$EU(L)-EU(L')=(1-p)(1-\zeta)[\alpha(y_Q-x_Q)-\alpha(y_R-x_R)-\beta(x_Q-y_Q)+\beta(x_R-y_R)]$$

I.e.

$$EU(L)-EU(L')=(1-p)(1-\zeta)(\alpha+\beta)[(y_Q-x_Q)-(y_R-x_R)]$$

Hence $|x_R-y_R| > |x_Q-y_Q|$ is here equivalent to $y_R-x_R > y_Q-x_Q$, so that

[24] Given the assumption of advantageous expected inequality in both lotteries, i.e. $px_P+(1-p)x_Q > py_P+(1-p)y_Q$ and $px_P+(1-p)x_R > py_P+(1-p)y_R$, the three allocations can only be in the advantageous domain if the three are in the same domain.

[25] The reverse case where $x_P < y_P$, $x_Q < y_Q$ and $x_R > y_R$ follows the same treatment.

[26] Once again the treatment of the case $x_P < y_P$, $x_Q > y_Q$ and $x_R > y_R$ is similar.

$EU(L') > EU(L)$ iff $-\alpha > \beta$ and $\zeta < 1$.

The other cases (when L and L' are not both in the favourable expected inequality domains) follow a similar logic. To make this explicit, the expressions of $EU(L) - EU(L')$ are provided in Table 14, they are obtained following the same calculation steps as above.

Exp. inequality for L	Exp. inequality for L'	$[EU(L) - EU(L')]/(1-\zeta)$
$px_P + (1-p)x_Q > py_P + (1-p)y_Q$	$px_P + (1-p)x_R > py_P + (1-p)y_R$	$(1-p)[A(x_Q, y_Q) - A(x_R, y_R) - \beta(x_Q - y_Q) + \beta(x_R - y_R)]$
$px_P + (1-p)x_Q < py_P + (1-p)y_Q$	$px_P + (1-p)x_R < py_P + (1-p)y_R$	$(1-p)[A(x_Q, y_Q) - A(x_R, y_R) - \alpha(y_Q - x_Q) + \alpha(y_R - x_R)]$
$px_P + (1-p)x_Q > py_P + (1-p)y_Q$	$px_P + (1-p)x_R < py_P + (1-p)y_R$	$(1-p)[A(x_Q, y_Q) - A(x_R, y_R) - \beta(x_Q - y_Q) + \alpha(y_R - x_R)] + p(\beta+\alpha)(y_P - x_P)$
$px_P + (1-p)x_Q < py_P + (1-p)y_Q$	$px_P + (1-p)x_R > py_P + (1-p)y_R$	$(1-p)[A(x_Q, y_Q) - A(x_R, y_R) - \alpha(y_Q - x_Q) + \beta(x_R - y_R)] + p(\beta+\alpha)(x_P - y_P)$

Table 14: *Expressions of* $EU(L) - EU(L')$

While the case of the second row follows the exact same logic as the case studied above, it is noteworthy to detail a bit the logic of the third and fourth row cases, which differs slightly. To do so, we consider the case when L is in the expected favourable inequity domain and L' in the expected unfavourable one (third row). Either P or R is in the unfavourable domain and either P or Q is in the favourable one, so case 1. of Proposition 2 cannot apply. Suppose now that $x_P > y_P$ and $x_Q > y_Q$ and $x_R < y_R$ (as in 2. of Proposition 2), we then have after replacement and simplification:

$$[EU(L) - EU(L')]/(1-\zeta) \quad = \quad p(\beta+\alpha)(y_P - x_P) < 0$$

And then the result 2. of Proposition 2 holds. Note that we obtain a symmetrical result in the mirror situation (also compatible with case 2.) when $x_P < y_P$ and $x_Q > y_Q$ and $x_R < y_R$:

$$[EU(L) - EU(L')]/(1-\zeta) \quad = \quad p(\beta+\alpha)(y_P - x_P) > 0$$

And L is preferred to L' (symmetric case).

Now consider the last possible case (situation 3. in the proof) where Q and R are in the same domain, for instance $x_P > y_P$ but $x_Q < y_Q$ and $x_R < y_R$.[27] We obtain:

$$\begin{aligned} [EU(L) - EU(L')]/(1-\zeta) &= (1-p)[\alpha(y_Q - x_Q) - \beta(x_Q - y_Q)] + p(\beta+\alpha)(y_P - x_P) \\ &= (\alpha+\beta)[(1-p)(y_Q - x_Q) + p(y_P - x_P)] \end{aligned}$$

By assumption $px_P + (1-p)x_Q > py_P + (1-p)y_Q$, hence $py_P - px_P + (1-p)y_Q - (1-p)x_Q < 0$ and $EU(L) < EU(L')$ iff $\alpha > -\beta$.

All other cases follow one or the other types of calculation and argumentation provided, QED.

[27] Again the symmetrical case where $x_P < y_P$ but $x_Q > y_Q$ and $x_R > y_R$ is treated in a similar way

Note that the requirement found in both propositions (namely that one allocation in the support of a lottery involves advantageous inequality and the other – disadvantageous inequality) for the lottery to be desirable results from (piecewise) linearity of the model. Under globally convex indifference curves mixing is typically expected to be attractive. In any case, under rather weak and quite plausible assumptions about social motives, availability of indifference curves for an individual allows testing the strength of *ex ante* social concerns, which is what we do in Stage 2.

Appendix B: Structural estimation

The parameters of the model displayed in Eq () based on all binary choices, and assuming a similar set of parameters for all individuals, the error term covering individual differences. We assume errors are of the Fechner type, meaning that noise applies to the level of utility. In our context, a Fechner error specification seems indeed more appropriate than pure tremble, or "unit error", (Moffatt and Peters, 2001): at the individual level, iMPL elicitation procedure promotes consistency of choice, more so than unstructured binary tasks) while at the aggregate level, equating individual differences solely with trembling errors of individuals assumed to have the same structural preferences is unlikely to capture heterogeneity. An alternative option would be to use a Random Preference (Loomes and Sugden, 1995) specification. In our case, that would mean that every structural parameter (δ and γ) is subject to noise (rather than the whole utility). Two reasons made us prefer the Fechner error in our setting: first, the differences are less prominent in our setting than in many others given the linearity of the model;[28] second, the parameters of the model are not theoretically independent so that the support of the pair of parameters is not given by the product of the support of each parameter, and specifying a cdf over such structural noise is problematic.

We estimate binary choices with logit or probit specification, the corresponding latent index being given by $I(i, \rho) = \frac{u_\rho(x_i^B, y_i^B) - u_\gamma(x_i^A, y_i^A)}{\mu}$, with $\rho = (\gamma, \delta, \mu)$, with i being a binary task as in Table 2 with (x_i^A, y_i^A) and (x_i^B, y_i^B) resp. the A and B allocations of task i.[29] The log-likelihood is given by:

$$ln\mathcal{L}(\rho) = \sum_{i=1}^{n} \Big[c_i.lnG\big(I(X_i, \rho)\big) + (1 - c_i).ln\Big(1 - G\big(I(X_i, \rho)\big)\Big)\Big]$$

with X_i denoting the vector of payoffs, c_i the observed choice (taking values 0 for the A choice, and 1 for the B choice, but also $\frac{1}{2}$ for indifference[30]), G either the standard normal cdf for the Probit model and the logistic cdf for the Logit one, and finally n the number of observed decisions.

[28] A random preference specification would mostly imply heteroskedasticity in x and y.

[29] Another option would be to use ratios rather than differences, for instance as in (Holt and Laury, 2002), this raises in our case the technical issue of negative values for utility, which are consubstantial to (some) of the social preference models).

[30] In the case the subject chose "Indifference" we assume a probability one half for each choice.

Appendix C: Analysis of hypothetical decisions

	Self-interest	Maximin	Quasi-Maximin or Altruism	Inequity Aversion	Spitefulness	Envy	Unclassified
No. and fraction of subjects, based on PW Linear estimates	21 20.79%	19 18.81%	32 31.68%	20 19.80%	0 0%	6 5.94%	0 0%
No. and fraction of subjects, based on MLE (Logit or Probit)	18 17.82%	6 5.94%	30 29.70%	33 32.67%	2 1.98%	3 2.97%	8 7.92%

Table 15: *Individual classification under hypothetical condition: number and percentage*

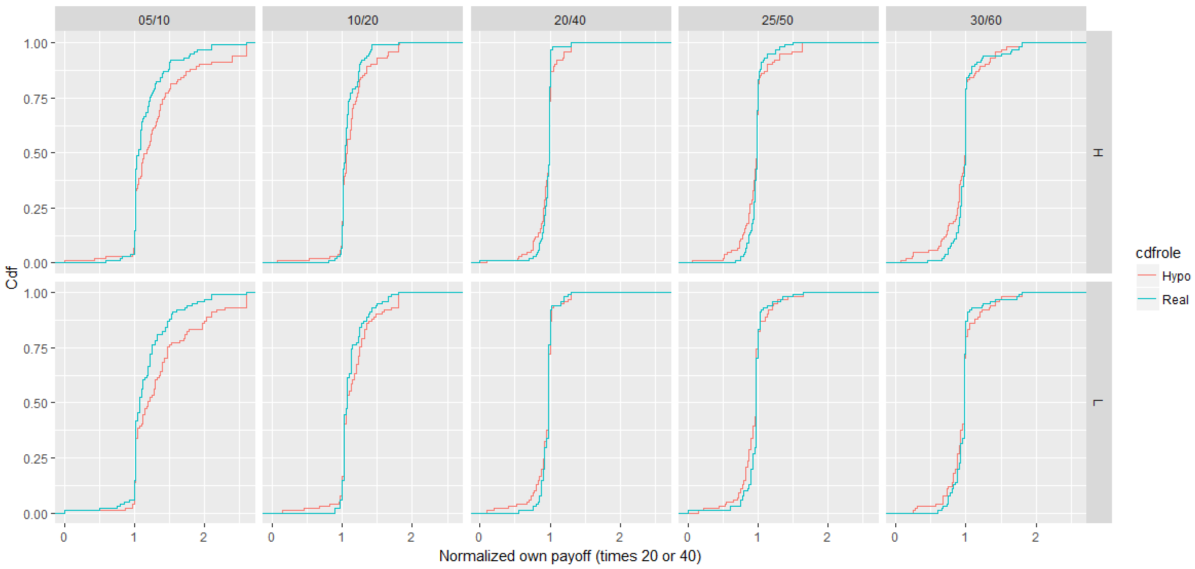


Figure 7: *Incentivized versus hypothetical indifference points' cumulative distribution*

	Maximin	Quasi-Maximin or Altruism	Inequity Aversion	Spitefulness	Envy
Gender (Male)	0.44 (0.66)	0.57 (0.62)	-0.89 (0.72)	0.65 (1.53)	0.02 (0.99)
Major≠ Economics	0.01 (0.78)	1.67** (0.67)	1.66** (0.73)	35.42*** (0.56)	1.17 (1.15)
Not a student	30.81*** (0.85)	32.67*** (0.59)	32.33*** (0.74)	0.89 (X)	33.45*** (0.83)
	n = 101, reference type=Selfish, AIC:340.0				

*=significant at the .10 level, **= significant at the .05 level, ***= significant at the .01 level

Table 16: *Hypothetical condition: Multinomial logisitic regression of types*

Dep. variable	s_u	s_f
Gender (Male)	0.120*	.001
	(.064)	(.069)
Major≠ Economics	-0.006	0.19**
	(0.067)	(.074)
Not a student	-0.161	0.09
	(0.107)	(.117)
Constant	-0.70	0.209**
	(.059)	(0.064)
$\bar{R}^2$	.147	.040
F(df)	1.965(3/97)	2.40 (3/97)
p	0.124	0.073
n	101	101

*= significant at the .10 level, **= significant at the .05 level, ***= significant at the .01 level.

Table 17: *Hypothetical condition: OLS estimates of slopes of indifference curves in the favourable and unfavourable domains*

Appendix D: Experimental Instructions

Welcome

Please switch off your mobile phone and remain quiet during the entire experiment. You are not allowed to contact other participants in any way. Participants who do not adhere to these rules may be excluded and receive no remuneration. If you have a question, raise your hand and wait for the experimenter.

You have already earned 5 zloty for being here on time. This amount will be added to whatever you make during the experiment. These earnings may depend on your decisions, decisions of other participants and luck. [...] Your total earnings will be paid out in cash, immediately after the end of the experiment.

The experiment is anonymous. It means that neither the experimenter nor other participants will be able to link your decision and your identity (and other participants will not even get to know what your earnings are).

Participants will be matched in pairs. You will not be able to identify the participant you are matched with. One participant in each pair will be selected randomly and assigned the role of "Person 1", while the other will be "Person 2". Each Person 1 will be asked to make choices between options A and B. These decisions will affect earnings of both herself and her matched Person 2. Person 2 will not be able to influence his earnings, but will be asked to get in the shoes of Person 1 and indicate what choices he would make if his decisions did matter. The experiment will consist of two parts. One of them will be randomly selected at the end of the experiment and one of the choices between options A and B, once again selected randomly, made by Person 1 will determine the final earnings of herself and her matched Person 2.

We will explain Part 1 now. Assume that you are Person 1, thus your choices will matter. The figure shows a typical screen that you may see (note that the actual numbers will be different; also the choices that have already been made only serve as illustration).

There are two options in each row: A and B. Each option involves certain payment in zloty for you [Polish: "dla Ciebie"] and the other participant ["dla innego uczestnika"]. For example, option B in row 5 means that you get 36 zloty and the other participant gets 18 zloty. In each row you have to decide whether you like option A better, option B better or you think both options are equally good (you are perfectly Indifferent) and click the corresponding button.

You may notice that Option A is the same in each row, while the payment you are getting in Option B increases as you go down the table (and the payment to the other person remains constant as well). It may thus be, for example, that you prefer A in rows 1-9, are indifferent in row 10 and prefer B in rows 11-13. Or you may prefer A in rows 1-8 and B in rows 9-13 etc. We will typically want to learn your preference in more detail. For example, if you chose A in rows 1-11 and B

Opcja A	Wybór A	Wybór I	Wybór B	Opcja B
70 dla Ciebie, 30 dla innego uczestnika	●	○	○	0 dla Ciebie, 18 dla innego uczestnika
70 dla Ciebie, 30 dla innego uczestnika	●	○	○	9 dla Ciebie, 18 dla innego uczestnika
70 dla Ciebie, 30 dla innego uczestnika	●	○	○	18 dla Ciebie, 18 dla innego uczestnika
70 dla Ciebie, 30 dla innego uczestnika	●	○	○	27 dla Ciebie, 18 dla innego uczestnika
70 dla Ciebie, 30 dla innego uczestnika	●	○	○	36 dla Ciebie, 18 dla innego uczestnika
70 dla Ciebie, 30 dla innego uczestnika	●	○	○	45 dla Ciebie, 18 dla innego uczestnika
70 dla Ciebie, 30 dla innego uczestnika	○	○	○	54 dla Ciebie, 18 dla innego uczestnika
70 dla Ciebie, 30 dla innego uczestnika	○	○	○	63 dla Ciebie, 18 dla innego uczestnika
70 dla Ciebie, 30 dla innego uczestnika	○	○	○	72 dla Ciebie, 18 dla innego uczestnika
70 dla Ciebie, 30 dla innego uczestnika	○	○	○	81 dla Ciebie, 18 dla innego uczestnika
70 dla Ciebie, 30 dla innego uczestnika	○	○	○	90 dla Ciebie, 18 dla innego uczestnika
70 dla Ciebie, 30 dla innego uczestnika	○	○	●	99 dla Ciebie, 18 dla innego uczestnika
70 dla Ciebie, 30 dla innego uczestnika	○	○	●	108 dla Ciebie, 18 dla innego uczestnika

OK

in rows 12-13, you will get another screen (we call it "second iteration"), whereby Option A will still be the same, but your payment in Option B will vary from 90 to 99. Just as in the "first iteration", you will have to indicate in each row whether you prefer A or B (or indicate indifference).

You will be asked to make such choices for 10 different tables (and possibly their "second iterations"). If this part of the experiment is randomly selected at the end of the experiment and if you indeed play the role of Person 1, one of your choices will be randomly selected to determine the earnings of yourself and Person 2 - each of you will get the amount prescribed by the option you chose in the relevant row. Although this decision is selected randomly, not every decision will be selected with identical probability. Indeed, your choices in the second iteration are generally less likely to be selected. These probabilities are calibrated in such a way that it does not pay e.g. to "strategically" "switch" from A to B in the very last rows of the table in each "first iteration" just to see attractive earnings for yourself in the "second iteration". On the contrary, assuming this part is selected, you should in each and every case choose the option that you like better, as if this question was to be selected for real payments. If you are perfectly indifferent, click on button 'I' and the computer will randomly select one option for you. If you want to see all the details of this procedure, ask the experimenter for additional explanation (this procedure is a bit tedious and, to the best of our judgment, learning it will not help you make a good decision in any way). There is no time limit. However, to make sure that the experiment goes smoothly, you should not dwell on any particular decision for more than a few minutes. You will see a clock on your screen count down starting from 180s (but you will still be able to answer if this time is past).

In Part 2 you will remain matched in the same pairs and the same selected participants will play the roles of Person 1 and Person 2. The details will be explained at the beginning of the second part.

There will be no other opportunity to make money in this experiment. In particular, no Person 2 will be able to increase his payoff - it will depend entirely on the randomly selected decision of his matched Person 1. If you understand the task and are ready to start, click the button on your screen to commence control questions.

Instructions: Part 2

[printed on a separate page and distributed at the end of Part 1]

As before, it is assumed in the instructions that you play the role of Person 1. If you are Person 2, please imagine that you are Person 1 and indicate in each case how you would behave if your choices were to matter indeed.

In this part you will make choices between pairs of options A and B. Any of them may involve a 'lottery' - one of two allocations of money between you and Person 2 would be randomly selected. A typical screen will look something like that (again, the numbers are for illustrative purposes only) [skipped to save space]

If you choose option A, computer will randomly select on of two possible allocations: with probability 50% you will receive 27 zloty and the other participant will obtain 13 zloty. Otherwise, you get 25 and he gets 35. If, on other hand, you go for option B, you will get 23 for sure, while the other will get 20. If you choose "I" (Indifference), computer will randomly select option A or B for you and proceed as before.

Again, as in Stage 1, some of your answers may affect the values you see in subsequent questions. Again, the procedure is such that you should answer each question taking into consideration your preference for A or B in this particular question only.

Just as in Stage 1, you will see a clock and you should try to answer each question before time is up (but you will still be able to answer the question when the clock gets to zero).

When this part is over we will randomly select the part and the question that will determine the earnings. As mentioned before, no matter which role you're playing, you will not be able to obtain any additional earnings in this experiment.

Post-experimental questionnaire on motivation

Motives to be rated on 0-7 scale in the post-experiment questionnaire

Stage 1:

1. maximizing your own payoff

2. maximizing other person's payoff
3. maximizing joint payoff
4. making sure that the other person at least gets some reasonable minimal payoff
5. maximizing your own payoff relative to the payoff of the other person
6. making sure that the other person did not have (much) more than you
7. making sure that the other person did not have (much) less than you
8. avoiding the impression that you are a selfish or greedy person
9. avoiding the feeling of guilt vis-a-vis the other person

Stage 2:

1. reducing inequality in ACTUAL payoff between you and the other person (that is to say, trying to avoid the situation in which, after the selected lottery is resolved, your payoff is much higher or much lower than the payoff of the other person)
2. reducing inequality in EXPECTED payoff between you and the other person (that is to say, trying to avoid the situation in which, before the selected lottery is resolved, your expected payoff is much higher or much lower than the expected payoff of the other person)
3. avoiding payoff uncertainty

TRAVAUX EN COURS MENTIONÉ DANS LA SYNTHÈSE ET DANS l'OEUVRE SCIENTIFIQUE CAR EN L'ETAT D'ÊTRE PUBLIÉ 2 : Stefan Grimm, Martin Kocher, Michal Krawczyk et Fabrice Le Lec (2017), "Sharing the pie... or gambling for it?", invitation à resoumettre pour **Experimental Economics**.

NB: La version présentée est celle ayant été invitée à resoumettre, initialement soumise à *Experimental Economics* en 2017.

Sharing or gambling?
On risk attitudes in social contexts[±]

Stefan Grimm[#], Martin G. Kocher[&], Michal Krawczyk[*], and Fabrice Le Lec[+]

Abstract

Decisions under risk are very often embedded in a social context that we usually abstract from when studying decision making in the laboratory. In contrast to that practice, our experiment investigates whether risk taking is affected by social comparisons. In particular, we focus on situations where some resource has to be allocated between two parties: either the resource can be shared, or a random device allocates the entire resource to one of the parties. We find that the social context of the decision matters strongly: When participants are in a disadvantaged initial position compared to the other party, they select the risky option much more often than in a purely individual decision, identical in all other respects. Overall, we find that individuals are relatively more risk-seeking in the socially unfavorable domain than in isolation, in contrast to the favorable one, where we find no or little change in elicited risk attitudes in comparison to an isolated decision.

JEL classification: A13, C91, D03, D81

Keywords: risk attitude, risk in social context, social comparison

This version: December 2017

[±] Financial support by the LMU Munich is gratefully acknowledged. Grimm acknowledges financial support by the German Research Foundation (DFG) through GRK 1928.

[#] LMU Munich, Department of Economics, Geschwister-Scholl-Platz 1, 80539 Munich, Germany, stefan.grimm@econ.lmu.de.

[&] Corresponding author. Institute for Advanced Studies, Vienna, Josefstaedter Str. 39, 1080 Vienna, Austria, martin.kocher@lrz.uni-muenchen.de; University of Vienna, Austria; and University of Gothenburg, Gothenburg, Sweden.

[*] University of Warsaw, 44/50 Dluga St, 00-241 Warsaw, Poland, mkrawczyk@wne.uw.edu.pl.

[+] University Paris-1 Panthéon-Sorbonne, Centre d'Economie de la Sorbonne UMR CNRS 8179; MSE 96-102 boulevard de l'Hôpital, 75647 Paris Cedex 13, France, fabrice.le-lec@univ-paris1.fr.

1 Introduction

Many – if not most – economic decisions take place in a social context. People observe other people's choices, and they are themselves observed when making decisions, they affect other people through their decisions, and they reflect on other people's situations when making a decision. This is also true for decisions under uncertainty. It is difficult to come up with examples of decisions under risk taken in situations that are totally free of social context: risky choices by managers have consequences for other organizational members; financial decisions within the family have an impact on all family members and will be influenced by similar peers' decisions; even at the roulette table or when playing lotteries social influences are very often present.

Models of decision under risk usually abstract from the social environment in which decisions are made: the typical situation studied by decision theory is one where the individual makes a choice with neither any influence on others nor any information on others' situations, choices, or outcomes. However, it may very well be that decisions under risk in social environments differ from the equivalent decisions taken in purely individual contexts. If that is the case, standard models would lack consideration for the social drivers of such risky decisions and ultimately lead to inaccurate behavioral predictions in many economic circumstances. At least two phenomena suggest an important role of the social context in risky decisions: first, broadly speaking, it has been shown that preferences depend heavily on theoretically 'irrelevant' aspects of the environment or context (Tversky and Simonson, 1993); second, there is ample evidence that individuals are sensitive in many ways to others' situations (Fehr and Schmidt, 1999; Frank, 2005).

Despite its potential relevance, the effect of social context in risk taking has only recently received attention in the empirical/experimental literature in economics: Following the burgeoning of studies on other-regarding preferences that focused on deterministic outcomes, empirical research has started to explore the issue of the interaction of risk and social concerns[1]. Some of the relevant studies explicitly focus on peer effects in decision making under risk (Cooper and Rege, 2011; Bursztyn, Ederer, Ferman, and Yuchtman, 2014; Cai, de Janry, and Sadoulet, 2015; Lahno and Serra-Garcia, 2015), while others have primarily looked at decision making about risk borne by others (Chakravarty, Harrison, Haruvy, and Rutstroem, 2011; Vieider, Villegas-Palacio, Martinsson, and Majia, 2015).

Most relevant for this paper is the literature on risk taking with payoff implications for oneself *and* another person. Brennan, Güth, Gonzalez, and Levati (2008) point towards only a small effect of another person's risk per se on own risky decision making. Bolton, Ockenfels,

[1]An early survey is provided in Trautmann and Vieider (2012). Another approach is to study the correlation of risk and social preferences on the individual level (e.g., Müller and Rau, 2016).

and Stauf (2015), on the other hand, indicate that individuals might become more risk averse when also being responsible for other people. Adam, Kroll, and Teubner (2014) show a decrease in risk taking if outcomes of lotteries of coupled participants in laboratory experiments are asymmetric, i.e. one player wins and the other one loses, compared to independent lotteries. Friedl, De Miranda and Schmidt (2014) observe lower insurance take-up when risks are positively correlated (albeit this was not replicated by Krawczyk, Trautmann, and van de Kuilen, 2016). Krawczyk and Le Lec (2010) report lowest giving in a dictator game with probabilistic, negatively correlated payoffs. Overall, although there are no unambiguous conclusions, the existing literature shows that individual differences in payoffs from lottery choices, and hence social comparisons, can often play a role in decision making under uncertainty.

However, very few empirical papers have explicitly focused on *the effects of social comparison on elicited risk attitudes*. The existing evidence, again, is not fully conclusive: Linde and Sonnemans (2012) find that decision makers are more risk averse when in a socially unfavorable situation (that is, when they are disadvantaged compared to another person that serves as a natural reference point) than in a socially favorable one. In contrast, Bault, Coricelli, and Rustichini (2008), Bolton and Ockenfels (2010), and Fafchamps, Kebede, and Zizzo (2015) observe that decision makers are less risk averse when the situation is unfavorable. Dijk, Holmen, and Kirchler (2014) find that investors on experimental asset markets performing below average favor positively skewed portfolios (those that have a small chance for very high returns), while those performing above average prefer negatively skewed portfolios. These effects occur independently of whether others' outcomes are payoff-relevant (tournament-based incentives) or not.

In this paper, we want to shed more light on risky decision making within a strong social comparison context. Does risk taking depend on whether somebody else's payoff is affected and on the relative position towards that other person? And how can risk taking patterns be defined, depending on how unequal the initial positions are? Our specific setting that we will look at is resource allocation. Consider a decision maker who can either implement a certain allocation of the resource between herself and a second individual (the 'receiver') or use a random device to allocate the entire resource to either herself or to the receiver. More specifically, the choice is between splitting the resource (dividing the pie into shares of x % for the decision maker and 100−x % for the receiver) and using a random draw to allocate it in one piece (whereby the chances to get the entire pie are x% and 100−x%, respectively). In our experimental protocol, x is varied across different decision tasks, allowing to test changes in risk taking related to the relative social situation of the decision-maker. Such a setup reproduces, in a simplified manner, important aspects of many situations that involve risk: a decision maker can either go for a given allocation (of financial resources, power, or positions) or gamble for the entire pie. For instance, a manager can accept the proposed split of available funding between her and another manager's project or argue that the company should focus on just one of them; a political leader may have a choice between accommodating the current division of power between herself and a party rival or go for a shootout that will leave just one of the two standing; a poker player in a cash game can leave the table with current possessions or continue playing until all is lost or won. In short, we capture a situation of competition for a resource, and by systematically varying the given x, we are able to analyze how risk attitudes are affected by the initial division of claims on the resource.

The different ranges for x correspond directly to the social standing of an individual. Socially favorable or advantageous situations are those where x is greater than 50: for instance with x = 70, the decision maker has to choose between (OPTION A) a deterministic division giving *70% of the resource* to herself and *30% of the resource* to the receiver, and (OPTION B) the gamble involving a *70 % chance of receiving the entire resource* for herself and the remaining *30 % chance* of losing the entire resource to the receiver. Symmetrically, unfavorable or disadvantageous situations are those in which x is smaller than 50.

The protocol is built in a way such that not only a risk-averse decision maker would choose the deterministic option for any x, but also that social preferences will either be neutral or reinforce this tendency. We consider two types of social consideration (in order to take into account the possibility that individuals trade-off between risks and payoffs). First, for ex ante comparisons (that is, having social consideration to the allocations of expected payoffs as in Trautmann, 2009; Trautmann and Wakker, 2010; Saito, 2013), the two options are equal in terms of expected payoffs. Hence, if people have other-regarding preferences over expected outcomes, they should not play a role in the decision between the deterministic and the risky alternative. The second type of social consideration is to focus on ex post situations (the allocation of payoffs between individuals). In this case, the effect depends on the type of social consideration, but all of those discussed in the literature either have no role, or reinforce the preference for the deterministic option over the risky option. Indeed, the gamble option always creates the maximal inequality (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), indicating that inequity aversion would rather lead to a preference for the deterministic option. A similar argument applies to a potential maximin motivation (Charness and Rabin 2002): choosing the risky option always worsens the situation of the less advantaged individual. Efficiency concerns do not play any role in the decisions because total payoff is fixed across the two options. And (concave) altruism would favor the worse-off individuals (in relative terms) and strengthen the tendency to choose the deterministic option.

Hence, after consideration of these different types of preferences we would not expect people to choose the gamble. In short, with typical linear or concave consideration for others, the decision-makers should prefer the deterministic option.

Moreover, if the social context, in the sense of the possibility of social comparisons, has no influence on risk-taking, then a typical decision-maker should choose the same option (deterministic or risky) when faced with the social lottery decision described above and when faced with the equivalent decision situation devoid of any opportunity for social comparisons. Typically, an individual that prefers splitting x% of the resource to herself and 100-x% of the resource for the receiver rather than gambling for it with the same odds proportion would be expected to choose x% of the resource for sure rather than x% chance of winning the resource. Any systematic change of choice in our main task (the social lotteries) and the individual lottery control can then be attributed to a change of elicited risk attitudes when social comparisons are possible.

Our main finding is that the fraction of risky choices is strongly affected by social context: subjects seem to be more risk-seeking when the deterministic option involves unfavorable inequity in comparison to the same task in isolation. In contrast, a favorable social context (when the deterministic option corresponds to favorable inequity) does not increase the willingness to take risks. The analysis of individuals' behaviors suggests that most of this asymmetry is driven by about two thirds of the subjects who very strongly exhibit this pattern of choices. This pattern is robust to various controls and sensitivity checks, and deeper inspection of the data suggests that a competitive element is at play in the participants' choices. Two specific explanations are compatible with our data: either the other participant's payoff plays the role of a (social) reference point, below which the decision-maker is risk-seeking (Kahneman and Tversky 1979, 1992), or individuals are attentive to social ranking in a way that being ahead ("gloating") is more intensively sought than standing behind is avoided.

The remainder of the paper is organized as follows: the next section presents the experimental design and procedures, section 3 shows our results, and section 4 discusses the results in light of the existing literature and existing theories of decision making under risk. Section 5 concludes the paper.

2 Experimental design and procedures

The experiment consisted of five short parts: a series of risky choices in a social context, two dictator game decisions, a series of tasks to elicit individual risk preferences and potential loss aversion (in decisions without a social context), a series of risky choices without a social context (but different than in the part before), and the so-called ring test (the incentivized social value orientation questionnaire) to measure social value orientation.

2.1 Decision making under risk in varying decision contexts

We use a within-subject design that allows us to compare decision making under risk in a social context with decision making under risk in a purely individual context. Since the decisions are identical with respect to the decision maker's payoffs and probabilities, differences in decision making between the context-free and the social-context-relevant decisions can be attributed to the context in which the decisions took place.

In **part 1** of the experiment, subjects faced tasks where 10 euros had to be allocated between the decision maker and an anonymous receiver, with both being present in the laboratory. Two options were available. The first one (OPTION A) is the deterministic (safe) option, which is the plain division of the 10 euros, i.e. the allocation (x, $10-x$) for a given x. The second one is the risky option (OPTION B), which is the social lottery where the decision maker has a probability of $x/10$ of getting the 10 euros and the receiver gets 0, and the receiver has a probability of $(10-x)/10$ of getting the 10 euros and the decision-maker receives 0. The chances of winning the 10 euros were mutually exclusive between the decision maker and the receiver. The amount x was systematically varied to obtain nine different tasks, with x ranging from 1 to 9 in steps of 1. Table 1 displays all tasks subjects faced in part 1. Participants were asked whether they preferred Option A (henceforth also referred to as 'the safe option') or Option B (henceforth also 'the risky option'). They could also indicate indifference (Option C). For that case, they were told that Option A or Option B would be implemented randomly with equal probability, realized through a draw of the computer. Each subject was asked to make one choice in each row of the table.

Task	Safe option (in euros)	Risky option (in chances of winning 10 €)
T1	1 for chooser, 9 for receiver	10% for chooser, 90% for receiver
T2	2 for chooser, 8 for receiver	20% for chooser, 80% for receiver
T3	3 for chooser, 7 for receiver	30% for chooser, 70% for receiver
T4	4 for chooser, 6 for receiver	40% for chooser, 60% for receiver
T5	5 for chooser, 5 for receiver	50% for chooser, 50% for receiver
T6	6 for chooser, 4 for receiver	60% for chooser, 40% for receiver
T7	7 for chooser, 3 for receiver	70% for chooser, 30% for receiver
T8	8 for chooser, 2 for receiver	80% for chooser, 20% for receiver
T9	9 for chooser, 1 for receiver	90% for chooser, 10% for receiver

Table 1: Part 1 – Social context tasks

In **part 4** of the experiment subjects faced a task equivalent to part 1 of the experiment; however, now the choice was individual. That is, they had to decide between a safe payoff of x euros and a lottery with probability x/10 of receiving ten euros and probability (10−x)/10 of receiving nothing. There was no other participant that was affected from the decisions taken in part 4.

By comparing decisions in part 1 and part 4 of the experiment, we can isolate attitudes towards risk in the social context and compare these to risk taking in the individual context. Social context here simply means that another participant's earnings were determined by the choices of the decision maker.

2.2 Experimental controls

The remaining parts of the experiment (parts 2, 3, and 5) aim at measuring social preferences, as well as risk attitudes and potential loss aversion. More precisely, in **part 2** of the experiment, subjects had to play two dictator games (Forsythe, Horowitz, Savin, and Sefton, 1994; Bolton, Zwick, and Katok, 1998). The first was a regular dictator game with 10 euros to be divided between the decision maker (the dictator) and the receiver. The second allocation decision consisted of dividing chances to win 10 euros (the 'competitive probabilistic dictator game' of Krawczyk and Le Lec, 2010). For example, the dictator could decide that with probability 70% she will win 10 euro, while the other participant would win nothing and otherwise the opposite would be implemented. Finally, participants had to indicate which of the two 'games' they preferred. The first game provides us with a control for outcome-based social concerns, while the second game speaks to preferences regarding procedural social concerns. Thus, we have a measure of subjects' concerns for others to potentially identify the role these concerns may have played in part 1 of the experiment.

In **part 3** of the experiment participants received a truncated and adapted Holt and Laury (2002) multiple choice list to estimate subjects' risk attitudes with stakes comparable to the ones used in the main part of our experiment. This three-question version of the standard choice list contains the relevant choices in which the vast majorities of experimental subjects usually switch from safe to risky lotteries. We also included three decisions that aim at measuring potential loss aversion. Table 2 lists all choices in part 3 of the experiment.

Task	Option A	Option B
R1	50%: 5 Euro, 50%: 4 Euro	50%: 9.50 Euro, 50%: 0.25 Euro
R2	60%: 5 Euro, 40%: 4 Euro	60%: 9.50 Euro, 40%: 0.25 Euro
R3	70%: 5 Euro, 30%: 4 Euro	70%: 9.50 Euro, 30%: 0.25 Euro
L1	0 Euro for sure	30%: -2.50 Euro, 70% 2.50 Euro
L2	0 Euro for sure	40%: -2.50 Euro, 60% 2.50 Euro
L3	0 Euro for sure	50%: -2.50 Euro, 50% 2.50 Euro

Table 2: Part 3 – Risk and loss aversion choices

Part 5 elicits the subjects' social value orientation with the so-called ring test (Offerman, Sonnemans, and Schram, 1996; van Lange et al., 1997; Brosig, 2002; van Dijk, Sonnemans, and van Winden, 2002). In this fully incentivized test, subjects have to make binary choices in 24 different allocation tasks (see Appendix A for details). In each task, a subject has to choose among two allocations that give money to herself and another (anonymous) recipient. The recipient stays the same in all 24 allocation tasks, and all 24 tasks are paid. Adding up the 24 decisions yields a total sum of money allocated to oneself (x-amount) and to the recipient (y-amount). Using the ratio (x/y) one can assign a subject to one of eight categories of social orientation (individualism, altruism, cooperation, competition, martyrdom, masochism, sadomasochism, and aggression).

2.3 Experimental procedures

The experiment was conducted at the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA) in summer 2015. The experiment was programmed using z-tree (Fischbacher, 2007), and recruitment of participants was done with ORSEE (Greiner, 2015). We ran six sessions with a total of 144 subjects; mainly students from the University of Munich. Subjects were allowed to participate in only a single session.

All subjects were asked to take all the choices described above. Their role in parts 1 and 2 of the experiment – either decision maker or receiver – was determined after the experiment, using the strategy method (Selten, 1967; Brandts and Charness, 2011). Resolution of uncertainty was implemented and outcome information was given only at the very end of the entire experiment. Instructions for parts 1 to 4 were distributed and read aloud at the beginning of the experiment, and upon finishing part 4, instructions for part 5 were read and distributed. Subjects knew that there were exactly five parts from the beginning of the experiment.

To determine payoffs, subjects were randomly matched with another participant at the end of the experiment. Always one subject in these matched pairs was randomly selected for the role of decision maker; the other was the receiver. For all pairs of participants, a random mechanism decided the payoff-relevant part from parts 1 to 4. If part 1 or 2 was chosen, another random mechanism then decided which specific task within the part was to be implemented for both participants. If part 3 or 4 was chosen, the specific task to be implemented was determined for both participants separately. In addition to the payoff from this single decision out of parts 1 to 4, all subjects received their earnings from the ring test in part 5, which consisted of their payoff from their own choices and the payoff from the choices of the matched participants. Matching of participants in part 5 of the experiment was independent of the matching in parts 1 and 2. On top of these earnings, participants received a fixed payment of 4 euros for showing up on time. On average, participants earned 13.40 euros, and a session took about 50 minutes.

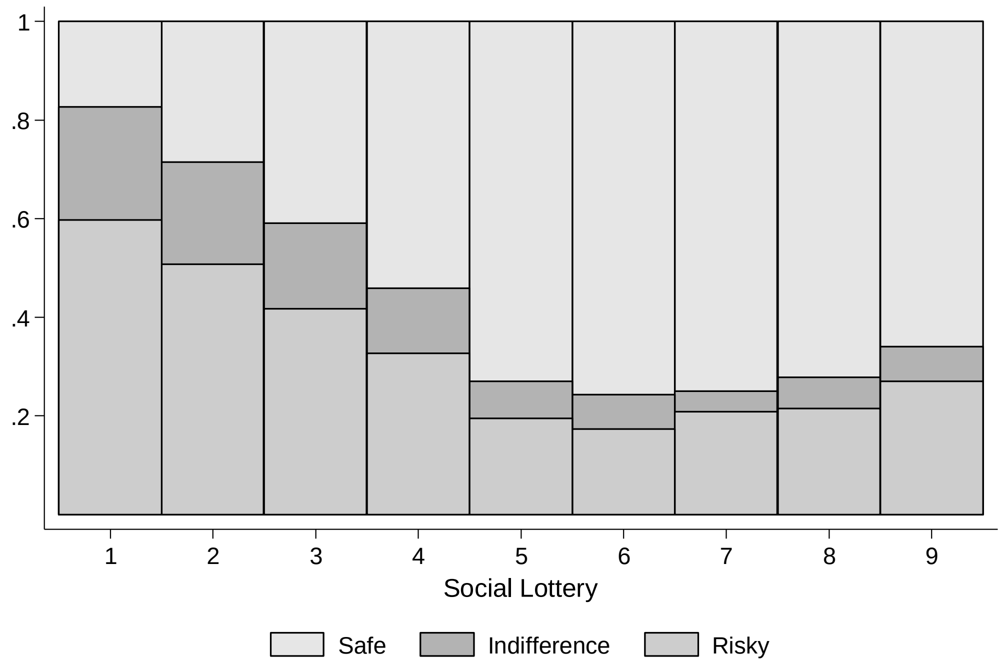
Participants were also asked to fill out a questionnaire after part 5 including a short description of motivations for decisions in the experiment and questions regarding socio-demographic characteristics. All design details and the procedural details were common knowledge among participants (see the instructions for all parts in Appendix B).

3 Experimental results

We will first have a look at aggregate results (section 3.1), before analyzing the data at an individual level and taking into account the heterogeneity in responses (section 3.2). Section 3.3 reports the results of a pilot experiment that provides further support for the robustness of our results.

3.1 Aggregate results

An overview of the results from decision making under risk in the social context (part 1 of the experiment) is shown in Figure 1. The aggregate pattern of risk taking is roughly L-shaped, with subjects willing to take considerably more risk in unfavorable tasks. The level of risk taking reaches its lowest value just above the equal split.



Notes: The y-axis denotes the fraction of subjects that chooses a certain option (risky, indifference or safe in different shades of grey) for a given lottery. The x-axis represents the different types of lotteries from T1 to T9 with unfavorable lotteries to the left (T1 to T4) and favorable one to the right (T6 to T9).

Figure 1: Distribution of choices in the social context

The asymmetry between favorable and unfavorable situations is statistically significant. Leaving the case of the equal split aside for the moment, all comparisons between tasks corresponding to sure payoffs adding up to 10 (T1 vs. T9, T2 vs. T8, T3 vs. T7, and T4 vs. T6) suggest that the risky option is relatively more appealing when the safe option implies unfavorable inequity: The differences are significant according to a Stuart-Maxwell test at the 1%-level.[2] If we pool indifference with the risky option or with the safe option, McNemar's tests remain significant at the 1%-level for either pooling version and for all comparisons.[3] Looking at the unfavorable situations only, statistical tests support increasing risk taking from the equal split towards the more unfavorable tasks. For all binary comparisons between the tasks in the unfavorable domain, choices move strongly towards more risk taking, the more unfavorable and risky the tasks become ($p < 0.01$ for Stuart-Maxwell tests for all comparisons). This pattern of choice is not necessarily indicative in itself of a change of behavior in the favorable and unfavorable social domains. It is overall equally compatible with an inverted-S transformation of probabilities as in cumulative prospect theory (Kahneman and Tversky 1992). As is well established (Wakker 2010 for instance), low

[2]The Stuart-Maxwell marginal homogeneity test is applied the same way as the McNemar test for testing marginal homogeneity, but it is used for variables with more than two categories ("safe", "indifference", and "risky"). For two categories, the two tests are equivalent.

[3]In the remainder we will only indicate the results of McNemar's tests grouping indifference with either safe or risky choices if they differ from the respective Stuart-Maxwell test.

probabilities of the good outcome are typically overweighed from 0 to roughly one third, while intermediate and large probabilities of the good outcome are usually underweighed. Hence, the asymmetry of choices could result from the typically observed probability transformation.

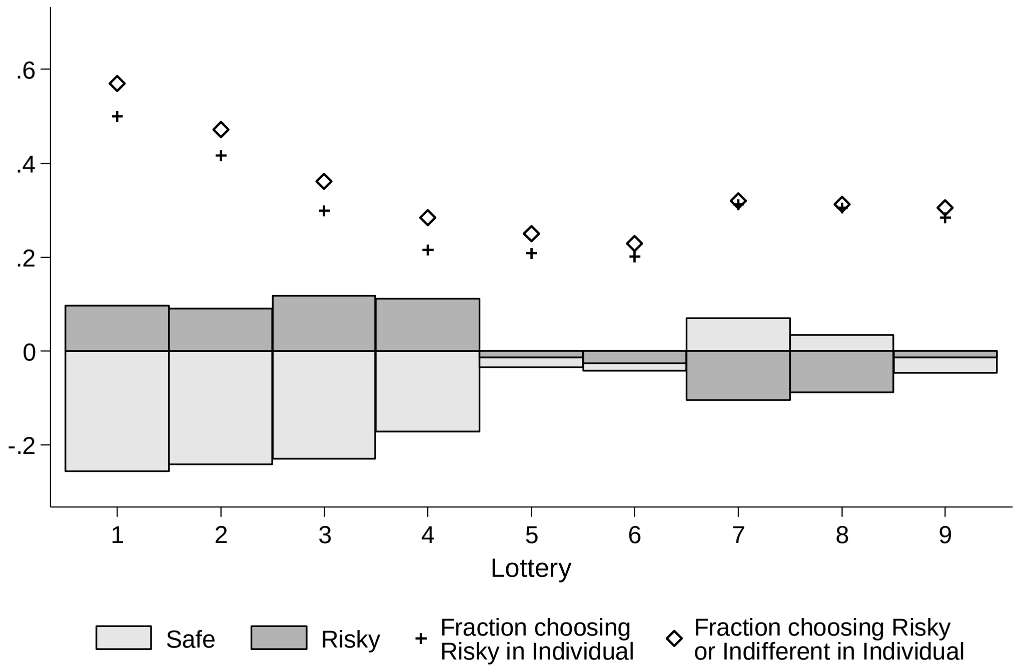
To test whether individuals' choices are actually driven by the social context of the decision, we compare the social tasks from part 1 with choices from part 4 of the experiment. The nine tasks in part 4 (henceforth T1i to T9i, where i stands for "individual") were the exact counterparts of T1 to T9 from part 1 in terms of payoffs and probabilities for the decision maker, but stripped from the social context, as there was no receiver. If choices are influenced by the social context, then we should observe differences in the frequencies of risky choices between the individual task and the social task. Comparisons are displayed in Figure 2, indeed suggesting systematic differences between decisions in social and individual contexts.

These differences are important in the unfavorable range: For T1 vs. T1i, T2 vs. T2i, T3 vs. T3i, and T4 vs. T4i, individuals take significantly more risk when facing the social lottery than in the equivalent individual task, and this difference is highly significant ($p < 0.01$ for all four Stuart-Maxwell tests)[4]. We observe, for T1-T4, that roughly 20 % of subjects move away from the safe lottery in the social context and about 10 % to the risky one. The percentage of changes is relatively constant for all the unfavorable social situations. The fact that already an important share of subjects chooses the risky option for low probabilities in the individual task (T1i-T3i) partly hides the extent of the change of choice between the individual and the social context. As an illustration, consider T1: as already 50.0 % the subjects chose the risky lottery in the individual task (T1i), a change towards a more risk-seeking decision can only be observed for the remaining half. Hence, for the subpopulation that chose the safe lottery in the individual task, exactly 60 % did not choose the safe option in the social case (moving to either indifference for 45 % of them or to the risky option for 55 %). This means that around one-half of subjects for whom it was possible to switch to a riskier option did so in T1. The share of subjects moving away from the safe option in the social context ranges from a quarter to a half from T4 to T1.

The pattern is less clear for the favorable range. For T7 vs. T7i and T8 vs. T8i, there is more risk taking in the individual tasks ($p = 0.01$ for Stuart-Maxwell tests). However, this result is not robust to using McNemar's tests and pooling indifference with risky choices,

[4]Here, if we pool indifference and safe option choices, McNemar's tests of marginal homogeneity result in differences that are significant only at the 10%-level for T1 vs. T1i and T2 vs. T2i and that are significant at the 5%-level for T3 vs. T3i and T4 vs. T4i. If we pool indifference with risky choices, all tests are significant at the 1%-level.

since many subjects simply switch from the risky choice in the individual to indifference in the social context. Other possible comparisons do not yield significant differences.



Notes: Bars denote the change in the fraction of subjects choosing the risky (safe) option when going from the individual to the social context. Positive values indicate that a higher fraction of subjects chose the respective option in the social context (part 1) in the specific lottery. Bars do not add up to zero, since the fraction of indifferent subjects changes simultaneously. The dots correspond to the fraction of subjects choosing risky (plus) and risky OR indifference (rectangular) in the individual lotteries to allow for a direct comparison to figure 1. The area above the rectangular dots (y-axis cut off here at 0.7) consequently refers to the fraction of subjects choosing safe in the individual lotteries.

Figure 2: Difference in choices between the individual and social context

Overall, we observe that decision makers seem to be affected by social context when making a risky decision, but not in asymmetric way: they unambiguously take more risk when the situation is unfavorable, but display about the same choices when it is favorable to them compared to a risk-equivalent individual context.

To test the robustness of our results, we ran an ordered probit model (column 1 of Table 3) on the choices made in all 18 lotteries (with and without social context). We use indicator dummies for the nine types of lotteries (*Type 1* to *Type 9*), without separating the social and individual tasks and dummies *Social 1*, *Social 2*, etc. for the task being social (T1, … T9) or not (T1i, T9i). Hence, coefficients on *Type 1* up to *Type 9* correspond to the average risk taking in the individual lottery tasks, while coefficients for *Social 1*, *Social 2*, etc. correspond to the additional risk taking in the social context (relatively to the individual one). The results are displayed in Table 3.

	Ordered Probit	Probit Baseline	Probit Indifference
Type 1	0.848***	0.812***	0.849***
	(0.146)	(0.150)	(0.146)
Type 2	0.616***	0.602***	0.605***
	(0.144)	(0.149)	(0.142)
Type 3	0.311**	0.284**	0.319**
	(0.136)	(0.142)	(0.134)
Type 4	0.0789	0.0240	0.106
	(0.117)	(0.120)	(0.118)
Type 5	- ref. -	- ref. -	- ref. -
Type 6	-0.0541	-0.0245	-0.0671
	(0.107)	(0.107)	(0.108)
Type 7	0.257**	0.323**	0.205
	(0.129)	(0.126)	(0.129)
Type 8	0.237*	0.304**	0.186
	(0.139)	(0.136)	(0.138)
Type 9	0.199	0.243*	0.166
	(0.145)	(0.145)	(0.144)
Social 1	0.431***	0.246*	0.765***
	(0.113)	(0.126)	(0.139)
Social 2	0.411***	0.228*	0.639***
	(0.107)	(0.120)	(0.120)
Social 3	0.457***	0.318**	0.584***
	(0.120)	(0.132)	(0.128)
Social 4	0.411***	0.338**	0.464***
	(0.116)	(0.136)	(0.121)
Social 5	0.0270	-0.0494	0.0642
	(0.123)	(0.136)	(0.123)
Social 6	-0.00237	-0.103	0.0451
	(0.141)	(0.146)	(0.143)
Social 7	-0.257*	-0.323**	-0.205
	(0.136)	(0.133)	(0.136)
Social 8	-0.171	-0.280**	-0.101
	(0.131)	(0.130)	(0.132)
Social 9	0.0449	-0.0414	0.0968
	(0.132)	(0.135)	(0.130)
Constant (Cut1)	0.644***	-0.812***	-0.674***
	(0.115)	(0.118)	(0.114)
Constant Cut2	0.875***		
	(0.117)		
Observations	2,592	2,592	2,592
Pseudo R-sq.	0.058	0.054	0.092

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: As dependent variable in the ordered probit regression in column 1 we use an ordinal scale for risk taking (0 = safe choice, 1 = indifference, 2 = risky choice) in the respective lottery. Columns 2 and 3 display results from regular probit regressions using an indicator variable for the choice being risky (column 2) and risky or indifferent (column 3). As independent variables we use dummies for all nine types of lotteries in general (Type 1 to 9), as well as nine dummies indicating whether the lottery is within a social context (Social 1 to 9) or not. Standard errors are robust and clustered at the subject level. Constant Cuts in column 1 are threshold parameters of the ordered probit model to differentiate low risk takers from indifferent and risky subjects (Cut1) and low risk takers and indifferent subjects from risky subjects (Cut2). They are estimated such that Pr(Safe)=Pr(Xb+u<Cut1), Pr(Indifferent)=Pr(Cut1<Xb+u<Cut2), and Pr(Risky)=Pr(Cut2<Xb+u).

Table 3: Risky choices in individual vs. social context

The regression results confirm the findings based on non-parametric tests: decision makers indeed take on average more risk in unfavorable social situations compared to the equivalent individual situations. All terms indicating the social context are positive and significant in the unfavorable domain (at the 1%-level). For favorable situations the effect is reversed. It seems like – if anything – individuals reduce risk taking in the social context for favorable situations compared to situations without such context. These results are robust to using an ordinary probit model. Both when taking an indicator variable for risky choices only (column 2) and when taking an indicator variable for risky and indifferent choices (column 3) as dependent variable, we obtain the same pattern of results.

In sum, we observe some variability in the proportion of risky choices in the individual tasks, likely related to a non-linear treatment of probabilities, but more relevantly here, we observe a strong effect of social comparisons in the unfavorable domain. In such tasks, participants take much more often the risky option than in the individual task. On the contrary, very little, if any, effect is found in the favorable domain.

3.2 Individual heterogeneity

We now turn to individual data, with three aims in mind. The first one is to check the robustness of our finding when taking into account possible sampling variations in other characteristics (social preferences, risk attitudes, etc.). The second one is that these characteristics can be correlated with the strength of the effect we observe and shed light on its psychological drivers. And the last one is to establish how heterogeneous the sample is with respect to the effect of social comparisons on risk taking. For that purpose, we can look at the individual characteristics elicited in parts 2, 3 and 5 of our experiment. Table 4 provides an overview of these characteristics.

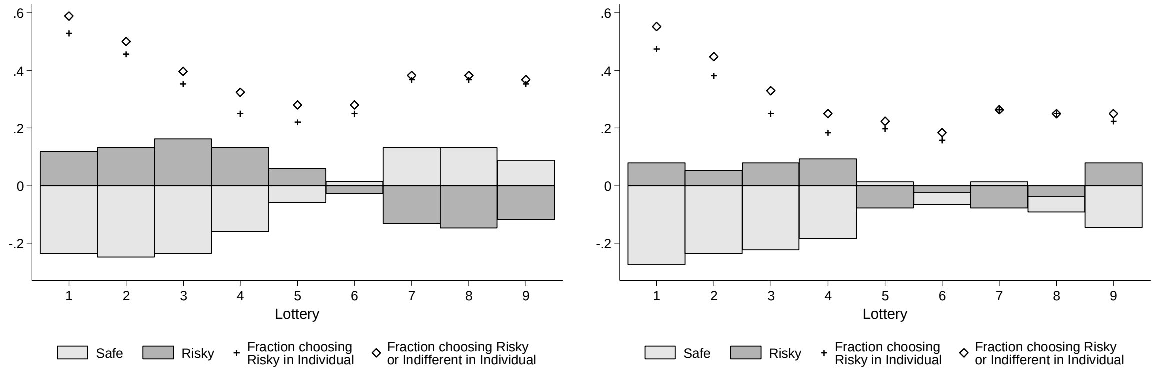
Dictator game	mean	median	10^{th}	25^{th}	75^{th}	90^{th}
Standard - Transfer	1.99	2	0	0	3	5
Probabilistic – Transfer (%)	14.04	10	0	0	27.5	40

Part 3 Lotteries	3 safe	2 safe	1 safe	0 safe	Inconsistent
Risk aversion (%)	31.94	43.75	11.11	12.5	1 obs.
Loss aversion (%)	22.92	40.28	14.58	19.44	4 obs.

Ring Test	Competitive	Individualist	Cooperative	Neg. angle	Pos. angle
Sample fraction	1.39	71.53	27.08	45.14	54.86

Table 4: Observables from parts 2, 3 and 5

One aspect in which subjects potentially differ is whether they are socially oriented, i.e. other-regarding (inequity averse, altruistic, etc.). Categorizing selfish and pro-social subjects on the basis of a median split in their offer in the dictator game in part 2 of the experiment provides us with additional insights.[5] Figure 3 shows the differences in choices for the two groups (for reasons of elucidation, call them "egoists" and "altruists") when going from the individual to the social context.



Notes: Same as figure 2, now only comparing the effects in two subsamples. The left part of the figure refers to the subjects with below-median dictator giving (in part 2 of the experiment), while the right part describes risk taking of above-median dictator giving subjects. The dots now also allow for level comparisons between the subsamples (left panel vs. right panel).

Figure 3: Choices by dictator giving (left: egoists, right: altruists) in the social vs. individual context

For both groups, decision making in the unfavorable range changes strongly from the individual to the social context. Still, the pattern of changes is slightly different: altruists in the dictator game (right panel) strongly switch from the safe option to mainly indifference (and some risky), while self-interested subjects (left panel) switch more to the risky option and less to indifference. In the favorable range, the difference between the groups becomes even more apparent. Selfish participants switch from risky to safe from the individual to the social context. Altruists, however, show a less clear-cut pattern of change; they often switch from safe and risky to indifference for T8i vs. T8 and from safe to risky and indifference for T9i vs. T9. These results remain roughly unchanged if we use generosity in the probabilistic dictator game for the sample split. This suggests that the effect of social comparisons is very widespread in the unfavorable domain while in the favorable one, it may only concern self-interested participants, and to a weaker extent.

To see how these effects depend on other personal characteristics and to check their robustness, we ran ordered probit models similar to the one in Table 3, now including interaction terms with the different types of personal characteristics. For that purpose, in contrast to Table 3, we now only use three dummies for the different types of lotteries. This limits the number of interaction terms and makes the interpretation of the results more straightforward. *Unfavorable* is a dummy indicating that the lottery has an expected value

[5]Roughly 47% of the dictators give nothing or 1 euro, while 53% give 2 euros or more.

below five (T1(i) to T4(i)); *Equal Split* indicates the equal split lottery (T5(i)); and *Favorable* stands for lotteries with an expected value for the decision maker larger than 5 (T6(i) to T9(i)). As before, we also include interaction terms for these lottery types with a dummy indicating a social context (*Social*). Column 1 shows results for this baseline specification with fewer dummy indicators than in Table 2. In columns 2-4, we then interact the six baseline variables with an indicator variable for below median dictator giving as in Figure 3 (column 2 of Table 5), for a negative angle in the ring test for social value orientation (column 3), and for low loss aversion (column 4) from part 3 of the experiment. This indicator variable is denoted *X*. A negative angle in the ring test implies that the decision maker in part 5 of the experiment chose such that the matched participant received a negative payoff from these choices. This is only possible if, at least at one point for the 24 tasks, the decision maker preferred to take money away from the matched participant, with no monetary benefit or possibly even at a cost for him- or herself.[6] However, as argued above, being classified as individualistic with a negative angle already implies some form of competitive preferences. Low loss aversion (column 4) means that subjects at least in all but one of the loss aversion decisions chose the option involving the chance of a loss. This is true for 49 of the subjects. The results are provided in Table 5.

The baseline regression results (column 1) again confirm the pattern found for the finer-grained lottery definitions in Table 2: Compared to the individual context, in the social context average risk taking increases for unfavorable tasks. Interestingly, this result is robust to all the specifications in Table 5. If we now look at the regression results including interaction terms, interesting patterns emerge. For altruists – according to our dictator giving measure (upper part of column 2) – the increase in risk taking due to the social context in unfavorable tasks is positive and significant. The interaction term for unfavorable lotteries in the social context for selfish participants (*Unfavorable * Social * X*) is small and not significant at conventional levels. This also holds for other specifications of the altruism indicator: None of the differences in the social context between altruistic and selfish participants are statistically significant if we consider positive (non-zero) transfers as altruistic in both the standard or probabilistic dictator game or if we define above-median giving in the probabilistic dictator game as altruistic behavior. Overall, pro-sociality as measured by generosity in a Dictator Game does not seem to be related to the tendency to take more risk in unfavorable social contexts.

[6]In our preferred specification, we refrain from using the strict classification into types (individualistic, competitive, cooperative, etc.) described in Appendix A, since we only have two subjects classified as purely competitive and a vast majority in the individualistic category. Median split dummies avoid estimation problems in interactions with continuous variables in ordered probit.

Dependent Var: Risk choice	(1)	(2)	(3)	(4)
Unfavorable	0.474***	0.463**	0.553***	0.649***
	(0.122)	(0.189)	(0.164)	(0.169)
Favorable	0.163	0.0616	0.100	0.0103
	(0.114)	(0.173)	(0.167)	(0.141)
Unfavorable * Social	0.417***	0.394***	0.255**	0.317***
	(0.0836)	(0.117)	(0.119)	(0.105)
Equal Split * Social	0.0270	-0.128	-0.0329	-0.0929
	(0.123)	(0.175)	(0.157)	(0.170)
Favorable * Social	-0.0961	0.0951	0.00301	0.00860
	(0.109)	(0.149)	(0.150)	(0.139)
		X: Low dict. giving	X: Non-cooperative	X: Low loss aversion
Unfavorable * X		0.172	-0.212	-0.110
		(0.169)	(0.169)	(0.178)
Equal Split * X		0.146	-0.0331	0.361
		(0.229)	(0.230)	(0.238)
Favorable * X		0.350*	0.106	0.736***
		(0.189)	(0.190)	(0.192)
Unfavorable * Social * X		0.0563	0.366**	0.292*
		(0.168)	(0.163)	(0.171)
Equal Split * Social * X		0.304	0.133	0.292
		(0.247)	(0.250)	(0.254)
Favorable * Social * X		-0.388*	-0.221	-0.251
		(0.219)	(0.218)	(0.236)
Constant Cut1	0.646***	0.717***	0.631***	0.780***
	(0.115)	(0.164)	(0.155)	(0.147)
Constant Cut2	0.871***	0.943***	0.857***	1.011***
	(0.117)	(0.164)	(0.158)	(0.144)
Observations	2,592	2,592	2,592	2,592
of which interaction	-	1,224	1,170	822
Pseudo R-squared	0.041	0.046	0.043	0.058

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable is an ordinal scale measure for risk taking (as in Table 2). As in column 1 of Table 2, the columns report results from an ordered probit regression with robust and clustered standard errors. Constant Cuts are threshold parameters of the ordered probit model to differentiate low risk takers from indifferent and risky subjects (Cut1) and low risk takers and indifferent subjects from risky subjects (Cut2) (also see Table 2). Lottery types are now grouped in three blocks: Unfavorable (T1(i) to T4(i)), equal split (T5(i)), and favorable (T6(i) to T9(i)). Columns (2) to (4) use different sample splits for the interaction terms. For each column, results in the upper part of the table refer to risk taking of subjects for which statement X of the corresponding column does not hold. The equal split is omitted for the individual context here. The three interaction terms refer to the effect of the social context on these subjects. The lower part describes whether risk taking by subjects for which X holds is different from that behavior. The first three coefficients refer to the difference in risk taking in the individual context. The last three coefficients describe the difference in risk taking in the social context for the subjects for which X applies compared to what can be expected by the effect of the social context on subjects for which X does not apply and by the effects of X applying in the individual context. A significant effect here indicates a particularly strong effect of the social context on subjects for which X applies.

Table 5: Heterogeneity in the effects of the social context

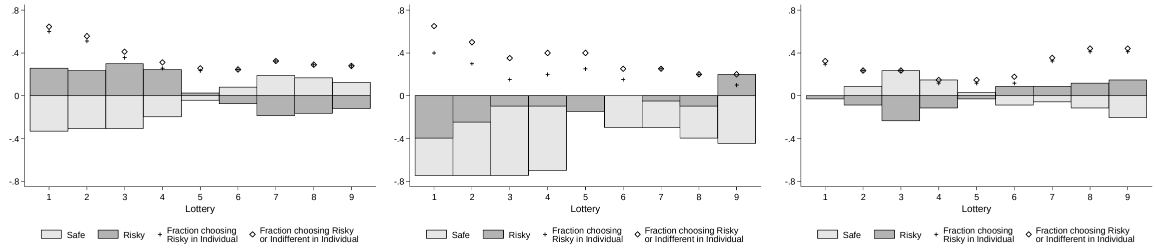
The differences are more clear-cut for the split based on ring test choices (column 3 in Table 5). For less competitive types in the upper part of the table, as for altruists in column 2, choices in the unfavorable range are affected by context. In this case, however, the more competitive types are clearly more strongly affected by the social context (significant at the 5%-level). Remember that the two measures for social preferences capture potentially different behavioral inclinations. Dictator giving is a proxy for altruism, whereas the ring test puts cooperative individuals against competitive individuals. The latter category cannot be captured by standard dictator giving decisions. It seems as if more competitive individuals show the strongest reaction to unfavorable situations in the social context. This line of reasoning is robust to a sample division into cooperative versus individualistic or competitive individuals, strictly based on the classifications described in the Appendix. In this regression, strictly cooperative types do not show an increase in risk taking in the unfavorable social domain. Instead, these subjects even take more risks in the favorable social domain. For the non-cooperative types based on this specification the reverse holds: they significantly take more risk in the unfavorable social domain while they take much less risk on average in the favorable social lotteries.

Column 4 looks at interactions with loss attitude: loss averse subjects, according to our measure, are, as in the overall pattern, affected by the social context in the unfavorable domain, but low loss averse subjects seem, if anything, disproportionally more strongly affected (significant at the 10%-level). This result, however, is not robust to defining low loss aversion as choosing all three lotteries involving losses or as choosing only at least one lottery involving a potential loss. Further, the coefficient on *Unfavorable * Social * X* is also insignificant if we consider indifference choices in the loss aversion tasks as taking the lottery involving the loss.

Finally, we ran a k-medians clustering[7] analysis on all social lotteries dividing the subjects into three clusters. This analysis resulted in the following characterization: the first and clearly largest class of decision makers is comprised of 90 (out of 144) individuals who exhibit a strongly domain-dependent pattern of risk attitudes in the social context (strongly risk-seeking in the unfavorable case, and risk-averse in the favorable one); the second cluster (20 subjects) is overall risk-averse and very often chooses indifference (especially for the unfavorable range); and the final cluster (34 individuals) shows increasing risk taking for the favorable range as well as for extremely unfavorable tasks.[8] The results from the cluster analysis by types are shown in Figure 4.

[7]The K-medians clustering partitions subjects into k groups by finding k centroids that minimize the overall distance between data points and the closest centroid. The use of medians rather than means (“k-means clustering”) is more appropriate for discrete data as is the case here.

[8]The exact clustering always depends on the random starting points for building the clusters. Hence, if we had chosen different starting points, we would have ended up with different clusters. However, varying the starting points leaves the overall conclusions largely unchanged.



Notes: See Figure 2 or 3 for an explanation of the bars and dots. The left-most panel represents choices of type-1 individuals, the middle panel refers to type-2 subjects, and the right-most panel represents type-3 individuals.

Figure 4: Choices by types (1-3) in the social vs. individual context

The categorization of subjects can help explain the aggregate pattern: the overall asymmetry in risky behavior across favorable and unfavorable situations seems to be mostly (but not only) driven by – the most prevalent – type-1 individuals. For these subjects the increase in risk taking in the unfavorable range when going from the individual to the social context is very pronounced, while they seem to reduce risk taking in the favorable range. Furthermore, the small surge in overall risk-seeking behavior when the situation becomes more and more socially favorable can almost entirely be accounted for by type-3 participants. These subjects, however, even reduce risk taking in the unfavorable range when they are in the social context. Type 2 individuals are most effectively characterized by their inclination towards indifference in the social lotteries – especially in the unfavorable domain. Statistical tests confirm this first impression (see Appendix C).

3.3 Additional evidence from a classroom experiment

A classroom experiment was conducted prior to the laboratory experiment described in detail above, and it inspired most of the latter’s design. In the classroom experiment, the social context tasks were the same as in the present study, even though stakes and payment procedures differed (only a subset of participants was selected for payments and the stake size was 50 euros instead of 10 euros). Next to the social context tasks, subjects also worked on an equivalent risk and loss aversion elicitation task, as well as on three individual context tasks (as opposed to all nine tasks in the above laboratory study) for comparison to the social tasks.

Due to the design differences, results in the two studies should only be compared with caution. Nonetheless, the conclusions from the classroom experiment and the laboratory experiment are strikingly similar. As in the present study, we also found much stronger risk taking in the unfavorable domain compared to the favorable situations, and risk taking in the unfavorable range increases towards more unfavorable situations. Comparing this behavior to choices in the individual context within the classroom experiments leads to similar conclusions as in the laboratory experiment: risk taking in the unfavorable decisions is clearly higher in the social context than in the individual one, while it is only weakly higher in the favorable range. The fact that it still is higher in the favorable range points towards the only difference between the two studies: In the earlier classroom experiment aggregate risk taking somewhat increases in the favorable range, too, such that the aggregate picture rather gives a U-shaped pattern of risk taking, whereas the laboratory experiment reveals a more L-shaped pattern. The details of the results from the classroom experiment are provided in Appendix E.

4 Discussion

Overall, our results suggest that individual attitudes towards risk are strongly affected by the social context: We observe systematic deviations in social situations from what decision makers decide in similar situations that do not allow for social comparisons. In the following, we deepen our discussion outlined in the introduction on potential explanations in the light of different utility functions or decision theories in turn.

A natural contender to explain a change in risk attitude in social situations is the role of (ex post) social preferences. For instance, it seems intuitive that more inequity averse individuals (or more spiteful subjects) would see the safe situation as more unattractive in the unfavorable domain than in the favorable domain and would consequently be willing to gamble rather than to stay go for the deterministic outcome in this disadvantageous situation. In fact, this intuition is erroneous. First, choosing the risky option, for instance in T1, means to end up, potentially with a very high probability (of 90% case of T1) in a situation even worse from the point of view of these ex post preferences. For the inequity averse individual, choosing the risky option means having a very high probability to ending in an even more disadvantageous situation. Second, and more formally, what determines the attitude towards risk in our social tasks is not the type of social motives (altruism, spitefulness, competitiveness, inequity aversion, etc.) but, leaving aside probability transformations, the curvature of the utility on the linear segment [(0,10), (10,0)]. Our results suggest that a substantial share of subjects have convex a convex utility function from (0,10) to (5,5) and a concave one from (5,5) to (10,0) (see Appendix D for the formal derivation). Said differently, the type of social ex post motivation that individuals have do not play an important role in determining their choice in our social lottery tasks (or if they do, they tend to favor the safe option). Likewise, ex ante (or procedural or process) fairness concerns cannot easily help in explaining our results (Trautmann, 2009; Trautmann and Wakker, 2010). Ex ante, both options provide the same expected payoff to both participants. Consequently, procedural inequity aversion preferences should not affect individual choices unless the decision maker has a preference for a stochastic allocation decision over a deterministic one. For instance, one could feel less responsibility for the stochastically implemented uneven distribution than for one that is implemented deterministically. Notice, however, that our subjects had the possibility to choose indifference and let a random mechanism decide. A part of the strong increase in indifference choices that we observe could be related to this, but that makes responsibility avoidance less of a plausible candidate for favoring the risky option.

Overall our data pattern is consistent with two explanations based on (i) a social reference point and (ii) a stronger willingness to being ahead of others compared to being behind. Regarding the first explanation, the other's payoff could play the role of a reference point in prospect theory, an idea developed by Linde and Sonnemans (2012). Gain and loss domains consequently would be defined through the earnings of the other participant, predicting more risk seeking in the loss domain (unfavorable situations) and more risk aversion in the gain domain (favorable situations). When in favorable situations, subjects in our experiment could mainly lose relative to the other participant when choosing the risky option. Instead, by selecting the safe option they can secure their relative social gain. In contrast to that, in unfavorable situations, subjects are not much affected by the prospect of getting (0,10) rather than (9,1) because of diminishing sensitivity (i.e. convex utility) in the (relative) loss domain. Gambling in this case means a large probability of a subjectively small loss but a small probability of a very large gain.

Such reasoning could also help to explain the data by Haisley, Mostafa, and Loewenstein (2008), who show that, when reminded of their low status, low income individuals were more likely to engage in risky purchase such as buying lottery tickets. It is also the reasoning of Schwerter (2013), indicating that decision makers indeed experience social losses and gains in a risk task when exposed to another participant receiving a varying fixed payment.

The second explanation relates to the strength of gloating, i.e. the utility of being ahead of the other. If gloating is more important than envy, i.e. the disutility of being behind, then we should observe the mirror effect in the social situation of what is observed under (individual) risk involving gains and losses because of loss aversion. Loss aversion, for lotteries involving gains and losses, implies a strong avoidance to risk in the individual situation (see Rabin 2000). Bault, Coricelli, and Rustichini (2008) argue that attitudes to gains and losses reverse in a social context: Whereas in its standard version, the theory implies that losses are valued more in absolute terms than gains, it may be that the opposite holds in social contexts; that is, relative gains may be subjectively valued more strongly than relative losses. Attitudes to gains and losses reversing in a social context may also in part explain the discrepancy between, on the one hand, our results and those of Bolton and Ockenfels (2010) and, on the other hand, the findings in Linde and Sonnemans (2012). In the latter paper, the authors did not use social lotteries that gave the decision maker the opportunity to switch relative positions with the receiver (from being behind to being ahead), but at best the possibility to reach the same level of payoffs. If being ahead is what is really prized by subjects, there is little motivation in Linde and Sonnemans's tasks to take risks, since it is impossible to earn more than the matched participant by choosing the risky option. An alternative version of this interpretation is that 'winning' – that is, earning more than the counterpart, independently of the absolute payoff difference – generates a psychological bonus: What is prized is not really the favorable difference between the decision maker and the receiver, but simply whether the decision maker has 'won'. In this case, there is no reason any more to take risks in favorable tasks, and such an explanation is consistent with the general pattern we observe.[9]

Our findings concerning individual heterogeneity are also in line with arguments in favor of a social reference point and a psychological bonus of winning. Those decision makers that reduce the other's payoff in the ring test exhibit the overall pattern more strongly than those that do not. Reducing the other's payoff can only be rationalized by making some form of relative comparisons with the matched participant and by a wish to earn more in relative terms (apart from pure forms of anti-social behavior). It is not surprising that these people then are more strongly affected by the social context. The cluster analysis also helps rationalizing the patterns. Type-1 individuals, who drive the aggregate pattern described above, are not only disproportionally less often categorized as cooperative, they also explicitly state motivations based on a social reference point story. In the subjects' comment section at the end of the experiment, where participants were supposed to elaborate on their motivation behind choices in the social context task, one type-1 subject explained switching to the risky option in the unfavorable cases by stating that "as long as I earned more than the other, I chose the certain amount". Another explicitly wanted to "get a higher payoff than the other". These statements are a specific characteristic of type-1 individuals.

In contrast to the large group of type-1 individuals, there seems to be something else driving behavior of type-2 and type-3 decision makers. Responsibility aversion is one potential explanation. Type-2 individuals are characterized by a switch towards indifference in unfavorable tasks in the social context – and slightly less pronounced in the favorable context. This might in part be driven by responsibility avoidance, which could have also lead to the results in Kircher, Ludwig, and Sandroni (2013). Subjects' comments provide some

[9]Such reasoning can be seen as a social version of aspiration level theory, developed by Diecidue and van de Ven (2008).

indication for such a conclusion: One subject, for example, explicitly stated that she did not want to make the decision herself, but rather leave it to luck.[10] Similar mechanisms could apply to type-3 subjects. Choosing the risky option more often in the favorable social decisions implies that in the end it is the random draw that establishes an uneven distribution and not the participant's choice directly. These subjects also more often state that they want to implement the probabilistic dictator game, instead of the deterministic version in part 2 of the experiment. For type-3 subjects procedural fairness concerns are another potential explanation. Even though procedural concerns (Trautmann, 2009; Bolton, Brandts, and Ockenfels, 2005; Saito, 2013) should play no role with equivalent expected outcomes for both options, it might still be that a subset of individuals perceives the risky lottery to be fairer in the favorable range. Giving a chance (even if small) to get the entire amount could be considered as more appropriate than implementing for sure a very unequal payoff structure. Participants' comments again are in line with both lines of reasoning. One subject explicitly stated that he or she chose out of fairness concerns and another said that probabilities reduce the responsibility and feeling of guilt. These motivations stand in stark contrast to type-1 individuals.

Is it possible that our results are driven partly by the experimental design? The order of the experimental parts (in particular between part 1 and part 4) was not randomized or varied, such that any within-subject treatment effects could potentially stem from this task sequence. However, it does not seem very plausible that the order of treatments would explain the change towards more risky choices in the socially unfavorable domain. First, uncertainty about the experimental payment as well as individual lotteries were only resolved at the very end of the experiment and no feedback of any sort was provided beforehand. Hence, there was no room for any type of income effects. Second, and most importantly, it is hard to see how order effects could explain the asymmetric effect observed from the social context in part 1 to the individual context in part 4. To stand as an explanation, order effects should have impacted the choices made by subject on T1i-T4i but not on T5i-T9i. Even more so, the documented differences for subsamples of our subjects and the contents of the comment section are very difficult to interpret based on the order of tasks.

5 Conclusion

Our data suggest that risk taking is influenced by the relative social situation of the decision maker: compared to equivalent situations without a social context, more risk is chosen in

[10]It also seems that these subjects are genuinely more altruistic: in both the deterministic and the probabilistic dictator game, on average, they give the most to the recipient.

unfavorable situations, while similar risk taking is observed for favorable social situations. A large share of our decision makers exhibits this pattern in a very pronounced way.

This observed behavioral pattern cannot be straightforwardly explained by extensions of models of outcome-based social preferences for stochastic environments. The overall asymmetric pattern rather points towards the importance of social reference points and/or a utility from winning, i.e. leaving the experiment with more money than the matched participant (the only available reference person).

Our experimental results suggest that the role of social context may be critical also in understanding organizational and financial risk taking. When subjects directly compete against each other (e.g., over resources or power), even without any explicit competition incentives such as tournament prizes, they might take excessive risks that they would not take absent information on outcomes of others. Information provision or the way this information is presented may affect managers and investors alike, and policy makers should take them into account when designing rules and regulations.

In a broader context, our study provides another piece of evidence for the idea that risk taking is strongly affected by the social environment in which decisions take place. Future studies could test specific theoretical models of excessive risk taking that embed the risky situation into a social environment. Further, the social situation could be varied in different dimensions (such as the level of competition, the size of the references group, the presentation of information, etc.), not only along the outcome dimension. We see our results as a first steps towards a better understanding of the influence of social comparisons in risk taking.

References

Adam, M. T., E. B. Kroll, and T. Teubner (2014): “A note on coupled lotteries”, Economics Letters, 124(1), 96–99.

Bault, N., G. Coricelli, and A. Rustichini (2008): “Interdependent utilities: How social ranking affects choice behavior,” PLoS ONE, 3(10), e3477.

Bolton, G., and A. Ockenfels (2010): “Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States: Comment,” American Economic Review, 100(1), 628–633.

Bolton, G. E., J. Brandts, and A. Ockenfels (2005): “Fair procedures: evidence from games involving lotteries,” Economic Journal, 115(506), 1054–1076.

Bolton, G. E., and A. Ockenfels (2000): “ERC: a theory of equity, reciprocity and competition,” American Economic Review, 90(1), 166–193.

Bolton, G. E., Ockenfels, A., and J. Stauf (2015): “Social responsibility promotes conservative risk behavior”, European Economic Review, 74, 109–127.

Bolton, G. E., R. Zwick, and E. Katok (1998): “Dictator game giving: rules of fairness versus acts of kindness,” International Journal of Game Theory, 27(2), 269–299.

Brock, M., A. Lange, E.Y. Ozbay (2013): “Dictating the risk – Experimental evidence on giving in risky environments,” American Economic Review, 103(1), 415–437.

Brandts, J., and Charness, G. (2011): “The strategy versus the direct-response method: a first survey of experimental comparisons,” Experimental Economics, 14(3), 375–398.

Brennan, G., W. Güth, L. Gonzalez, and M. V. Levati (2008): “Attitudes toward private and collective risks in individual and strategic choice situations,” Journal of Economic Behavior and Organization, 67(1), 253–262.

Brosig, J. (2002): “Identifying cooperative behavior: Some experimental results in a prisoner’s dilemma game,” Journal of Economic Behavior and Organization, 47(3), 275–290.

Bursztyn, L., F. Ederer, B. Ferman, and N. Yuchtman (2014): “Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions,” Econometrica, 82(4), 1273–1301.

Cai, J., A. de Janry, and E. Sadoulet (2015): “Social networks and the decision to insure,” American Economic Journal: Applied Economics, 7(2), 81–108.

Chakravarty, S., Harrison, G. W., Haruvy, E. E., and E.E. Rutström (2011): “Are you risk averse over other people's money?” Southern Economic Journal, 77(4), 901–913.

Charness, G., and M. Rabin (2002): “Understanding social preferences with simple tests,” Quarterly Journal of Economics 117(3), 817–869.

Cooper, D., and M. Rege (2011): “Misery loves company: Social regret and social interaction effects in choices under risk and uncertainty,” Games and Economic Behavior, 73(1), 91–110.

Diecidue, E., and J. van de Ven (2008): “Aspiration level, probability of success and failure and expected utility,” International Economic Review, 49(2), 683–700.

Dijk, O., M. Holmen, and M. Kirchler (2014): “Rank matters – The impact of social competition on portfolio choice,” European Economic Review, 66, 97–110.

Fafchamps, M., B. Kebede, and D.J. Zizzo (2015): “Keep up with the winners: Experimental evidence on risk taking, asset integration, and peer effects,” European Economic Review, 79, 59–79.

Fehr, E., and K. M. Schmidt (1999): “A theory of fairness, competition, and cooperation,” Quarterly Journal of Economics, 114(3), 817–868.

Fischbacher, U. (2007): “z-Tree: Zurich Toolbox for ready-made Eeonomic experiments”, Experimental Economics 10(2), 171–178.

Forsythe, R., J. Horowitz, N. Savin, and M. Sefton (1994): “Fairness in simple bargaining experiments,” Games and Economic behavior, 6(3), 347–369.

Frank, R. H. (2005): “Positional externalities cause large and preventable welfare losses,” American Economic Review, 95(2), 137–141.

Friedl, A, K. L. de Miranda, and U. Schmidt (2014): “Insurance demand and social comparison: An experimental analysis,” Journal of Risk and Uncertainty, 48(2), 97–109.

Greiner, B. (2015): “Subject pool recruitment procedures: Organizing experiments with ORSEE,” Journal of the Economic Science Association, 1(1), 114–125.

Haisley, E., R. Mostafa, and G. Loewenstein (2008): “Subjective relative income and lottery ticket purchases,” Journal of Behavioral Decision Making, 21(3), 283–295.

Holt, C., and S. Laury (2002): “Risk aversion and incentive effects,” American Economic Review, 92(5), 1644–1655.

Kircher, P., Ludwig, S., and A. Sandroni (2013): “On the difference between social and private goods”, The BE Journal of Theoretical Economics, 13(1).

Krawczyk, M., and F. Le Lec (2010): “’Give me a chance!’ An experiment in social decision under risk,” Experimental Economics, 13(4), 500–511.

Krawczyk, M. W, S. T. Trautmann, and G. van de Kuilen (2016) “Catastrophic risk: Understanding insurance decisions,” Theory and Decision

Lahno, A. M., and M. Serra-Garcia (2015): “Peer effects in risk taking: Envy or conformity?" Journal of Risk and Uncertainty, 50(1), 75–93.

Linde, J., and J. Sonnemans (2012): “Social comparisons and risky choices,” Journal of Risk and Uncertainty, 44(1), 45–72.

Müller, S., H.A. Rau (2016): ”The relation of risk attitudes and other-regarding preferences: A within-subjects analysis,” European Economic Review, 85(1), 1–7.

Offerman, T., J. Sonnemans, and A. Schram (1996): “Value orientations, expectations and voluntary contributions in public goods”, Economic Journal, 106(437), 817–845.

Rabin, M. (2000): “Risk aversion and expected-utility theory: A calibration theorem,“ Econometrica, 68(5), 1281–1292.

Saito, K. (2013): “Social preferences under risk: equality of opportunity versus equality of outcome”, American Economic Review, 103(7), 3084–3101.

Selten, R. (1967): „Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments,“ in: H. Sauermann (Ed.), Beiträge zur experimentellen Wirtschaftsforschung (pp. 136–168). Tübingen: Mohr

Schwerter, F. (2013): “Social reference points and risk taking”, Bonn Econ Discussion Papers, 11/2013.

Sutter, M., Haigner, S., and M. G. Kocher (2010): “Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations”, Review of Economic Studies, 77(4), 1540-1566.

Trautmann, S. (2009): “A tractable model of process fairness under risk,” Journal of Economic Psychology, 30(5), 803–813.

Trautmann, S. T., and F. M. Vieider (2012): “Social influences on risk attitudes: Applications in economics,” in: S. Roeser, R. Hillerbrand, P. Sand, and M. Peterson (Eds.), Handbook of risk theory. Epistemology, decision theory, ethics, and social implications of risk (pp. 575–600). Amsterdam: Springer.

Trautmann, S. T., and P. Wakker (2010): “Process fairness and dynamic consistency,” Economics Letters, 109(3), 187–189.

Tversky, A., and D. Kahneman (1992): “Advances in prospect theory: Cumulative representation of uncertainty,” Journal of Risk and Uncertainty, 5(4), 297–323.

Tversky, A., and I. Simonson (1993): “Context-dependent preferences,” Management Science, 39(10), 1179–1189.

Van Dijk, F., J. Sonnemans, and F. van Winden (2002): “Social ties in a public good experiment,” Journal of Public Economics, 85(2), 275–299.

Van Lange, P., E. De Bruin, W. Otten, and J. A. Joireman (1997): “Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence,” Journal of Personality and Social Psychology, 73(4), 733–746.

Vieider, F., Villegas-Palacio, C., Martinsson, P., and M. Mejía (2015): "Risk taking for oneself and others: A structural model approach", Economic Inquiry, 54(2), 879–894.

Wakker, P. (2010): Prospect theory: For risk and ambiguity. Cambridge University Press.

Appendix:

Sharing or gambling?
On risk attitudes in social contexts

Stefan Grimm, Martin G. Kocher, Michal Krawczyk, and Fabrice Le Lec

Appendix A: Social Orientation Questionnaire

The design and description of the social orientation questionnaire is based on Sutter et al. (2010). The questionnaire consists of 24 choices (see Table A1) between two own-other payoff allocations in constant and anonymous pairs of subjects. The two options in all 24 choices assign an amount of money to the subject herself (x) and a certain amount to the matched player (y). Subjects knew that everybody received the same questionnaire, and there was no feedback given about the matched player's choices during filling in the questionnaire. For all payoff allocation $r^2 = 15^2 = x^2 + y^2$ holds, such that each option represents a vector in a Cartesian plane lying on a circle with radius r centered at the origin.

Table A.1: 24 choices for own-other payoff allocations

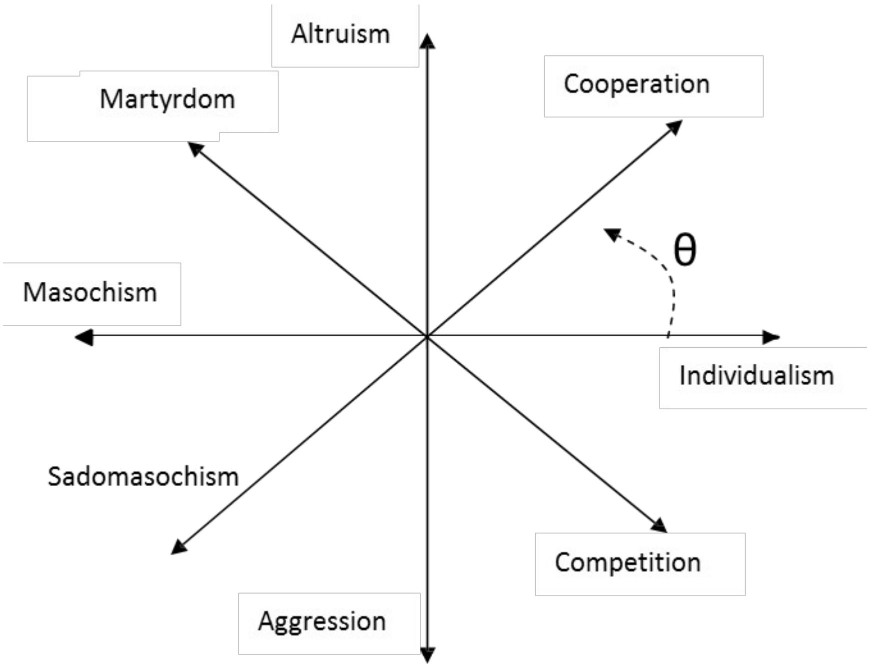
Question number	self (x)	other (y)	self (x)	other (y)
1	15	0	14.5	-3.9
2	13	7.5	14.5	3.9
3	7.5	-13	3.9	-14.5
4	-13	-7.5	-14.5	-3.9
5	-7.5	13	-3.9	14.5
6	-10.6	-10.6	-13	-7.5
7	3.9	14.5	7.5	13
8	-14.5	-3.9	-15	0
9	10.6	10.6	13	7.5
10	14.5	-3.9	13	-7.5
11	3.9	-14.5	0	-15
12	14.5	3.9	15	0
13	7.5	13	10.6	10.6
14	-14.5	3.9	-13	7.5
15	0	-15	-3.9	-14.5
16	-10.6	10.6	-7.5	13
17	-3.9	-14.5	-7.5	-13
18	13	-7.5	10.6	10.6
19	0	15	3.9	14.5
20	-15	0	-14.5	3.9
21	-7.5	-13	-10.6	-10.6
22	-13	7.5	-10.6	10.6
23	-3.9	14.5	0	15
24	10.6	-10.6	7.5	-13

By adding up x and y of all 24 choices, the motivational vector M can be constructed, yielding an angle θ of vector M (x on the x-axis and y on the y-axis, see Figure A1). With this angle subjects can then be classified into one of the following eight categories based on their social motivation: individualism, altruism, cooperation, competition, martyrdom, masochism, sadomasochism, and aggression.

The classification of subjects can be seen in Figure A1:

Subjects with a θ between 0° and 22.5° or 337.5° and 0° are classified as "individualistic"; subjects with an angle between 22.5° and 67.5° as cooperative. More infrequent types are altruism (between 67.5° and 112.5°), martyrdom (between 112.5° and 157.5°), masochism (between 157.5° and 202.5°), sadomasochism (between 202.5° and 247.5°), aggression (between 247.5° and 292.5°), and competitive (between 292.5° and 337.5°).

Figure A1: Vectors defining the basic social motivation



Subjects' earnings in Part 5 were given by the sum of choices made by the subject herself (sum of own x) and by the sum of choices made by the matched player (sum of other's y).

Appendix B: Instructions for all parts

B1: Instructions before start of the experiment

Please do not talk to other participants anymore and remain silent throughout the entire experiment. For simplicity we will use masculine terms in the following. These will refer to both male as well as female participants.

General information regarding procedures

The experiment aims at investigating decision making. You can earn money which will be paid out at the end of the experiment in private and in cash.

The entire experiment will last around 45 minutes. It consists of two completely independent parts in which you have to make decisions. The first part is divided into 4 blocks. In block 1 and 2 your earnings can depend on the decisions of another participant, who will be randomly assigned to you. In block 3 and 4 your earnings will be solely determined by your own decisions. In the second part of the experiment your earnings will again depend on your own decisions and the decisions of another participant. For this purpose, you will again be randomly assigned to another person. We will not use the same pairs as in part 1, but make new random pairs. After part 2 we will ask you to answer a general questionnaire.

While you make your decisions a clock will run down in the upper right corner of the screen. This provides guidance for how much time you can use for your decisions. If the clock is down to zero, please come to a decision. However, you can still complete your decisions with the clock down to zero.

If you still have questions after the instructions or during the experiment, please raise your hand or press the red button on your keyboard. One of the experimenters will then come to your seat and answer your question in private. If the question is of interest to all participants, we will repeat the question and answer publicly.

Anonymity

None of the other participants will be able to reconstruct your decisions in the experiment. Moreover, the data from the experiment will be analyzed anonymously. For accounting reasons you have to sign a receipt for your earnings at the end of the experiment. Your name cannot be linked to your decisions in the experiment.

The Experiment – Part 1

Block 1 and 2

In block 1 and 2 you will be randomly assigned a role: active or passive participant. Your decisions will only be relevant for your earnings and the earnings of your matched participant if you are active participant. Decisions of the passive participant have no impact on earnings. However, your role will only be revealed at the end of the experiment. For that reason please assume for these decisions that you are the active participant. Otherwise decisions might be implemented that you want to avoid.

In block 1 you will make decisions for 9 scenarios. In block 2 there is one scenario.

Block 3 and 4

Assigned roles are irrelevant in block 3 and 4. Your potential earnings only depend on your own decisions. In block 3 you will make decisions for 6 scenarios. In block 4 there will be 9 scenarios.

Payment

For all decisions in part 1 all potential earnings will be stated in Euro. Since you will make many different decisions in these blocks, the computer will randomly draw one single decision at the end of the experiment, which will be relevant for your earnings. The procedure is as follows: Only one out of the 4 blocks is relevant. This relevant block will be randomly determined by the computer. Within this block, one specific decision (scenario) will again be determined randomly to be payoff relevant. If the chosen decision involves uncertain payments (probabilities) the computer will again determine randomly which probabilistic event will be realized. Further, the computer will randomly assign roles of active and passive participants for all randomly matched pairs of participants. This role will only be relevant for your earnings if block 1 or 2 is relevant for your earnings. Let us assume you are assigned the active role and block 1 was determined to be payoff relevant. Based on the randomly chosen scenario you and your matched participant will receive earnings based on your decision in this scenario. If you were assigned the passive role and block 1 was determined relevant, you will receive earnings based on the decision of the matched participant in the respective scenario. Every decision in blocks 1 to 4 can be relevant for your earnings. Choose your answers carefully.

The Experiment – Part 2

Upon finishing part 1 of the experiment you will start with part 2. This part is completely independent of part 1. Here, you will again be randomly assigned to one other participant. The pairs, however, will be randomly drawn anew. After part 1 you will be provided with more information on part 2.

Your total earnings in today's experiment hence will consist of the described earning from part 1 and the earnings from part 2. In addition, you receive 4€ for showing up on time.

B1: Instructions before part 2 of the experiment (distributed and read out after part 1)

In part 2 you will again be randomly assigned to one other participant. You will make multiple decisions which affect your own payoff as well as the payoff of your matched participant. There will be no roles in this part of the experiment. That is, both your decisions as well as the decisions of your matched participant will be implemented. Both you and the other will remain anonymous.

You will make 24 decisions with 2 options each (Option A and Option B). Each option assigns a certain amount of the experimental currency "Taler" to your account ("Your Payoff") and a certain amount to the account of your matched participant ("Other's Payoff").

An example:

	Option A	Option B
Your Payoff	15,00	14,50
Other's Payoff	0,00	-3,90

If you choose option A, 15 Taler will be transferred to your account and zero Taler to the account of the other participant. If you choose option B you receive 14,50 Taler and the other participant will reveice -3,90 Taler (3,90 Taler will be deducted from his account).

Your total earnings of part 2 will be the sum of „Your Payoff" of your 24 decisions. The payoff for the other participant based on your decisions is the sum of "Other's Payoff". That is, every single decision in this part of the experiment will affect your own and the other's earnings.

Your matched participant makes decisions for exactly the same choices. Hence, in addition to the sum of "Your Payoff" of your own decisions you will receive the sum of "Other's Payoff" of the decisions of your matched participant. Similarly, next to the earnings from your decisions, the other participant receives a payment based on his own decisions, too.

The resulting total earnings in “Taler“ will then be converted to Euros and represent your earnings from part 2 of the experiment. The exchange rate is: 10 Taler = 1.50 Euro.

During the experiment you will not receive feedback on any decision of your matched participant. Only at the end of the experiment will you see the sums of “Your Payoff”, “Other’s Payoff” and “Other’s Payoff” of your matched participant, as well as your total earnings from part 2.

Potential negative earnings in single parts of the experiment will be offset by earnings from the other part and the 4€ received for showing up on time such that total earnings of the experiment will always be positive.

If you have any questions please raise your hand now. We will then come to your seat and answer your questions in private.

Appendix C: Additional results for the cluster analysis

The categorization of subjects can help in explaining the aggregate pattern: Remember that for type-1 subjects the increase in risk taking in the unfavorable range when going from the individual to the social context is very pronounced, while there is a reduction in risk taking in the favorable range. Type-3 subjects increase overall risk-seeking behavior and reduce risk taking in the unfavorable range, when they are in the social context. Type 2 individuals are characterized by their leap towards indifference in the social lotteries – especially in the unfavorable domain. Statistical tests confirm this first impression from the cluster analysis.

For type-1 individuals, risk taking very strongly and significantly increases in the unfavorable range ($p < 0.001$ for Stuart-Maxwell test) for all lotteries when going from the individual context to the social context. The reverse is true for the favorable situations. Here, type-1 individuals even reduce risk taking in the social context. For lotteries T7(i) to T9(i) this difference is significant (at the 5%-level using the McNemar test[11]). For type-3 subjects most comparisons do not result in significant differences, most probably due to a lack of statistical power, given the much smaller number of decision makers than in the type-1 cluster. There is at least some tentative evidence that risk taking increases from T9i to T9 ($p < 0.1$ for Stuart-Maxwell test) and that risk taking decreases in the unfavorable range at least from T3i to T3 ($p < 0.01$ for McNemar's test), in contrast to the aggregate pattern. As indicated before, the change in pure risk-taking behavior for type-2 individuals is less clear cut due to the large number of indifference choices. This trend towards more indifference however is clearly significant (mostly at the 1%-level) in a McNemar test, grouping indifference against risky and safe choices for all lotteries except T5 vs T5i and T9 vs. T9i.

[11] Since there are no indifference choices for type-1 individuals, we can only look at McNemar's test. This also holds for type-3 individuals in T3 and T3i.

Appendix D: Theoretical model

The decision subjects have to make is to choose (for all tasks) between a deterministic outcome (safe) L_s =(1*;* ($x\pi$,(1 − x) π)) and a lottery L_r =(x, (π, 0)*;* 1 – x, (0, π), with π the pie and x the probability (or the share of pie). The function V represents the individual's preferences over lotteries.

D1: Ex post social preferences

We assume here that individuals have social concerns that apply only to final allocations (i.e. the probabilistic distribution of the outcomes does not play a role as for ex ante social preferences, see below).

For a given individual, we set that her preferences over final allocations are represented by u: (x_1; x_2) → u(x_1; x_2), with x_1 being the payoff of the decision maker and x_2 the payoff of the recipient.

Under the usual assumption of expected utility maximization we have for any x:

$$V(L_s) = u(x\pi, (1-x)\pi)$$

and

$$V(L_r) = xu(\pi,0) + (1-x)u(0;\pi)$$

First, note that $(x\pi, (1-x)\pi)$ is a convex combination of (π, 0) and (0, π):

$$(x\pi, (1-x)\pi) = (x\pi + (1-x) \times 0; (1-x)\pi + x \times 0)$$

$$= x.(\pi,0) + (1-x).(0,\pi)$$

We observe experimentally that a significant share of individuals $V(L_s) > V(L_r)$ for large x (greater than or equal to $\frac{1}{2}$), but that for small x, the opposite holds $V(L_s) > V(L_r)$. This implies that u cannot be concave on (0, 1) since for x small:

$$u(x.(\pi, 0) + (1-x).(0,\pi)) < xu(\pi,0) + (1-x)u(0,\pi)$$

Nor can it be convex, since for x≥ $\frac{1}{2}$

$$u(x.(\pi, 0) + (1-x).(0,\pi)) > xu(\pi,0) + (1-x)u(0,\pi)$$

And for the same reasons it cannot be linear either.

Consider the function $f(x) = u(x.(\pi, 0) + (1-x).(0,\pi))$. It is continuous (if u is); hence by the intermediate value theorem there exists $\bar{x} \in (0,1)$ such that

$$f(\bar{x}) = u(\bar{x}.(\pi, 0) + (1-\bar{x}).(0,\pi)) = \bar{x}u(\pi,0) + (1-\bar{x})u(0,\pi)$$

Assume for the sake of simplicity that   is unique.[12]

For $0 < x < \bar{x}$, $u(x.(\pi, 0) + (1 - x).(0, \pi)) < xu(\pi,0) + (1 - x)u(0, \pi)$

Or denoting $\pi_1 = (\pi, 0)$ and $\pi_2 = (0, \pi)$ for the sake of conciseness:

For $0 < x < \bar{x}$, $u(x.\pi_1 + (1 - x).\pi_2) < xu(\pi_1) + (1 - x)u(\pi_2)$

Given that any point on the line $[\bar{x}\,\pi_1 + (1 - \bar{x})\,\pi_2; \pi_1]$ is such that $0 < x < \bar{x}$, it ensures that for any convex combination $\alpha.(\bar{x}\,\pi_1 + (1 - \bar{x})\,\pi_2) + (1 - \alpha).\pi_1$ that:

$$u[\alpha.(\bar{x}.\pi_1 + (1 - \bar{x}).\pi_2) + (1 - \alpha).\pi_1] < \alpha\, u(x.\pi_1 + (1 - x)\pi_2)) + (1 - \alpha)u(\pi_1)$$

Hence u is convex on $[\bar{x}, \pi_1 + (1 - \bar{x}).\pi_2, \pi_1]$.

For $\bar{x} < x < 1$, $u(x.(\pi, 0) + (1 - x).(0, \pi)) > xu(\pi,0) + (1 - x)u(0, \pi)$

And the same reasoning as for $x < \bar{x}$ yields:

$$u[\alpha.(\bar{x}.\pi_1 + (1 - \bar{x}).\pi_2) + (1 - \alpha).\pi_2] < \alpha\, u(x.\pi_1 + (1 - x)\pi_2)) + (1 - \alpha)u(\pi_2)$$

Hence u is concave on $[\pi_2; \bar{x}, \pi_1 + (1 - \bar{x}).\pi_2]$.

To conclude, for any ex post social preference, there exists a share $\bar{x}$ such that the utility is concave for $x > \bar{x}$ and is convex for $x < \bar{x}$. Said differently, independently of the model of social preference under consideration (altruism, inequity aversion, maximin/efficiency, spitefulness, selfishness), a change in the curvature (from convexity to concavity) is required to observe a change of choice between the safe option and the social lottery.

The observation that a significant share of individuals chooses the social lottery for x small (when they chose the safe individual option for the same x) implies a change in the curvature of their utility function. This change of behavior is independent of their pro-social or anti-social motivations.

This also applies to a self-interested utility maximizer: a risk averse individual ($u'' < 0$) always chooses the safe option, and a risk seeking individual ($u'' > 0$) always chooses the risky one. Note here that the variance of the lottery in terms of individual payoff (which increases as x gets further away from $\frac{1}{2}$) does not play any role.

[12]The multiplicity of x does not change the argumentation, it suffices to take the minimum and maximum of those multiple $\bar{x}$ to obtain the same result.

D2: Ex ante social preferences

Under the assumption that the ex ante term of social preferences is based on expected payoff, the ex ante fairness of the safe option and the social lottery is constant for all decisions. Hence it cannot play a role.

When considering a mix of ex ante and ex post social consideration, the same is true and the only relevant driver for the switch from a risky to a safe social lottery can be, as in the case of ex post social preferences, the curvature of the utility function. Once again, independently of the motives under consideration.

Appendix E: Evidence from a classroom experiment

E1: Experimental Design

The experiment was divided into three parts: a series of risky choices in the social context (similar to part 1 of our lab experiment), two dictator games (see our part 2), and a series of individual decisions under risk (see our part 4). The first part, as before, is the core of the study and aims at measuring how risk attitude is affected by social contexts, whereas the latter two again provide a control for social concerns and risk attitude in a purely individual context.

In the first part of the experiment, subjects faced tasks where fifty euros had to be allocated (either deterministically or randomly) between the decision maker and the receiver. This is equivalent to our design described in the paper. However, in the classroom experiment, 50 Euros instead of 10 Euros had to be divided with expected payoffs for the decision maker ranging from 5 Euros ("T5") to 45 Euros ("T45") in steps of 5 Euros. Order effects were controlled for by presenting the choices in ascending orders to half the subjects and in descending order to the other half.

Parts 2 and 3 aim at measuring social preferences in a risk-free environment and risk preferences in an individual setting (without social context). Part 2 was equivalent to part 2 of our lab experiment in that subjects had to play the two dictator games. Here again, 50 Euros were to be distributed. Part 3 consisted of a series of nine binary decisions under risk. The first three were a truncated and adapted Holt and Laury (2002) procedure to estimate subjects' risk attitudes with stakes comparable to the one used in the main part of our experiment (see first half of part 3 in the lab experiment). The next three tasks were aimed at measuring loss aversion (second half of part 3 above), and the last three tasks were risky binary choices that were exactly equivalent to three of the tasks in part 1, but without any social component (part 4 above). Hence, in contrast to the lab experiment, we only have 3 equivalent individual tasks to compare to risk taking in the social context. One of these tasks was in the unfavorable range (expected payoff for the decision maker of 15 Euros, "T15i"), one was in the favorable range (expected payoff of 35 Euros, "T35i") and one was payoff equivalent to the equal spilt task (expected payoff of 25 Euros, "T25i"). Finally, as in the lab experiment, subjects were asked to provide some socio-demographic characteristics.

E2: Experimental Procedures

The design described above was implemented as a classroom experiment with 82 undergraduates in economics at the University of Munich. Their role – either decision maker or receiver - was only determined after the experiment. Decision sheets and instructions were first distributed for parts 1 and 2 together, and upon finishing, also for part 3. Subjects knew that there were three parts of the experiments already at the beginning. For payment, four randomly selected decision makers were matched with four randomly selected receivers. For each pair one of the ‘social’ tasks (parts one and two, including the question regarding their preference for the regular dictator game or the probabilistic one) was randomly selected for payment. In addition, four participants were randomly picked for payment in the individual lottery part, where one task was once again randomly picked to be implemented. Payments were provided individually and confidentially. All design details and the procedural details were common knowledge among participants.

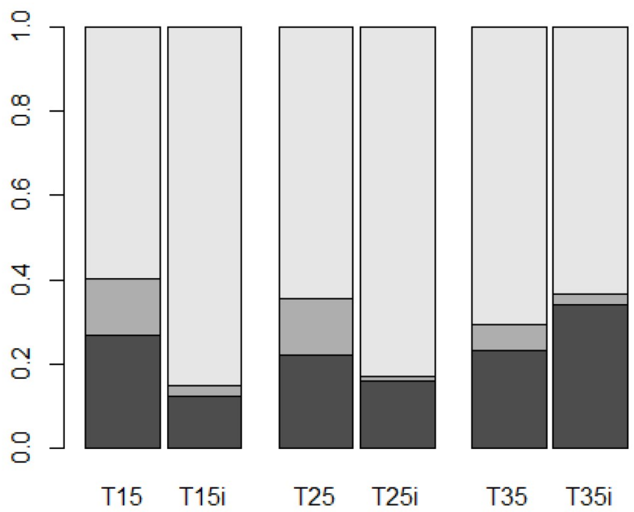
These design and procedure details result in three major differences between the classroom and lab experiment, apart from the obvious differences in the setting: First and most importantly, in the classroom experiment we did not collect data on all choices in the individual context. Second, we only paid a small fraction of subjects while the amounts to be shared were much higher. Third, in the lab experiment we included the ring test to measure social value orientation to have a better individual control for social preferences.

E3: Results

In the social decisions under risk (part 1), supporting the results from our lab experiment, subjects clearly become more risk taking the more unfavorable the tasks become in the unfavorable range. The proportion of subjects taking the risky option in T30 is significantly lower than in T45 ($p < .001$ for the Maxwell-Suart’s marginal homogeneity test), although conventional levels of significance are not reached with T40 and T35 vs. T45. In contrast to the results in the main part of the paper (Figure 1), however, risk taking in the social context is more U-shaped. That is, risk taking also increases in the favorable range towards the extremely favorable decision T45. Nevertheless also here, the U-shaped pattern seems to be asymmetric: the number of risky choices appears higher in the case of unfavorable inequity for the decision maker than in the case of favorable inequity. Leaving the case of the equal split aside for the moment, all comparisons between tasks corresponding to sure payoffs adding up to 50 (T5 vs. T45, T10 vs. T40, T15 vs. T35, and T20 vs. T30) suggest that the risky option is relatively more appealing when the sure option implies unfavorable inequity: The differences are significant according to Maxwell-Stuart’s tests at the 5%-level. By the same token, comparing the number of times decision makers have chosen the risky option in the four favorable situations against the same number in the four unfavorable situations yields a significant difference (using a two-sided Wilcoxon signed ranks test for matched observations; $p = 0.02$).

Core to our analysis, however, is the comparison between otherwise identical decisions in the social and individual context. Comparing the three individual tasks from part 3 of the experiment (T15i, T25i, T35i) with the social context counterparts T15, T25 and T35 yields very much similar results as in the lab experiment (see Figure D1). In T15 vs. T15i, where the social situation is unfavorable to the decision maker, individuals take significantly more risk than in the equivalent individual lottery and this difference is strongly significant ($p < .001$ for Maxwell-Stuart's test). However, in case of a favorable social context (T35 vs. T35i) the difference is not significant at conventional significance levels. The effect also points in the opposite direction: in the favorable range of the social context decision makers, if anything, seem to reduce risk taking compared to the equivalent individual decision. That is, as in the lab experiment, decision makers seem to be affected by social context when making a risky decision, but not in a homogeneous way: they take more risk when the situation is unfavorable or equal, but less when it is favorable to them. Different to the results in the main part of the paper, we can also observe a difference between the contexts for the equal situation (T25 vs. T25i), where subjects take more risk in the social lottery than in the individual one ($p < 0.01$ for Maxwell Stuart).

Figure D1: Choices in social versus individual context (Risky: black; Safe: light grey)



As in the main body of the paper, we can also look at individual patterns and heterogeneity between participants. If we split the sample based on the median offer in the dictator game in part 2 of the classroom experiment, we can see that the difference between the contexts in the unfavorable range seems to be driven by selfish subjects only ($p < 0.01$ for Maxwell Stuart and not significant difference for pro-social subjects). We also ran a k-medians cluster analysis on all social lotteries dividing the subjects into three clusters. This leads to the following characterization here: 20 individuals (type 1) exhibit a very dichotomous pattern of risk attitude in the social context (strongly risk-seeking in the unfavorable case, and risk-averse in the favorable one), 41 subjects (type 2) are rather overall risk-averse, and 21 individuals (type 3) show a relatively stable attitude towards risk, except in the case of high probabilities of winning (T35, T40, T45), where they strongly increase risk taking. Comparing the effect of the decision context for the different types, we can draw similar conclusions as in the main body of the paper: Type 1 subjects seem to be most strongly affected by social context ($p < 0.001$ for Maxwell Stuart for T15 vs. T15i and $p < 0.05$ for T35 vs. T35i), increasing risk taking in the unfavorable range and decreasing risk taking in the favorable range. The same pattern also holds for type 2 subjects, even though the differences are less pronounced (not significant in the favorable range). There is no significant effect for type 3 subjects. That means that also for these subjects in this experimental setting, a clear majority of subjects (roughly three quarters) exhibit the pattern observed in the laboratory. They tend to take more risks in unfavorable situations than in equivalent individual contexts, and they – if anything – seem to be more risk-averse in socially favorable situations.

DOCUMENT DE TRAVAIL 1 : Fabrice Le Lec, Astrid Matthey et Ondrej Rydval (2019),“Punishing the weakest-link – Voluntary Sanctions and Efficient Coordination in the Minimum Effort Game”, version actualisée de “Efficiency and Punishment in a Coordination Game”, CERGE-EI Working Paper 526.

NB: La version présentée est la dernière version, remaniée et qui devrait être soumise sous peu, et non le document de travail initial, qui a existé sous un autre titre.

Punishing the weakest-link

Voluntary Sanctions and Efficient Coordination in the Minimum Effort Game

Fabrice Le Lec, Ondřej Rydval and Astrid Matthey*

January 16, 2018

Abstract: Using a laboratory experiment, we examine whether voluntary monetary sanctions induce subjects to coordinate more efficiently in a repeated minimum effort coordination game. While most groups first experience Pareto inferior coordination in a baseline treatment, the level of effort increases substantially once *ex post* sanctioning opportunities are introduced, that is, when one can assign costly punishment points to other group members in order to reduce their payoffs. We compare the effect of this voluntary punishment possibility with the effect of *ex post* costless communication: in contrast to the punishment treatment, the latter only temporarily increases efforts and fails to bring the players to higher payoff equilibria permanently. This suggests that decentralized sanctions can play an important role as a coordination device in Pareto-ranked coordination settings, such as teamwork in firms and other organizational contexts. It also suggests that the motivations behind voluntary sanctions (e.g. reciprocity) may be more general than usually put forth in the literature on cooperation games.
Keywords: coordination, minimum effort, order-statistic game, punishment, sanction, weakest link
JEL classification: C72, C91, D01, D03

*Le Lec: Université Catholique de Lille, Anthropo-Lab, 60bis, rue du Port, Lille, France and Université Paris-1 Panthéon Sorbonne, Centre d'Economie de la Sorbonne UMR CNRS 8174, Maison des Sciences Economiques - 106-112 Boulevard de l'Hôpital - 75647 Paris cedex 13 - France, fabrice.le-lec@univ-paris1.fr (corresponding author)
Matthey: German Federal Environment Agency, and Max Planck Institute of Economics, Jena, Germany, e-mail: astrid.matthey@uba.de
Rydval: CERGE-EI, a joint workplace of Charles University in Prague and the Economics Institute of the Academy of Sciences of the Czech Republic, Politickych veznu 7, 111 21 Prague, Czech Republic, e-mail: ondrej.rydval@cerge-ei.cz. While working on this paper, Rydval was supported by the Czech Science Foundation grant P402/12/G130.

1 Introduction

Coordination issues arise routinely in economic circumstances. In microeconomics, the ubiquity of coordination problems within firms, organizations and even industrial branches has been widely acknowledged (e.g., Becker and Murphy, 1992). One prominent game theoretic description of coordination issues is given by the minimum effort game, also known as the weakest-link game: a group member's payoff depends on her own effort (i.e., action) as well as the minimum effort of the group. The higher the minimum effort, the higher every member's payoff. In contrast to cooperation games, any common effort level chosen by all group members is an equilibrium, so it is in no one's interest to deviate upward or downward from the common effort. Hence reaching the payoff-dominant equilibrium is a problem of coordination of actions rather than the result of a conflict between individual interest and social good. Many economic and organizational contexts feature situations where agents must coordinate on a common action with the group's success depending on the least favorable action of a team member. Among canonical examples are teams of assembly-line workers whose overall productivity depends on the least productive member, teams of construction workers whose ability to proceed to the next construction step hinges on every member having completed a task, law firm cases that are only as sound as their weakest part, and even collaboration on scientific projects. Camerer and Knez (1994) have underlined many ways weakest-link coordination games can account for within-firm interactions.

Numerous experimental studies have been carried out to determine whether agents are able to collectively coordinate on Pareto dominant outcomes. For minimum effort games in particular, ample evidence from various contexts has documented a widespread failure to coordinate on the payoff maximizing equilibria on a long-term basis, starting with the seminal work of Van Huyck, Battalio, and Beil (1990). Various features to Pareto-improve outcomes have been tested experimentally and several of those were found to partly achieve this goal. Among these are smaller groups (Van Huyck, Battalio, and Beil, 1990); higher incentives in the form of exogenous bonuses (Brandts and Cooper, 2007) or lower effort costs (Goeree and Holt, 2005); more refined action space (Van Huyck, Battalio, and Rankin, 2007); communication opportunities including pre-play cheap talk (Blume and Ortmann, 2007); *ex post* disapproval messages (Dugar, 2010) or centralized communication by a team leader (Brandts and Cooper, 2007); group composition (either through socio-demographic homogeneity, Engelmann and Normann 2010, or through endogenous decisions Riedl, Rohde, and Strobel 2015 or Kopányi-Peuker, Offerman, and Sloof forthcoming). Except in these very specific settings, there usually appears to be a gradual and pronounced failure to coordinate on the payoff-maximizing equilibrium, even when the stage game is repeated with the same subjects. As highlighted by Weber (2006) or Kopányi-Peuker, Offerman, and Sloof (forthcoming), the frequency with which failures to coordinate efficiently is observed in the lab contrasts with the numerous examples of relatively efficient coordination achieved routinely in the field by firms, working teams or communities. Our purpose is to test whether voluntary and decentralized sanctions can function as an efficient coordination mechanism.

Indeed, voluntary sanctions have been established to be a powerful force to foster cooperation in public goods games (Fehr and Gaechter, 2000; Masclet, Noussair, Tucker, and Villeval, 2003; Masclet, Noussair, and Villeval, 2013), implying that decentralized, informal sanctions may explain successful cooperation in the field (Ostrom, Walker, and Gardner, 1992). We hypothesize that a similar mechanism may be at work in coordination contexts. For instance, in team projects similar to the examples above, workers may have opportunities to sanction low-effort individuals to ensure coordination on the payoff-maximizing equilibrium. An important difference between voluntary sanctions and most coordinating devices put forth in the experimental literature (see above) is that, to a large extent, sanctions are endogenous to the group and do not require an external change. In most external situations with coordination issues, individuals often have the opportunity to punish other members of the group (either formally or informally). This contrasts with mechanisms which require a structural change in the group (e.g., group composition) or in the game (e.g., exclusion, centralization or change of the payoff matrix).

Even though perhaps intuitively appealing, it is not straightforward that individuals will use punishment in a coordination game: in contrast to cooperation games, the individual motivation for punishment is less clear in coordination games where choosing low efforts penalizes oneself to a certain extent. Reciprocity, which has been found to be a powerful driver of such sanctioning behavior (Falk, Fehr, and Fischbacher, 2008a), does not necessarily lead to punishment of low efforts in coordination games. The " shirker" primarily hurts her/himself, and lack of coordination does not stem, at least not necessarily, from a lack of concern for the others' situations. Alternatively, if voluntary and decentralized punishment occurs as a way to enforce a social norm (for instance efficiency or equity), individuals may employ sanctions against deviation from the norm in the minimum effort game. The second question is to know, would sanctions be voluntarily used by subjects, whether it helps coordinating to payoff-maximizing equilibria and to which extent. In particular, it is of interest to observe whether sanctions have an added-value in this respect, in particular in comparison with the generally Pareto-improving effect of communication. Whether punishment opportunities will be used in this context and whether they increase efficiency – in particular compared to mere disapproval communication opportunities – are the empirical questions we aim to shed light on.

To do so, we set up an experiment of the minimum effort game based on the work of Van Huyck, Battalio, and Beil (1990). At the beginning of each round, subjects, in groups of eight, choose an integer effort level between 1 and 7. They receive anonymous feedback on the effort choices of their fellow group members and, depending on the treatment, can assign points to them. In the *Disapproval* treatment, these points simply act as a communication device signaling disapproval, with no monetary consequence, as tested by Dugar (2010). In the *Punishment* treatment, assigning points imposes a *fine* on the punished group member, but also comes at a *fee* to the punisher, with the fine being twice as large as the fee. As in Masclet, Noussair, Tucker, and Villeval (2003), comparing the punishment and disapproval treatments allows us to disentangle what part of the punishment effect is due to implicit *ex post* communication, e.g., expression of disapproval, and what is due to punishment *per se*. To provide a strong test, subjects in all treatments first complete eight rounds of play in the baseline minimum effort design without punishment, likely creating a history of low efficiency to be overcome in the next eight rounds with disapproval or punishment opportunities. A similar setup with an initial baseline phase has been used, for instance, by Brandts and Cooper (2006) to study the effect of *ex ante* communication, Romero (2015) to examine variation in effort cost, and Fatas, Neugebauer, and Perote (2006) to assess the magnitude of a pure 'restart' effect between two successive identical baseline stages. Based on these studies, we expect to find strong path-dependence and, at best, a mild positive restart effect, hence facilitating a strong test of the viability of *ex post* monetary punishment and cheap-talk disapproval as coordination devices. This initial baseline stage distinguishes our *Disapproval* treatment from an otherwise similar disapproval treatment conducted in Dugar (2010). The purpose of our two-stage design is to assess more explicitly the efficiency-enhancing strength of punishment and disapproval by submitting the coordination devices to more adverse conditions.[1]

Our results show that, even after a history of coordination on inefficient equilibria, the possibility to punish others in the minimum effort game brings groups to (or very close to) Pareto-optimality in about a third of cases and lead to considerably Pareto superior equilibria in another third of cases, even without much punishment being implemented. By contrast, only temporary improvements are observed in the payoff-neutral disapproval treatment, and only a very limited restart effect takes place in a baseline treatment with no communication nor sanctioning device. This suggests that punishment provides a powerful coordination device, similar to its effect in public goods games, and superior to the effect of an *ex post* communication device alone.

The remainder of the paper is organized as follows. The next section considers potential effects of punishment on coordination and set hypotheses. The third one presents the experimental design and procedures. Results are described in section 4, and discussed with some concluding remarks in the last section.

2 Motivation and theoretical considerations

In Pareto-ranked coordination games, the general finding is a tendency to inefficient coordination. Even though, as Ortmann and Devetag (2007) put forth, a plethora of ways to engineer efficient coordination have been found in the experimental literature, the default result seem to be a general dynamics towards the coordination to Pareto inferior levels. In part, these efficiency-enhancing mechanisms may explain how agents may reach coordination on relatively efficient outcomes. Yet, an alternative mechanism may be the use of decentralized and voluntary

[1] One may also interpret this as an investigation into whether, in a team exhibiting coordination failures, organizational changes that render punishment and disapproval possible have any positive effect, for instance a change in the availability of information regarding other team members' effort.

sanctions against individuals that would deviate from the payoff-maximizing equilibrium. Indeed voluntary sanctions have been found to have a profound effect in cooperation games: a massive literature has shown that players often engage in voluntary sanctions even at a cost. As a consequence cooperation levels increase substantially in the presence of punishment opportunities and seems stable through repetitions. A lot is known about the effect of such sanctions in the case of a public goods game: the relative effect of varying punishment technology and costs (Carpenter, 2007), the long-term effect of punishment for repeated games ((Fehr and Gaechter, 2000), the motivation behind the use of punishment (Carpenter and Matthews, 2012), the effect of the possibility of antisocial punishment (Herrmann, Thöni, and Gächter, 2008), the relative merits of punishment relative to other cooperation-enhancing features (Chaudhuri, 2011), the existence of third-party punishment (Fehr and Fischbacher, 2004b), etc. Overall, this prolific literature provides solid grounds to account for the high level of cooperations observed in many human societies. An interesting question is then to know whether voluntary sanctions can influence behaviors in other types of games. In particular, Pareto-ranked coordination games provide a relevant field of study: they include a social dilemma dimension in the sense that some of their potential outcomes are socially superior to others but the attainment of such socially desirable outcomes is not guaranteed by pure individual rationality.

However, the fact that players voluntarily sanction others is not granted straightforwardly based on some generalization of what happens in public goods game: in Pareto-ranked coordination games, the main issue is for players to coordinate, and not to counterbalance individually rational deviations from the socially desirable outcome. Whether, in the eyes of individuals, a deviation from payoff-maximizing equilibrium strategies is a reason enough to punish some player is not fully clear: would the the low-effort individuals have known others were coordinating on the payoff-maximizing equilibrium, it would have been in their own interest to comply. More generally, the question of the effect of punishment in the minimum effort game is two-fold: firstly, do individuals resort to punishment voluntarily in the case of a failure to coordinate efficiently? And secondly, if they do, do these sanctions lead to a Pareto improvement?

2.1 Do subjects use sanctions in Pareto ranked coordination games?

Whether one should expect individuals to use sanctions in the minimum effort game largely depends on whether the motivations for punishment in cooperation games are present in the coordination game. There are at least three reasons highlighted in the literature to account for voluntary punishment in cooperation games: distributional preferences such as inequity aversion (Fehr and Schmidt, 1999), reciprocity-based preferences (Falk and Fischbacher, 2006), and the enforcement of a (social) norm (Fehr and Fischbacher, 2004a; Lopez-Perez, 2008). To discuss whether individuals have grounds to use sanctioning opportunities in the minimum effort game, we consider each of these motivations and whether they may apply to this type of games.

The clearest case is with plain distributional preferences as inequity aversion: the motivation for punishment extends immediately from public goods games to the minimum effort games (at least in their symmetric versions). The players with the lower efforts get a higher payoff than those who chose higher efforts. If sufficiently inequity-averse, the latter may want to use sanctions to reduce inequity. The effect may be weaker in the minimum effort game because in the typical matrix payoff the difference in payoffs is smaller – since in the minimum effort game, the lowest effort player does not benefit from deviating from higher levels). Although popular, these models of inequity aversion (and more generally of distributional preferences as the source of punishment) have faced empirical challenges, many experiments showing that on the one hand sanctions are still used when the inequity aversion would predict otherwise (Falk, Fehr, and Fischbacher, 2005) and on the other hand intentions and not only outcomes seem to matter to trigger sanctions (Falk, Fehr, and Fischbacher, 2008b). Yet, even if not the main driver of punishing behaviors in cooperation settings, they may still explain part of the effect and hence may be present in the minimum effort game.

Reciprocal preferences (Rabin, 1993; Charness and Rabin, 2002; Falk and Fischbacher, 2006) explain punishment in public goods game on the following basis: people tend to reward individuals who they believe have behaved kindly to them, and to punish those who they believe have behaved meanly. The source of punishing behavior is seen in negative reciprocity, that is the tendency to behave unkindly towards those perceived as malevolent. In this approach, the crux of the matter is to know whether high-effort players interpret low efforts from other as unkind or unfair. This critically depend on the reasons behind low effort when the vast majority of other players choose a high effort. There are several possibilities. The main reason usually evoked is a divergence of beliefs about what the group of players will do: if a given player believes that at least one other player will choose a low effort and chooses accordingly, while another one thinks that others will all choose the highest effort and does the same, there is little reason to interpret that the shirker's behavior as aggressive, unkind or malevolent *per se*. In this case, one should not expect reciprocal behavior.

But there are some other possibilities that may justify some sanctions for low effort players. A few of those reasons originate in a more general notion of reciprocity. Firstly, a player may believe that there is a fair chance that all others will choose a high effort, but also some non negligible probabilities that at least one of them will play low. Because of risk aversion, she may choose a risk-dominant strategy (instead of the payoff-dominant one), hence a low effort, but doing so she neglected others in order to pursue her own interest, once risk aversion is taken into consideration. She simply dislikes individual risk more than she cares about the social outcome. In a slightly more sophisticated way due to the fact that risk preferences come into play, she does nothing really different from not cooperating in a public good game, where players favor their individual benefit rather than the social good. This gives room to have others interpreting her action as unfair, and may be reason enough to want to see her punished. Another possible motive to deviate from a high effort equilibrium can be spitefulness or competitive preferences. There is no ambiguity that such an action is a voluntary decision to hurt the others, and as such it may trigger negative reciprocity. A last reason related to reciprocity for sanctioning in the minimum effort game is that reciprocity may be fueled by responsibility rather than intentions. Even if low efforts are viewed as a plain mistakes – otherwise, the "shirkers" would have chosen higher efforts, if only for selfish reasons–, individuals have to be taken accountable for the results of their actions. An action leading to a departure from a fair standard – be it in terms of equity or in terms of efficiency– could be seen as unfair or unkind and trigger negative reciprocity, in part *independently of the underlying reasons* for this action, e.g. erroneous beliefs, inattention, etc.[2] Some evidence of such a phenomenon is provided by Falk and Fischbacher 2006 on the basis of an empirical survey where individuals still assess actions resulting in an unfair outcome, without the presence of an intention from others, as unkind, as well in Bartling and Fischbacher's (2011) study where responsibility, and not just intentions, seems to matter.

Another account for the existence of voluntary punishment relies on the idea that societies are organized with social norms, that individuals tend to enforce with more or less informal sanctions (Bicchieri, 2006). One relatively frequent norm could be a cooperation norm and could explain the resort to voluntary sanctions towards those who do not abide by the norm. Based on the existence of a cooperation norm, not only direct reciprocity – as in the models referred above – can be explained but also but also indirect reciprocity, that is the fact that third parties – external to the initial interactions – may punish non-cooperators (Fehr and Fischbacher, 2004a). A general model for norm-breaking is provided by Lopez-Perez (2008), where social norms take the form of strategies that leads to equal or efficient outcomes in interactions. In the minimum effort game, where deviations from the Pareto superior equilibrium results in less efficiency and more inequity, failure to coordinate with the majority of others could be seen as a transgression of the norm and trigger sanctions from those who abide by this norm. Explanations of sanctioning behaviors on the basis of norms has received recently quite some experimental support (Carpenter and Matthews, 2009; Capraro and Rand, forthcoming; Kimbrough and Vostroknutov, 2016).

To summarize, one can put three main reasons why individuals would engage in punishment in the case of a minimum effort game: plain distributional preferences, an extended notion of reciprocity and the enforcement of social norms. In these three cases, it seems fair to expect that punishment behavior will be less prevalent and less intense than in cooperation games, given that some features that are known to take part in the motivation for punishment, e.g. intentions, lack in the coordination game.

[2]To some extent, this would be in the spirit of Falk and Fischbacher's (2006) and Charness and Rabin's (2002) where the "fairness" of another player's action is seen as how it results in a deviation from some social standard (possibly weighted by an intention or or "demerit" factor). Note that, to the letter, the models do not apply to the minimum effort case, either because as in Charness and Rabin 2002, the model is left unspecified for actions leading to Pareto inferior outcomes, or because equilibrium play is assumed.

2.2 Do sanctions work as an efficient coordination device?

Assuming that individuals use punishment, this will not necessarily lead to Pareto superior outcomes, even though sanctions possess features that may be helpful in this respect. First, they can work as "costly talk": not only have sanctions a cost for the punisher and may be interpreted as a costly way to signal her willingness to commit to a high-ranking equilibrium, but also it may convey a normative message to the "shirkers" if sufficiently many individuals do punish him. In this sense, punishment can be expected to have more strength than *ex post* communication, e.g. disapproval messages, if individuals are willing to bear the cost of this form of communication.

Moreover, the monetary consequences of sanctions may provide additional incentives to low effort players to follow the majority and increase their effort. As it does for cooperation games, subsequent punishment change the structure of incentives: if high effort players are numerous and willing to punish low-effort ones, the plain appeal of low effort may disappear. Typically, one of the reasons why a player chooses a low effort is the risk-dominance of such a strategy: by changing the final payoffs associated with such strategies, the risk-dominant actions in the original game are no more risk-dominant. They may simply lead to worst payoff in all circumstances. In this sense, sanctions may not help coordinate beliefs by players – or only indirectly – but by changing dramatically the structure of payoffs, they make the high effort action the risk-dominant one, in addition to be the payoff dominant. In sum, sanctions may not so much work as a coordination device *per se* but as a transformation of the payoff matrix of the situation, hence its strategic structure.

Yet, at least two phenomena may impair the efficacy of punishment as an efficiency improving device in a coordination game. First, as is classically the case for cooperation games, sanctions have a social cost and the positive effect in terms of efficiency within the game may not overcome the costs, both to the punisher and the punished. Second, the setting can generate a second-order coordination problem – on top of the usual second-order cooperation problem that is also present in the public goods game: if a majority of subjects do not agree on the normatively appealing equilibrium, they may engage in endless war of punishment, counter-punishment, etc. or more simply that may demotivate them to simply try to coordinate highly. Some subjects may hence feel undeservedly punished and may be reluctant to make any subsequent effort to coordinate on Pareto superior equilibria. The main difference here with cooperation games is that it that what constitutes a kind or fair action is much more ambiguous. This is particularly true if the main motivation for sanctioning is linked to social norms: social norms seem to maintain their appeal only to the extent that they are shared by a vast majority of the group (Schram and Charness, 2015).

In sum, although the usual reciprocity models may not predict sanctions there are other plausible reasons to expect that some individuals will use punishment in the minimum effort game. The extent of these voluntary sanctions as well as the impact on the efficiency of coordination is an open question that our experiment aims at shedding light on.

3 Experiment

The participants played 16 rounds of the minimum effort game, split into two stages of eight rounds each. At the beginning of the experiment, we handed out the instructions for the first stage and announced that there would be a second stage of unspecified nature (experimental instructions are available in the on-line Additional Material). Subjects also knew that only one of the two stages would be chosen at random to be paid.

In Stage 1, all the treatments featured a baseline design closely resembling the seminal one of Van Huyck et al. (1990). Groups consisting of eight players were formed randomly prior to the first round and remained the same for the entire experiment, this matching scheme being known to the subjects. In each round, players simultaneously chose an integer effort level between 1 and 7. Each player's payoff depended on her effort choice and the lowest effort choice in her group. In particular, let $N = \{1, 2, 3, ..., 8\}$ be the group of players and $E = \{1, 2, .., 7\}$ be the set of effort levels, each player choosing effort $e_i \in E$. With $s = (e_i)_{i \in N}$ being the strategy profile of all players in the group, the payoff (in euros) of player i in a given round is

$$\pi_i(e_i) = 0.4 \times \min_{j \in N}(e_j) - 0.2e_i + 1.2 \quad (1)$$

Table 1 shows the corresponding payoff matrix. This payoff matrix with seven Pareto-ranked equilibria along the main diagonal was used with minor modifications by Van Huyck et al. (1990), Blume and Ortmann (2010), Dugar (2010) and many others.

		minimum effort choice in the group						
		7	**6**	**5**	**4**	**3**	**2**	**1**
	7	2.60	2.20	1.80	1.40	1.00	0.60	0.20
	6		2.40	2.00	1.60	1.20	0.80	0.40
own	**5**			2.20	1.80	1.40	1.00	0.60
choice	**4**				2.00	1.60	1.20	0.80
	3					1.80	1.40	1.00
	2						1.60	1.20
	1							1.40

Table 1: *Payoff matrix of the minimum effort game (in euros)*

After all group members had made their effort choices, the feedback screen displayed the player's effort choice and payoff for the current round and her cumulative payoff for Stage 1. The same screen showed a table with the current effort choices and payoffs of fellow group members, ordered from the lowest to the highest effort. This feedback format closely resembles the one used by Engelmann and Normann (2010) and Dugar (2010). Since this feedback is necessary for treatment variations in Stage 2, it was used in both stages for reasons of symmetry and comparability. A player's total payoff for Stage 1 consisted of the sum of her round payoffs plus an initial endowment of 4 euros.

After Stage 1, subjects received the instructions for Stage 2 in which the design differed across treatments. In the *Baseline* treatment, Stage 2 was identical to Stage 1. In the *Punishment* treatment, after receiving the feedback on effort choices and payoffs, subjects could (but were not required to) assign punishment points to fellow group members. Each point inflicted a cost of 10 cents on the punisher and 20 cents on the punished subject. After all players had assigned points, the feedback screen showed the sum and costs of points assigned by and to the subject in the current round, her resulting payoff for the current round, and her cumulative payoff for Stage 2. Since effort choices and payoffs of fellow group members were ordered from the lowest to the highest effort in each round and hence players' identity was concealed, counter-punishment or punishment of past effort choices was not possible. We chose this form of post-punishment feedback to parallel the one used in public goods games with punishment (e.g., Fehr and Gaechter, 2000, and Anderson and Putterman, 2006).

Punishment was limited by the punishing player's own cumulative payoff up to the previous round. In order to give players the opportunity to punish in the very first round of Stage 2 independently of their earnings in that round, subjects received an initial endowment of 4 euros. The endowment meant that a player with an effort choice of 7 facing seven other group members choosing effort level 1 was able to almost equalize the profit of all members in the first round (i.e., by assigning 6 points to each fellow group member). This ensures comparability between rounds and limits the effect of past earnings on punishment decisions. For reasons of symmetry, the 4 euro endowment was granted in both stages of all treatments.

The simultaneous choice of punishment points in any given round generates a second order public goods game where players may free-ride on others carrying out punishment. This problem is magnified by the fact that punishment points could only reduce other members' game payoff from the current round to zero at most, so some of the assigned points may be "wasted" in case they were to reduce her game payoff to below zero. While subjects of course did not know *ex ante* how many points other members would assign, the full cost of assigning points had to be borne *ex post*. This results in a round payoff for player i of

$$\pi(s)_{Punish} = \max\left(0; 0.4\min_{j\in N}\{e_j\} - 0.2e_i + 1.20 - 0.2\sum_{j\in N} P_{ji}\right) - 0.1\sum_{j\in N} P_{ij} \quad (2)$$

where P_{ij} denotes the punishment points that player i assigns to player j.

To compare the effect of monetary and non-monetary sanctions, we ran a third treatment called *Disapproval*. The procedure in this treatment was as similar as possible to the one in *Punishment*, with the important difference that disapproval points did not inflict monetary costs on either the disapproving or the disapproved group member. The points were merely a means of communicating one's opinion about others' behavior. After receiving the feedback on effort choices and payoffs, a player could assign between zero and six points (only integer) to each other group member, with six points expressing the maximum disapproval. Thereafter, the displayed information matches the post-punishment feedback provided in the *Punishment* treatment. This differs slightly from the disapproval treatment in Dugar (2010) where subjects could in addition observe the sum of points assigned to their fellow group members. Other differences vis-a-vis Dugar's design are the number of group members and the number of rounds per stage - in both cases 8 in ours and 10 in Dugar's. Judged from the literature surveys of Devetag and Ortmann (2007) and Engelmann and Normann (2010), minor variation in these design features appears to have little or no (consistent) effect on coordination outcomes.

A likely more important design difference is the absence of Stage 1 in Dugar's experiment. There are at least two reasons for including the initial baseline Stage 1 in all our treatments. First, we wished to examine the effect of our treatment manipulation after a history of inefficient effort choices (anticipated on the basis of the findings of previous studies with similar design features), which arguably allows us to draw stronger conclusions regarding the hypothesized positive effect of monetary and non-monetary sanctions. The second reason for including Stage 1 is that it permits a difference-in-differences comparison of behavior across treatments. In other words, in addition to the standard contemporaneous across-treatment comparison of behavior in Stage 2, we are able to compare treatments in terms of between-stage *changes* in behavior, thereby assessing the overall effect of treatments with more power as well as the effect at the individual and group level.

Eight lab sessions of 32 subjects were run for a total of 256 subjects composing eight groups (2 sessions) for *Baseline*, 12 groups (3 sessions) for *Disapproval*, and 12 groups (3 sessions) for *Punishment*.[3] The experiment was programmed and conducted in Z-Tree (Fischbacher, 2007) and lasted on average 80 minutes. Including a 2.50 euro show-up fee, the average earnings in the experiment were 18.18 euros (at that time around 24 USD), ranging between 7.10 and 27.30 euros. Participants were paid privately in cash, according to their performance. They were recruited among students of various disciplines at the local university using the ORSEE software (Greiner, 2004). In each session, gender composition was approximately balanced and each subject took part in only one session. In order to verify the subjects had understood the instructions, subjects were asked to answer several control questions. After all subjects had answered the questions correctly, the experiment started.

[3]Another session was run with the baseline condition (32 subjects, 4 groups), but due to a technical problem the second stage of the experiment could not be run. The results of this session are not reported here, but are very similar to what is observed in the first stage of the experiment in all treatments.

4 Results

We present the results along three main dimensions: the effect of *Punishment*, *Disapproval* and *Baseline* treatments on the level of coordination and efficiency in the game, punishing and disapproving behaviors, and the effects of the treatments on the net efficiency.

4.1 Coordination and effort levels

To study the effect of treatments on coordination and effort levels, we first focus on the first stage to ensure comparability, then on the impact of the experimental manipulations in Stage 2, to eventually analyze the whole data in terms of difference-in-difference.

4.1.1 Stage 1: similar results accross treatments

For each treatment, Figure 1 shows the evolution of the average effort, and Figure 2 displays the evolution of average minimum effort (i.e., the average of groups' minimum effort). For Stage 1, both figures suggest little or no across-treatment differences. In all treatments, the average effort is initially around 5 and gradually falls to 2. Average minimum effort starts off at about 2 and does not diverge much from that level throughout the stage. At the end of the stage, the average effort is only marginally above the average minimum effort, which implies a low within-group variance. Figure 3 presents a more disaggregate look at effort choices, minimum efforts by team and levels of coordination (through the individual difference to team's minimum effort). In all treatments, the highest effort level is initially the most frequent choice and the lowest effort level is chosen by less than a tenth of subjects. Throughout Stage 1and for the the three treatments, efforts choice decrease so that most teams have a minimum effort of 1 or 2 after a few rounds, yet with some teams managing to coordinate on the highest effort levels (in *Disapproval* and *Baseline*) or intermediate effort levels (in *Punishment*). Regarding the dispersion of effort choices, the average individual difference of effort choice and minimum effort in the team, decreases steeply to reach almost in all cases 0, that is perfect coordination. The general pattern is very similar for the three treatments as would be expected by the fact that in Stage 1 the design is strictly identical, and consistent with the typical finding of similar experiments.

This conclusion is supported by statistical tests. We first compare effort choices by the Mann-Whitney U test applied to average efforts at the group level.[4] The across-treatment differences were not significant either overall or in individual rounds, reflecting the identical design setup across treatments. Parametric tests provide a similar degree of support for the across-treatment differences: Wald tests from ordered probit estimation indicate that effort does not significantly differ

[4]Depending on the type of comparison, groups' average efforts in a given treatment are calculated for each stage or each round. In round 1, we apply the test directly to effort choices since these are independent unlike in subsequent rounds.

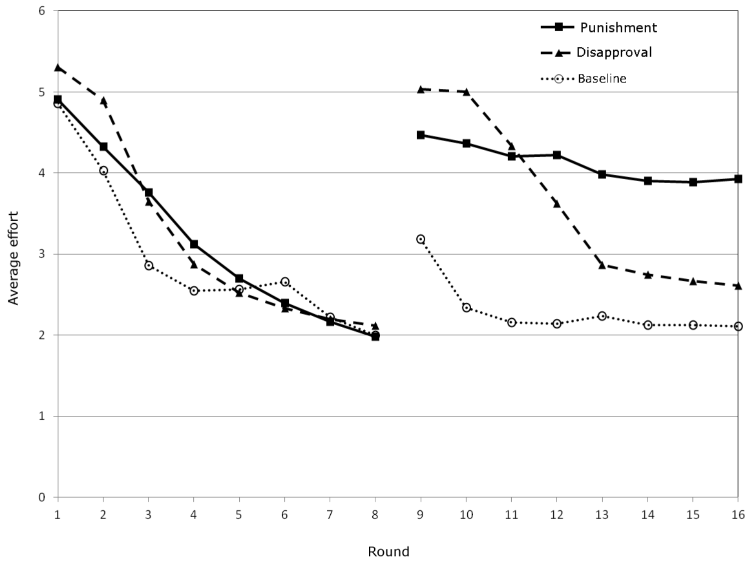


Figure 1: *Average individual effort per round and treatment*

across treatments in Stage 1 overall, nor in each round.[5] As an exception, effort in round 1 is significantly higher in *Disapproval* compared to both *Punishment* and *Baseline* ($p < 0.10$). Finally, turning to minimum effort instead of average effort, groups' minimum efforts do not significantly differ across treatments in Stage 1 either overall or in individual rounds, according to both the Mann-Whitney U test and the Wald test.[6] In sum, the patterns observed in Stage 1 are similar in all treatments and suggest that any substantial difference in Stage 2 is highly likely to stem from the treatments rather than from sampling variations.

4.1.2 Stage 2: *Punishment* increases efforts in the long run

Turning to Stage 2, Figure 1 shows that the average effort jumps upwards to 4.5 in *Punishment* and to 5.0 in *Disapproval* in the restart round 9, almost reaching the initial levels of round 1. Subsequently, the *Punishment* effort stays almost stable to reach 3.9 in the final round, whereas the *Disapproval* effort falls much faster from round 11 onwards to eventually reach 2.6. These dynamics can be compared

[5] We regress effort choices on treatment dummies interacted with a stage dummy or round dummies (for across-treatment comparison at the stage level or the round level, respectively). The estimations are based on a panel of 256 subjects with 16 rounds of effort choices each. We use the cluster-robust estimator of variance allowing for intra-group correlation of effort choices. The number of clusters (i.e., groups) seems sufficient given the perfectly balanced cluster sizes (e.g., Kezdi (2004); Rogers (1993)). The results are unaffected if including a second level of clustering at the subject level, or instead including group and individual-level random effects. Wald tests from ordered logit models and t-tests from linear probability models yield very similar results in terms of significance levels, as do separate estimations for round 1 performed without group clustering (since effort choices are independent).

[6] We regress groups' minimum efforts on treatment dummies interacted with a stage dummy or round dummies. The estimations are based on a panel of 32 groups with 16 rounds of minimum efforts each. As above, we use the cluster-robust estimator of variance allowing for intra-group correlation of observations. The results are unaffected if one includes instead group-level random effects. Other estimation details are identical to the estimation for effort choices.

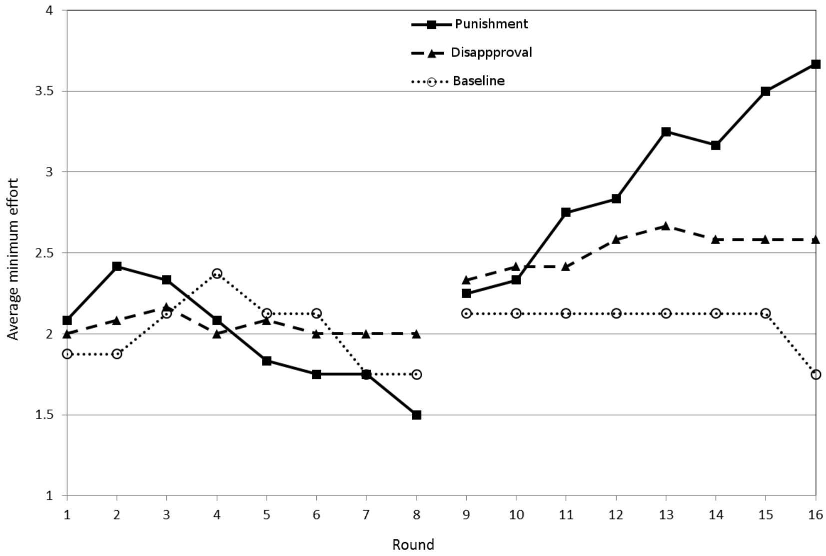


Figure 2: *Average minimum effort in groups per round and treatment*

with the pure restart effect in *Baseline* where the average effort jumps up much less in round 9 and then almost immediately falls back to the lowest level of 2.1 reached at the end of Stage 1. These results also hold for average minimum effort. For *Baseline*, Figure 2 indicates a small positive restart effect for average minimum effort that remains at 2.1 throughout Stage 2, except for a slight drop in the final round. In *Punishment*, minimum effort rises markedly to eventually reach 3.7, whereas in *Disapproval* minimum effort increases only mildly for several rounds, thereafter remaining around 2.6. Overall, these descriptive statistics indicate that *Punishment* seems to lead to higher levels of effort, despite a smaller initial effect on individual efforts at the beginning of Stage 2 than *Disapproval*, and despite a lower minimum effort at the end of stage 1.

Statistical tests confirm these findings: for the Mann-Whitney U test, groups' average efforts were significantly higher in *Punishment* compared to *Baseline* both overall ($p < 0.05$) and in the first six rounds ($p < 0.05$ in rounds 10-12 and 14; $p < 0.10$ otherwise), and also significantly higher in *Disapproval* compared to *Baseline* both overall ($p < 0.05$) and in the first five rounds ($p < 0.05$ in round 10; $p < 0.10$ otherwise). Parametric tests lead to similar conclusions. Individual efforts are significantly different between *Punishment* and *Baseline* both overall ($p < 0.10$) and in the first five rounds ($p < 0.05$ in round 10; $p < 0.10$ otherwise). The same holds between *Disapproval* and *Baseline* in the first three rounds ($p < 0.01$ in round 10; $p < 0.05$ otherwise). Although, the difference between *Punishment* and *Disapproval* in terms of individual effort, though substantial, are not statistically significant, the effect on groups' minimum effort is significant at the end of Stage 2. For both the Mann-Whitney U and the Wald test, minimum effort is significantly higher in *Punishment* compared to *Baseline* in the final round 16 ($p < 0.10$).

In terms of coordination, almost all groups have managed to coordinate on particular equilibria at the end of Stage 2. This can be seen on Figure 3 where the

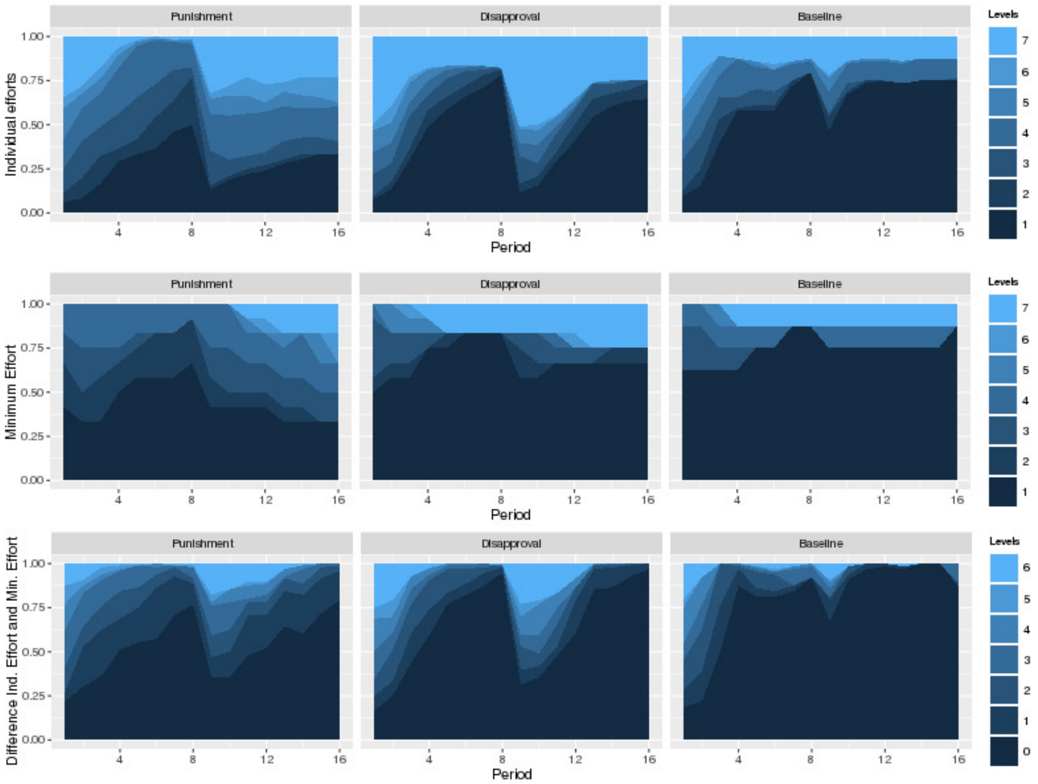


Figure 3: *Distribution of individual efforts, team minimum efforts, and individual distance to the team minimum effort, per round and treatment*

second row displays the distribution of minimum efforts by team, complemented by the distributions of the differences between individual effort and the team's minimum effort, in the third row. Figure 3 broadly confirms the finding that the attained equilibria overall involve more efficient effort levels in *Punishment* compared to *Disapproval* and especially to *Baseline*. In particular, there is a much stronger shift towards Pareto superior equilibria in *Punishment* where the percentage of subjects choosing the two highest effort levels rises from 2 to 36 percent between the last rounds of each stage, while the percentage choosing the two lowest effort levels falls from 70 to 33 percent (compared with a decrease from 80 to 72 percent in *Disapproval*). Likewise, the minimum efforts in *Punishment* seem to increase steadily during stage 2, while, after an initial mild increase, seem to plateau in *Disapproval*.

In sum, Stage 2 generates across-treatment differences in the posited direction. From about the same aggregate starting point at the end of Stage 1, the efficiency gains in Stage 2 are initially slightly larger in *Disapproval* than in *Punishment* – perhaps reflecting subjects' initial hopes of the effectiveness of the cheap-talk communication device – but these hopes fade rather quickly and the efficiency gains are eventually considerably larger in *Punishment* than in *Disapproval*. Except for a small positive restart effect, Stage 2 brings about no efficiency gains in *Baseline*.

4.1.3 Between-stage comparisons: A stronger positive effect of *Punishment*

In order to test the robustness of these effects, we next compare between-stage *changes* across treatments. The results are provided in Table 2. The first row (i.e., block of results) displays effort changes and their statistical significance between Stages 1 and 2, both overall and for each round-pairs (i.e., rounds 1 and 9, 2 and 10, etc.). From the same effort level of about 3 in Stage 1, the average effort in Stage 2 increases by 0.95 (30 percent) in *Punishment* and 0.38 (12 percent) in *Disapproval*, whereas it decreases by 0.66 (22 percent) in *Baseline*. The overall change in efforts in *Punishment* as well as in *Baseline* are significant for both the ordered probit Wald test and the Wilcoxon signed-rank test applied to groups' average efforts. All treatments feature an initial average-effort decrease in the first round-pair, reflecting the dynamics towards less efficient equilibria described above in Stage 1 and commonly observed in the experimental literature. Only in *Punishment* is this followed by an increase that becomes larger over time: The effort increases in the last five round-pairs are significant only for this treatment. A pattern of initial average-effort decreases followed by increases also occurs in *Disapproval*, but the increases fade after the fifth round-pair and subsequently remain much smaller than in *Punishment*. Moreover, changes are not significant in any round-pair. *Baseline* generally features average-effort decreases of declining magnitude (except for a small increase in the last round-pair) which are significant in the first three round-pairs and in the fifth and sixth round-pairs.

	Treatment	Stage 1-2	Round 1-9	Round 2-10	Round 3-11	Round 4-12	Round 5-13	Round 6-14	Round 7-15	Round 8-16
Average effort change	Punishment	0.95 $^{ww,ss}_{bbb}$	-0.44 $_{bb}$	0.04 $_{bbb}$	0.45 $_{bb}$	1.09 $^{ww,s}_{bb}$	1.28 $^{www,s}_{bbb}$	1.51 $^{www,s}_{bbb}$	1.72 $^{www,ss}_{bbb,d}$	1.95 $^{www,ss}_{bbb,d}$
	Disapproval	0.38 $_{bb}$	-0.27 $_{bbb}$	0.10 $_{bbb}$	0.69 $_{bb}$	0.75	0.34	0.42	0.47	0.50
	Baseline	-0.66 ^{ww,ss}	-1.67 ^{www,ss}	-1.69 ^{www,ss}	-0.70 ^{w,ss}	-0.41	-0.33 ^{ss}	-0.53 ^{w,ss}	-0.09	0.11
Average minimum-effort change	Punishment	1.00 $^{www,sss}_{bb}$	0.17	-0.08	0.42	0.75 $_{b}$	1.42 $^{www,ss}_{bbb}$	1.42 $^{www,ss}_{bbb}$	1.75 $^{www,sss}_{bb,d}$	2.17 $^{www,sss}_{bbb,dd}$
	Disapproval	0.48	0.33	0.33	0.25	0.58 $_{b}$	0.58	0.58 $_{b}$	0.58	0.58 $_{b}$
	Baseline	0.08	0.25	0.25	0.00	-0.25	0.00	0.00	0.38	0.00
Fraction of groups with an average-effort increase	Punishment	0.54 $_{bbb}$	0.33 $_{bb}$	0.50 $_{bb}$	0.58	0.58	0.58 $_{bb}$	0.58 $_{bb}$	0.58 $_{b}$	0.58
	Disapproval	0.36 $_{bb}$	0.42 $_{bb}$	0.58 $_{bbb}$	0.67 $_{b}$	0.42	0.17	0.17	0.25	0.25
	Baseline	0.11	0.13	0.13	0.13	0.13	0.00	0.00	0.13	0.25
Fraction of groups with a minimum-effort increase	Punishment	0.45 $_{bb,d}$	0.33	0.17	0.42	0.42 $_{b}$	0.50 $_{bb,d}$	0.50 $_{bb}$	0.58 $_{b,dd}$	0.67 $_{bbb,dd}$
	Disapproval	0.24	0.33	0.33	0.33	0.25	0.17	0.17	0.17	0.17
	Baseline	0.06	0.13	0.13	0.13	0.00	0.00	0.00	0.13	0.00

The "w", "t" and "s" superscripts denote a significant difference across stages or across a round-pair (see the top row), using an appropriate ordered probit Wald test, t-test, and Wilcoxon signed-rank test, respectively. The "b" resp. "d" symbols denote a significant difference across stages or across a round-pair for the changes between the treatment directly above the symbol and *Baseline* or *Disapproval* respectively, using an appropriate ordered probit Wald test (in the first and third blocks) or Mann-Whitney U test (in the second and fourth blocks). Significance levels are 1%, 5% resp. 10% for three, two resp. one superscripts or symbols of a kind in a given cell.

Table 2: *Between-stage and between-round effort changes in each treatment*

Analyzing our data in terms of difference-in-difference reinforces these conclusions. Comparing changes in effort accross treatments[7], we observe (first block of Table 2) a positive and significant treatment effect between *Punishment* and *Baseline* both overall and in each round-pair, while the positive treatment effect between *Disapproval* and *Baseline* is significant overall and only in the first three round-pairs. The positive treatment effect between *Punishment* and *Disapproval* is significant at the .10 level in the last two round-pairs. These conclusions remain largely true when focusing instead on the average minimum-effort changes (block 2), on the the fraction of groups with an average effort increase (block 3), or on the fraction of groups with a minimum effort increase (block 4). These analyses support the fact that *Punishment* at the end of Stage 2 leads to a significant increase in efforts, unparalleled in the other treatments. *Disapproval* tends to yield significant gains in efforts only temporarily since these gains seem to vanish and lose significance by the end of Stage 2.

Overall, the results yield a consistent picture. *Baseline* replicates the typical findings in the literature on experimental Pareto-ranked games, namely gradual convergence to low effort coordination and a very small and temporary improvement in the restart stage. Both *Disapproval* and *Punishment* bring about a substantial towards Pareto superior outcomes following the restart, but only in *Punishment* does this positive effect persist throughout the restart stage and becomes stronger over time in terms of the outcome of the game, i.e., minimum effort. The strong positive effect of *Punishment* vis-a-vis the other treatments is evident not only in the plain between-subject comparison in Stage 2, but especially in the tighter within-subject and within-group comparison of efficiency gains between stages. Voluntary monetary sanctions in seem capable of persistently increasing coordinated effort levels, even in groups that previously converged to very low ranked coordination outcomes. By contrast, the effect of *ex post* cheap talk in *Disapproval* does not seem strong enough to stabilize coordination at a substantially higher ranked level compared to *Baseline*.

4.2 Punishment and disapproval behavior

The strong effect of *Punishment* on the outcomes of the game suggest that subjects do resort to sanctions. We observe indeed that 657 points are assigned overall – 80 percent in the first half of Stage 2 – inflicting a total cost of 65.7 euros on the punishers and 131.4 euros on the punished (i.e., about 9 percent of punishment points were not actually implemented because they would have decreased a punished subject's round payoff to below zero). Figure 4 shows that the percentage of punishers and punished in all rounds of Stage 2, as well as the number of punishment points used per punisher. Resorting to sanctions is quite a frequent behavior: 44 percent of subjects sanction others in the first round

[7]We regress within-subject effort-choice changes on treatment dummies, and their interaction with round-pair dummies whenever performing separate tests for each round-pair. The estimations are based on a panel of 256 subjects with eight effort-choice changes each (i.e., changes between rounds 1 and 9, 2 and 10, etc.). As above, we use the cluster-robust estimator of variance allowing for intra-group correlation of observations. Other estimation details are identical to the estimation for effort choices.

and more than 60 percent of subjects used sanctions throughout the 8 rounds. Although not directly comparable to public goods experiments, especially when other features than the structure of the stage game changes, this figure is not very far from those observed in cooperation games, for instance around 84 percent for Fehr and Gächter (2005), especially since the usual punishment cost/fine ratio is usually higher. Symmetrically, the percentage of punished is also relatively high, and starts off at 53 percent in the first round. The proportion of punishers and punished declines gradually to just nine percent in the final round. Each punisher initially assigns four points on average. This figure declines gradually to below two points in the penultimate round and then jumps back to four points in the final round. On average each active punisher sent 11.14 points throughout the 8 rounds of *Punishment*. The average punishment sent to a given punished is of 3.30 points, i.e. a cost of around 0.3 euros and a fine of 0.6 euros, for stakes in the stage game ranging from 1.40 to 2.60. Overall punished subjects received on average a penalty, when counting all the punishers active at a given round, of 3.6. Such a high number, which dwarfs the possible payoff in the stage game, clearly sets strong incentives for shirker to increase their effort. Overall, 72.9 percent of players in *Punishment* have been sanctioned at least once, receiving on average 9.39 points overall throughout the 8 periods. It seems fair to say that the use of sanctions is considerable and affects directly almost all the players (only 6.25 percent of them have never received nor sent sanctions).

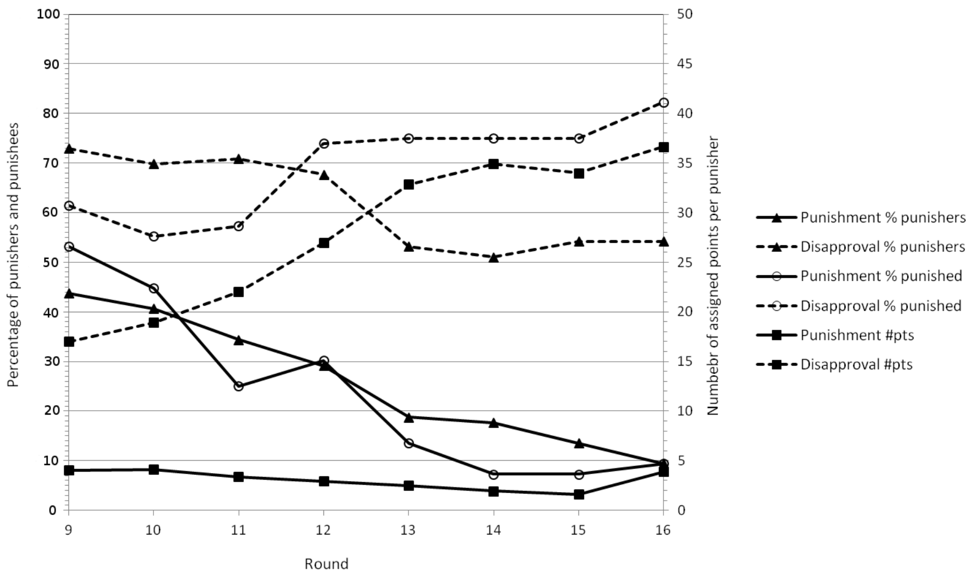


Figure 4: *Punishment and disapproval points assigned*

In *Disapproval*, 12,766 points are assigned overall – 45 percent in the first half of Stage 2 – which is almost 20 times higher than in *Punishment*. The percentage of disapprovers starts off at 73 percent and is still at 54 percent at the end, while the percentage of disapproved begins at 61 percent and eventually reaches 82 percent. Each disapprover initially assigns 17 points on average, and this figure steadily rises to eventually reach 37 points, rather close to the maximum of 42 points.

Thus disapproval is much more widespread than punishment and the gap widens over time.

		punished subject's effort level							Row	disapproved subject's effort level							Row
		1	2	3	4	5	6	7	total	1	2	3	4	5	6	7	total
	1	3.2	0	0	0	0	0	1.1	4.7	49.9	2.8	1.6	0.3	0.4	0	0.4	55.5
punisher's or	2	0	0	0	0	0	0	0	0	6.8	3.2	1.5	0.3	0.1	0	0.2	12.0
	3	3.7	4.6	0.8	0.3	0	0	0	9.3	3.4	1.7	2.0	0.7	0.5	0.0	0.3	8.6
disapprover's	4	5.5	2.9	5.0	0.8	0.2	0.2	0.9	15.4	1.5	0.6	0.9	0.2	0.1	0	0.1	3.4
	5	3.5	1.5	2.0	3.0	0.2	0.5	0	10.7	1.7	0.6	1.4	0.7	0.2	0	0	4.6
effort level	6	1.5	0.6	1.4	5.3	4.0	0.2	0.2	13.1	0	0	0.1	0.2	0.1	0	0	0.5
	7	19.6	1.2	6.5	11.9	2.9	3.5	1.2	46.9	6.2	1.0	3.7	1.9	1.4	0.8	0.3	15.3
Col. Total		37.0	10.8	15.7	21.8	7.2	4.3	3.3	657 pts	69.6	9.9	11.7	4.3	2.8	0.9	1.3	12,766 pts

Table 3: *Percentage of points assigned in* Punishment *and* Disapproval

Table 3 displays the distribution of punishment and disapproval points, aggregated across Stage 2, conditional on effort choices of the subjects by whom and to whom the points were assigned in any given round. In *Punishment*, punishers assign 90 percent of points to group members with a lower effort than theirs, i.e., the assigned points appear below the main diagonal. The most populated bottom-left cell contains points of punishers with effort level 7 assigned to subjects with effort level 1. As could be expected, punishment points are overwhelmingly targeted at relative "shirkers". By comparison, very few sanctions point are ever sent on or over the diagonal, that is when the punishers have chosen identical or lower efforts than the punished.

In sharp contrast, only 35 percent of points are assigned by disapprovers to group members with a lower effort than theirs in *Disapproval*. Half disapproval points are assigned from shirkers to other shirkers choosing the same effort level 1. These points are assigned in two-thirds of *Disapproval* groups that converge to the Pareto lowest equilibrium, mostly in the last rounds when the groups had already reached or almost reached the equilibrium. Even if one leaves out this rather special category of disapproval behavior, disapproval points are generally less consistently targeted at shirkers compared to *Punishment*, especially towards the end where group coordination outcomes are more or less settled. Such a behavior probably reflects some failed attempts to coordinate upward, or perhaps plain frustration to be collectively stuck in the Pareto lowest equilibrium. This tends to show that in the adverse conditions of a history of Pareto inferior coordination, plain ex post communication rather ends up in a "blame game", where most seem to disapprove of most, with little hope to coordinate upward.

The use of sanctions against relative shirkers can reflect some form of reciprocity or norm enforcement, but also some purely instrumental strategic motivation. Given that sanctions are most efficient at the beginning of Stage 2, since once a high-ranked equilibrium is reached it sustains itselfs by the incentive structure of the game, the dynamics of use of sanctions in *Punishment* may shed some light on the motivation behind sanctions. Figure 4 suggests that the use of sanctions weakens with time, both the intensity of sanctions and the share of punishers/punished in the group. This pattern would be expected if the main motivation for punishment is instrumental. On the other hand, the dynamics towards higher-ranked outcomes may mechanically explain the decrease in punishment: the opportunities to use sanctions may simply vanish as the group coordinates better with rounds. When taking this effect into account, there seems to be a slight decay in the intensity of punishment, but not as much as one would have expected in the case of a purely strategic motivation. Table 4 presents the average punishment points received by a shirker depending on rounds and on the difference between the effort chosen by the and the median effort: there is a clear issue of comparability because strong shirkers basically disappear after period 11. Nevertheless, when comparing the level of punishment for moderate shirkers, there seems to be a rather small decrease in the level of punishment received. For the most comparable case of a difference of 1 unit in efforts, the difference between average punishment in the four first rounds versus the four last ones appears significant with a t-test ($p < .001$), giving support to the idea that sanctioning behavior is in part motivated by strategic considerations. It is worth noting though that the difference only corresponds to 20 percent of the original punishment levels (from 2.84 to 2.28 points), suggesting that a large part of the motivation for punishing is not directly related to instrumental motives. The increase in punishment points given in the last round to 4 points on average, from 2 in the 15^{th} round (see Figure 4), to levels similar to those of the first round where punishment was possible, also suggest that at least some of the subjects have other motivations for sanctioning others than influencing future play.

Diff. to the median effort	-6	-5	-4	-3	-2	-1
Periods 9-10	9.17		0	6.85	5.44	3.00
	(6)		(1)	(13)	(9)	(23)
Periods 11-12	1.80	6.40	9.67	4.75	2.80	2.69
	(5)	(5)	(3)	(4)	(5)	(26)
Periods 13-14				8.00	5.00	2.38
				(2)	(2)	(21)
Periods 14-15					1.67	2.14
					(3)	(14)

Number of observations in parenthesis.

Table 4: *Average punishment points received by a lower effort player by rounds, in* Punishment*.*

4.3 Net efficiency

The fact that *Punishment* leads to better ranked outcomes in the game does not guarantee that overall efficiency is improved: the losses due to punishment (to both parties involved) may exceed efficiency gains in the stage game. Figure 5 shows for each treatment the evolution of average payoff as a fraction of the maximum achievable payoff (i.e., 2.60 euros per subject achieved if everyone chose the highest effort level 7 in a given round). In *Punishment*, we distinguish between *payoff* in the stage game and *profit*, i.e., the payoff minus the cost of punishing and being punished. For the other treatments, payoff and profit are obviously equal. Starting off at about 40 percent in all treatments, the average payoff rises throughout Stage 1 to eventually reach 60 percent in *Disapproval*, 58 in *Baseline* and 54 in *Punishment*. The upward trend and the magnitude of the average payoff for all treatments reflect the improving individual coordination but on low ranked equilibria.

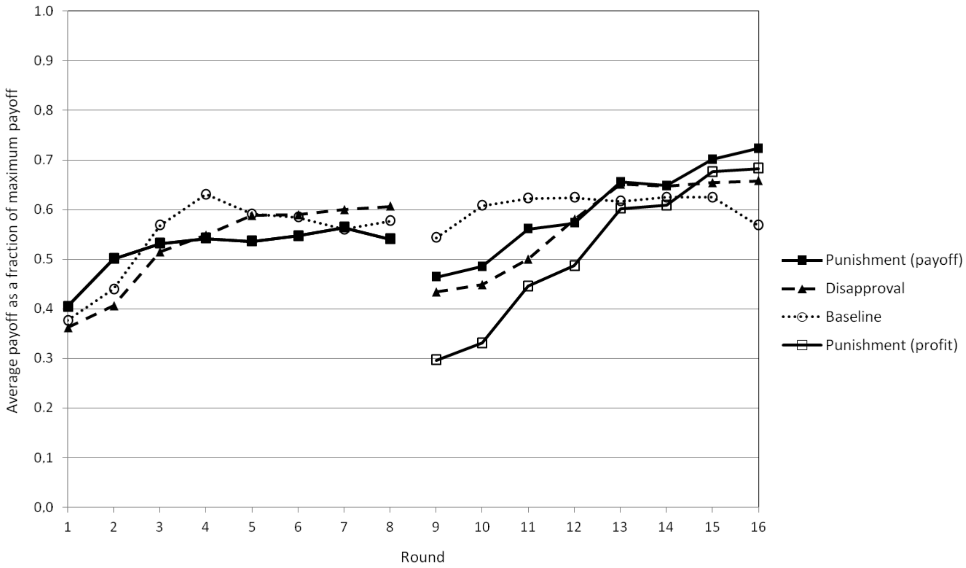


Figure 5: *Payoffs and profits per round and treatment*

In Stage 2, the average payoff in *Baseline* initially rises slightly above the level reached at the end of Stage 1 and then stays at that level. This reflects that individual and collective outcomes in *Baseline* remain at, or quickly return to those attained at the end of Stage 1. The average payoff in *Punishment* and *Disapproval* initially drops substantially to slightly above the initial round 1 level, subsequently rising steadily and surpassing the *Baseline* average payoff in the second half of Stage 2. The initial drop is due to the extensive attempts in both treatments to move the collective outcome updwards, with negative consequences for coordination. The subsequent upward trend in average payoff stems from the gradual individual coordination improvements as well as from about half the groups in *Punishment* and two groups in *Disapproval* improving their minimum efforts.

Figure 5 further shows that at the beginning of Stage 2, the average profit in *Punishment* is only at 30 percent of the maximum achievable payoff and 17 percentage points below the average payoff. Clearly, the consequences of the cost of punishment on overall efficiency are considerable. They seem to decrease over time as both the average payoff and average profit eventually reach about 70 percent, which is higher compared to *Disapproval* and *Baseline*. Nonetheless, the across-treatment efficiency differences at the end of Stage 2 are minor compared to the effort differences observed in Figures 1 and 2. Indeed, the Mann-Whitney U test suggests that average payoff is significantly higher in *Punishment* compared to *Baseline* in the final round 16 ($p < 0.05$), while other across-treatment differences are not significant. When including the cost of punishment, profit is significantly lower in *Punishment* compared to *Baseline* in the first two rounds of Stage 2 ($p < 0.10$ in round 9; $p < 0.01$ in round 10) but significantly higher in the final round 16 ($p < 0.10$). Results from t-tests in OLS estimation lead to similar conclusions.[8]

	Treatment	Stage 1-2	Round 1-9	Round 2-10	Round 3-11	Round 4-12	Round 5-13	Round 6-14	Round 7-15	Round 8-16
Average payoff change (in p.p.)	Punishment (payoff)	$8.07^{tt,ss}$	5.93^{t}	-1.60_{bb}	2.96	3.13	$11.94^{tt,ss}_{bb}$	$10.18^{tt,ss}$	$13.70^{ttt,sss}$	$18.35^{ttt,sss}_{bbb,dd}$
	Punishment (profit)	-0.46	$-10.82^{tt,ss}_{bbb,dd}$	$-17.03^{ttt,ss}_{bbb,ddd}$	-8.61_{b}	-5.49	6.61	6.21	$11.18^{ttt,sss}$	$14.30^{tt,ss}_{bbb}$
	Disapproval	4.49	7.21	4.33	-1.44	3.21	6.33^{s}	5.77^{s}	5.37^{ss}	5.13
	Baseline	$6.31^{ttt,ss}$	16.71^{tt}	$16.83^{ttt,ss}$	5.41	-0.72	$2.52^{tt,ss}$	$4.09^{tt,ss}$	6.49^{s}	-0.84
Fraction of groups with an average-payoff increase	Punishment (payoff)	0.73	0.67_{b}	0.58_{bb}	0.67	0.75	0.83	0.75	$0.75_{b,dd}$	$0.83_{bbb,dd}$
	Punishment (profit)	0.53	$0.33_{bbb,d}$	$0.17_{bbb,dd}$	0.42	0.58	0.58	0.67	$0.75_{b,dd}$	$0.75_{bbb,d}$
	Disapproval	0.50	0.58	0.58	0.50	0.58	0.50	0.42	0.50	0.33_{b}
	Baseline	0.55	0.88	0.88	0.88	0.25	0.50	0.63	0.38	0.00

The "t" and "s" superscripts denote a significant difference across stages or across a round-pair (see the top row), using an appropriate t-test and Wilcoxon signed-rank test, respectively. The "b" resp. "d" symbols denote a significant difference across stages or across a round-pair between the treatment directly above the symbol and *Baseline* resp. *Disapproval*, using a t-test (in the first block), or Mann-Whitney U test (in the second block). Significance levels are 1%, 5% and 10% for three, two and one superscripts or symbols of a kind in a given cell.

Table 5: *Between-stage and between-round welfare changes in each treatment*

The first block of results in Table 5 displays *payoff changes* between Stages 1 and 2. From about the same level of 52-54 percent of the maximum achievable payoff in Stage 1, average payoff in Stage 2 increases by 8.07 percentage points (0.21 euros) in *Punishment*, 4.49 percentage points (0.12 euro) in *Disapproval*, and 6.31 percentage points (0.16 euro) in *Baseline*. The overall welfare gains in *Punishment* and *Baseline* are significant by both the t-test described above and the Wilcoxon signed-rank test. In *Punishment*, the overall profit (including punishment costs) decreases insignificantly by 0.5 percentage points (0.01 euro). Because of high punishment costs, *Punishment* initially features relatively large profit decreases that are significant in the first two round-pairs. The pattern eventually reverses: In the last four round-pairs, the total gains in *Punishment* reach over 10 percentage points and are significant by the t-test as well as the Wilcoxon signed-rank test applied to groups' average payoffs. *Disapproval* generally features small gains throughout, while *Baseline* features large and significant gains in the first two round-pairs.

The second block in Table 5 complements the first one by displaying the fraction of groups with an average payoff increase between stages. The figures tend to

[8] As for the effort-choice comparisons in section 3.2, we regress individual payoffs on treatment dummies interacted with a stage dummy or round dummies. The estimations are based on a panel of 256 subjects with 16 rounds of payoffs each. As above, we use the cluster-robust estimator of variance allowing for intra-group correlation of observations. Other estimation details are identical to the estimation for effort choice changes.

confirm the general picture from the first block. Initially, the treatment effect between *Punishment* and *Baseline* and between *Punishment* and *Disapproval* is significantly negative, at least in the first two round-pairs. However, in the last two round-pairs, the gains are significantly higher in *Punishment*.

Overall, sanctions have a substantial effects on the minimum effort attained within the game, in comparison to both alternative treatments and Stage 1, yet their effect on overall efficiency is mixed. Their direct monetary cost in the first rounds of Stage 2 and the miscoordination implied by attempts to shift the outcome of the game upward leads to significant losses in overall efficiency. This is reversed for the final rounds of Stage 2, that is, once subjects have managed to coordinate on new and Pareto superior equilibria, sanctions have almost disappeared, a mild positive effect on overall efficiency is observed. This pattern is very similar to what is observed in public goods game.

5 Concluding remarks

On a general note, our experiment reveals that, first, voluntary sanctions are widely used by subjects in the context of a Pareto ranked coordination game. The use of sanctions seem to be in part due to strategic considerations but a careful look at the last rounds suggest that some participants are motivated by other reasons than simply influencing other's future moves. Second, sanctions move coordination upwards in terms of Pareto ranked outcomes in a robust and stable way, both in comparison to Stage 1 and to the other treatments in Stage 2. Third, the efficiency gains associated with introducing the sanctioning mechanism are initially negative (partly due to the high punishment costs) but ultimately turn out significantly positive. After the initial episode of miscoordination and adjustment to the new conditions, the sanctions can substantially improve coordination outcomes. In comparison, *Disapproval* seems to have only a transient and limited effect, even though subjects massively disapprove of others but with little effects. It seems that the change implied by the possibility of sanctions does not have more effect in initially raising efforts that mere disapproval communication, but it appears essential to maintain efforts at these higher levels on a longer range.

Our findings raise several issues. First, our results suggest that communication – more precisely, *ex post* disapproval communication – may not be a strong enough efficiency-enhancing coordination device in particularly adverse conditions, such as when there are large groups, anonymous actions, or a history of inefficient coordination. By contrast, punishment opportunities seem more powerful under the same conditions, despite the fact that they imply a monetary cost to their user, unlike cost-free disapproval. In this sense, our findings resemble the effect of punishment found in cooperation games, thus possibly contributing to an explanation of why efficient coordination arises in real economic settings.

A related methodological point would be that it may generally be more appropriate to test the (relative) power of efficiency-enhancing coordination devices after allowing for a history of low-efficiency coordination, as has been demonstrated in cooperation settings (e.g., Fehr and Gaechter, 2000). In coordination settings such as ours, the reasons for doing so are even more pronounced due to a possibly stronger path-dependence typically observed in coordination games. Without initially allowing for a low-efficiency coordination, a rather mild initial nudge provided by an otherwise weak coordination device may be sufficient to improve efficiency and to sustain it. Hence the relative strength of different mechanisms may be hard to assess, since small and not easily observed differences in initial conditions, e.g. subject's expectations, may have a great empirical impact.

Although we are clearly able to identify a positive effect of sanctions on efficient coordiantion in the minimum effort games, much is left unknown among which the effect of sanctions in a Pareto ranked game when the game is not repeated, or when a stranger matching design is used, the role of different punishment technology and so on. Basically, all the questions that arose in the case of public goods and were subsequently addressed by numerous experiments seem to be valid lines of inquiry. To give just an example, we observe that with a punishment efficiency of 2 to 1, the total welfare effects of punishment are close to zero, while they only start to be significantly positive in the last two rounds of Stage 1. In public good games, a similar threshold is found : in their systematic study of the total welfare effects of punishment, Nikiforakis and Normann (2008) find that sanctions have a positive effect from an level of efficiency of 3 to 1. In addition, more investigation on the motivation for sanctioning in the case of a Pareto ranked coordination game may shed some light about the roots of punishment in cooperation contexts.

Finally, we suggest a future line of research about the effect of punishment opportunities in more general coordination games, i.e., not Pareto-ranked coordination games. If the motive for punishment is purely instrumental (inducing others to raise their effort) or based on group-level reciprocity or the enforcement of social norms, it is possible that sanctioning opportunities may improve the stability of any arbitrary equilibrium: if players who occasionally or randomly deviate from an equilibrium face sanctions, such deviations may become less frequent. Deviations, even if only erroneous, are made much more costly than in the absence of such punishment. This may have interesting consequences for the formation of conventions or social norms and their decentralized enforcement – a question critical to economics in many respects (Knack and Keefer, 1997) – even though the roots of their transgression may not actually be malevolent or ill-intended. The specific conditions under which informal sanctions help coordination at a broader level are of course an empirical question.

References

Anderson, C., and L. Putterman (2006): "Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism," *Games and Economic Behavior*, 54(1), 1–24.

Bartling, B., and U. Fischbacher (2011): "Shifting the blame: On delegation and responsibility," *The Review of Economic Studies*, p. rdr023.

Becker, G., and K. Murphy (1992): "The division of labor, coordination costs, and knowledge," *The Quarterly Journal of Economics*, 107(8), 1137–1160.

Bicchieri, C. (2006): "The grammar of society: The emergence and dynamics of social norms," .

Blume, A., and A. Ortmann (2007): "The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria," *Journal of Economic Theory*, 132, 274–290.

Brandts, J., and D. Cooper (2006): "A change would do you good: An experimental study on how to overcome coordination failure in organizations," *American Economic Review*, 96(3), 669–693.

——— (2007): "It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure," *Journal of the European Economic Association*, 5(6), 1223–1268.

Camerer, C. F., and M. Knez (1994): "Creating 'expectational assets' in the laboratory: 'Weakest-link' coordination games," *Strategic Management Journal*, 15, 109–109.

Capraro, V., and D. G. Rand (forthcoming): "Do the Right Thing: Preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality," *Judgement and Decision Making*.

Carpenter, J. (2007): "The demand for punishment," *Journal of Economic Behavior and Organization*, 62(4), 522–542.

Carpenter, J., and P. H. Matthews (2009): "What norms trigger punishment?," *Experimental Economics*, 12(3), 272–288.

Carpenter, J. P., and P. H. Matthews (2012): "Norm enforcement: anger, indignation, or reciprocity?," *Journal of the European Economic Association*, 10(3), 555–572.

Charness, G., and M. Rabin (2002): "Understanding social preferences with simple tests," *Quartely Journal of Economics*, 117, 817869.

Chaudhuri, A. (2011): "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature," *Experimental Economics*, 14(1), 47–83.

Devetag, G., and A. Ortmann (2007): "When and why? A critical survey on coordination failure in the laboratory," *Experimental Economics*, 10(3), 331–344.

Dugar, S. (2010): "Nonmonetary sanctions and rewards in an experimental coordination game," *Journal of Economic Behavior and Organization*, 73(3), 377–386.

Engelmann, D., and H.-T. Normann (2010): "Maximum effort in the minimum-effort game," *Experimental Economics*, 13(3), 249–259.

Falk, A., E. Fehr, and U. Fischbacher (2005): "Driving forces behind informal sanctions," *Econometrica*, 73(6), 2017–2030.

——— (2008a): "Testing theories of fairness: Intentions matter," *Games and Economic Behavior*, 62(1), 287–303.

——— (2008b): "Testing theories of fairnessIntentions matter," *Games and Economic Behavior*, 62(1), 287–303.

Falk, A., and U. Fischbacher (2006): "A theory of reciprocity," *Games and Economic Behavior*, 54, 293–315.

Fatas, E., T. Neugebauer, and J. Perote (2006): "Within-team competition in the minimum effort coordination game," *Pacific Economic Review*, 11(2), 247–266.

Fehr, E., and U. Fischbacher (2004a): "Social norms and human cooperation," *Trends in Cognitive Sciences*, 8, 185–190.

——— (2004b): "Third-party punishment and social norms," *Evolution and human behavior*, 25(2), 63–87.

Fehr, E., and S. Gächter (2005): "Altruistic punishment in humans," *Nature*, 415, 137–140.

Fehr, E., and S. Gaechter (2000): "Cooperation and punishment in public goods experiments," *American Economic Review*, 90, 980–994.

Fehr, E., and K. M. Schmidt (1999): "A theory of fairness, competition, and cooperation," *Quarterly Journal of Economics*, 114, 817–868.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments.," *Experimental Economics*, 10, 171–178.

Goeree, J. K., and C. A. Holt (2005): "An experimental study of costly coordination," *Games and Economic Behavior*, 51(2), 349–364.

Greiner, B. (2004): "An online recruitment system for economic experiments," in *Forschung und wissenschaftliches Rechnen 2003*, ed. by K. Kremer, and V. Macho. Gttingen : Ges. fr Wiss. Datenverarbeitung,.

Herrmann, B., C. Thöni, and S. Gächter (2008): "Antisocial punishment across societies," *Science*, 319(5868), 1362–1367.

Kimbrough, E. O., and A. Vostroknutov (2016): "Norms make preferences social," *Journal of the European Economic Association*, 14(3), 608–638.

Knack, S., and P. Keefer (1997): "Does social capital have an economic payoff? A cross-country investigation," *The Quarterly Journal of Economics*, 112(4), 1251–1288.

Kopányi-Peuker, A., T. Offerman, and R. Sloof (forthcoming): "Team production benefits from a permanent fear of exclusion," *European Economic Review*.

Kezdi, G. (2004): “Robust standard error estimation in fixed-effects panel models.,” *Hungarian Statistical Review Special*, 9, 96–116.

Lopez-Perez, R. (2008): “Aversion to norm-breaking: A model,” *Games and Economic Behavior*, 64, 237–267.

Masclet, D., C. Noussair, S. Tucker, and M.-C. Villeval (2003): “Monetary and nonmonetary punishment in the voluntary contributions mechanism,” *American Economic Review*, 93, 366–380.

Masclet, D., C. N. Noussair, and M.-C. Villeval (2013): “Threat and punishment in public good experiments,” *Economic Inquiry*, 51(2), 1421–1441.

Nikiforakis, N., and H.-T. Normann (2008): “A comparative statics analysis of punishment in public-good experiments,” *Experimental Economics*, 11(4), 358–369.

Ostrom, E., J. Walker, and R. Gardner (1992): “Covenants with and without a sword: Self governance is possible,” *American Political Science Review*, 86, 404–417.

Rabin, M. (1993): “Incorporating fairness into game theory and economics,” *American Economic Review*, 83, 1281–1302.

Riedl, A., I. M. Rohde, and M. Strobel (2015): “Efficient coordination in weakest-link games,” *The Review of Economic Studies*, p. rdv040.

Rogers, W. H. (1993): “sg17: Regression standard errors in clustered samples.,” *Stata Technical Bulletin*, 13, 19–23.

Romero, J. (2015): “The effect of hysteresis on equilibrium selection in coordination games,” *Journal of Economic Organization and Behavior*, 111, 88–105.

Schram, A., and G. Charness (2015): “Inducing social norms in laboratory allocation choices,” *Management Science*, 61(7), 1531–1546.

Van Huyck, J., R. Battalio, and R. Beil (1990): “Tacit coordination games, strategic uncertainty, and coordination failure,” *American Economic Review*, 80(1), 234–248.

Van Huyck, J., R. Battalio, and F. Rankin (2007): “Evidence on learning in coordination games,” *Experimental Economics*, 10(3), 205–220.

Weber, R. A. (2006): “Managing Growth to Achieve Efficient Coordination in Large Groups,” *The American Economic Review*, 96(1), 114.

PUBLICATION 4: Serge Macé et Fabrice Le Lec (2011), “On fatalistic long-term health behaviour”, ***Journal of Economic Psychology***, 32:434-439.

Journal of Economic Psychology 32 (2011) 434–439

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

On fatalistic long-term health behavior

Serge Macé *, Fabrice Le Lec [1]

Institut Catholique de Lille, Faculté Libre de Sciences Economiques et de Gestion – LEM CNRS UMR 8179, 41 rue du Port, 59 016 Lille Cedex, France

A R T I C L E I N F O

Article history:
Received 18 May 2009
Received in revised form 11 February 2011
Accepted 24 February 2011
Available online 5 March 2011

JEL classification:
I12
D03

PsycINFO classification:
3365

Keywords:
Hedonic adaptation
Health
Aging
Prospect theory

A B S T R A C T

Many adults have an overly pessimistic view of old age because they fail to correctly predict their ability to hedonically adapt to old-age health related problems. A standard utility model where the marginal utility of health is higher at a lower level of health predicts that this overly pessimist view raises the incentive for healthy behavior. But this is at odds with empirical research that indicates that people with more negative aging stereotypes tend to adopt less healthy practices, transforming this negative view into a self-fulfilling prophecy. The aim of this note is to show that this fatalistic behavior can be explained through prospect theory by modelling this overly pessimistic view of old age as a failure to predict the change in the reference point due to hedonic adaptation. Given the diminishing sensitivity in the loss domain, people undervalue the future marginal value of health investment and may therefore underinvest in health as long as loss aversion is not too strong.

1. Introduction

Many people have an overly pessimistic view of old age. They tend to systematically associate it with languor, frailty, illnesses, dependence or asexuality. In fact, those traits are far from being as frequent as people believe as ageism surveys indicate (see in particular Palmore, 1998). And they are not experienced as negatively as expected because people partially adapt hedonically to them. Both hedonic adaptation to old age and its misprediction have been now well documented. It is indeed a classic result of psychology that people tend to adapt hedonically to new conditions (Frederick & Loewenstein, 1999) and to deteriorated health conditions in particular (Albrecht & Devlieger, 1999; Ashby, O'Hanlon, & Buxton, 1994; Brickman, Coates, & Janoff-Bulman, 1978; Oswald & Powdthavee, 2008; Schulz & Decker, 1985; Wu, 2001). Young people however are not fully aware of their future ability to adapt. They typically tend to underestimate it. For instance, people experiencing chronic illness, long-term treatment, and disability report greater happiness or quality of life ratings than what healthy people predict they would under similar circumstances (Buick & Petrie, 2002; Lacey, Fagerlin, et al., 2006; Riis et al., 2005; Smith, Sherriff, Damschroder, Loewenstein, & Ubel, 2006; Ubel, Loewenstein, Schwarz, & Smith, 2005; Walsh & Ayton, 2009). Hedonic adaptation and the failure to fully anticipate it also help explain the "old age paradox", that is the fact that old people are happier than expected, and even sometimes happier than younger people (Blanchflower & Oswald, 2008; Lacey, Smith, & Ubel, 2006; Yang, 2008).

* Corresponding author. Tel.: +33 3 20 13 41 01.
E-mail addresses: serge.mace@icl-lille.fr (S. Macé), fabrice.lelec@icl-lille.fr (F. Le Lec).
[1] Tel.: +33 3 20 13 40 92.

0167-4870/$ - see front matter
doi:10.1016/j.joep.2011.02.016

This negative old age perception is likely to affect health behaviors. It may lead some to make huge efforts to delay these effects by choosing a healthier lifestyle (sports, healthy goods, cosmetic creams or surgery…) but may result for others in adopting a fatalistic behavior by doing nothing and enjoying life to the maximum before getting old. In the latter case, the negative perception of old age gives rise to a self-fulfilling prophecy. Available studies suggest that such an effect could be predominant. Using the Baltimore Longitudinal Study, Levy, Zonderman, Slade, and Ferruci (2009) showed for example that people aged between 19 and 49 holding negative age stereotypes had a greater likelihood of experiencing cardiovascular events up to 38 years later than individuals with more positive age stereotypes after controlling for the main risk factors. Levy and Myers (2004) found that older people with more positive perceptions of aging adopted on average a more preventive approach like eating a balanced diet, exercising and following directions for taking prescribed medications, over the next two decades after controlling for age, education, functional health, gender, race, and self-rated health. Other studies also observed that seniors with low expectations regarding further aging are less likely to take part in physical activities (Sanchez, Torres, & Mena, 2009; Sarkisian, Prohaska, Wong, Hirsch, & Mangione, 2005), to have a primary care provider or to receive vaccinations (Goodwin, Black, & Satish, 1999; Sarkisian, Hays, & Mangione, 2002), again after adjusting for socio-demographic characteristics. The consequence is that people with more positive perceptions of aging report better health, when health is measured by functional abilities (Levy, Slade, & Kasl, 2002) or by longevity (Levy et al., 2002).[2]

The previous observations raise a major theoretical issue since a fatalistic behavior cannot be straightforwardly accounted for in the standard utility framework in which the marginal utility of health is higher at a lower level of health. In other words, in this framework, a negative view of old age should lead people to pay more attention to their long-term health, not less. The aim of this note is to show that such a fatalistic health behavior can be explained with prospect theory (Kahneman & Tversky, 1979; Thaler, 1980, 1985) by modelling the overly pessimistic view of old age as a failure to predict the change in the reference point due to hedonic adaptation. The core idea is that a given improvement in future health from a low level may be coded as a reduction of a loss when not anticipating hedonic adaptation, but as a gain when anticipating it. Given the diminishing sensitivity in the loss domain, the failure to predict adaptation when old may lead individuals to strongly underestimate the future marginal value of a moderate health improvement at that age. Loss aversion, which captures the fear of health deterioration as a consequence of aging, plays as a countervailing force though. This is so since the steeper utility for losses, the larger the benefits from investments in health at an older age. Hence, those most likely to underinvest in health are people who do not anticipate hedonic adaptation and are least loss averse.

2. A prospect theory approach to health valuation under hedonic adaptation

Consider for the sake of simplicity an individual living two periods: period 1 (young) and period 2 (old). The first-period health capital h_1 is given and the second-period one h_2 depends on the health investment made during the first period. If the individual does not invest in her health capital, she will reach when old a minimum level $h_{min} > 0$, which allows her to live. Furthermore, whatever efforts she makes to preserve her health and to "stay young", she ends up less healthy when old ($h_2 < h_1$). If I represents any given health improvement starting from the minimum health level, these assumptions can be summarized as:

$$h_2 = h_{min} + I \quad \text{with} \quad I < h_1 - h_{min} \tag{1}$$

When the individual is young and has to decide whether or not to engage in healthy or unhealthy practices, she does so on the basis of the anticipated value of these practices. In prospect theory, people subjectively evaluate situations in terms of losses and gains compared to a reference point. But hedonic adaptation may have a strong influence on this evaluation. Following Frederick and Loewenstein's (1999) idea of "shifting adaptation", we model hedonic adaptation to a negative event as a decrease in the reference point. Here, people adapt hedonically to old-age health related problems by reducing their health reference point, that is the health level below which they consider being in the loss domain. As long as they do not anticipate hedonic adaptation however, they base their evaluation on a higher health reference point. For the sake of simplicity, suppose that h_{min} is the health reference state that the individual should have chosen had she correctly predicted hedonic adaptation but that the individual uses her present health state h_1 as the reference point to measure the future consequences of her action.[3] Compared to the minimum level h_{min}, any health improvement I is thus coded as a gain after taking into account hedonic adaptation when old, but as a loss reduction when young, that is without anticipating hedonic adaptation. Suppose for the sake of simplicity and following Grossman (1972) that the health level h can be measured on a cardinal scale. And let V be a standard prospect theory S-shaped value function defined over gains and losses in health with respect to a reference state r. A simple specification of V is given by:

[2] Based again on the Baltimore Longitudinal Study, the authors showed that people with positive perceptions of aging live up to 7.5 years longer over a 23 years period than those with less positive perceptions of aging when age, gender, socioeconomic status, loneliness, and functional health are included as covariates.

[3] Both are simplifications. First the health reference state after hedonic adaptation may be a higher than h_{min}. Second, it is likely that people are capable of partially anticipating their future adaptation, thus basing their choices on a lower reference point than h_1. As long as the gap between both reference points is large enough, it does not change the consequences of the model.

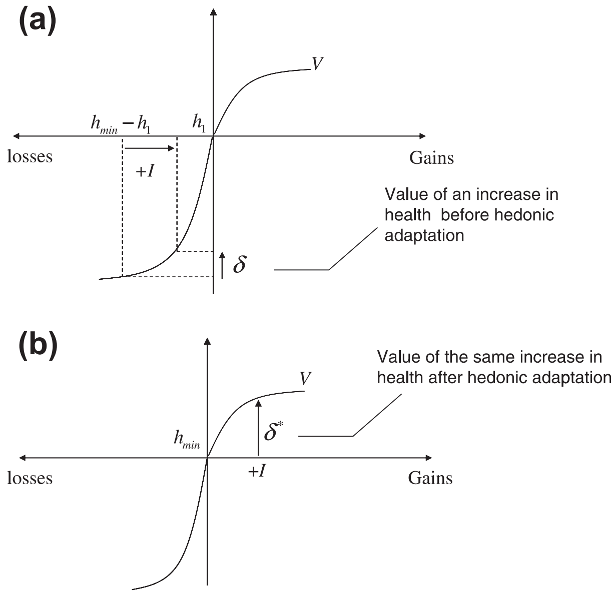


Fig. 1. Value of health improvement before and after adaptation.

$$V(h,r) = \begin{cases} v(h-r) & \text{if } h > r \\ 0 & \text{if } h = r \\ -\lambda v(r-h) & \text{if } h < r \end{cases} \quad (2)$$

with v defined on R_+, $v(0) = 0$, $v' > 0$, $v'' < 0$, and $\lambda \geqslant 1$ to capture loss aversion.[4] Given (2), the additional subjective value brought by any health improvement I beyond h_{min} is given by $V(h_{min} + I, r) - V(h_{min}, r)$. It depends on the reference state r and so on the anticipation of hedonic adaptation. When the individual correctly predicts adaptation and takes h_{min} as the health reference state, the value ascribed to some health improvement I calculated from h_{min}, denoted δ^*, is given by $\delta^* = V(h_{min} + I, h_{min}) - V(h_{min}, h_{min})$ or after replacement by:

$$\delta^* = v(I) \quad (3)$$

When the individual does not predict hedonic adaptation and takes h_1 as the health reference state, the predicted value ascribed to the same health improvement I, denoted δ, corresponds to $\delta = V(h_{min} + I, h_1) - V(h_{min}, h_1)$ or given (2) and after rearrangement by:

$$\delta = \lambda[v(h_1 - h_{min}) - v(h_1 - h_{min} - I)] \quad (4)$$

The two values δ and δ^* have no particular reasons to be identical in the general case. In particular, as illustrated in Fig. 1, it is possible for some value of I that $\delta < \delta^*$.

In this figure, the diminishing sensitivity in the domain of losses makes the value given to I lower before than after adaptation. In many circumstances, this low valuation of I makes health underinvestment possible. Suppose for instance that the individual has the choice between two lifestyles, healthy and unhealthy. The healthy style is not more expensive but implies a hedonic sacrifice (exercising, going on a diet, not smoking or drinking, etc.), which can be viewed as an investment that gives a health improvement I at the second period. The decision maker has to compare the hedonic cost of the healthy lifestyle at the first period, denoted c, to the estimated additional value given by health improvement for the second period δ. If $\delta < c < \delta^*$, she chooses the unhealthy lifestyle whereas she would have chosen the healthy one had she predicted hedonic adaptation. An overly pessimistic view about old-age health related problems thus turns into the kind of fatalistic behavior that seems to be observed empirically.

In the general case, we can state the following propositions (see Appendix A for the proofs):

Proposition 1. *In the absence of loss aversion to health deterioration, the individual always undervalues any future health improvement $I < h_1 - h_{min}$ when not anticipating adaptation.*

[4] Note that Eq. (2) assumes the same curvature of utility for health gains and losses. There is some evidence that the utility function is more curved in the gain than in the loss domain though differences are not massive (see estimates in Fox & Poldrack, 2008). This does not dramatically affect our results (see footnote 7).

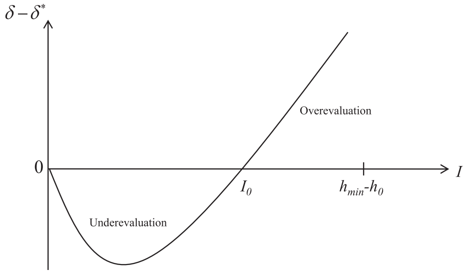


Fig. 2. Difference in the valuation of health improvement before and after adaptation for various levels of health improvement (with $\lambda > 1$).

Without loss aversion, a positive gain and a reduction of a loss are identically valued only for $I = h_1 - h_{min}$.[5] Since we assume that the individual cannot maintain her health level between the two periods ($I < h_1 - h_{min}$), the value given to I is lower if it is considered as a reduction of a loss than as a gain.

Consider now that $\lambda > 1$. Given Eqs. (3) and (4), the value ascribed to a marginal health improvement from h_{min} is equal to $v'(0)$ after hedonic adaptation but to $\lambda v'(h_1 - h_{min})$ before. Define as *moderate* any level of loss aversion $\lambda > 1$ such that $v'(0) > \lambda v'(h_1 - h_{min})$ or simply $1 < \lambda < v'(0)/v'(h_1 - h_{min})$.[6]

Proposition 2. *If loss aversion is moderate, there exists $I_0 \in (0, h_1 - h_{min})$ such that any health improvement $I < I_0$ is undervalued when the individual does not anticipate her adaptation to health deterioration.*

When loss aversion is moderate, there is always an area of undervaluation for relatively low health improvements.[7] In this zone, loss aversion does not totally counteract the diminishing sensitivity of losses.[8] Yet, for $I > I_0$, the individual *overvalues* health improvement, that is considers I more favorably than she would have done had she predicted her adaptation. Hence, in the general case of moderate loss aversion, the individual simultaneously undervalues mild heath improvements and overvalues higher health improvements as illustrated in Fig. 2.

It is worth noting that even when the individual overvalues any health improvement $I > I_0$, whether or not the individual will overinvest in health will depend first on whether possible health improvements higher than I_0 can be reached, and second on the corresponding opportunity cost. A fatalistic behavior happens only when these high future health levels cannot be reached or only at a prohibitive opportunity cost.

Proposition 3. *The higher loss aversion, the lower the undervaluation and the smaller the undervaluation area.*

Eq. (4) shows that a higher loss aversion increases the value given to any health improvement in the domain of losses and for this reason reduces I_0, the threshold under which the individual undervalues any health improvement. Therefore, loss aversion, which captures here the fear of health deterioration, acts as a countervailing force against the diminishing sensitivity to losses. It protects people from undervaluing either too much or systematically any health improvement.

3. Conclusion

Many adults do not seem to pay much attention to their long-term health. Usual behavioral explanations include preference for immediate gratification, myopia or a feeling of invulnerability. This article focuses on another possible explanation: People who fail to correctly predict hedonic adaptation to old age-related health problems may be led to underinvest in their long-term health because of the diminishing sensitivity to health losses, though loss aversion plays as a countervailing force. This explanation goes further than alternative ones by putting forward the role of incorrect old age health perceptions, a phenomenon that is widely documented empirically. In policy terms, it suggests that fighting negative stereotypes about aging might have a potentially large impact on long-term health practises.

[5] Eqs. (3) and (4) indicate that for $I = h_1 - h_{\min}$ we have $\delta = \delta^* = v(h_1 - h_{\min})$.

[6] Technically, this condition is automatically satisfied with some standard specification used in prospect theory like the power function $v(I) = I^\alpha$ with $0 < \alpha < 1$ since $\lim_{I \to 0} v'(I) = +\infty$.

[7] No moderate loss aversion would mean that a marginal health improvement from $h_{\min}$ is more valued when it is considered as the reduction of a loss than as a gain. Given that the marginal value of health improvement is increasing with I in the domain of losses, it will also be the case for any higher values of I so that no area of undervaluation appears.

[8] This proposition has been established by assuming the same curvature in the utility function for gains and losses (Eq. (2)). But for an individual to undervalue health investment, what only matters is that there is enough diminishing sensitivity in the domain of losses when she is closed to $h_{\min}$. Assume instead that $V = v^+(h - r)$ and $V = -\lambda v^-(r - h)$ with v^+ and v^- concave but not necessarily identical. Proposition 1 does not hold unless we restrict it to some particular specifications of v^+ and v^-. However, the second proposition still holds, but only now if there is enough convexity in the domain of losses to ensure that there exists $\lambda \geqslant 1$ such that the value ascribed to a marginal health improvement from $h_{\min}$ after adaptation, now equal to $v^{+\prime}(0)$, is higher than the same value ascribed before adaptation $\lambda v'(h_1 - h_{\min})$.

Appendix A. Proofs of propositions 1–3

Proposition 1. *In the absence of loss aversion to health deterioration, the individual always undervalues any future health improvement* $I < h_1 - h_{min}$ *when not anticipating adaptation.*

Proof. Denote $m(I) = \delta(I) - \delta^*(I)$, the difference in valuation of health improvement before and after hedonic adaptation for any health improvement $I < h_1 - h_{min}$. Given Eqs. (3) and (4), we have:

$$m(I) = \lambda[v(h_1 - h_{min}) - v(h_1 - h_{min} - I)] - v(I) \tag{A.1}$$

with v defined on R_+, $v(0) = 0$, $v' > 0$, $v'' < 0$, and $\lambda \geqslant 1$ to capture loss aversion.

If $\lambda = 1$, $m(I) = v(h_1 - h_{min}) - v(h_1 - h_{min} - I) - v(I)$. Let $y = h_1 - h_{min}$ and $I = ty$. Then $m(I) = v(y) - v((1-t)y) - v(ty)$. Since by strict concavity of v we have $v(ty) + v((1-t)y) > v(y)$, $m(I) < 0$, which is equivalent to say that the individual undervalues health improvement. □

Proposition 2. *If loss aversion is moderate* $(1 < \lambda < v'(0)/v'(h_1 - h_{min}))$, *there exists* $I_0 \in (0, h_1 - h_{min})$ *such that any health improvement* $I < I_0$ *is undervalued when the individual does not anticipate her adaptation to health deterioration.*

Proof. Start again from the difference in the valuation of health improvement m as defined by Eq. (A.1). Note first that m is strictly convex. Indeed, $m''(I) = -\lambda v''(h_1 - h_{min} - I) - v''(I)$ is strictly positive since $v'' < 0$ by assumption.

Let us show that m is negative for low levels of I. First $m(0) = 0$ straightforwardly, and second $m'(0) = \lambda v'(h_1 - h_{min}) - v'(0) < 0$ because of moderate loss aversion. This implies that for low levels of I, m takes negative values. In plain words, for low levels of I, the valuation is lower before than after adaptation. But for the maximum possible health improvement, *i.e.* $h_1 - h_{min}$, the value before adaptation is greater to that after since

$$m(h_1 - h_{min}) = \lambda v(h_1 - h_{min}) - v(h_1 - h_{min}) > 0 \tag{A.2}$$

Given the continuity of V, there exists some level of health improvement I_0 for which $m(I_0) = 0$. Since m is strictly convex, m is negative on $(0, I_0)$ and positive on $(I_0, h_1 - h_{min}]$. □

Proposition 3. *The higher loss aversion, the lower the undervaluation and the smaller the undervaluation area.*

Proof. The first part of the proof is straightforward from Eqs. (3) and (4).

Now, define $I_0 \in (0, h_1 - h_{\min})$ as in Proposition 2 such that

$$\lambda[v(h_1 - h_{min}) - v(h_1 - h_{min} - I_0)] - v(I_0) = 0 \tag{A.3}$$

Note that $m'(I_0) = \lambda v'(h_1 - h_{min} - I_0) - v'(I_0) > 0$.

Differentiating (A.3) with respect to I_0 and λ implies:

$$[v(h_1 - h_{min}) - v(h_1 - h_{min} - I_0)]d\lambda + [\lambda v'(h_1 - h_{min} - I_0) - v'(I_0)]dI_0 = 0 \tag{A.4}$$

Rearranging gives:

$$\frac{dI_0}{d\lambda} = \frac{\overbrace{-[v(h_1 - h_{min}) - v(h_1 - h_{min} - I_0)]}^{<0}}{\underbrace{\lambda v'(h_1 - h_{min} - I_0) - v'(I_0)}_{m'(I_0)>0}} < 0 \tag{A.5}$$

An increase in loss aversion reduces the threshold I_0 under which the individual undervalues any health improvement. □

References

Albrecht, G. L., & Devlieger, P. J. (1999). The disability paradox: High quality of life against all odds. *Social Science and Medicine, 48*, 977–988.

Ashby, J., O'Hanlon, M., & Buxton, M. J. (1994). The time trade-off technique: How do the valuations of breast cancer patients compare to those of other groups? *Quality of Life Research, 3*, 257–265.

Blanchflower, D. G., & Oswald, A. J. (2008). Is well-being U-shaped over the life cycle? *Social Science & Medicine, 66*, 1733–1749.

Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology, 36*, 917–927.

Buick, D. L., & Petrie, K. J. (2002). "I know just how you feel": The validity of healthy women's perceptions of breast cancer patients receiving treatment. *Journal of Applied Social Psychology, 32*, 110–123.

Fox, C. R., & Poldrack, R. A. (2008). Prospect theory and the brain: Studies on the neuroeconomics of decision under risk. In P. Glimcher, C. Camerer, E. Fehr, & R. Poldrack (Eds.), *Handbook of neuroeconomics* (pp. 145–174). New York: Elsevier.

Frederick, S., & Loewenstein, G. (1999). Hedonic adaptation. In D. Kahneman, E. Diener, & N. Schwartz (Eds.), *Scientific perspectives on enjoyment, suffering, and well-being*. New York: Russell Sage Foundation.

Goodwin, J., Black, S., & Satish, S. (1999). Aging versus disease: The opinions of older black, Hispanic, and non-Hispanic white Americans about the causes and treatment of common medical conditions. *Journal of the American Geriatrics Society, 47*, 973–979.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy, 80*, 223–255.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.

Lacey, H., Fagerlin, A., Loewenstein, G., Smith, D., Riis, J., & Ubel, P. (2006). It must be awful for them: Healthy people overlook disease variability in quality of life judgments. *Judgment and Decision Making, 1*, 146–152.

Lacey, H., Smith, D., & Ubel, P. (2006). Hope I die before I get old: Mispredicting happiness across the adult lifespan. *Journal of Happiness Studies, 7*, 167–182.

Levy, B., & Myers, L. (2004). Preventive health behaviors influenced by self-perceptions of aging. *Preventive Medicine, 39*, 625–629.

Levy, B., Slade, M., & Kasl, S. (2002). Longitudinal benefit of positive self-perceptions of aging on functional health. *Journal of Gerontology, 57B*, 409–417.

Levy, B., Zonderman, A., Slade, M., & Ferruci, L. (2009). Age stereotypes held earlier in life predict cardiovascular events in later life. *Psychological Science, 20*, 296–298.

Oswald, A., & Powdthavee, N. (2008). Does happiness adapt? A longitudinal study of disability with implications for economists and judges. *Journal of Public Economics, 92*, 1061–1077.

Palmore, E. (1998). *The facts on aging quiz*. New York: Springer.

Riis, J., Loewenstein, G., Baron, J., Jepson, C., Fagerlin, A., & Ubel, P. (2005). Ignorance of hedonic adaptation to hemodialysis: A study using ecological momentary assessment. *Journal of Experimental Psychology, 134*, 3–9.

Sanchez, P., Torres, T., & Mena, B. (2009). Negative aging stereotypes and their relation with psychosocial variables in the elderly population. *Archives of Gerontology and Geriatrics, 48*, 385–390.

Sarkisian, C., Hays, R., & Mangione, C. (2002). Do older adults expect to age successfully? The association between expectations regarding aging and beliefs regarding healthcare seeking among older adults. *Journal of American Geriatrics Society, 50*, 1837–1843.

Sarkisian, C., Prohaska, T., Wong, M., Hirsch, S., & Mangione, C. (2005). The relationship between expectations for aging and physical activity among older adults. *Journal of General Internal Medicine, 20*, 911–915.

Schulz, R., & Decker, S. (1985). Long-term adjustment to physical disability: The role of social support, perceived control, and self-blame. *Journal of Personality and Social Psychology, 48*(S), 1162–1172.

Smith, D. M., Sherriff, R. L., Damschroder, L. J., Loewenstein, G., & Ubel, P. A. (2006). Misremembering colostomies? Former patients give lower utility ratings than do current patients. *Health Psychology, 25*, 688–695.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization, 1*, 39–60.

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science, 4*, 199–214.

Ubel, P., Loewenstein, G., Schwarz, N., & Smith, D. (2005). Misimagining the unimaginable: The happiness gap and healthcare decision making. *Health Psychology, 24*, S57–S62.

Walsh, E., & Ayton, P. (2009). What would It be like for me and for you? Judged impact of chronic health conditions on happiness. *Medical Decision Making, 29*, 15–22.

Wu, S. (2001). Adapting to heart conditions: A test of the hedonic treadmill. *Journal of Health Economics, 20*, 495–508.

Yang, Y. (2008). Social inequalities in happiness in the United States, 1972 to 2004: An age-period cohort analysis. *American Sociological Review, 73*, 204–226.

PUBLICATION 5: Fabrice Le Lec et Serge Macé (2012), “Investissement santé, prudence et adaptation hédonique à un risque santé”, ***Economie Publique/Public Economics***, 28-29:305-316.

Économie publique/Public economics

28-29 (2012/1-2)
Varia

Fabrice Le Lec et Serge Macé

Investissement santé, prudence et adaptation hédonique à un risque santé

Référence électronique
Fabrice Le Lec et Serge Macé, « Investissement santé, prudence et adaptation hédonique à un risque santé », *Économie publique/Public economics* [En ligne], 28-29 | 2012/1-2, mis en ligne le 20 décembre 2012, consulté le 23 décembre 2012. URL : http://economiepublique.revues.org/8975

Éditeur : Institut d'économie publique (IDEP)
http://economiepublique.revues.org
http://www.revues.org

Document accessible en ligne sur : http://economiepublique.revues.org/8975
Ce document est le fac-similé de l'édition papier.

public economics

économie publique

Revue de **l'Institut d'Économie Publique**

Deux numéros par an

nº 28-29 – 2012/1-2

idep

Investissement santé, prudence et adaptation hédonique à un risque santé

Fabrice Le Lec *
Serge Macé **

Résumé

L'investissement des individus dans leur santé dépend de la perception qu'ils ont des conséquences d'une variation de leur santé sur leur bien-être. De nombreux travaux en psychologie montrent cependant que cette perception est biaisée en raison de la difficulté des individus à anticiper leur faculté d'adaptation hédonique à un changement de leur santé. En utilisant un modèle à deux périodes, nous montrons que lorsque l'individu s'attend à une détérioration de sa santé objective, ce biais de perception l'amène à augmenter son investissement santé pour limiter une détérioration qui, subjectivement, ne se serait pas avérée aussi douloureuse à vivre, une fois passée la période d'adaptation. En présence d'un risque d'espérance nulle sur l'ampleur de cette détérioration, un nouvel effet survient : les individus tendent aussi à exagérer les conséquences en termes de bien-être de ce risque. La hausse de l'investissement santé est alors amplifiée si l'individu est prudent mais réduite, voire inversée si l'individu est imprudent.

*. Lille Économie et Management UMR CNRS 8179, Faculté Libre de Sciences Économiques et de Gestion, Université Catholique de Lille.

**. Faculté Libre de Sciences Économiques et de Gestion, Université Catholique de Lille.
Correspondance : fabrice.lelec@icl-lille.fr et serge.mace@icl-lille.fr.
Nous remercions David Crainich, Florence Jusot et un rapporteur anonyme pour les commentaires très utiles qu'ils ont apportés aux versions préliminaires de ce travail.

Summary

Individuals base their investment in health on how they perceive the consequences of a change in health on their welfare. When they underestimate their ability to hedonically adapt to a deterioration of health, individuals tend, under general assumptions, to increase their level of effort to reduce a deterioration that they perceive more painful than it is actually, once adapted to it. When we add a zero-mean risk on the magnitude of this deterioration, a new effect occurs: The individual also overestimates the welfare consequences of this risk. As a consequence, the increase in health investment is amplified if the individual is prudent, but reduced, if not reversed, if he is non-prudent.

Mots-clés : Prudence, adaptation, investissement santé.

Keywords: Prudence, Hedonic Adaptation, Health Investment.

J.E.L. : I12 – D91

1. Introduction

Les individus qui souffrent d'un handicap, d'une maladie chronique ou qui suivent un traitement contraignant ne sont généralement pas aussi malheureux que les personnes en bonne santé ne l'imaginent. La principale raison en est que les individus s'adaptent en partie à une dégradation de leur état physique, mais qu'ils sous-estiment *ex ante* cette faculté d'adaptation. L'adaptation hédonique définie par Frederick et Loewenstein (1999) comme « la réduction de l'intensité affective d'événements positifs ou négatifs » ainsi que sa sous-estimation ont fait l'objet de nombreux travaux ces dernières années. Il a ainsi été montré que les individus bien portants tendaient à sous-estimer le bien-être des individus souffrant d'insuffisance rénale (Riis *et al.*, 2005), de problèmes cardiaques (Wu, 2001), d'arthrites (Hurst *et al.*, 1994), du cancer du sein (Ashby *et al.*, 1994), ou de handicaps variés (Albrecht et Devliger, 1999 ; Oswald et Powdthavee, 2008). Ces travaux alimentent aujourd'hui un débat à la fois méthodologique et normatif sur les biais des QALY (Quality-Adjusted Life Years) et les critères à retenir en matière de santé publique (Dolan et Kahneman, 2008 ; Loewenstein et Ubel, 2008 ; Tessier, 2009). Comme ce biais modifie la perception que les individus ont de la détérioration de leur santé, il est susceptible d'affecter leurs efforts d'investissement pour la limiter. L'objectif de cet article est de montrer dans quel sens et par quels mécanismes.

D'un point de vue théorique, en l'absence d'adaptation hédonique, l'investissement santé dépend d'abord de la détérioration du niveau de santé objectif pour un effort donné. Plus celle-ci est importante, plus l'incitation à investir augmente. Dardanoni et Wasgaff (1990) ont cependant montré que l'investissement santé dépendait aussi du risque que cette détérioration soit plus faible ou plus forte qu'attendue. Ce comportement a été relié de manière plus récente dans la littérature au concept de prudence introduit par Kimball (1990) dans le cas d'un risque financier et appliqué dans des contextes différents aux problématiques de santé notamment par Bui *et al.* (2005) et Courbage et Rey (2006). Pour une même espérance de santé future, et face à un risque d'espérance nulle sur le niveau de détérioration de la santé, l'individu prudent tendra à accroître son investissement santé de la même manière qu'il augmenterait son épargne de précaution si le risque sur sa richesse future augmentait. S'il est imprudent, il réduira cet d'investissement.

L'incapacité des individus à prédire correctement leur adaptation hédonique est donc susceptible d'affecter l'investissement santé par ces deux canaux. Elle conduit à exagérer les conséquences négatives pour le bien-être, d'une part de la détérioration attendue de la santé, d'autre part du risque que cette détérioration soit plus faible ou plus forte qu'attendu. La présente note a pour objectif de distinguer clairement ces deux effets dans le cadre d'un modèle d'investissement santé à deux périodes où la sous-estimation de l'adaptation hédonique à une détérioration future de la santé est modélisée comme une baisse du niveau de santé subjectif prédit. À notre connaissance, peu de travaux économiques ont exploré le lien théorique entre l'adaptation hédonique et l'investissement santé à l'exception de Gjerde *et al.* (2005) dans un cadre dynamique qui ne fait cependant pas référence aux conséquences sur le comportement - prudent ou pas - des individus. L'approche en termes d'investissement santé adoptée est similaire à celle de Dardanoni et Wasgaff (1990) et Picone *et al.* (1998), mais adaptée à un modèle à deux périodes. Elle se distingue donc des modèles de prévention primaire qui modélisent les déterminants des efforts réalisés pour réduire la probabilité d'apparition des maladies (Eeckhoudt, Godfroid et Marchand, 1998). Elle se distingue aussi des travaux qui étudient le lien entre comportement de prudence et épargne de précaution face à un risque de *dépenses* de santé future (Palumbo, 1999 ; Courbage et Rey, 2006 ; Calvo et Arrondel, 2008). Nous montrons que la sous-estimation de l'adaptation hédonique à une détérioration de son état de santé conduit l'individu à augmenter son investissement santé. En présence d'un risque d'espérance nulle sur l'ampleur de cette détérioration, cette hausse est amplifiée si l'individu est prudent mais réduite voire inversée si l'individu est imprudent.

La partie 2 ci-dessous introduit le formalisme utilisé pour modéliser l'adaptation hédonique à une détérioration de la santé et sa sous-estimation. La partie 3 présente les implications dans un modèle d'investissement santé à deux périodes. La dernière partie conclut.

2. L'adaptation hédonique à la détérioration de la santé

Considérons un individu dont la fonction d'utilité est donnée par $u = u(c, h)$ où c désigne la consommation et h le capital santé, mesuré pour simplifier sur une échelle cardinale. L'utilité marginale des deux biens est décroissante. La fonction d'utilité est concave ($u_{cc}u_{hh} > u_{ch}$) et au moins trois fois dérivable par rapport à chacun de ses arguments. L'individu vit deux périodes (1 et 2). Le niveau de santé de la deuxième période est plus faible que celui de première période ($h_2 < h_1$) mais l'individu s'adapte partiellement à cette diminution de son capital santé. Suivant Groot (2000) sur ce point, on peut modéliser l'adaptation partielle à une détérioration de la santé en supposant que tout se passe comme si le niveau subjectif de santé [1] de la seconde période après adaptation h_2^{α}, ne décroît pas aussi rapidement que le niveau objectif de santé h_2. Pour simplifier, nous adoptons la transformation linéaire suivante introduite par Macé (2012) :

$$h_2^{\alpha} = h_2 + \alpha(h_1 - h_2) \tag{1}$$

où $\alpha \in [0, 1]$ mesure le degré d'adaptation de l'individu. Lorsque $\alpha = 0$, le niveau de santé subjectif décroît aussi rapidement que le niveau objectif et donc $h_2^{\alpha} = h_2$. Dans le cas extrême d'une adaptation parfaite, $\alpha = 1$, et la dégradation de l'état de santé n'a plus d'effet de sorte que $h_2^{\alpha} = h_1$. Dans le cas normal intermédiaire, α est compris entre 0 et 1 de sorte $h_1 < h_2^{\alpha} < h_2$.

Comme nous l'avons exposé précédemment, non seulement les individus tendent à s'adapter hédoniquement en partie à un changement durable de leur santé, mais ils sous-estiment en général cette capacité d'adaptation. Dans les termes de Loewenstein, O'Donoghue et Rabin (2003), les individus sont victimes d'un biais de projection au sens où la prédiction de leurs préférences futures est biaisée par leurs préférences présentes. Afin de modéliser ce biais, supposons que l'individu ne prédise qu'une partie $(1 - m) \in]0, 1[$ de son degré d'adaptation, le niveau de santé subjectif prédit pour la seconde période, noté $\hat{h}_2$, est maintenant donné par :

$$\hat{h}_2 = h_2 + (1 - m)\alpha(h_1 - h_2) \tag{2}$$

1. Par santé subjective, nous entendons dans la suite de l'article une mesure de la qualité de la santé telle qu'elle est évaluée subjectivement par l'individu.

Le paramètre m mesure le degré de sous-estimation de la capacité d'adaptation. Plus il est élevé, et plus l'individu sous-estime son niveau de santé subjective future. Pour $0 < m < 1$, les équations (1) et (2) impliquent que $h_2^\alpha > \hat{h}_2$. L'individu ne se trompe pas sur son niveau de santé objectif futur h_2, qui est donné, mais seulement sur la valeur subjective qu'il lui attribue. En conséquence, l'individu sous-estime l'utilité totale associée à tout niveau de santé $u(c,\hat{h}_2) < u(c,h_2^\alpha)$ mais surestime l'utilité marginale $\partial u(c,\hat{h}_2)\big/\partial h_2 > \partial u(c,h_2^\alpha)\big/\partial h_2$. Autrement dit, une dégradation de sa santé objective augmente son utilité marginale mais moins qu'il ne le prédit.

Introduction d'une composante stochastique sur la détérioration de la santé

L'équation (2) suppose que l'individu ne prédit pas correctement l'écart entre sa santé future subjective après adaptation h_2^α et son niveau objectif de santé h_2, connu de manière certaine, pour simplifier. En pratique, la détérioration du niveau de santé objectif possède une composante stochastique importante. Il existe toujours un risque que la détérioration de la santé $(h_1 - h_2)$ soit un peu plus faible mais aussi plus élevée que ce qui est attendu en moyenne (pour un effort donné). Ou pour le dire autrement, il existe toujours un risque que le niveau de santé objectif futur (h_2)soit plus élevé ou plus faible qu'attendu. Afin d'incorporer ce risque, on considère maintenant que le niveau de santé de la deuxième période, désormais noté $\tilde{h}_2$, est donné par :

$$\tilde{h}_2 = h_2 + \tilde{\varepsilon} \tag{3}$$

où $\varepsilon \in \left[\underline{\varepsilon},\bar{\varepsilon}\right]$ avec $\underline{\varepsilon} < 0$ et $\bar{\varepsilon} > 0$, $h_2 + \underline{\varepsilon} > 0$, $E[\tilde{\varepsilon}] = 0$ et $\mathrm{Var}(\tilde{\varepsilon}) = \sigma^2$. Sous ces hypothèses, h_2 représente maintenant le niveau attendu de santé objective $h_2 = E\left[\tilde{h}_2\right]$. Il est utile pour l'interprétation de voir que l'équation (3) est équivalente à poser l'hypothèse d'un bruit blanc égal à $-\tilde{\varepsilon}$ sur la *détérioration* attendue de la santé objective puisque $h_1 - \tilde{h}_2 = h_1 - (h_2 + \tilde{\varepsilon}) = (h_1 - h_2) - \tilde{\varepsilon}$. En intégrant l'équation (3) dans l'équation (2), le capital santé subjectif prédit après adaptation devient donc aussi une variable aléatoire, notée $\hat{\tilde{h}}_2$ et donnée par :

$$\hat{\tilde{h}}_2 = \tilde{h}_2 + (1-m)\alpha\left(h_1 - \tilde{h}_2\right) \tag{4}$$

Une nouvelle fois, dans cette équation, l'individu anticipe correctement la distribution de $\tilde{h}_2$. Cependant, en raison du biais de projection, il surestime systématiquement le niveau de satisfaction associé aux différents niveaux de santé possibles [2]. Pour comprendre comment l'adaptation hédonique, et en sens inverse

2. En réalité, les individus peuvent également se tromper sur la distribution de probabilité de leur niveau santé objectif futur, par exemple, parce qu'ils perçoivent mal les progrès médicaux futurs dont ils bénéficieront. Par souci de clarté de l'exposé, nous n'introduisons pas cette possibilité.

la sous-estimation de celle-ci, affecte la perception subjective des conséquences d'un risque d'espérance nulle sur la détérioration de la santé, considérons le cas où $m = 0$, quand l'individu anticipe pleinement son adaptation future et son niveau de santé subjectif après adaptation devient $\hat{\tilde{h}}_2 = \tilde{h}_2 + \alpha\left(h_1 - \tilde{h}_2\right)$. Il apparaît qu'en présence d'adaptation, tout se passe comme si un risque objectif donné $\tilde{\varepsilon}$ était transformé en un risque plus faible $(1-\alpha)\tilde{\varepsilon}$. Comme l'individu s'adapte à la fois aux changements positifs et négatifs de sa santé, il devient plus indifférent au risque que la détérioration de sa santé objective soit plus faible ou plus forte qu'attendue. Il s'*adapte* à ce risque. Cependant, lorsqu'il sous-estime sa capacité d'adaptation ($m > 0$), tout se passe comme si, maintenant, il exagérait ce risque.

3. La décision d'investissement santé dans un modèle à deux périodes avec sous-estimation de l'adaptation hédonique à la détérioration de la santé

Considérons maintenant un modèle à deux périodes sans épargne dans lequel la santé se détériore entre les deux périodes à un rythme qui dépend négativement des efforts réalisés pour la préserver. Dans ce modèle, l'individu alloue son revenu exogène de première période y_1 à la consommation (c) ou à l'investissement santé (i). Ce dernier correspond aux dépenses de santé (soins médicaux, visites chez le docteur…) mais aussi de manière plus générale au coût d'opportunité du temps qui y est consacré. Par simplicité, les dépenses de santé n'affectent que la santé future [3]. La fonction de production de santé objective est donnée par :

$$\tilde{h}_2 = h_2(i) + \tilde{\varepsilon} \tag{5}$$

avec $dh_2/di > 0$, $dh_2/di \leq 0$et $h_2(0)+ > 0$. On suppose aussi par ailleurs que $h_2(y_1) < h_1$: Même si l'individu choisit d'allouer toutes ses ressources de première période à la préservation de sa santé, en l'absence de circonstances aléatoires particulièrement favorables, il ne peut maintenir son niveau de santé initial. En remplaçant h_2 par $\tilde{h}_2$ dans l'équation (4), le capital santé subjectif prédit peut être exprimé comme une fonction de *i :*

$$\hat{\tilde{h}}_2(i) = \tilde{h}_2(i) + (1-m)\alpha\left(h_1 - \tilde{h}_2(i)\right) \tag{6}$$

3. En pratique, une grande partie des activités de santé affecte également la santé présente, en particulier, si le capital santé est interprété de manière extensive pour inclure le capital esthétique. Mais dans la mesure où l'on s'intéresse seulement aux efforts qui impliquent des bénéfices de santé à long terme marginaux, mêmes faibles, l'individu doit toujours réaliser un arbitrage entre sa satisfaction présente et future de sorte que l'on peut simplifier l'analyse en considérant que les bénéfices en termes de santé ne seront observés qu'à la seconde période.

En l'absence d'épargne, la consommation de la seconde période est égale au revenu de la seconde période y_2. Après avoir intégré la contrainte $(y_1 = c + i)$, le programme de maximisation de l'individu est donné par :

$$\underset{i}{Max}\, \hat{U}(i) \equiv u(y_1 - i, h_1) + \delta \cdot E\left[u\left(y_2, \hat{\tilde{h}}_2(i)\right)\right] \tag{7}$$

Si on adopte les notations allégées $u_h\left(y_2, \hat{\tilde{h}}_2(i^*)\right) = u_h$, $u_{hh}\left(y_2, \hat{\tilde{h}}_2(i^*)\right) = u_{hh}$ et $u_{cc} = u_{cc}(y_1 - i^*, h_1)$, on obtient les conditions de premier et second ordre suivantes :

$$-u_c(y_1 - i^*, h_1) + \delta E\left[\hat{\tilde{h}}_2'(i^*) u_h\right] = 0 \tag{8a}$$

$$u_{cc} + \delta E\left[\hat{\tilde{h}}_2''(i^*) u_h + \left(\hat{\tilde{h}}_2'(i^*)\right)^2 u_{hh}\right] < 0 \tag{8b}$$

Et l'effet d'un changement marginal de m sur i^* est donné par[4] :

$$\frac{di^*}{dm} = \frac{-\delta\alpha E\left[\tilde{h}_2{}'(i^*) u_h - \hat{\tilde{h}}_2'(i^*)(h_1 - h_2(i^*) - \tilde{\varepsilon}) u_{hh}\right]}{\underbrace{u_{cc} + \delta E\left[\hat{\tilde{h}}_2''(i^*) u_h + \left(\hat{\tilde{h}}_2'(i^*)\right)^2 u_{hh}\right]}_{<0}} \tag{9}$$

Notons, tout d'abord, que par hypothèse pour $i^* \in [0, y_1]$, $h_1 - h_2(i^*) > 0$. En l'absence de risque sur le niveau de santé futur objectif $(\tilde{\varepsilon} = 0)$ ou, ce qui revient au même, sur la magnitude de la détérioration de la santé objective, il s'en suit directement que $di^*/dm > 0$. Comme l'individu sous-estime sa faculté d'adaptation hédonique à une détérioration de la santé, il investit un niveau de ressources i^* dans sa santé supérieur à celui qu'il aurait choisi s'il avait parfaitement prédit l'évolution de ses préférences.

En présence de risque sur l'ampleur de la détérioration de la santé objective, la valeur de di^*/dm dépend à la fois de la taille du risque et de la réaction de l'individu face au risque. Supposons dans un premier temps que le risque soit limité au sens où l'individu n'a jamais la possibilité de maintenir son niveau initial de santé même dans le cas le plus favorable $(\bar{\varepsilon} < h_1 - h_2(i^*)\ \forall i^* \in [0, y_1])$. On peut penser ici à la détérioration de la santé consécutive au processus de vieillissement où, quels que soient les efforts réalisés par l'individu pour préserver sa santé et « rester jeune », son état physique général se détériore inexorablement entre la période jeune et la période âgée. Dans cette situation, on a toujours $di^*/dm > 0$. La conclusion est

4. Le calcul du numérateur apparaît plus facilement si, étant donnée l'équation (6), on écrit d'abord la condition de premier ordre (8a) sous la forme plus développée : $-u_c(y_1 - i^*, h_1) + \delta E\left[(1 + \alpha - m\alpha)\tilde{h}_2'(i) u_h(y_2, \tilde{h}_2(i) + (1 - m)\alpha(h_1 - \tilde{h}_2(i)))\right] = 0$.

qualitativement analogue au cas sans risque. Cependant, il est possible d'établir la proposition suivante :

Proposition – *Considérons un individu qui sous-estime sa capacité d'adaptation hédonique à une détérioration de sa santé telle que décrite dans l'équation (6). S'il n'a jamais la possibilité de maintenir son niveau initial de santé, même dans le cas le plus favorable où la détérioration de sa santé est minimale,* $(\bar{\varepsilon} < h_1 - h_2(i^*)\ \forall i^* \in [0, y_1])$, *un individu qui sous-estime sa capacité d'adaptation à un changement de son état de santé augmente son investissement santé. Cette hausse est amplifiée si l'individu est prudent* $(u_{hhh} > 0)$ *mais réduite lorsque l'individu est imprudent* $(u_{hhh} < 0)$.

Pour comprendre ce résultat, nous pouvons repartir de l'équation (4). Si on dérive $\hat{\hat{h}}_2$ par rapport à m, nous obtenons après réarrangement [5] :

$$\partial \hat{\hat{h}}_2 \big/ \partial m = \underbrace{-\alpha(h_1 - h_2)}_{\text{Baisse prédite dans le capital santé}} + \underbrace{\alpha\tilde{\varepsilon}}_{\text{hausse du risque perçu}} \qquad (10)$$

Cette équation montre que, pour tout niveau d'effort i donné, deux effets indépendants surviennent lorsque l'individu sous-estime sa capacité d'adaptation hédonique à une détérioration de la santé :

i) une baisse de la santé subjective attendue,

ii) mais aussi, maintenant, une augmentation subjective du risque d'espérance nulle que cette détérioration soit plus faible ou plus importante que prévue [6].

Le premier effet, comme on l'a établi, incite l'individu à accroître son investissement santé. En extrapolant le résultat désormais classique de Kimball (1990) sur l'épargne de précaution en présence d'un risque sur la richesse future à l'investissement santé en présence d'un risque de moyenne nulle sur la santé future, on peut montrer que la direction du second effet dépend du signe de u_{hhh}(voir annexe pour la démonstration). Si $u_{hhh} > 0$, c'est-à-dire si l'individu est prudent, l'accroissement à moyenne constante du risque subjectif l'incitera à augmenter son investissement santé. Au contraire, si l'individu est imprudent $(u_{hhh} < 0)$, il réduira son investissement santé (voir aussi Dardanoni et Wasgaff, 1990 sur ce point). Pour résumer, deux effets surviennent lorsque l'individu sous-estime son adaptation à une détérioration de la santé : une diminution du niveau subjectif de la santé et une augmentation de la « peine » (Crainich et Eeckhoudt, 2005) associée au risque d'espérance nulle que cette détérioration soit plus importante ou plus faible que

5. Le calcul de la dérivée est facilité si l'on note que : $\hat{\hat{h}}_2 = h_2 + (1 - m)\alpha(h_1 - h_2) + (1 - \alpha + \alpha m)\tilde{\varepsilon}$.

6. Rappelons que l'individu perçoit correctement le risque que la détérioration de santé soit plus faible ou plus élevée que prévue mais qu'il tend à exagérer les variations en termes de bien-être provoquée par ce risque. Compte tenu du formalisme utilisé pour modéliser l'adaptation, cette erreur de perception est équivalente à une exagération du risque.

prévue. Les deux effets jouent dans la même direction lorsque $u_{hhh} > 0$. Lorsqu'au contraire ils jouent de façon opposée, la réaction de l'individu à la hausse perçue du risque tempère l'incitation de l'individu à augmenter son investissement santé.

Notons que cette conclusion vaut seulement dans le cas où l'on suppose que l'individu n'a jamais la possibilité de maintenir son niveau initial de santé, même si la détérioration de la santé objective s'avère plus faible qu'attendue en moyenne. Supposons, au contraire, que l'individu ait une chance d'atteindre un niveau de santé en période 2 supérieur au niveau de santé en période 1. On peut envisager le cas d'un individu qui souffre d'une maladie de longue durée pour laquelle il existe des chances significatives de guérison comme dans certaines formes de cancer. Son état futur de santé ne dépend pas que de l'évolution exogène de la maladie mais de la qualité de suivi du traitement et de l'hygiène de vie adoptée correspondant ici à son investissement santé. Si l'individu est imprudent, le signe de di^*/dm n'est plus systématiquement positif pour n'importe quel degré d'erreur m de sous-estimation de l'adaptation. Il devient alors possible que, pour certaines valeurs de m, l'incapacité de l'individu à prédire correctement son adaptation l'incite à *réduire* et non augmenter son investissement santé. Cette indétermination de l'effet d'un risque santé d'espérance négative sur l'investissement santé d'un individu imprudent existe en fait indépendamment de tout processus d'adaptation hédonique. Un individu imprudent qui s'attend en moyenne à une détérioration de la santé mais qui a une petite chance de voir celle-ci s'améliorer peut choisir de réduire son investissement santé. L'exagération de la peine provoquée par la variabilité de la détérioration de la santé objective lorsque l'individu imprudent sous-estime son adaptation hédonique peut simplement rendre cette diminution de l'effort santé plus probable.

4. Conclusion

L'investissement des individus dans leur santé dépend de la perception qu'ils ont des conséquences d'une variation de leur santé sur leur satisfaction. De nombreux travaux empiriques ont établi ces dernières années que cette perception était souvent biaisée en raison de la difficulté pour les individus à percevoir leur adaptation hédonique à des changements de santé. Le modèle précédent montre que, sous certaines conditions générales, cette difficulté conduit les individus à exagérer la variation d'utilité consécutive à toute détérioration de sa santé mais aussi au risque que celle-ci soit plus faible ou plus importante que prévue. Tant que ce risque est limité, au sens où la santé ne peut s'améliorer entre les deux périodes, les individus augmentent leur investissement santé. Ils font davantage d'efforts pour leur santé que ce qu'ils auraient fait s'ils avaient correctement prédit l'évolution de leurs préférences. En outre, cette hausse est amplifiée si l'individu est prudent mais tempérée si l'individu est imprudent.

Il est toutefois dangereux de conclure que l'individu « surinvestit » alors dans sa santé. D'une part, la hausse des ressources consacrées à la santé ne vaut qu'indépendamment d'autres mécanismes très largement documentés dans la littérature qui, en pratique, favorisent plutôt le sous-investissement dans la santé (tendance à la procrastination, incohérence temporelle des décisions, évaluation incorrecte des risques, etc.). D'autre part, il n'y a pas de consensus sur le critère normatif à adopter en présence d'adaptation hédonique pour juger de l'efficacité d'une décision (Lowenstein et Ubel, 2008). Car, même si les individus s'adaptent à la détérioration de leur santé au sens où ils indiquent un niveau de satisfaction similaire aux individus en bonne santé, ils sont aussi souvent prêts à faire des sacrifices importants en terme d'espérance de vie pour revenir à leur niveau de santé initial, celui qui précédait, par exemple, la colostomie (Smith *et al.* 2006) ou la dégradation fonctionnelle de leurs reins (Torrance, 1976).

Le modèle précédent a cependant une implication directe pour la communication des pouvoirs publics en matière de prévention des risques de santé. Il montre que l'insistance sur les aspects les plus contraignants des maladies peut favoriser l'investissement santé des individus si elle entretient la myopie partielle des individus sur leur capacité d'adaptation hédonique. Et cela sera d'autant plus efficace que les individus sont prudents.

Références

Albrecht, G. L., Devlieger, P. J. 1999. "The Disability Paradox: High Quality of Life Against all Odds". *Social Science and Medicine*, 48, 977–988.

Ashby, J., O'Hanlon, M., Buxton, M. J. 1994. "The Time Trade-off Technique: How do the Valuations of Breast Cancer Patients Compare to Those of Other Groups?", *Quality of Life Research*, 3, 257–265.

Bui, P., Crainich, D., Eeckhoudt, 2005. "Allocating Health Care Resources under Risk: Risk Aversion and Prudence Matter". *Health Economics*, 14, 1073-1077.

Calvo, H., Arrondel, L. 2008. « Les français sont-ils prudents ? Patrimoine et risque sur le marché du travail », *Économie et Statistique*, 417(1), 27-43.

Courbage, C., Rey, B. 2006. "Prudence and Optimal Prevention for Health Risks", *Health Economics,* 15, 1323–1327.

Crainich, D., Eeckhoudt, L. 2005. « La notion économique de prudence. Origine et développements récents », *Revue Économique*, 56 (5), 1021-1032.

Dardanoni, V., Wagstaff, A. 1990. "Uncertainty on the Demand for Medical Care", *Journal of Health Economics*, 9, 23-38.

Dolan, P., Kahneman, D. 2008. "Interpretations of Utility and their Implications for the Valuation of Health". ***The Economic Journal, 118***, 215-234.

Eeckhoudt L., Godfroid P., M. Marchand, 1998. «Risque de santé, médecine préventive et médecine curative», *Revue d'Économie Politique*, 108, 3, 321-337.

Frederick, S., Loewenstein, G. 1999. "Hedonic Adaptation" *in* D. Kahneman, E. Diener, and N. Schwartz (eds). *Scientific Perspectives on Enjoyment, Suffering, and Well-Being.* New York. Russell Sage Foundation.

Gjerde, J., Grepperudb, S., Kverndokk, S., 2005. "On Adaptation and the Demand for Health". *Applied Economics*, 37, 1283-1301.

Groot, W. 2000, "Adaptation and Scale of Reference Bias in Self-assessments of Quality of Life". *Journal of Health Economics*, 19, 403-420.

Hurst, N. P., Jobanputra, P., Hunter, M., Lambert, M., Lockhead, A., Brown, H. 1994. "Validity of EuroQolF: A Generic Health Status Instrument for Patients with Rheumatoid Arthritis". *British Journal of Rheumatology*, 33, 655–662.

Kimball, S. 1990. "Precautionary Saving in the Small and in the Large". *Econometrica*, 58 (1), 53-73.

Loewenstein, G., O'Donoghue, T., Rabin, M. 2003. "Projection Bias in Predicting Future Utility". *Quarterly Journal of Economics*, 118(4), 1209–48.

Loewenstein, G., Ubel, P. A. 2008. "Hedonic Adaptation and the Role of Decision and Experience Utility in Public Policy". *Journal of Public Economics*, 92(8), 1795-1810.

Macé, S., (2012) "Saving and Perceived Health Risks". *Review of Economics of the Household.* Forthcoming, DOI : 10.1007/s11150-012-9160-y.

Oswald, A., Powdthavee, N. 2008. "Does Happiness Adapt? A Longitudinal Study of Disability with Implications for Economists and Judges". *Journal of Public Economics.* 92(5-6), 1061-1077.

Palumbo, M. G. 1999. "Uncertain Medical Expenses and Precautionary Saving Near the End of the Life Cycle". *The Review of Economic Studies*, 66(2), 395-421.

Picone, G., Uribe, M., Wilson, R.M .1998. "The Effect of Uncertainty on the Demand for Medical Care, Health Capital and Wealth". *Journal of Health Economics*, 17, 171-185.

Riis, J., Loewenstein G., Baron J., Jepson C., Fagerlin A., Ubel P. 2005. "Ignorance of Hedonic Adaptation to Hemodialysis: A Study Using Ecological Momentary Assessment", *Journal of Experimental Psychology*, 134(1), 3-9.

Smith, D.M., Sherriff, R.L., Damschroder, L.J., Loewenstein, G., Ubel, P.A. 2006. "Misremembering Colostomies? Former Patients Give Lower Utility Ratings than do Current Patients", *Health Psychology*, 25 (6), 688–695.

Tessier, P. 2009. Harsanyi, Sen ou Bentham. « Quelle perspective adopter pour l'évaluation du bien-être en santé ? », *Revue Économique*, 60 (6), 1309-1334.

Torrance, G.W., 1976. "Toward a Utility Theory Framework for Health Status Index Models", *Health Service Research*, 11, 349–369.

Wu, S. 2001. "Adapting to Heart Conditions: A Test of the Hedonic Treadmill", *Journal of Health Economics*, 20, 495–508.

Annexe

A. L'effet d'une hausse de la variabilité du niveau subjectif de santé futur sur l'investissement santé

Le capital santé subjectif prédit est donné dans le cas général par

$$\hat{\tilde{h}}_2(i) = \tilde{h}_2(i) + (1-m)\alpha\left(h_1 - \tilde{h}_2(i)\right) \tag{A.1}$$

ou encore

$$\hat{\tilde{h}}_2(i) = h_2(i) + (1-m)\alpha(h_1 - h_2(i)) + (1-\alpha)\tilde{\varepsilon} + m\alpha\tilde{\varepsilon} \qquad \text{(A.2)}$$

Notons $H(i) = h_2(i) + (1-m)\alpha(h_1 - h_2(i))$ et $\tilde{H}(i) = H(i) + (1-\alpha)\tilde{\varepsilon}$. Avec ces notations, l'équation (A.1) peut être réécrite sous la forme :

$$\hat{\tilde{h}}_2(i) = \tilde{H}(i) + m\alpha\tilde{\varepsilon} \qquad \text{(A.3)}$$

Le terme $m\alpha\tilde{\varepsilon}$ isole l'effet propre de la sous-estimation de l'adaptation hédonique au risque santé pour i donné sur la variabilité du niveau subjectif de santé future.

En l'absence de cet effet, le consommateur maximise uniquement le programme :

$$\underset{i}{Max}\, \tilde{V}(i) \equiv u(c-i, h_1) + E\left[u\left(c, \tilde{H}(i)\right)\right] \qquad \text{(A.4)}$$

Soit i^{**} la solution de ce programme. La condition de premier ordre implique :

$$u_c(c - i^{**}, h) = H'(i^{**})E\left[u_h\left(c, \tilde{H}(i^{**})\right)\right] \qquad \text{(A.5)}$$

Dans le cas général où l'on tient aussi compte de la sous-estimation de l'adaptation hédonique au risque santé, le programme est donné par l'équation (7) du texte principal ou en utilisant la notation de l'équation (A.3) par :

$$\underset{i}{Max}\, \hat{U}(i) \equiv u(y_1 - i, h_1) + \delta \cdot E\left[u\left(y_2, \tilde{H}(i) + m\alpha\tilde{\varepsilon}\right)\right] \qquad \text{(A.6)}$$

Soit i^* la solution de ce programme. Si l'on dérive la fonction d'utilité $\hat{U}(i)$ en i^{**}, il va

$$\hat{U}'(i^{**}) = -\, u_c(c - i^{**}, h) + H'(i^{**})E\left[u_h\left(c, \tilde{H}(i^{**}) + \alpha m\tilde{\varepsilon}\right)\right] \qquad \text{(A.7)}$$

ou étant donné (A.5) par

$$\hat{U}'(i^{**}) = H'(i^{**})\left(E\left[u_h\left(c, \tilde{H}(i^{**}) + \alpha m\tilde{\varepsilon}\right)\right] - E\left[u_h\left(c, \tilde{H}(i^{**})\right)\right]\right) \qquad \text{(A.8)}$$

Il s'ensuit que $\hat{U}'(i^{**}) > 0$ et donc $i^* > i^{**}$ si

$$E\left[u_h\left(c, \tilde{H}_2(i^{**}) + \alpha m\tilde{\varepsilon}\right)\right] > E\left[u_h\left(c, \tilde{H}_2(i^{**})\right)\right],$$

c'est-à-dire si u_h est convexe par rapport à h ou $u_{hhh} > 0$. Au contraire, $i^* < i^{**}$si $u_{hhh} < 0$.

PUBLICATION 6: Fabrice Le Lec et Serge Macé (2018), “The Curse of Hope”, **Theory and Decision**, 84(3):429-451.

Theory Dec.
DOI 10.1007/s11238-017-9621-0

The curse of hope

Fabrice Le Lec[1] · **Serge Macé**[2]

Abstract In Kőszegi and Rabin's (Q J Econ 1133–1165, 2006, Am Econ Rev 97:1047–1073, 2007) reference-dependent model of preferences, the chance of obtaining a better outcome can reduce an agent's expected utility through an increase in the stochastic reference point. This means that individuals may prefer stochastically dominated lotteries. In this sense, hope, understood as a small probability of a better outcome, can be a curse. While Kőszegi and Rabin focus on a linear specification of the utility function, we show that this effect occurs more broadly. Using fairly plausible assumptions and parameter values, we specify the conditions under which it occurs, as well as the type of lotteries in which this should be expected. We then show that while a simple subjective transformation of probability into weights of the reference point may in some cases mitigate the issue, in others, it can intensify it or even generate new ones. Finally, we extend the model by adding the individual's current reference point (status quo) to the stochastic reference point. We show that this modification can reconcile Kőszegi and Rabin's model with the apparent empirical infrequency of stochastically dominated choices while maintaining its main qualitative results.

Keywords Adaptation · Multiple reference point · Loss aversion · Stochastic dominance

✉ Serge Macé
serge.mace@edhec.edu

Fabrice Le Lec
fabrice.le-lec@univ-paris1.fr

1 Université Paris 1, Centre d'Economie de la Sorbonne, umr CNRS 8174, 106-112 Boulevard de l'Hôpital, 75013 Paris, France

2 Edhec Business School, 24 rue Gustave Delory, 59057 Roubaix, Lille, France

Published online: 01 July 2017

> I have observed that not the man who hopes when others despair, but the man who despairs when others hope, is admired by a large class of persons as a sage.
>
> John Stuart Mill

1 Introduction

Consider an individual who has just suffered a major financial setback or a serious health shock like a sudden handicap or the beginning of a long-term disease. Under normal circumstances, she should value any small positive probability p of getting her money back or of healing. However, because a positive probability of an improvement also increases expectations and the magnitude of the disappointment if they are not met, she may prefer the certainty of the loss rather than a small probability of a return to the initial situation. The model developed by Kőszegi and Rabin (2006, 2007) allows for this possibility. In this model, an individual facing the prospect of an uncertain future updates her reference point to her (rational) expectations: her new stochastic reference point mimics the future lottery. She then evaluates this future prospect according to its expected utility, with the utility of each outcome being the weighted average of how it feels, relative to each possible realization of this stochastic reference point. The model is then able to capture the effect of a higher probability of obtaining a good outcome on the expected utility of the individual, while taking into account the influence it has on her expectations. Kőszegi and Rabin (hereafter, KR) show, under some assumption of linearity, that an individual may prefer the worst outcome of a given lottery for sure, compared to the lottery itself. This implies a violation of first-order stochastic dominance, underlined by KR (Proposition 7) and more recently by Masatlioglu and Raymond (2016). Hence, an individual may consider that she will be better off when she is certain that her situation will not improve, than when some hope exists that she will end up in a better situation. We call this effect the 'curse of hope'.

In this paper, we show that in a more general setting, with concave intrinsic utility and diminishing sensitivity in the distance to the reference point, this curse of hope is likely to occur because the marginal effect of the intrinsic term is plausibly small in many circumstances. This leads to the reference-dependent component being dominant. The effect corresponds to a strong violation of stochastic dominance, and one for which empirical evidence is scarce.[1] Furthermore, intuition suggests that a lab exper-

[1] Violations of first-order stochastic dominance are observed in the experimental literature, but to the best of our knowledge, none directly relates to the role of expectations in the formation of reference points. For instance, Butler et al. (2014) link such violations to the complexity of the presentation format of the lotteries, and Birnbaum (2005) to cognitive processes (attention transfers). Relatedly, Leland (1998) shows the role played by the presentation format of lotteries to obtain such violations. Another instance of stochastically dominated choices is provided by Gneezy et al. (2006). But, for their findings, dominated-choices disappear in the within-subject treatment and the replication attempts have given mixed results (Simonsohn 2009; Rydval et al. 2009; Keren and Willemsen 2009). Perhaps the closest evidence for violations related to reference point is given by Loomes et al. (1992). Indeed, as stated by KR in their 2007's paper, there exists a close link between their model and regret theory. Nevertheless, Loomes et al. (1992) find only mixed support for regret theory based on the violations observed. Generally speaking, none of the experiments referred to

iment is not needed to conclude that virtually nobody will invoke the fear of being disappointed to turn down a 1% chance of a substantial monetary gain in the future. In other domains, in particular for health, there may be some indirect evidence that such an effect plays a role (Macé 2016). For instance, it has been observed that some cancer patients, at an advanced stage of their illness, deliberately refuse a chemotherapy treatment even if it gives them a small chance of remission. The main explanation lies in the constraints and the side effects of the treatment that are weighed against the low probability of healing and the weak quality of additional life years. But, there may also be a downward adjustment of expectations, some acceptance of the situation and the will to live more peacefully during the remaining years or months of their lives (Sharf et al. 2005). It is possible to make a parallel here with the period of relief and apparent improvement which can sometimes be observed in depressive individuals once they have decided to commit suicide (Eastridge et al. 2012).

Except for these particular and indirect cases, there is little evidence of a strong violation of stochastic dominance of the kind implied by KR's model. This raises some questions, given that the model has become a reference for referent-dependent preferences, and that there is growing evidence that expectations influence reference points (Abeler et al. 2011; Crawford and Meng 2011; Card and Dahl 2011; Gill and Prowse 2012; Bartling et al. 2015; Ericson and Fuster 2011).[2] Our objective here is, therefore, twofold: (i) to investigate the general conditions under which this phenomenon and other related effects occur, and (ii) to suggest explanations for the fact that so few individuals reject a small probability of obtaining a better outcome. On the first point, using essentially binary lotteries, we show that the curse of hope is predicted to occur under very plausible specifications and in relatively frequent settings. In addition, it implies that an increase in the probability of the good outcome may reinforce the individual's motivation to choose the less attractive outcome for sure. On the second point, we amend the model by relaxing the simplifying assumption that the weights of the possible outcomes in the multiple reference point correspond exactly to their probabilities. This may relate to psychological phenomena such as defensive pessimism (Norem 2001) and the projection bias (Loewenstein et al. 2003). Doing so, we obtain that the curse of hope can be mitigated, but also that a new effect occurs, the "blessing of fear": in this case, an individual may prefer a little uncertainty over future outcomes rather than the certainty of the best outcome. Eventually, we consider the case when individuals update their reference point, but with some inertia. The idea is that at the time of the decision-making, the individual will use a stochastic reference point composed of the future lottery and her current situation. If the weight put on her current situation (inertia) is non-negligible as suggested by the vast literature on the status quo *bias* and some recent experimental investigations on

Footnote 1 continued
here concern future lotteries but present ones, none of them relies on lagged expectations. As will be clear in the second section, the type of violations predicted by KR concerns choices between simple (binary) lotteries, for which dominance is *transparent*, i.e., when the framing and the choice set make it evident to the decision-maker that one lottery dominates the other (Fishburn 1978).

[2] It should be noted however that the results by Abeler et al. (2011) and Ericson and Fuster (2011) are not replicated by Camerer et al. (2016). Some other experiments do not find much, if any, effect of expectations, for instance (Heffetz and List 2014).

the definition of the reference point (Baillon et al. 2017), then the set of situations where the curse of hope can occur shrinks, and if inertia is large enough, the effect disappears completely.

The remainder of the paper is organized as follows. In the second section we present the general situations in which the curse of hope may occur. Section 3 analyses the effect of a subjective transformation of the probability of outcomes on the stochastic reference point. In Sect. 4, we introduce the idea that agents only partly incorporate expected outcomes in their stochastic reference point, at the moment of making the decision. Section 5 concludes. All proofs are set out in the Appendix.

2 The curse of hope

Following KR (2007), the utility of a deterministic outcome x is given by

$$u(x|r) = m(x) + \mu(m(x) - m(r)) \tag{1}$$

The utility is the sum of two components. The intrinsic utility m corresponds to the traditional utility function. It is increasing in x and concave in the general case. The second component μ captures the assumption of prospect theory (Kahneman and Tversky 1979) that there is a gain–loss utility, that is an additional sensation of loss or gain generated by x departing from the reference point r. The function μ is continuous on the real line with $\mu(0) = 0$. It is twice differentiable but at 0. Loss aversion for large deviations from the reference point is captured by the assumption that if $t > s \geq 0$, then $\mu(t) - \mu(s) < \mu(-s) - \mu(-t)$, and for small deviations by $\mu'_-(0)/\mu'_+(0) = \lambda > 1$ with μ'_- and μ'_+ the left and right derivative. We also assume, in coherence with prospect theory, some diminishing sensitivity in the domains of losses and gains, that is, $\mu''(z) < 0$ for $z > 0$ and $\mu''(z) > 0$ for $z < 0$.

In an uncertain environment, KR assume that the individual evaluates any lottery, defined here on discrete outcomes, $L = (L_1, L_2, \ldots; p_1, p_2, \ldots)$, according to its expected utility, with the utility of each outcome being the average of how it feels relative to each possible realization of a stochastic reference point $R = (R_1, R_2, \ldots; \rho_1, \rho_2, \ldots)$:

$$U(L|R) = \sum_j \sum_i p_i \rho_j u(L_i|R_j) \tag{2}$$

If the individual fully adjusts her anticipated reference point to the future lottery and forms rational expectations (that is, she has an accurate view of the distribution), then $R = L$. The expected utility of an individual facing the lottery L in the future is then given by $U(L|L)$.

In the following, and for the sake of simplicity, we focus on simple binary lotteries.[3] Consider the lottery $L_p = (y, x; p, 1 - p)$ with $y > x$, p being the probability of the good outcome. For conciseness, we denote $L(y, x; 0, 1)$ as x and $L(y, x; 1, 0)$

[3] Some of our results can be generalized to more complex discrete lotteries, and a section of the Appendix is dedicated to this issue.

as y. Denote $\Delta m = m(y) - m(x)$, $\mu^+ = \mu(\Delta m)$ and $\mu^- = \mu(-\Delta m)$. After rearrangement,[4] we have

$$U(L_p|L_p) = m(x) + p\Delta m + p(1-p)\mu^+ + (1-p)p\mu^- \quad (3)$$

Equation (3) shows that when an individual has a probability $p > 0$ to get y instead of x for sure, three effects occur. It gives the individual the following: (i) a probability p of obtaining an increase in intrinsic utility equal to Δm; (ii) a probability p of benefiting from an additional sensation of gain equal to $(1-p)\mu^+$, corresponding to the fact that the individual can get y, whereas she had a probability $(1-p)$ to have x; and (iii) a probability $(1-p)$ of experiencing a loss sensation equal to $p\mu^-$ if she gets x while expecting y with a probability p. The first two effects play positively while the last one enters the total utility negatively. In the absence of loss aversion (ii) and (iii) cancel out. But if loss aversion is strong enough, (iii) may more than offset effects (i) and (ii). Specifically, $x \succ L_p$ iff $U(L_p|L_p) - m(x) = p\Delta m + p(1-p)(\mu^+ + \mu^-) < 0$, that is after rearrangement if $-(\mu^+ + \mu^-) > \Delta m/(1-p)$, which is satisfied for some values of the parameters. We can summarize this in the following proposition:

Proposition 1 (The curse of hope) $x \succ L_p$ *iff* $-(\mu^+ + \mu^-) > \Delta m/(1-p)$.

In the presence of loss aversion, the individual may prefer the certainty of having x to the lottery $L_p = (y, x; p, 1-p)$. Having a positive probability to obtain y instead of x is normally a good thing, but only as long as it does not inflate too much the expected loss sensation if y does not occur. Otherwise, hope could be "a curse". The crucial role is played by the magnitude of $\mu^+ + \mu^-$, denoted now $\Delta\mu$, the absolute net sensation of loss. Proposition 1 indicates that $x \succ L_p$ if and only if $-\Delta\mu > \Delta m/(1-p)$, which implies that $-\Delta\mu > \Delta m$ is a necessary condition for $x \succ L_p$. Indeed, if $-\Delta\mu > \Delta m$, then exists $\bar{p} = (\Delta m + \Delta\mu)/\Delta\mu > 0$ for which the individual is indifferent between x and L_p. Above $\bar{p}$, the individual prefers the lottery; below, she prefers x for sure. It should be noted, however, that an increase in the probability p of having the good outcome does not always automatically make the curse less likely. The opposite may actually be true:

Proposition 2 *If the condition of Proposition 1 holds, then $U(L_p|L_p)$ decreases with p on* $[0, \bar{p}/2]$.

Starting from a situation in which the individual prefers x to L_p with $p < \bar{p}/2$, Proposition 2 indicates that a marginal increase of the probability of getting y can reduce her expected utility. Not only is a small positive probability of getting the favorable outcome not always a good thing *ex ante*, but in some cases, a marginal increase of this probability just reinforces the individual in her choice. More generally, it means that the effect is not only restricted to a "certainty effect", but may also occur with non-degenerated lotteries. Consider L_p and $L_{p'}$ with $p < p'$. If both are lower than $\bar{p}/2$, then L_p is preferred to $L_{p'}$ since the overall expected utility is decreasing on $[0, \bar{p}/2]$. To put it simply, there are some cases in which the individual prefers to get

[4] Using Eqs. 1 and 2, $U(L_p|L_p) = pu(y|L_p) + (1-p)u(x|L_p)$ with $u(y|L_p) = m(y) + (1-p)\mu(m(y) - m(x))$ and $u(x|L_p) = m(x) + p\mu(m(x) - m(y))$.

only 1% chance of getting a good outcome, rather than 3%, because the latter would give her "too much" hope.

Now, an important question concerns the frequency of the curse of hope for common specifications and parameters of the utility functions. Proposition 1 states that a sufficient condition is $-\Delta\mu > \Delta m/(1-p)$, which, for $0 < p < \bar{p}$, is equivalent to $-\Delta\mu > \Delta m$. It is likely to be encountered for some small differences between x and y. The reason is that in the neighborhood of the reference point, we have $-\mu^- \simeq \lambda\mu^+$ and so $\Delta\mu \simeq (\lambda-1)\mu^+$. Hence the condition becomes $\mu^+ > \Delta m/(\lambda-1)$ or assuming a conventional value of 2 for λ, $\mu^+ > \Delta m$. The additional gain sensation given by an unexpected change dominates the intrinsic utility change. The following proposition summarizes this result for small differences between x and y in the general case:

Proposition 3 *Assume $\mu'_+(0) > \frac{1}{\lambda-1}$. Then $x \succ L_p$ for p small enough. Moreover, for p and p' with $p' > p$ small enough, $L_p \succ L_{p'}$.*

In plain words, if the reference-dependent component dominates (in marginal terms) the intrinsic utility (up to a threshold determined by the level of loss aversion), then the curse of hope necessarily occurs when p is small enough. Virtually any specification of μ satisfies this condition, up to a constant factor.[5]

It is worth noting that the concavity of m plays an important role on the range of stakes, i.e. $y-x$, for which the effect occurs. When m is linear, as in KR, the effect exists on the same range of stakes independently of x. This means that if the individual prefers the certainty of having x rather than a small probability of having $y = x + \delta$ to avoid harmful expectations, this will be true whatever the level of x. When m is concave, as x and y are shifted upwards, the effect is necessarily present for larger differences between x and y. The intuition is straightforward: as x and y shifts upwards, Δm decreases and gets closer to 0 where the effect (under the condition of Proposition 3) is certain to occur.

Proposition 4 *Suppose the curse of hope occurs for x and $y = x + \delta$ with $\delta > 0$. Then for any $x' > x$, it occurs for x' and $y' = x' + \delta'$ with $\delta' > \delta$.*

The concavity of m implies that the range of stakes $(y-x)$ for which the effect occurs gets larger as the worst outcome increases. If $m'(x)$ tends to 0 as x takes arbitrarily large values (as is the case for the most usual concave functions), the range of outcomes for which the effect appears can be arbitrarily large. Said differently, for any stake $y-x$, there exists some x large enough for the effect to occur. This means that not only are the conditions for the curse of hope to occur easily satisfied for some specific stakes (Proposition 3), but that it will occur for all sufficiently large wealth levels and for arbitrarily large stakes.

To illustrate our points, we consider a specification where the intrinsic component is a log function while the reference-dependent one is an exponential one: $m(x) = ln(x)$ and $\mu(z) = 1-e^{-\beta z}$ for $z \geq 0$ and $\mu(z) = \lambda(1-e^{-\beta' z})$ otherwise. We set $\beta = \beta' = 3$ and $\lambda = 2$. The specifications and the parameters are plausible. For $x = 100{,}000$ and

[5] Since by assumption, $\mu' > 0$, there exists some factor a large enough for $a\mu'_+(0) > \frac{1}{\lambda-1}$. This is what Proposition 7 of KR (2007) also shows when m is linear. In the linear case, see also Masatlioglu and Raymond (2016) for conditions on loss aversion that avoid violations of stochastic dominance.

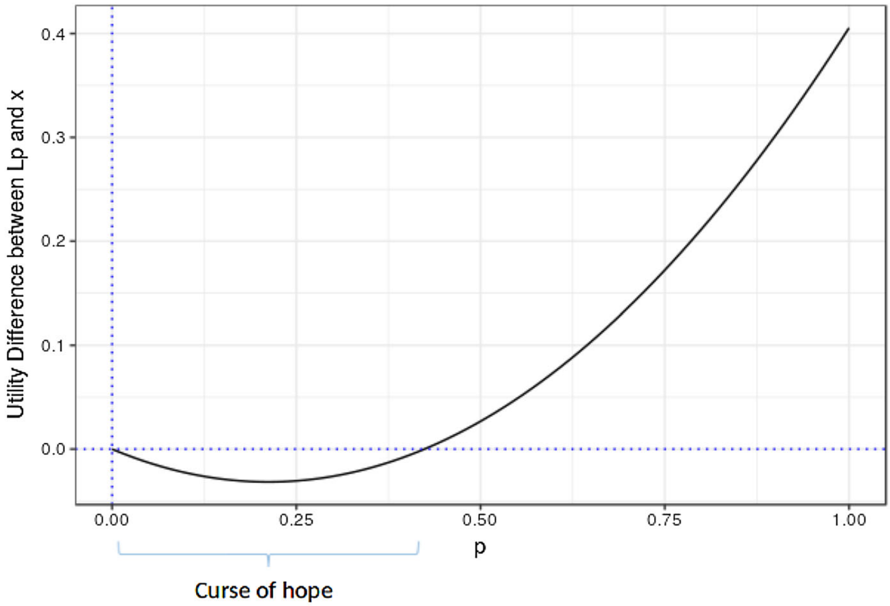


Fig. 1 The curse of hope: $U(L_p|L_p) - U(x|x)$ as a function of p

$y = 150{,}000$, the values of $U(L_p|L_p) - U(x|x)$ depending on p are displayed on Fig. 1. The difference is decreasing in p from 0 to $\bar{p}/2 \simeq .212$ and negative from 0 to $\bar{p} \simeq .424$. The curse of hope, in this case, concerns a large range of lotteries, and the stakes are quite large. Moreover, the function displayed in Fig. 1 is the same for all x and y with $y = 1.5x$, because of the specificities of the log function. Taken at face value, an individual endowed with this utility specification associated with these parameters would reject a one-third probability to see her income increase by 50% in the future.

Hence, our results suggest that in the KR model, the curse of hope is not a limit case restricted to small stakes or very low probabilities. A possibly large share of individuals would make stochastically dominated choices in relatively frequent circumstances. Given the (quasi-)absence of evidence of such choices in practice, this prediction weakens the relevance of the model.

One possibility to escape this apparent contradiction,[6] explored below, is that individuals do not adjust their stochastic reference point to be strictly identical to the

[6] Another possibility is to assume that the reference-dependent component is weaker than the intrinsic utility. Indeed, as stated indirectly by Proposition 3, to fully suppress the curse of hope, it is enough that $\mu'_+(0) \leq \frac{1}{\lambda - 1}$. Nevertheless, we see two weaknesses in this solution. First, it imposes restrictions on the strength of the reference-dependent component and the intensity of loss aversion in a rather ad hoc manner. Second, because the gain–loss component captures the standard properties of prospect theory (relevance of the reference point, loss aversion, diminishing sensitivity in the domains of loss and gains), it would also regrettably limit the ability of the model to explain other well-documented phenomena like strong risk-aversion for small stakes, the disposition effect, the endowment effect, etc. when expectations do not play a role in the formation of the reference point, i.e., for "surprise lotteries".

lottery. Even though individuals may form accurate beliefs about the lotteries they will face in the future, their stochastic reference point may not directly mirror the objective probabilities of future outcomes. Hereafter, we consider two modifications of the model: first, we allow probabilities for the stochastic reference point to be subjectively transformed. Second, we introduce some inertia in the behavior of the individual allowing them to update only partly their reference point, in view of the coming lottery.

3 Subjectively transformed probabilities for the reference point

When an individual faces a lottery with two possible outcomes, as KR argue, these two possible outcomes constitute two natural reference points that the individual can use to evaluate the outcome that will be observed, generating sensations of gains and losses. KR's model of a stochastic reference point implies that (i) the possible reference points only correspond to the possible outcomes, and (ii) the weights given to each possible outcome correspond exactly to their probabilities, to capture the intuitive idea that the additional sensations of gains and losses of having one outcome should be more intense if the individual had a high probability of obtaining the other.

These two assumptions may not always hold. First, it is clear that there may be other reference points than x and y. In the domains of wealth, health or wages, the empirical literature generally puts forth several potential reference points: the present level (because of a *status quo* bias, see Sect. 4), a weighted mean of past levels, the level that the individual takes as an objective or which is considered normal in the group of people to whom she compares herself.[7] Second, even if (i) holds, the weights given to each possible outcome may also differ significantly from their probability for various psychological reasons. For instance, consider a young individual whose health suddenly deteriorates. Her level of health falls from y (good health) to x (bad health) but she keeps a small probability p to return to y. Given that y is not only a possible outcome but also corresponds in this case to her past level and to the normal level for someone of her age, it makes sense, psychologically, to assume that y will be her dominant reference point. This idea can be captured by assuming that $\pi > p$, with π being the subjective weight applied to y as a reference point.

In addition, there may be threshold effects that operate in the mind of people. As noted by Gilbert and Ebert (2002), when an individual has no hope, she is forced to accept her situation, which is a necessary condition to adapt to it, to go through a "grieving process" and to stop to allocate her attention to what she had and will not have anymore. In some circumstances, a small probability of getting a good outcome (healing, getting a large amount of money, finding one's missing child) may impede that adaptation process. The individual does not necessarily overestimate the probability, she just cannot escape thinking about this positive outcome, giving it more weight in her comparisons. One way to capture this idea is to assume again that $\pi > p$ for $p > 0$.

[7] For incomes or wealth, these normal levels refer, for instance, to the median levels of income or wealth in a reference group, like the family, other workers at the individual's place of employment, people in the same neighborhood or region, etc. (Clark et al. 2008). For health, the reference level may correspond to the health level that individual considers "acceptable" given their age category (Wouters et al. 2015).

Furthermore, beyond a certain probability p^* of obtaining the favorable outcome y, say $p^* = 70\%$, one may also imagine that this favorable outcome becomes the unique reference point, in which case $\pi = 1$ for $p \geq p^*$.

By contrast, people may adopt some form of pessimism to prepare themselves psychologically to the possible occurrence of bad outcomes. They may set low expectations, not by changing their beliefs, i.e., reducing their subjective probability p of the occurrence of the good outcome, but by voluntarily directing their attention to the bad outcome, thus putting less emphasis and weight on the good outcome ($\pi < p$). In the psychological literature, this refers to a form of "defensive pessimism" (Norem and Cantor 1986; Norem 2001) that helps the individual reduce the loss sensation in the unfavorable case and increase the gain sensation in the favorable case.[8] In decision-making under risk or uncertainty, the idea that objective probabilities are subjectively weighted or transformed is widely acknowledged, and the previous reasoning shows that there is little a priori reason to consider that this could not occur for the stochastic reference point.

To take into account these possibilities, we now hypothesize that the individual subjectively transforms the probability (p) of outcomes into weights (π) for the stochastic reference point. Given that we focus on binary lotteries, there is no need to specify further this transformation. We simply assume that the sum of weights equals 1. In the simple case of binary lotteries, we only need to consider how π relates to p. In the case of the binary lottery $L_p = (y, x; p, 1 - p)$, this means that the individual gives a weight π to the comparison with y and $1 - \pi$ to the comparison with x (with $0 \leq \pi \leq 1$). By analogy with a random variable, we denote this weighted multiple reference point: $R_\pi = (y, x; \pi, 1-\pi)$. If $p = \pi$, we have $R_\pi = L_p$. The utility of an individual facing the lottery L_p and having the reference point R_π is now given by:[9]

$$U(L_p|R_\pi) = m(x) + p\Delta m + p(1 - \pi)\mu^+ + (1 - p)\pi\mu^- \tag{4}$$

If the individual gives a higher weight to y in her multiple reference point ($\pi > p$), her expected utility will be lower, i.e. $U(L_p|R_\pi) < U(L_p|L_p)$. As a result, the curse of hope is still present but under less restrictive conditions. Loss aversion in particular is no longer a necessary condition as stated by the following proposition (see the specific conditions in the Appendix):

[8] According to the authors, defensive pessimism not only helps individuals to cushion the potential blow of a bad outcome, but motivates them to work hard to prepare for the situation in which they can influence it, making their prediction potentially self-defeating. Given that we restrict our analysis to exogenous probabilities in this paper, we do not explore this second consequence.

[9] One may wonder why the objective probabilities are not transformed. In fact, it is very likely that they are (following rank-dependent models or cumulative prospect theory and the extensive empirical evidence of an inverted-S shaped transformation). Here we focus on the specific transformation that individuals may apply to their stochastic reference point: the type of transformation is conceptually different and in the absence of empirical assessment of this particular transformation and its interaction with the usual lottery probabilities, we treat the simple case where objective probabilities of occurrences are not transformed, while reference point weights are, as a first pass.

Proposition 5 *Suppose that $\pi > p$. Then, even in the absence of loss aversion, the individual may prefer the certainty of having x for sure to the lottery $L_p = (y, x; p, 1 - p)$.*

Proposition 5 states that, if $\pi > p$, then, even in the absence of loss aversion, the individual may prefer the certainty of having x rather than a probability p to get more than x. More generally, the curse of hope is more likely to occur when π is high compared to p, that is when the individual adopts a "higher" multiple reference point for a given p.[10] Coming back to our example of an individual who has just suffered a sudden deterioration of her health, even without loss aversion, a very small probability to heal may be a bad thing if it raises too much attention, too much hope and forbid acceptation and adaptation by the individual to the most probable state x.

By contrast, if the individual puts less weight on favourable outcomes, i.e. $\pi < p$, it contracts the range of situations in which the curse of hope occurs, quite straightforwardly: while objective probabilities are not affected, the (expected) loss sensation decreases as the subjective weight given to the higher reference point y decreases. In the limit case, i.e. π is null, the effect completely disappears. Yet, allowing π to be less than p gives rise to the mirror effect of the curse of hope: the "blessing of fear". It becomes possible that the individual prefers the uncertainty of the lottery L_p, to the certainty of obtaining its favorable outcome y for sure. To see why, note first that $m(x) + p\Delta m = m(y) + (1 - p)(-\Delta m)$. Replacing in Eq. (4) and subtracting $m(y)$, we have

$$U(L_p|R_\pi) - m(y) = (1 - p)(-\Delta m) + (1 - p)\pi\mu^- + p(1 - \pi)\mu^+ \tag{5}$$

To interpret Eq. (5), suppose that the individual initially owns y for sure, giving her the utility level $m(y)$ and she is now facing $L_p = (y, x; p, 1 - p)$. This uncertainty implies that she now has the following: (i) a probability $(1 - p)$ of suffering from a decrease in intrinsic utility; (ii) a probability $(1 - p)$ of suffering from an additional sensation of loss corresponding to the fact that she can only get x, while y is possible; and (iii) a probability p of obtaining an additional sensation of gain if the individual stays at y and compares her situation with outcome x. The fear of a loss is also a chance of being surprised positively.

Formally, the net expected additional sensation of gain/loss utility is given by $(1 - p)\pi\mu^- + p(1 - \pi)\mu^+$. Given that $|\mu^-| > \mu^+$ in the presence of loss aversion, this net additional sensation is positive if $(1 - p)\pi < p(1 - \pi)$ and so if $\pi < p$. It is maximal for p fixed when $\pi = 0$. In some cases, this expected additional positive sensation can even offset the expected decrease in intrinsic utility $(1 - p)(-\Delta m)$. For this to be the case, it must be that $U(L_p|L_p) - m(y) > 0$, which is satisfied, given Eq. 3, if $\pi\mu^- + p\frac{1-\pi}{1-p}\mu^+ > \Delta m$. This is summarized in the following proposition:

Proposition 6 (The blessing of fear) *Consider $L_p = (y, x; p, 1 - p)$. When the individual gives a weight π to the good outcome y that is lower than its probability p, she may prefer not to be certain of having y.*

[10] In this paper, $R_{\pi_i} = (y, x; \pi_i, 1 - \pi_i)$ is said to be *higher* than $R_{\pi_j} = (y, x; \pi_j, 1 - \pi_j)$ if $\pi_i > \pi_j$.

Hence, if $\pi < p$, the model allows an individual to prefer a probability, say, 99% of obtaining y instead of x, rather than the certainty of getting y. Given that $\pi\mu^- + p\frac{1-\pi}{1-p}\mu^+$ is a linear combination of μ^- and μ^+, the magnitude of loss aversion, however, makes this second effect less likely (contrary to the curse of hope). To examine the intuitive relevance of this theoretical possibility, consider the case where an individual owns some wealth y and has at each period a 1% chance of seeing her wealth or health fall to x. This risk obviously has a cost: 1% of the time on average, the individual will lose $y - x$. But it also gives the individual an additional sensation of experiencing a gain, 99% of the time. It is unlikely that many people accept a 1% chance of a loss without a compensation. But if the individual cannot avoid facing the lottery, she may react by adopting a form of defensive pessimism, by voluntarily focusing on the worst case, in order to prepare herself for the occurrence of a bad outcome. Psychologically, this emphasis on the possible bad outcome may lead her to anticipate that she will be better off *ex ante* with the risk, rather than being without it. This form of risk-seeking is stronger than the usual risk-seeking behavior explained by the convexity of the utility function that excludes such preferences.

To illustrate the "blessing of fear" numerically, consider the same specifications and parameters as in Sect. 2. We set that $\pi(p) = \frac{p^c}{(p^c+(1-p)^c)^{\frac{1}{c}}}$, which is the specification used by Tversky and Kahneman (1992) for the subjective probability transformation in the context of cumulative prospect theory. Since we have no empirical guidance about weights for stochastic reference points, using the transformation for probabilities in risk decisions seems a rather safe bet. We use $c = .69$. The difference $U(L_p|L_p) - U(y|y)$ is displayed on Fig. 2. As put forth in Proposition 6, for a small range near 1, this difference is positive, meaning $U(y|y) < U(L_p|L_p)$. Once again, the parameters are plausible, and this would happen for large stakes ($x = 100{,}000$ and $y = 150{,}000$). An individual endowed with such a utility function would choose to bear a 5% risk of losing a third of her income.

Overall, the introduction of subjective weights does little to limit the violations of stochastic dominance observed in the case where weights are exactly objective probabilities. When the weight of the favorable outcome is greater than its probability, it reinforces the curse of hope, making it more likely and possible even without loss aversion. When it is lower, the curse is mitigated, but it may give rise to the mirror violation of stochastic dominance (the blessing of fear), that no longer seems appealing.

4 Inertia in the updating of the reference point

Kőszegi and Rabin (2007) discuss the distinction between expected lotteries (or changes) and surprise lotteries. In particular, they assume that in the case of a surprise lottery, the reference point will not be updated with respect to expectations, and so the status quo will be the reference point (for instance in the lab where individuals have little, if any, prior knowledge of the possible outcomes). This distinction implicitly relies on the assumption that the adjustment of the reference point has a time dimension: surprise lotteries are evaluated relative to the status quo, only because this adjustment does not occur immediately. The adjustment of the reference point to future

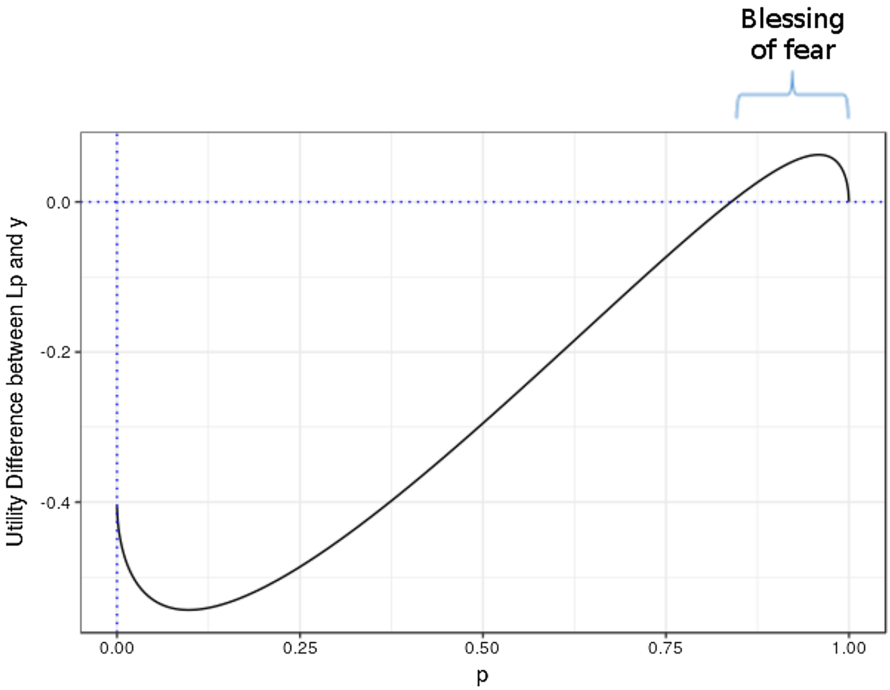


Fig. 2 The blessing of fear: $U(L_p|L_p) - U(y|y)$ as a function of p

expectations, given by the future lotteries the individual will face, takes time. When dealing with future (not surprising) lotteries, the assumption of rational expectations used by KR corresponds to two specific features. First, the adjustment is *complete*, that is the adjustment process of the reference point converges, with sufficient time, to the lottery that will be the new stochastic reference point. Second, the anticipation of the future adjustment is *accurate*, that is the individual correctly anticipates her future reference point. To illustrate these implicit assumptions, consider an individual who learns in t_0 that she will face a new lottery L_p in t_1. KR makes two assumptions: First, the reference point will change to be equal to L_p in t_1 at the time the lottery is resolved (the adjustment of the reference point is complete). Second, the individual perfectly predicts in t_0 that her reference point will be L_p in t_1 (the anticipation of this adjustment is correct).[11]

Regarding the first assumption (partial adjustment), one can envision that in many situations, the influence of the initial situation (in t_0) on the new reference point will persist in t_1 despite the anticipation of the new lottery. For instance, suppose that a worker learns in t_0 that she will be made redundant in one year, which will reduce her monthly income from 5000 per month to $y = 2000$ (or to $y = 2000$ or $x = 0$ with equal

[11] Note here that although time is present with respect to its influence on the reference point, intertemporal preferences and discounting do not matter because there is no trade-off between periods. The comparison is made between lotteries that are resolved at the same future period.

probabilities). Most of the time, she will not fully adjust her reference point to 2000 (or to a multiple stochastic point in which 0 and 2000 have an equal weight) despite knowing the odds in advance. Instead, she will keep her initial situation (an income of 5000) as part of her new multiple reference point. Indeed, the initial situation has long been the natural reference point considered in the literature, and recent studies seem to maintain an important role for this initial situation in the formation of the reference point (Baillon et al. 2017; Lien and Zheng 2015; Heffetz and List 2014).

Regarding the second assumption (*underestimation of reference adjustment*), the literature on the projection bias (Loewenstein et al. 2003) has shown that in many circumstances individuals tend to project their current situation onto future ones, even though it is obvious that they will differ. The change in the reference point actually reflects a process of adaptation called "shifted adaptation" by Frederick and Loewenstein (1999). It works for both gains and losses. And in a deterministic setting, numerous studies have shown that the anticipation of adaptation is incomplete, most of the time. People typically underestimate the extent to which they will adapt to positive as well as negative shocks (Gilbert et al. 1998). This is well documented, in particular in the area of health.[12] In a reference-based model of preferences, this means that individuals, although aware of the direction of the future adjustment of their reference point, can underestimate the magnitude of this adjustment. The worker mentioned above for instance may fail to predict in t_0 that having only 2000 in t_1 instead of 5000 will not be as painful as it may seem initially, once the adaptation process takes place. This corresponds to a violation of rational expectations, not about what the future lottery will be, but about what will happen to their own reference point. It is noteworthy that KR themselves pointed to such a limit of their own models.[13] In the presence of uncertainty, it is plausible that individuals, although aware of the future adjustment of their reference point because of expectations, tend to underestimate its extent.

To take into account these two possible effects (imperfect anticipation of the adjustment, and partial adjustment), we suggest treating them as a form of inertia in the projected reference point situation. The reference point an individual facing a future lottery L_p will be in part determined by this lottery, and in part by her current reference point. We assume that her projected multiple reference point is a convex combination of the initial reference point W and the future lottery L: $R(W, L) = qW + (1-q)L$ with q the inertia parameter. Now, denote w the current situation of the individual and by extension/simplification her original reference point ($W = w$), it gives

$$R(w, L_p) = qw + (1-q)L_p = (w, y, x; q, (1-q)p, (1-q)(1-p)) \tag{6}$$

with $q \in [0, 1]$ an individual parameter indicating the inertia with which an individual updates her reference/anticipates her adjustment. For $q = 0$, the individual fully

[12] People suffering from chronic illnesses or disability, report better mood, happiness or Quality of Life (QoL) ratings than what healthy people predict they would feel if facing similar circumstances (see for instance Riis et al. 2005 concerning dialysis), and for this reason, they have different preferences regarding what is an acceptable level of health (Brouwer et al. 2005; Wouters et al. 2015).

[13] People may also overattend to losses and gains because they underestimate how quickly they will adapt to these changes. On both of these accounts, the nature and scope of reference-dependent choices seems to reflect mistakes our fully rational model does not capture (KR 2006).

adjusts her stochastic reference point to the future lottery (KR's assumption) and for $q = 1$, she does not update her reference point at all (pure *status quo* bias). We assume that q is the same for any lottery or initial reference point. Hence, when the individual exhibits some inertia in the updating of her reference point, her expected utility for L_p is given by

$$U(L_p|R(w, L_p)) = qU(L_p|w) + (1-q)U(L_p|L_p) \tag{7}$$

And for the certain outcome x:

$$U(x|R(w, x)) = qU(x|w) + (1-q)U(x|x) \tag{8}$$

In the KR framework, the curse of hope occurs when $U(L_p|L_p) < U(x|x)$. Under the assumption that the reference point incorporates the former reference point w because of inertia (Eq. 6), the curse of hope now occurs if $U(x|R(w, x)) > U(L_p|R(w, L_p))$. Given that the individual gives more weight to her past situation w, the condition is stricter and the curse of hope less likely. Specifically, we can establish the following proposition:

Proposition 7 *$x \succ L_p$ iff $-\Delta\mu > \frac{\Delta m}{1-p} + A(q)$ where $A(0) = 0$ and $A'(q) > 0$.*[14] *Moreover, as q increases, the range of probabilities on which $x \succ L_p$ shrinks.*

When $q = 0$, this proposition is equivalent to Proposition 1. In the trivial case where $q = 1$, we have $U(x|R(w, x)) = U(x|w)$ and $U(L_p|R(w, L_p)) = U(L_p|w)$. Since first-order stochastic dominance is never violated when the reference point is held constant, $U(x|w) < U(L_p|w)$ so that the curse of hope completely disappears. For $0 < q < 1$, the higher the weight the individual gives to her present reference point w, the less likely this condition will be satisfied.

This result holds for all w, and in particular when w corresponds to one of the final outcomes. When $w = x$, the reference point is $R(x, L_p) = (y, x; \pi, 1-\pi)$ with $\pi = (1-q)p < p$. The individual gives more weight and attention to the unfavorable outcome because it corresponds to the present situation, which makes the curse of hope less likely as in Sect. 3. When $w = y$, we have $R(y, L_p) = (y, x; \pi', 1-\pi')$ with $\pi' = q + (1-q)p > p$. The agent now gives more weight to the favorable outcome. However, in this model with inertia, and contrary to Proposition 5, this does not make the curse of hope more likely: this higher weight reduces the expected utility of the lottery L_p, i.e., $U(L_p|R(y, L_p) < (L_p|L_p)$, but it does so less than it reduces the utility of choosing x for sure, which is now given by $U(x|R(y, x)) = m(x) + q\mu^-$ where $q\mu^-$ corresponds to the additional sensation of loss of having x for sure when y was the previous present situation.[15]

A critical question is whether a sufficiently strong inertia can make the curse of hope fully disappear. Put differently, the issue is to know whether the curse of hope

[14] More in detail (see Appendix), $A(q) = \frac{q(\Delta m + \Delta\mu_w)}{(1-p)(1-q)}$. Note that it is only defined for $q \neq 1$ and $p \neq 1$.

[15] Indeed, we can easily verify that $U(L_p|R(y, L_p)) - U(x|R(y, x)) > U(L_p|L_p) - U(x|x)$, that is the relative interest of the lottery (compared to x for sure) is higher in the model with inertia even when y is the initial situation.

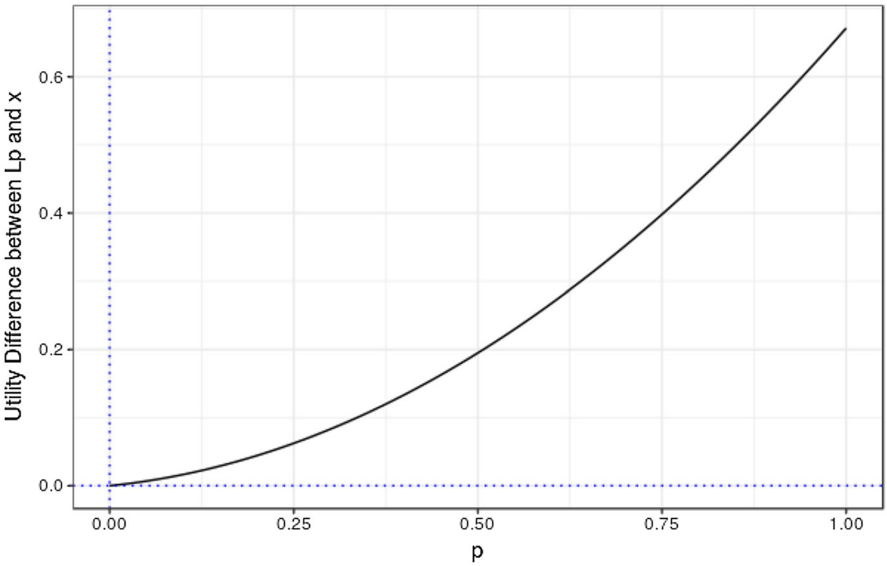


Fig. 3 The absence of the curse of hope for $q = .2$: $U(L_p|R(w, L_p)) - U(x|R(w, x))$ as a function of p

is still present for q large enough, but lower than 1. This will mean that most of KR's qualitative results are still true, while violation of stochastic-dominance disappears. We show that if the reference-dependent component of U does not dominate without bounds the intrinsic term in marginal terms (that is μ' is bounded), then a large enough inertia—yet still less than 1—can make the curse of hope disappear:

Proposition 8 *If μ' is bounded and q is large enough, then $U(L_p|R(w, L_p)) > U(x|R(w, x))$ for all p, x, y and w.*

In other words, under the assumption that the marginal effect of the reference-dependent component is bounded,[16] it is always possible to find an inertia parameter q large enough, yet less than 1, so that the curse of hope never occurs. There is a threshold value, $\bar{q}$, so that for any inertia parameter greater than this threshold the curse of hope is not predicted to occur. Proposition 9 extends this result to non-degenerated lotteries:

Proposition 9 *If μ' is bounded and q is large enough, then $U(L_p|R(w, L_p)) > U(L'_p|R(w, L'_p))$ for all $p > p'$, x, y and w.*

Most usual specifications of μ satisfy this bounded derivative condition.[17]

To pursue with our example, Fig. 3 displays the difference $U(L_p|R(w, L_p)) - U(x|R(w, x))$ for a relatively small inertia parameter $q = .2$ and an intermediate

[16] Note that this is necessarily the case if $\mu'_-(0) < +\infty$, given that μ' decreases in the distance to the reference point, and that $\mu'_-(0) \geq \mu'_+(0)$.

[17] A counter-example is given by power functions for which $\mu'_-(0) = \infty$. In this case, the curse of hope cannot disappear fully (or trivially by setting $q = 1$), but a positive q shrinks the range of stakes $(y - x)$ where it applies.

$w = 120{,}000$. The difference is monotonically increasing in p and is always positive: the curse of hope has vanished. The threshold value of q, as calculated in the proof of Proposition 9 that guarantees the curse of hope never occurs for this specification is $\frac{2}{3}$. Note though that it is a higher bound of the exact threshold, given that some positive terms are dropped in the proof.

5 Conclusion

We have argued that the curse of hope, a violation of stochastic dominance implied by KR's model, is likely to be a theoretically widespread phenomenon. This is at odds with the scarce empirical evidence that does not point to frequent violations of stochastic dominance. While a simple subjective transformation of probability into weights of the reference point may in some cases mitigate the issue, in others, it can intensify it or even generate new ones in the form of a "blessing of fear."

These paradoxical effects appear because in a sense, the KR model gives too much weight to anticipations in the reference point and so to the magnitude of the possible sensations of hope and disappointment. This paper shows that a modified version of the KR model which would take into account some inertia in the updating of the reference point (due to an imperfect anticipation of the future adjustment or to a partial adjustment) is able to prevent the curse, for a large set of specifications. This may help reconcile the idea of reference points being influenced by expectations with the lack of empirical evidence for stochastically dominated choices in contexts where the model predicts it should occur. This modified version has the advantage of maintaining an influence of expectations on the reference point without its most problematic prediction. This comes at the price of making the model slightly less parsimonious in the sense that it adds a new parameter. Arguably, this modification is rather frugal, as it involves a single parameter that should be estimated empirically or experimentally. Most importantly, this modification does not change the main qualitative results established by KR: it maintains that risk aversion should be large for small stakes, and mild for large stakes.

In the end, this amended version of KR's model raises a welfare question. Indeed, the two reasons behind inertia that we have put forward (incomplete adjustment, inaccurate anticipation) lead to two very different consequences in terms of expected welfare in the future. When individuals do not fully adjust to new situations (or lotteries that will determine these situations), their decision to choose the stochastically dominant lottery is the rational one in terms of expected welfare. The utility function that determined their choice in the initial situation exactly mirrors their situation (and their reference point) at the moment the lottery will be resolved. In contrast, an inaccurate anticipation corresponds to the failure of individuals to foresee fully their future adjustment. Individuals in this case choose the stochastically dominant lottery, but doing so they do not always maximize their future expected welfare. They fail to predict the adjustment of their reference point and suffer a curse of hope in terms of welfare. They would sometimes have been better off in expectation by choosing the dominated lottery. The curse of hope may not exist in the domain of choices but may very well be present in terms of welfare: people may indeed engage in risky activities with stochastically-dominant consequences in the future and yet end up dissatisfied because of a strong disappointment.

Appendix

Proof of Proposition 2

$$\frac{\partial U(L_p|L_p)}{\partial p} = \Delta m + (1-2p)\mu^- + (1-2p)\mu^+$$

It is positive if $p < \frac{\Delta m + \Delta\mu}{\Delta\mu}$, that is if $p < \frac{\bar{p}}{2}$. Proposition 2 ensues.

We complement the proof by some additional elements on the behavior of $U(L_p|L_p)$.

$$\frac{\partial^2 U(L_p|L_p)}{\partial p^2} = -2[\mu^- + \mu^+]$$

Because of loss aversion, for any $t > 0$ we have

$$[\mu(-t) + \mu(t)] < 0$$

Hence

$$\frac{\partial^2 U}{\partial p^2} = -2[\mu^- + \mu^+] > 0$$

So $U(p)$ is convex. Since $\frac{\partial U(L_p|L_p)}{\partial p}(\frac{\bar{p}}{2}) = 0$, we have $U(L_p|L_p)$ decreasing in p up to $\bar{p}/2$ and increasing afterwards, reaching 0 at $\bar{p}$. This means that $p \leq \bar{p}$ not only implies $L_p \succ x$ but that the two are equivalent.

Proof of Proposition 3

For infinitesimal differences between x and y, the condition becomes

$$\lim_{\Delta m \to 0^+} \frac{-\Delta\mu}{\Delta m} > 1$$

That gives

$$\lim_{\Delta m \to 0^+} \frac{-\mu^+ - \mu^-}{\Delta m} > 1$$

Hence

$$-\mu'_+(0) + \mu'_-(0) > 1$$

That is

$$\mu'_+(0) > \frac{1}{\lambda - 1}$$

The limit of $\bar{p}$ is given by

$$\lim_{\Delta m \to 0^+} \bar{p} = \lim_{\Delta m \to 0^+} \left(1 + \frac{\Delta m}{\Delta \mu}\right) = \lim_{\Delta m \to 0^+} \left(1 + \frac{1}{1/(\Delta\mu/\Delta m)}\right)$$
$$= 1 + \frac{1}{(1-\lambda)\mu'_+(0)}$$

Proof of Proposition 4

$m(r+\delta) - m(r)$ is less than $m(r'+\delta) - m(r')$ because $m'' < 0$. Hence by continuity there exists $\delta' > \delta$ with $m(r'+\delta') - m(r') = m(r+\delta) - m(r)$. Given that the condition for the curse of hope to occur at r and $r+\delta$ is that $\mu(m(r+\delta) - m(r)) + \mu(m(r) - m(r+\delta)) > m(r+\delta) - m(r)$, it holds that $\mu(m(r'+\delta') - m(r')) + \mu(m(r') - m(r'+\delta')) > m(r'+\delta') - m(r')$; hence the curse of hope occurs at r' and $r'+\delta'$.

Proof and conditions of Proposition 5

$x \succ L_p$ iff $m(x) - U(L_p|R_\pi)$, that is, using Eq. (4) and rearranging, if

$$-(\pi((1-p)/p)\mu^- + (1-\pi)\mu^+) > \Delta m$$

In the absence of loss aversion, $-\mu^- = \mu^+$ and the inequality boils down to $((\pi - p)/p)\mu^+ > \Delta m$, which is satisfied for some values of the parameters, in particular when π is large enough.

Proof of Proposition 7

We first note that

$$U(x|R(w,x)) = qU(x|w) + (1-q)U(x|x) = m(x) + q\mu(m(x) - m(w))$$

and

$$\begin{aligned} U(L_p|R(w,L_p)) &= qU(L_p|w) + (1-q)U(L_p|L_p) \\ &= q[p(m(y) + \mu(m(y) - m(w))) \\ &\quad + (1-p)(m(x) + \mu(m(x) - m(w)))] \\ &\quad + (1-q)[m(x) + p\Delta m + p(1-p)(\mu^+ + \mu^-)] \\ &= m(x) + p\Delta m + qp\Delta\mu_w + q\mu(m(x) - m(w)) \\ &\quad + (1-q)p(1-p)\Delta\mu, \end{aligned}$$

where $\Delta\mu_w = \mu(m(y) - m(w)) - \mu(m(x) - m(w)) \geq 0$, since $y > x$.

A curse of hope still exists if $U(x|R(w,x)) > U(L_p|R(w,L_p))$, that is, after replacement:

$$-p\Delta m - qp\Delta\mu_w - (1-q)p(1-p)\Delta\mu > 0$$

That is

$$\begin{aligned}
-\Delta\mu &> \frac{\Delta m + q\Delta\mu_w}{(1-q)(1-p)} \\
&> \frac{\Delta m}{1-p} + \frac{q\Delta m}{(1-q)(1-p)} + \frac{q\Delta\mu_w}{(1-q)(1-p)} \\
&> \frac{\Delta m}{1-p} + \frac{q(\Delta m + \Delta\mu_w)}{(1-q)(1-p)}
\end{aligned}$$

Set $A(q) = \frac{q(\Delta m+\Delta\mu_w)}{(1-q)(1-p)}$, $A(0) = 0$ and $A'(q) > 0$.

Now compare the value of the threshold probability below which the curse of hope occurs in the absence of inertia ($\bar{p}$) and the threshold probability for some value of q, denoted $\bar{p}_q$. As already noted, $\bar{p} = \frac{\Delta m+\Delta\mu}{\Delta\mu}$. For $q > 0$, we have

$$-\Delta m - q\Delta\mu_w - (1-q)(1-\bar{p}_q)\Delta\mu = 0$$

After rearrangement:

$$\bar{p}_q\Delta\mu = \Delta\mu + \frac{\Delta m + q\Delta\mu_w}{1-q}$$

That is,

$$\bar{p}_q\Delta\mu = \frac{(1-q)\Delta\mu + (1-q)\Delta m + q\Delta m + q\Delta\mu_w}{1-q}$$

$$\bar{p}_q = \frac{\Delta\mu + \Delta m}{\Delta\mu} + \frac{q}{1-q}\frac{\Delta m + \Delta\mu_w}{\Delta\mu}$$

In the end,

$$\bar{p}_q = \bar{p} + \frac{q}{1-q}\frac{\Delta m + \Delta\mu_w}{\Delta\mu}$$

Note that $\frac{\Delta m+\Delta\mu_w}{\Delta\mu} < 0$ and $\frac{q}{1-q}$ is increasing in q and is unbounded for q arbitrarily close to 1. Hence, $\bar{p}_q$ decreases in q and the interval $[0, \bar{p}_q]$ on which the curse of hope occurs shrinks as q increases. Moreover, for any μ_w, i.e. for any w, exists $q < 1$ such that $\bar{p}_q = 0$.

Proof of Proposition 8

We have:

$$U(L_p|R(w, L_p)) - U(x|R(w, x)) > p\Delta m + (1-q)[p(1-p)\mu^+ + p(1-p)\mu^-]$$

The right-hand side is of the same sign as

$$\Delta m + (1-q)[(1-p)\mu^+ + (1-p)\mu^-]$$

A sufficient condition for this to be positive for any p is that

$$\Delta m + (1-q)[\mu^+ + \mu^-] > 0$$

Define f on $\mathbb{R}_+$ by $f(\Delta m) = \Delta m + (1-q)[\mu^+ + \mu^-]$; then its derivative is given by

$$f'(\Delta m) = 1 + (1-q)[\mu'(\Delta m) - \mu'(-\Delta)m]$$

Because μ' is bounded, exists $b > 0$ such that $\mu'(\Delta m) - \mu'(-\Delta) > -b$. That is

$$f'(\Delta m) > 1 - (1-q)b$$

Taking $q > 1 - \frac{1}{b}$ ensures that $f'(\Delta m) > 0$ for all $\Delta m \geq 0$. Moreover, $f(0) = 0$; hence f has non-negative values.

To conclude, $q > 1 - \frac{1}{b}$ implies that $\Delta m + (1-q)[\mu^+ + \mu^-] > 0$, which implies $U(L_p|R(w, L_p)) - U(x|R(w, x)) \geq 0$.

Proof of Proposition 9

We have for any binary lottery

$$\begin{aligned} U(L_p|R(w, L_p)) = \; & pm(y) + (1-p)m(x) \\ & + q[p\mu(m(y) - m(w)) + (1-p)\mu(m(x) - m(w))] \\ & + (1-q)p(1-p)[\mu(m(x) - m(y)) + \mu(m(y) - m(x))] \end{aligned}$$

Then its derivative is given by

$$\begin{aligned} \frac{\partial U(L_p|R(w, L_p))}{\partial p} = \; & m(y) - m(x) + q[\mu(m(y) - m(w)) - \mu(m(x) - m(w))] \\ & + (1-q)(1-2p)[\mu(m(x) - m(y)) + \mu(m(y) - m(x))] \end{aligned}$$

And

$$\frac{\partial^2 U(L_p|R(w, L_p))}{\partial p^2} = -2q[\mu(m(x) - m(y)) + \mu(m(y) - m(x))] > 0$$

Hence $U(L_p|R(w, L_p))$ is convex.

The sign of the first-order derivative is given by the sign of

$$\Delta m + q[\mu(m(y) - m(w)) - \mu(m(x) - m(w))] \\ + (1 - q)(1 - 2p)[\mu(\Delta m) + \mu(-\Delta m)]$$

Given that $\mu(m(y) - m(w)) - \mu(m(x) - m(w)) > 0$, this is positive for any p if

$$\Delta m + (1 - q)[\mu(\Delta m) + \mu(-\Delta m)] > 0$$

Which is the exact same condition studied in the proof of Proposition 8.

A generalization to non-binary lotteries

First suppose the effect exists for some $L_p = (p, y; 1 - p, x)$ with $0 < p < 1$, i.e. the condition of Proposition 1 is satisfied. We then proceed by induction.

Suppose the effect occurs for n fixed. That is, there exists

$$L_n = (p_1, p_2, \ldots, p_n; x_1, x_2, \ldots x_n)$$

with $x_1 < x_2 < x_2 < \cdots < x_n$ and for all $k = 1, \ldots, n$, $p_k > 0$. L_n is such that $U(x_1|x_1) > U(L_n|L_n)$. We want to show that there exists

$$L_{n+1} = (q_1, q_2, \ldots, q_n, q_{n+1}; x_1, x_2, \ldots, x_{n+1})$$

with for all $k = 1, \ldots, n+1$, $q_k > 0$ and x_{n+1} different from any x_k for $k = 1, \ldots, n$.

For $\epsilon < p_1$, denote $L_\epsilon = L_{n+1} = (p_1 - \epsilon, p_2, p_3, \ldots, p_n, \epsilon; x_1, x_2, \ldots, x_n, x_{x+1})$. It is straightforward that the mapping f defined for any $\epsilon \in [0, p_1)$ by $f(\epsilon) = U(L_\epsilon|L_\epsilon)$ is continuous because of the continuity of U with respect to probabilities, and that $f(0) = U(L_n|L_n)$. This is true independently of x_{n+1}. Given that $f(0) = U(L_n|L_n) < U(x_1|x_1)$, by continuity exists $\epsilon > 0$ small enough so that

$$f(\epsilon) = U(L_\epsilon|L_\epsilon) < U(x_1|x_1)$$

Hence the existence of L_ϵ implies the existence of L_{n+1}. It obtains: if exists L_n such that $U(L_n|L_n) < U(x_1|x_1)$, then there exists L_{n+1} with $U(L_{n+1}|L_{n+1}) < U(x_1|x_1)$.

By assumption and Proposition 1, L_2 exists. Hence exists L_k for any k with $U(L_{k+1}|L_{k+1}) < U(x_1|x_1)$.

References

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *The American Economic Review*, *101*, 470–492.

Baillon, A., Bleichrodt, H., & Spinu, V. (2017). Searching for the reference point. Erasmus University of Rotterdam, Mimeo.

Bartling, B., Brandes, L., & Schunk, D. (2015). Expectations as reference points: Field evidence from professional soccer. *Management Science*, *61*, 2646–2661.

Birnbaum, M. (2005). A comparison of five models that predict violations of first-order stochastic dominance in risky decision making. *Journal of Risk and Uncertainty*, *31*, 263–287.

Brouwer, W., Van Exel, N., Rohde, K., & Wouters, S. (2005). Acceptability of less than perfect health states. *Social Science and Medicine, 60*(2), 237–246.

Butler, D., Isoni, A., Loomes, G., & Tsutsui, K. (2014). Beyond choice: Investigating the sensitivity and validity of measures of strength of preference. *Experimental Economics*, *17*, 537–563.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*, 1433–1436.

Card, D., & Dahl, G. B. (2011). Family violence and football: The effect of unexpected emotional cues on violent behavior. *The Quarterly Journal of Economics*, *126*, 103.

Clark, A., Frijters, P., & Shields, M. (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature*, *46*, 95–144.

Crawford, V. P., & Meng, J. (2011). New york city cab drivers' labor supply revisited: Reference-dependent preferences with rational expectations targets for hours and income. *The American Economic Review*, *101*, 1912–1932.

Eastridge, B. J., Mabry, R. L., Seguin, P., Cantrell, J., Tops, T., Uribe, O., et al. (2012). Death on the battlefield (2001–2011): Implications for the future of combat casualty care. *Journal of Trauma and Acute Care Surgery*, *73*, S431–S437.

Ericson, K. M. M., & Fuster, A. (2011). Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. *The Quarterly Journal of Economics*, *126*, 1879–1907.

Fishburn, P. (1978). On Handa's "New theory of cardinal utility" and the maximization of expected return. *Journal of Political Economy*, *86*, 321–324.

Frederick, S., & Loewenstein, G. (1999). Hedonic adaptation. In D. Kahneman, E. Diener, & N. Schwartz (Eds.), *Scientific perspectives on enjoyment, suffering, and well-being*. New York: Russel Sage Foundation.

Gilbert, D. T., & Ebert, J. E. (2002). Decisions and revisions: The affective forecasting of changeable outcomes. *Journal of Personality and Social Psychology*, *82*, 503.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *75*, 617.

Gill, D., & Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *The American Economic Review*, *102*, 469–503.

Gneezy, U., List, J. A., & Wu, G. (2006). The uncertainty effect: When a risky prospect is valued less than its worst possible outcome. *The Quarterly Journal of Economics, 121*(4), 1283–1309.

Heffetz, O., & List, J. A. (2014). Is the endowment effect an expectations effect? *Journal of the European Economic Association*, *12*, 1396–1422.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291.

Keren, G., & Willemsen, M. C. (2009). Decision anomalies, experimenter assumptions, and participants' comprehension: Revaluating the uncertainty effect. *Journal of Behavioral Decision Making*, *22*, 301–317.

Kőszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics, 71*(4), 1133–1165.

Kőszegi, B., & Rabin, M. (2007). Reference-dependent risk attitudes. *The American Economic Review*, *97*, 1047–1073.

Leland, J. (1998). Similarity judgments in choice under uncertainty: A reinterpretation of the predictions of regret theory. *Management Science*, *44*, 659–672.

Lien, J. W., & Zheng, J. (2015). Deciding when to quit: Reference-dependence over slot machine outcomes. *The American Economic Review*, *105*, 366–370.

Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Projection bias in predicting future utility. *The Quarterly Journal of Economics*, *118*, 1209–1248.

Loomes, G., Starmer, C., & Sugden, R. (1992). Are preferences monotonic? Testing some predictions of regret theory. *Economica*, *59*, 17–33.

Macé, S. (2016). Un faible espoir de guérison est-il toujours une bonne chose? *Revue française d'économie, XXXI*, 153–179.

Masatlioglu, Y., & Raymond, C. (2016). A behavioral analysis of stochastic reference dependence. *The American Economic Review*, *106*, 2760–2782.

Norem, J. K. (2001). Defensive pessimism, optimism, and pessimism. In E. C. Chang (Ed.), *Optimism and pessimism: Implications for theory, research, and practice* (pp. 77–100). http://www.apa.org/pubs/books/431754A.aspx.

Norem, J. K., & Cantor, N. (1986). Defensive pessimism: Harnessing anxiety as motivation. *Journal of Personality and Social Psychology*, *51*, 1208.

Riis, J., Loewenstein, G., Baron, J., Jepson, C., Fagerlin, A., & Ubel, P. A. (2005). Ignorance of hedonic adaptation to hemodialysis: A study using ecological momentary assessment. *Journal of Experimental Psychology: General*, *134*, 3.

Rydval, O., Ortmann, A., Prokosheva, S., & Hertwig, R. (2009). How certain is the uncertainty effect? *Experimental Economics*, *12*, 473–487.

Sharf, B. F., Stelljes, L. A., & Gordon, H. S. (2005). 'A little bitty spot and I'm a big man': Patients' perspectives on refusing diagnosis or treatment for lung cancer. *Psycho-Oncology*, *14*, 636–646.

Simonsohn, U. (2009). Direct risk aversion evidence from risky prospects valued below their worst outcome. *Psychological Science*, *20*, 686–692.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.

Wouters, S., Van Exel, N., Rohde, K., & Brouwer, W. (2015). Are all health gains equally important? An exploration of acceptable health as a reference point in health care priority setting. *Health and Quality of Life Outcomes, 13*(1), 13–79.

TRAVAUX EN COURS MENTIONÉ DANS LA SYNTHÈSE ET DANS l'OEUVRE SCIENTIFIQUE CAR EN L'ETAT D'ÊTRE PUBLIÉ 3: Fabrice Le Lec et Joel Santos, "On attitude towards extreme risks".

On attitude toward extreme risks

Fabrice Le Lec*, Joel Santos†

May 22, 2018

Abstract: This article studies attitudes toward extreme risks, defined as rare occurrences of dramatic consequences. We propose a model-free measure of attitude toward extreme risks based on the value of the risk premium in a setting where individuals can fully insure against an extreme risk. More specifically, we study how generalized expected utility models impact this attitude. We show the critical role played by the derivative of the probability transformation function. In particular, for rank-dependent expected utility models, the usual empirically-based specifications imply that individuals are hyperaverse to extreme risks, i.e. their willingness to pay for not bearing an extreme risk can be arbitrarily larger that the expected loss. This is at odds with the empirical findings that the response to extreme risk is bimodal, with a large share of individuals simply discarding such risks. Applying this measure to cumulative prospect theory, a polarization of attitudes emerges: under this theory, individuals can be either hyperaverse to extreme risks or insensitive to them. Moreover, we show that, for prospect theory, some apparently secondary feature of the decision, namely the distance to the reference point, plays a major role. It may lead individuals to be insensitive to extreme risks for an insurance purchase, as often observed, and hyperaverse to extreme risks in an investment setting.

JEL Classification : D03, D81

Keywords: extreme risk, low probability, massive consequence, catastrophe, disaster

1 Introduction

Many risks people face belong to the category of extreme risks: they are rare but of dramatic consequences. This category includes individual bankruptcies, fatal diseases, car accidents, house fires or flood, or, at a collective level, financial crashes, macroeconomic disasters, nuclear meltdown or major terrorist attacks. In all these examples, there is a very low probability of occurrence but if a bad outcome was to occur, it would have massively negative consequences. Extreme risks have attracted a lot of attention from the point of view of insurance behavior, in particular since the seminal work of Kunreuther and Slovic (1978). Many insurance decisions relate to some form of extreme risk in the sense that the covered outcomes are rather rare and their consequences generally important. For most insured consumers, serious health conditions, car accidents, house fires, floods, tornadoes are relatively rare events. The topic of extreme risk has gained renewed interest in recent years through natural or ecological catastrophes (Rheinberger and Treich, 2016; Krawczyk, Trautmann, and van de Kuilen, 2016; Eeckhoudt, Schieber, and Schneider, 2000) or the macroeconomic or financial impact of catastrophes in a broad sense (Barro, 2006; Gabaix, 2012; Farhi and Gabaix, 2016). Most of these works are carried out in the

*Université Paris 1, Centre d'Economie de la Sorbonne, umr CNRS 8174, 106-112 Boulevard de l'Hôpital, 75013 Paris, France, fabrice.le-lec@univ-paris1.fr

†IESEG School of Management, Catholic University of Lille, Lille, France and EDHEC Business School, Lille, France, j.santos@ieseg.fr

framework of expected utility:[1] individuals' attitudes toward such risks can only be accounted for through the curvature of the utility function. Because of the linear treatment of probabilities, whether the risk is extremely rare or just infrequent only affects individual behavior proportionally. The main relevant parameter is the magnitude of the negative event. Aumann and Kurz (1977) defined a measure coined "Fear of ruin" to account for attitude toward extreme risk: the "fear of ruin" depends only on the utility function and, as a consequence, the willingness to pay to avoid the risk of ruin is simply linear in the probability of occurrence. In this sense, the linear treatment of probabilities implies that fear of ruin essentially measure attitudes to great losses (or dramatic outcomes). Small probabilities *per se* play a minor part.

Yet, most of the empirical studies dedicated to extreme risks highlight that such risks do not seem to be treated through the usual expected utility framework. Numerous studies have shown that individuals do not react to the possibility of extreme events as would be predicted by expected utility models. Moreover, the literature seem do be quite divided about how individuals' behaviors depart from this framework. Starting from the seminal work of Kahneman and Tversky (1979), the view has developped that individuals can either be very sensitive towards low probability-large consequences events, or may simply neglect them. The empirical findings echoes this dichotomy: some studies find that individuals underpurchase insurance against natural catastrophes (Kunreuther and Slovic, 1978; Schade, Kunreuther, and Kaas, 2004), some estimate on the contrary that individuals strongly overweight rare events based on real data on insurance purchases and deductibles (Barseghyan, Molinari, O'Donoghue, and Teitelbaum, 2013), and eventually other studies find a rather bimodal distribution of behaviors (McClelland, Schulze, and Coursey, 1993; Botzen and van den Bergh, 2012) where many individuals seem to neglect such extreme risks while those who do not are willing to pay a large risk premium. Many reasons have been put forth to explain either the neglect of extreme risks, or the coexistence of contradictory attitudes, among which wrong beliefs about the probability of occurence of such events and the inability of individuals to handle accurately very small probabilities (Botzen, Kunreuther, and Michel-Kerjan, 2015), the role of the risk format presentation and cognitive salience (Erev, Glozman, and Hertwig, 2008) or the role of ambiguity (Hogarth and Kunreuther, 1989) – see also Camerer and Kunreuther (1989) for an early survey on the topic. Our purpose is to study systematically the theoretical implications of generalized expected utility models such as rank-dependent expected utility and cumulative prospect theory to clarify their implications and assess whether these models lead to predictions congruent with the empirical evidence.

To do so, we propose to study attitudes toward extreme risk in a general framework that allows the probability of occurrence of the bad consequence to play a role: we generalize fear of ruin that assumes a linear treatment of probability to a coefficient of aversion to extreme risk that takes into account subjective probability distortion. The critical role of the subjective probability transformation in the case of low probability events has been highlighted in Bleichrodt and Eeckhoudt (2006) in the context of health decisions and the value of a statiscal life. Our results show that attitude toward extreme risk depends critically on the behavior of the derivative of the probability transformation function (the marginal probability sensitivity) at the boundaries. More specifically, the aversion to extreme risk coefficient can be shown to be the product of the derivative of the probability transformation and Aumann and Kurz's fear of ruin. Given that in most empirically based specifications, this derivative is unbounded at boundaries, our developments suggest that, would rank-dependent expected utility models describe accurately individuals' preferences, *hyperaversion* to extreme risks should be widespread. This result is at odds with the empirical literature, especially with respect to insurance against catastrophic events: as reported above, extreme risk neglect is a relatively frequent attitude.

[1] There is a recent study interested in insurance behavior for low probability-large loss risks using prospect theory, but it maintains the assumption of linearity in probabilities (Schmidt, 2016).

While, more recent results point to a bi-modal distribution of sensitivities, with insensitivity and strong aversion to extreme risks both present (McClelland, Schulze, and Coursey, 1993; Schade, Kunreuther, and Kaas, 2004; Botzen and van den Bergh, 2012), no empirical study gives unambiguous support to the conclusion that people are very averse to extreme risks. As a consequence, rank-dependent expected utility cannot account for a lack of concern toward extreme risk as seems very present in the empirical literature.

A possible congruence with the empirical stylized facts may be obtained with cumulative prospect theory (Tversky and Kahneman, 1992), through its emphasis on the role of the reference point. Cumulative prospect theory differs from rank-dependent expected utility by the critical importance of the marginal utility in the neighborhood of the reference point. This leads the model to two cases: the individuals exhibit either hyperaversion to extreme risk or plain insensitivity, with little room left for an intermediate attitude. This is more consistent with the empirical bi-modal distribution often observed. Interestingly, cumulative prospect theory implies hyperaversion to extreme risk once the outcomes are shifted from the reference point. In other words, the individual can be insensitive to extreme risk exactly at the reference point, but may exhibit extreme risk hyperaversion everywhere else. It allows an individual not to purchase an insurance for some catastrophic risk, given that an insurance purchase takes place at the current level of wealth, i.e. at the individual's reference point. But for a financial decision, where the returns of a safe assets shifts the level of wealth away from the reference point, her cost of engaging into an extreme risks investment is very high. These findings may help reconcile the low level of catastrophe insurance coverage and the existence of a large premium for assets that have an extreme risk component (e.g., "equity premium puzzle"). More generally, we show that, under cumulative prospect theory, the location of the outcomes to which the protection cost applies relative to the reference point matters critically, leading to the possibility of both extreme risk hyperaversion and insensitivity for the same individual.

The remainder of the paper is organized as follows: the next section defines attitudes toward extreme risk and recalls Aumann and Kurz's fear of ruin. The third section studies attitude to extreme risk in the context of generalized expected utility models and show that under common specification it leads to hyperaversion to extreme risks. The fourth section is dedicated to cumulative prospect theory and highlights the possibility of insensitivity to extreme risk. The fifth section presents a more general framework to approach extreme risk under prospect theory and shows that while individuals can be insensitive at the reference point, they may exhibit hyperaversion everywhere else. The sixth section presents numerical applications under plausible parameter values to assess the economic relevance of our results. Eventually section 6 concludes.

2 Attitude Toward Extreme Risk: Definition

We define an extreme risk as a prospect where the decision maker faces a very small probability of a massive loss: an individual is endowed with a level of wealth W and faces probability ϵ of a large loss L. The corresponding lottery is hence: W with probability $1-\epsilon$ and $W-L$ with probability ϵ, which we denote as $\mathcal{L}(\epsilon) = (1-\epsilon, W; \epsilon, W-L)$.

The agent's willingness to pay not to bear this risk, denoted z, provides a model-free measure of her attitude toward such risks. Said differently, z is the maximum premium the individual is willing to pay to get a full coverage insurance. The introduction of fixed deductibles or partial coverage has no major impact on the results.[2]. Let V be the value function representing

[2]It is easy to check that partial coverage or deductibles simply impact the "fear of ruin" component of our results.

the individual's preferences under risk. The willingness to pay z satisfies: $V\big((1, W - z(\epsilon))\big) = V(\mathcal{L}(\epsilon))$ with $(1, W - z(\epsilon))$ being the certainty of keeping one's wealth minus the willingness to pay. A convenient measure of the attitude toward extreme risk[3] is given by the ratio $\zeta(\epsilon)$ of the willingness to pay z over the expected loss: $\zeta(\epsilon) = \frac{z(\epsilon)}{\epsilon L}$. This coefficient depends on the size of the loss and as a consequence is unique for an individual for a given loss. We use a ratio rather than a difference as is usual in risk aversion because it is more appropriate to study extreme risks for which ϵ approaches 0. The usual risk premium, i.e. $z(\epsilon) - \epsilon L$ here, will have a limit value of 0 and as consequence is not very informative. Moreover, this ratio will reveal itself quite convenient to compare theories.

Since extreme risks are defined on the basis of very small probabilities for an adverse consequence, a natural measure is to consider the limit of this ratio when ϵ approaches 0. We hence refer to the *attitude to extreme risk coefficient of* as ζ given by: $\zeta = \lim_{\epsilon \to 0} \zeta(\epsilon)$. It provides a simple and unified measure that is independent of the underlying model of preferences under risk: when ζ is greater than 1, the individual is averse to extreme risks, that is, is willing to pay more than the expected loss to be covered. If ζ is less than 1 she is extreme risk seeking. Neutrality toward extreme risks corresponds to a ζ of exactly 1. The empirical literature, as well as some non-expected utility models as we will show, suggests the possibility of radical attitudes to extreme risk: we define *hyperaversion to extreme risk* as the case where ζ is infinite and *insensitivity to extreme risk* as the one where ζ is zero.[4] It is worth noting that if z is differentiable at 0, ζ is given by $\zeta = \lim_{\epsilon \to 0} \zeta(\epsilon) = \frac{z'(0)}{L}$. The critical element of our extreme risk coefficient is hence the derivative of the willingness to pay at 0.

The use of such a limit coefficient to account for the attitude toward extreme risks was earlier proposed within the notion of "fear of ruin" (Aumann and Kurz, 1977; Foncel and Treich, 2005) in the context of expected utility theory. When the utility function u is differentiable, the main result is the following:[5]

Proposition 1 (Aumann and Kurz 1972; Foncel and Treich 2002) *For an expected utility maximizer, the willingness to pay not to bear an extreme risk is given by:*

$$z(\epsilon) \simeq \epsilon \frac{u(W) - u(W - L)}{u'(W)} \tag{1}$$

with $\frac{u(W)-u(W-L)}{u'(W)} = F_{EU}$ *the* fear of ruin *coefficient. The corresponding attitude to extreme risk coefficient is:* $\zeta = F_{EU}/L$.

For expected utility, the willingness to pay to avoid a risk of ruin is smooth with respect to the probability of the loss: ϵ has an approximately proportional effect on the willingness to pay. The fear of ruin coefficient depends only on the curvature of u: for a risk-averse individual ($u'' < 0$), it is greater than L and her willingness to pay is greater than ϵL while the opposite holds for a risk-seeking individual. Or equivalently, ζ is greater than 1 for a risk-averse individual while lower than 1 for a risk-seeking one.

[3]A note of terminology is in order: we use the usual yet perhaps misguided "risk-averse" or "risk-seeking" terminology to refer to the curvature, concave or convex, of the utility function and not to "probability risk aversion" as is usual in generalized expected utility framework, given that the probability transformation has a very specific role in our framework. Moreover, we refer to the *aversion to extreme risk* as the general attitude toward the type of lotteries defined here, which as we will see is for a large part independent from the shape of the utility function.

[4]We prefer to use the term insensitivity to extreme risk rather than extreme risk hyperseekingness even though it would be more symmetrical with the usual economic denomination regarding risk attitudes. Extreme risk seekingness may be misconstrued due to the two meaning of the word risk: on the one hand the traditional economic use – uncertainty but with known probabilities – and the more common meaning that refers to an uncertain hazard, which in the phrase "extreme risks" may dominate.

[5]All formal derivations and calculations are in the Appendix.

3 Extreme Risk Attitude for Generalized Expected Utility Models

Ample empirical evidence has accumulated that shows that individuals do not necessarily treat probabilities linearly (Starmer, 2000). Generalized expected utility models have flourished to account for this empirical regularity. A generic version of such models, referred to as *generic weighted expected utility*, can be written for a finite lottery $L = (p_1, x_1; p_2, x_2; ..., p_i, x_i; ..., p_n, x_n)$ as:

$$V(L) = \sum_{i=1}^{n} \tau_i u(x_i)$$

With τ_i the decision weight associated to x_i that depends on x_i and p_i and possibly on the whole probability distribution over outcomes as for rank dependent models. Such models generally assume in addition that $\sum_{i=1}^{n} \tau_i = 1$. This generic model includes weighted expected utility (Chew, 1983), decision weight expected utility (Handa, 1977) and rank-dependent expected utility (Quiggin, 1982). For an extreme risk, let τ be the weight associated with the consequence $W - L$ and the probability ϵ. Since the structure of the extreme risk lottery $\mathcal{L}_\epsilon$ is fixed, when $W - L$ is held constant, τ only depends on ϵ. As a consequence, $z(\epsilon)$ is defined by:

$$u\Big(W - z(\epsilon)\Big) = \tau(\epsilon)u(W - L) - \big(1 - \tau(\epsilon)\big)u(W)$$

Under the assumptions that τ and u are differentiable, we have:

Proposition 2 (Generic weighted expected utility) *For a generic weighted expected utility maximizer, the willingness to pay not to bear an extreme risk is given by:*

$$z(\epsilon) \simeq \tau'(0)F_{EU} \tag{2}$$

The attitude to extreme risk coefficient is $\zeta = \tau'(0)F_{EU}/L$.

A weighted expected utility maximizer's attitude toward extreme risk is the same as an expected utility maximizer *up to a factor equal to the derivative of the weight at 0.*

This gives a richer spectrum of possible attitudes to extreme risk, in the sense that if τ is not linear in probabilities, the willingness to pay of the agent is substantially impacted by the weight of the probability of the loss. For instance, it is possible for τ' large at 0 to have a situation where a generally risk-seeking individual ($u'' > 0$) would exhibit aversion to extreme risk (that is $\zeta > 1$).

One well-known issue with the generic weighted model is that it does not necessarily satisfy first-order stochastic dominance. When the satisfaction of dominance, among other reasonable properties, is required, the generic weighted expected utility model reduces to the rank-dependent expected utility model (Diecidue and Wakker, 2001), hereafter RDU. The general formulation of rank-dependent expected utility models is given for a finite lottery $L = (p_1, x_1; p_2, x_2; ..., p_i, x_i; ..., p_n, x_n)$ with $x_1 > x_2 > ... > x_n$ by:

$$V(L) = \sum_{j=1}^{n} [\mu(\sum_{k=1}^{j} p_k) - \mu(\sum_{k=1}^{j-1} p_k)]u(x_j)$$

The rank-dependent transformation of probability[6] $\mu : [0, 1] \to [0, 1]$ is assumed to be differentiable on $(0, 1)$ and monotonically increasing ($\mu' > 0$) with $\mu(0) = 0$ and $\mu(1) = 1$. The

[6] We use the term *decision weight*, or weight for short, for τ, that is the applied weight to ponder the utility of an outcome, and *probability transformation function* for μ, that is sometimes referred to as probability weight in the literature.

RDU model is a special case of the generic weighted utility model, where weights τ depends on the ranks of the outcome and μ. In our case, $\tau(\epsilon) = 1 - \mu(1-\epsilon)$ and $1 - \tau(\epsilon) = \mu(1-\epsilon)$. As a consequence, $z(\epsilon)$ is then defined here as:

$$u(W - z(\epsilon)) = \mu(1-\epsilon)u(W) + (1 - \mu(1-\epsilon))u(W - L)$$

Provided that μ is differentiable at 1 and u' bounded on $[W - L, W]$, the attitude to extreme risk for rank-dependent expected utility is a special case of Proposition 2:

Corollary 1 (Rank-dependent expected utility) *If $\lim_{p\to 1} \mu'(p) \neq \infty$, then the willingness to pay not to bear an extreme risk is given by:*

$$z(\epsilon) \simeq \epsilon\mu'(1)F_{EU} \tag{3}$$

The corresponding attitude to extreme risk coefficient is $\zeta = \mu'(1)F_{EU}/L$

This formulation highlights the respective roles of the utility function through fear of ruin on the one hand and the subjective transformation of probabilities on the other hand. Decision-makers are often characterized (Wakker, 2010) as optimistic ($\mu'' > 0$) or pessimistic ($\mu'' < 0$). In this case, Corollary 1 indicates that the attitude toward extreme risk results from the interaction between utility-based risk-aversion and probabilistic sensitivity, which can be seen as local optimism/pessimism. Pessimism, which by construction implies $\mu'(1) > 1$, will reinforce risk-aversion toward more extreme-risk aversion and a pessimistic agent will be more averse to extreme risk than a plain expected utility maximizer with the same utility function. The opposite holds for optimistic agents for whom $\mu'(1) < 1$: they are less averse to extreme risks than an equivalent plain risk-averse expected utility maximizer. The other cases (risk seeking pessimistic agents and risk averse optimistic agents) will be undetermined in the general case and will depend on the relative strength of the two terms of the product.

Most specifications aiming at accounting for empirical decisions under risk imply a very large value for μ' around 1. It is in fact unbounded for the most prominent specifications (Tversky and Kahneman, 1992; Prelec, 1998), i.e. $\lim_{p\to 1} \mu'(p) = +\infty$ (see Appendix for more details). In this case, one can lower-bound the willingness to pay when facing an extreme risk, and show that such individuals display hyperaversion to extreme risks:

Proposition 3 (Hyperaversion to extreme risk for RDU) *If $\lim_{p\to 1} \mu'(p) = +\infty$, then for any $M > 0$, exists $\bar{\epsilon}$ small enough such that for all $\epsilon < \bar{\epsilon}$, it holds that: $z(\epsilon) > M\epsilon$.*
Or equivalently: $\zeta = +\infty$

A RDU agent with such a probability transformation is hyperaverse to extreme risks: she would be, at the limit, infinitely more averse to extreme risk than a plain expected utility maximizer. Most specifications found in the literature allow μ' to diverge at 1, and empirical estimates of parameters often imply this divergence (Tversky and Kahneman, 1992; Prelec, 1998; Gonzalez and Wu, 1999). Note though that we use specifications for μ that were mostly developed for Cumulative Prospect Theory in the gain domain. To the best of our knowledge, there does not exist RDU-based specifications for the probability transformation function (details in Appendix about the derivation of the limits of μ').

Hyperaversion to extreme risks can be construed as an artefact from specifications of probability transformation function whose purpose is not primarily to properly describe sensitivity to very small probabilities. Indeed, such specifications and their estimations have been widely used as tools to investigate the subjective weighting of probabilities within the range of .01 to .99 most of the time. Extrapolating to situations outside the boundaries of the validity domain of such specifications can hence be misleading. For instance, a simple piecewise linear specification – as proposed by Wakker (2010) – can account for the most commonly observed behaviors without implying divergence for the derivative of the probability transformation. Yet, we still view these limit cases as informative, and if to be taken with caution, indicative *in qualitative terms* of attitude to extreme risks: individuals may not be hyperaverse to extreme risks as defined here, but nevertheless strongly averse to extreme risks. For instance, consider Wakker's piecewise linear model: the derivative of the probability transformation will be large but finite for $p < 0.01$ for instance. Using Kahneman and Tversky's (1992) specification, a probability of 0.01 is weighed around 0.05, hence the slope of the probability transformation function near 0 will be of a magnitude of 5. This means that under the rather conservative assumption that the probability transformation derivative is constant between 0 and 0.01, this still implies that aversion to extreme risk is 5 times stronger under rank-dependent expected utility than under expected utility.

4 Hyperaversion and Insensitivity to Extreme Risk under Cumulative Prospect Theory

One of the limits of the theoretical results above is that they appear at odds with the empirical evidence gathered on individuals' attitudes toward extreme risks. If a significant share of individuals seem to be 'overly' concerned with rare though dramatic events, a no less substantial share simply discards such events (Viscusi and Evans, 1990; Kunreuther, 1996; Camerer and Kunreuther, 1989).[7] The latter case cannot be accounted for in our framework unless one assumes that the individual is extremely optimistic, specifically that $\mu'(1) = 0$; an assumption that finds very little support in the experimental literature (Abdellaoui, L'Haridon, and Zank, 2010). Cumulative prospect theory (Tversky and Kahneman, 1992; Wakker, 2010), hereafter CPT, may provide a theoretical account for extreme risk insensitivity. Indeed, as we will show, under the most common specifications of both the probability transformation function and the value function, attitudes toward extreme risks are very polarized with virtually only hyperaversion or insensitivity to extreme risks present in the population.

Cumulative prospect theory differs from standard rank-dependent expected utility in several ways. First, the model distinguishes gains and losses, relative to a reference point, where agents are respectively risk-averse and risk-seeking; second, the ranks are determined as for RDU in the gain domain but in the reverse order for the loss domain; and finally, losses are felt more negatively than gains are positively. Typically, cumulative prospect theory relies on a motivation function U which depends on the payoff x and the reference point r as follows:

$$U(x,r) = \begin{cases} u_+(x-r) = u(x-r) & \text{if } \ x \geq r \\ u_-(r-x) = -\lambda u(r-x) & \text{if } \ x < r \end{cases}$$

With $\lambda \geq 1$ the loss aversion parameter, $u' > 0$ and $u'' < 0$.[8] As already mentioned, another difference with rank-dependent expected utility is that the ranks are not treated in the same way for gains (where gain-ranks are similar) and for losses (where "loss-ranks" are reversed for CPT compared to RDU). This is of course important for our lottery where all outcomes are in the

[7]Other explanations have been put forth, in particular in the context of a lack of insurance against natural catastrophes: misrepresentation of the risks and the available insurance policies, a belief in the intervention of the state in case of natural disasters, skewed intertemporal preferences, etc. See Camerer and Kunreuther (1989) or Kunreuther, Michel-Kerjan, Doherty, Grace, Klein, and Pauly (2009).

[8]We chose the symmetric version of CPT for its direct applicability. Using the more general form, i.e. with u_+ and u_- for gains and losses with non-constant loss aversion, does not change the results, but the simplified symmetric definition of the motivation function shows more straightforwardly the absence of effect of loss aversion.

loss domain, using the current level of wealth as the reference point. For the sake of simplicity and without loss of generality, we normalize u by setting that $v(x) = \frac{u(x)}{u(L)}$. This immediately implies $v(L) = 1$. Clearly (μ, u) and (μ, v) represent the same preferences. Assuming that the probability transformation μ is the same for losses and gains,[9] we have z defined by:

$$U(W - z(\epsilon), W) = \tau(W - L)U(W - L, W) + \tau(W)U(W, W)$$

with $\tau(W - L)$ the weight applied to outcome $W - L$, namely $\mu(\epsilon)$ since $W - L$ is in the loss domain, and $\tau(W) = \mu(1 - \epsilon)$ if W is considered in the gain domain. After replacement and simplification, this gives:

$$u(z(\epsilon)) = \mu(\epsilon)u(L)$$

This highlights two features: first, only the weight associated with the loss matters, second, loss aversion does not play any role because both the willingness to pay and the massive but improbable loss L are in the loss domain.

Under the assumptions that v' (or u') is bounded on $\mathbb{R}^+$, one can show the following proposition:

Proposition 4 (Cumulative Prospect Theory) *If* $\lim_{t \to 0} v'(t) < \infty$ *and* $\lim_{p \to 1} \mu'(p) < \infty$, *a CPT individual's attitude to extreme risks is characterized by:*

$$z(\epsilon) \simeq \epsilon \frac{\mu'(0)}{v'(0)} \quad \textit{and} \quad \zeta = \frac{\mu'(0)}{Lv'(0)} \tag{4}$$

In this situation, CPT differs from RDU (Corollary 1) by the fact that the behavior of μ' in the neighborhood of 0 will matter instead of that of 1. This comes from the fact that for losses, the ranks are ordered in the opposite direction as for gains.

An important point is that the behavior of the value function at the reference point – through $v'(0)$ – can have a large impact: whereas for RDU the usual concave value function implies that the marginal effect of the value function will be relatively low, or more exactly far from reaching its maximum, in CPT, the marginal effect is the largest at the reference point. In many specifications, the marginal value is very large (if not infinite, as for the very common power function) and may compensate the marginal effect of the probability transformation function near 0. These two combined effects account for the fact that the spectrum of attitudes to extreme risk in CPT is broader than for expected utility or rank-dependent expected utility. A CPT individual with a standard value function such as the power function but who treats probabilities linearly will not only be extreme-risk seeking (because she is risk-seeking in the loss domain) but will be insensitive to extreme risk, behaving as if she discarded very low probabilities.

For many specifications, both $\lim_{x \to 0} v'(x)$ and $\lim_{p \to 0} \mu'(p)$ are unbounded. What will hence matter is not just the behavior of the derivative of the probability transformation function at the boundary but the combination of this derivative with that of the value function in the neighborhood of 0. The divergence of v' at 0 is in many settings an undesirable property, but here, it may explain some of the phenomena observed in the empirical literature on extreme risks.

[9] It is not necessarily the case, but the generic properties – inverted S shape – are usually the same in the loss as in the gain domains. Distinguishing μ as μ_+ and μ_- for gains and losses does not change the qualitative results we highlight here.

Proposition 5 (General attitude to extreme risk for CPT) *If* $\lim_{p\to 0}(v^{-1}\circ\mu)' \neq \infty$*, then the willingness to pay not to bear an extreme risk for a CPT agent is given by:*

$$z(\epsilon) \simeq \epsilon \lim_{p\to 0}(v^{-1}\circ\mu)'(p) \quad \text{and} \quad \zeta = \lim_{p\to 0}(v^{-1}\circ\mu)'(p)/L \tag{5}$$

The behavior of the decision maker facing extreme risks depends critically on the limit of the derivative of the composite of v^{-1} and μ when x approaches 0. Intuitively, it means that the marginal effects of the utility and the probability transformation function "compete" to determine the individual's attitude toward extreme risk: the stronger the marginal effect of the utility (that is the weaker the derivative of its inverse function), the lower the willingness to pay, and in contrast, the stronger the marginal effect of the probability transformation function, the higher the willingness to pay.

An interesting case arises when $\lim_{p\to 0}(v^{-1}\circ\mu)' = 0$, that is when the marginal effect of the utility function in the neighborhood of 0 is "large" enough (in comparison to that of μ). In this case, the individual exhibits *insensitivity to extreme risk*:

Corollary 2 (Insensitivity to extreme risk for CPT) *If* $\lim_{p\to 0}(v^{-1}\circ\mu)' = 0$*, then* $\zeta = 0$ *and the individual exhibit insensitivity to extreme risks.*

This condition holds in the simple case where $\lim_{x\to 0} v'(x) = +\infty$ and $\lim_{p\to 0}\mu'(p) \neq \infty$. A CPT individual whose utility function would be a power function and that would treat probabilities linearly would exhibit insensitivity to extreme risk: because of its proximity to the reference point, the utility function has a very large marginal effect, whereas the marginal effect of the probability of the adverse consequence is limited. This also suggests that the apparent lack of concern with which many individuals deal with extreme risks may not stem from a tendency to discard low probabilities, as initially suggested by Kahneman and Tversky themselves (1979) in their first prospect theory model (through the discontinuity of the probability weights at boundaries), or more generally under some general "low probability neglect". But it may come from the subjective cost to protect oneself, be it monetary or in effort, which is, in marginal terms, extremely large. This explanation for extreme risk insensitivity has the major advantage of reconciling some empirical stylized facts where small probabilities seem to be strongly distorted and the widespread lack of insurance against catastrophes.

Yet, for most specifications and corresponding empirical estimates (see Table 1), the effect of the probability transformation function seems to "dominate", in the sense that $\lim_{p\to 0}(v^{-1}\circ\mu)' = +\infty$, which directly leads to extreme risk hyperaversion:

Proposition 6 (Extreme risk hyperaversion for CPT) *If* $\lim_{p\to 0}(v^{-1}\circ\mu)' = +\infty$*, then* $\zeta = +\infty$*.*
Equivalently, for any $M > 0$*, exists* $\bar{\epsilon}$ *small enough such that for all* $\epsilon < \bar{\epsilon}$*, the willingness to pay to avoid an extreme risk for a CPT individual, is such that* $z(\epsilon) > \epsilon M$*.*

Table 1 presents the attitudes to extreme risk for common specifications with a power utility function.[10] A notable point, at least for Kahneman and Tversky (1992) and Gonzalez and Wu's (1999) specifications, is that they predict polarized attitudes toward extreme risk: individuals will be either extreme risk hyperaverse or extreme risk insensitive, depending on the relative

[10]Note that one prominent alternative, the exponential function, leads to hyperaversion too without allowing for extreme risk insensitivity: the typical utility function (or motivation function here) is given by $u(x) = 1 - e^{-bx}$, that gives $u'(x) = be^{-bx}$ and $u'(0) = b$. Note that the normalization to v only changes this result by a constant factor. Hence if μ' is unbounded at 0, the extreme-risk attitude is automatically hyperaversion.

Specification	Model	Behavior
Kahneman and Tversky (1992)	$\mu(p) = \frac{p^\gamma}{(p^\gamma+(1-p)^\gamma)^{\frac{1}{\gamma}}}$ $u_-(x) = -\lambda(-x)^\alpha$ $v_-(x) = -\lambda\frac{(-x)^\alpha}{L^\alpha}$	if $\alpha = \gamma$, $z'(0) = L$ and $\zeta = 1$ if $\alpha > \gamma$, $z'(0) = +\infty$ and $\zeta = +\infty$ if $\alpha < \gamma$, $z'(0) = 0$ and $\zeta = 0$
Prelec (1998)	$\mu(p) = exp(-\beta(-ln(p))^\sigma)$	$z'(0) = +\infty$ and $\zeta = +\infty$
Gonzalez and Wu (1999)	$\mu(p) = \frac{\delta p^\gamma}{\delta p^\gamma+(1-p)^\gamma}$	if $\alpha = \gamma$, $z'(0) = L$ and $\zeta = 1$ if $\alpha > \gamma$, $z'(0) = +\infty$ and $\zeta = +\infty$ if $\alpha < \gamma$, $z'(0) = 0$ and $\zeta = 0$
Rieger and Wang (2006)	$\mu(p) = \frac{3-3b}{a^2-a+1}(p^3 - (a+1)p^2 + ap) + p$	$z'(0) = 0$ and $\zeta = 0$

For all cases, $u_-(x) = -\lambda(-x)^\alpha$ and $v_-(x) = -\lambda\frac{(-x)^\alpha}{L^\alpha}$, dropped after KT92

Table 1: *Extreme risk attitudes for usual CPT specifications*

magnitude of the probability transformation parameters and the utility parameter. More generally, all specifications virtually imply only radical attitudes to extreme risks in line with the polarized findings of the empirical literature although not all are compatible with insensitivity. This may explain why in empirical studies, authors tend to find a bimodal response to extreme risks: a non-negligible proportion of individuals seem to overly protect themselves against very unlikely events, while others seem to simply discard such risks.

Once again, the divergence of both v' and μ' at 0 and 1 may be seen as an artefact stemming from the use of specifications outside their domain of validity, and we agree that it may be the case as empirical studies focused on the behahvior of such functions in these domains are scarce. Nevertheless, as already argued for RDU, the *qualitative* result does not change: instead of the limit cases of hyperaversion and insensitivity, would the two derivatives be finite, we would observe strong aversion – if μ' “dominates” strongly v' – or strong extreme risk seekingness – if v' dominates strongly μ'. As an illustration, consider two CPT individuals whose functions μ and v are assumed to be locally linear respectively near 1 and 0. For both individuals, we assume a reasonable marginal utility of money of 2 between 0 and $L\epsilon$ monetary units. For the first individual, we consider a typical S-inverted transformation: As argued in the RDU case (section 3), a slope of 5 near 1 for the probability transformation is plausible. Then her extreme risk coefficient will be 2.5: the individual is willing to pay 2.5 times the expected loss to protect herself. Now consider another individual, which only departs from the former through her probability transformation which is linear over $[0, 1]$. In this case, the latter will only purchase insurance if it is less than 0.5 times the expected loss. Once again, the limit cases put forth in our formal results may well be limit cases, but we believe they qualitatively capture the mechanisms at play.

5 Insensitivity at the Reference Point and Hyperaversion Everywhere Else

Insensitivity with CPT critically depends on being the decision being in the neighborhood of the reference point. It is only because the marginal utility of money at the reference point is unbounded that it counter-balances the marginal transformation of probability near 0. This seems relevant when studying an insurance decision: the most natural setting is an agent who has some wealth she wishes to insure and needs to pay, from her current level of wealth, an insurance fee. But, for some decisions, there may be, intertwined with the extreme risk, a change in the wealth level. In this case, this change will plausibly imply that the subjective cost will not be borne exactly at the reference point, but at some shifted point. This idea relies on the assumption that the reference point is the current level of wealth, and leaves aside the role of expectations in the forming of reference points, as in Koszegi and Rabin (2006) for instance.[11] The marginal utility will be bounded since $v'(x) < \infty$ for $x \neq 0$, and will not compensate the unbounded marginal transformation of the probability. Consider the same setting as in the previous section, with a change δ in the agent's wealth. The willingness to pay z not to bear the extreme risk is given by:

$$U(W + \delta - z(\epsilon), W) = \mu(\epsilon)U(W - L + \delta, W) + \mu(1 - \epsilon)U(W + \delta, W)$$

Note that for mixed lotteries (involving gains and losses), the sum of weights is not necessarily 1, as in the previous equation. Denote ζ_δ the extreme risk coefficient for a given level of δ, the change in wealth. An extreme risk insensitive CPT individual as defined in the previous section will exhibit extreme risk hyperaversion everywhere but at $\delta = 0$:

Proposition 7 (Insensitivity at the reference point and hyperaversion everywhere else) *For a CPT individual, if* $\lim_{p\to 0}(v^{-1} \circ \mu)' = 0$ *with* $\lim_{p\to 0} \mu'(p) = +\infty$*, then:*

(i) If $\delta = 0$*,* $\zeta_0 = 0$

(ii) If $\delta \neq 0$*,* $\zeta_\delta = +\infty$

This means that it is only in the immediate neighborhood of the reference point that the agent will exhibit insensitivity to extreme risk. Everywhere else, the agent will be extreme risk hyperaverse. This highlights a feature of extreme risk insensitivity: it depends critically on the cost, e.g. an insurance fee, to be borne exactly at the reference point (here the wealth level). Said differently, individuals, according to some CPT specifications, would be hyperaverse to extreme risk everywhere but at their current reference point where they exhibit extreme risk insensitivity. An apparently contradictory behavior in two different settings can be explained within the framework of cumulative prospect theory.

This may have immediate consequences in the context of an investment decision where the individual can expect an increase of one's wealth. Consider a situation where exists a safe investment that will automatically shift the wealth of the agent upward thanks to interests. In such a situation, hyperaversion to extreme risk is predicted. This may help explain the "equity premium puzzle". This puzzle is the empirical regularity that stocks (the risky investment) give historically much higher returns than safer assets and this difference would require implausible

[11] This simplifying assumption may be based on three grounds: first, there is at best mixed empirical evidence that reference points incorporate expectations in usual decision under risk (see for instance Baillon, Bleichrodt, and Spinu 2017); second, there is no model of reference-dependent preferences incorporating expectations that include a probability transformation – and it is far from trivial to do so, because two probability measures are to be taken into account, the outcome probability and the expectation probability; and third, the decision-maker may only partially update her reference point in the direction of expectations, what will result in a shifted reference point, but still different from the favourable consequence.

levels of risk aversion in the context of expected utility (Mehra and Prescott, 1985). Following an original idea from Rietz (1988), recent studies have focused on the role of rare macroeconomic/financial disasters to explain this puzzle (Barro, 2006; Gabaix, 2012; Fahri and Gabaix, 2016). As is usual in this strand of literature, the agent is supposed to maximize expected utility. As a consequence, to account for the order of magnitude of the equity premium, the models require not so rare disasters: Barro and Ursua (2008b) – and Barro and Ursua 2008a – estimate, based on historical data, the probability of a disaster at 1.7%. Typically, in these studies, the calibration assumes a few percent risk of a disaster.[12] We suggest here another approach: the subjective weight given to low probabilities can be the main driver of the investors' concern rather than a higher than usually thought frequency of such disasters.

To formalize this intuition, suppose that the agent has to make an investment decision. She has her all wealth W to invest. We assume a very simple situation where there are only two options: either a safe asset that yields a return of i percent or a risky asset that yields a return $i+\Delta_i$ with very high likelihood but leads to a large loss L, e.g. due to a stock market crash, with a very low probability ϵ. This setting is clearly a gross simplification of any realistic investment decision, but we believe it captures the effect of the presence of an extreme risk in an investment decision. First, certain realistic features, if anything, would reinforce our results: If one takes into account the uncertainty on the return of the risky investment, by allowing the return of the risky asset to be a random variable, a usual risk aversion argument implies the expected return to be larger if random rather than in our simplified setting where it is considered as certain, conditionally on the extremely bad consequence not occurring. Second, we leave aside portfolio composition, that is which share of W is put in the two types of investments.[13] The situation can be construed in the following way: the risky asset in the model is a portfolio of risky assets, with little remaining variance outside of the extreme risk. The extreme risk can arguably be considered as non-hedgeable given that it refers to a macroeconomic or financial disasters that can, by nature, affect all risky assets. We aim to determine the spread Δ_i that makes the agent indifferent between the safe and the risky asset, or said differently, the spread threshold above which the agent would invest in the risky asset. As for a standard extreme risk studied before, Δ_i is a function of ϵ.

We study this problem for both RDU (and as a special case EU) and CPT. For a rank-dependent expected utility maximizer (including an expected utility maximizer when μ is linear), Δ_i is given by:

$$\mu(1-\epsilon)u((1+i+\Delta_i(\epsilon))W) + (1-\mu(1-\epsilon))u(W-L) = u((1+i)W)$$

[12] Arguably, these probabilities are high, and the whole point of Barro and Ursua is to show that they are higher than usually thought, at least to a level that may explain the equity premium under the assumption of the maximization of expected utility.

[13] Although no full-fledge portfolio composition theory under CPT exists to the best of our knowledge, it is possible to calculate, for a given share of W invested in the risky asset, the required δ_i for the investor to be indifferent with investing everything in the safe asset. In this case, z is defined, for a given share s by:

$$v(iW) = \mu(1-\epsilon)v((i+\Delta_i(\epsilon)s)W) - \lambda\mu(\epsilon)v(sL)$$

Using the same calculation as below it gives:

$$s\Delta_i'(\epsilon) = \frac{\mu'(1-\epsilon)v((i+\Delta_i(\epsilon)s)W)) + \lambda\mu'(\epsilon)v(sL)}{\mu(1-\epsilon)v'((i+\Delta_i(\epsilon)s)W))}$$

That is, when ϵ approaches 0, we have:

$$\lim_{\epsilon\to 0}\Delta_i'(\epsilon) = \frac{v(iW)}{sv'(iW)}\lim_{\epsilon\to 0}\mu'(1-\epsilon) + \frac{\lambda v(sL)}{sv'(iW)}\lim_{\epsilon\to 0}\mu'(\epsilon)$$

The qualitative finding stays true.

Differentating wrt ϵ and rearranging gives :

$$\Delta_i'(\epsilon) = \mu'(1-\epsilon)\frac{u((1+i+\Delta_i(\epsilon))W) - u(W-L)}{\mu(1-\epsilon)u'((1+i+\Delta_i(\epsilon))W)}$$

Then:

$$\lim_{\epsilon\to 0}\Delta_i'(\epsilon) = \lim_{\epsilon\to 0}\mu'(1-\epsilon)\frac{u((1+i)W) - u(W-L)}{u'((1+i)W)}$$

We find a case similar to that of the plain extreme risk,[14] the required spread depends critically on the behavior of μ' near 0. In the case of a plain expected utility maximizer $\mu'(p) = 1$ for all p, and the spread is given by a ratio similar to fear of ruin. Said differently, the attitude toward an investment decision with some extreme risk component can be straightforwardly approached with the fear of ruin measure: a strong sensitivity to extreme risk for expected utility then requires very large levels of risk aversion. But, for $\lim_{p\to 1}\mu'(p) = +\infty$ as is the case in prominent specifications, the ratio of the spread over the probability of the negative consequence is arbitrarily large when ϵ approaches 0. The behavior of an agent when facing an extreme-risk investment and a safe one is similar to the behavior of the same agent when facing an extreme risk insurance decision.

For a CPT individual, the situation is slightly different. Using the notations of the previous section, we have Δ_i given by:

$$v(iW) = \mu(1-\epsilon)v((i+\Delta_i(\epsilon))W) - \lambda\mu(\epsilon)v(L)$$

since W is the assumed reference point, that is, her level of wealth when she makes her decision). This implies:

$$\Delta_i'(\epsilon) = \frac{\mu'(1-\epsilon)v((i+\Delta_i(\epsilon))W) + \lambda\mu'(\epsilon)v(L)}{\mu(1-\epsilon)v'((i+\Delta_i(\epsilon))W)}$$

When ϵ approaches 0, we have:

$$\lim_{\epsilon\to 0}\Delta_i'(\epsilon) = \frac{v(iW)}{v'(iW)}\lim_{\epsilon\to 0}\mu'(1-\epsilon) + \frac{\lambda v(L)}{v'(iW)}\lim_{\epsilon\to 0}\mu'(\epsilon)$$

The notable difference with the situation studied in the previous section is that the denominator, $v'(iW)$, is bounded for $i > 0$, so the behavior of the agent does not depend on the composition of the inverse of v with μ but simply on the behavior of μ' near the boundaries. The relevant derivative of v is not at the 0 point (where it is the largest and possibly unbounded) but at iW where it is lower and bounded. Generally speaking, the effect of the derivative of the probability transformation function is the same as for the insurance extreme risk setting but magnified, since it is not counter-balanced by the high marginal utility near 0. Moreover, the loss aversion parameter, greater than 1, comes into play in the second term.

In line with Proposition 7, a CPT individual may exhibit insensitivity to extreme risk in an insurance setting, that is where the reference point is W, and yet have the subjective cost of an extreme risk investment being arbitrarily large relatively to the expected loss. An agent can discard extreme risk insurance (as observed in the empirical literature) and yet be extreme risk hyperaverse in the context of an investment. A straightforward example is obtained with Kahneman and Tversky's (1992) specification (see Table 1) when $\alpha < \gamma$. In this case, the extreme risk coefficient ζ_0 will be 0 while in the financial setting above the required spread Δ_i is arbitrarily larger than the expected loss.

[14] In fact it suffices to set $W' = W(1+i)$, $L' = L + iW$ and $z = \Delta_i W$ to obtain exactly the same situation.

6 Numerical Applications with Commonly Estimated Parameters

An open question remains: up to now, we only have focused on limit behavior. Although our developments predict, especially in the case of CPT, interesting phenomena seemingly in line with some empirical stylized facts, nothing guarantees that the domain (near 0) for which these effects kick in is of economic relevance. In particular, it is undetermined whether the willingness to pay would be non-negligible for economically credible situations, especially for the apparently prevailing hyperaversion case. The fact that ζ is infinite does not guarantee that for low but sill plausible probabilities and large losses, the effect of extreme risk hyperaversion would be substantial. To check this, we run numerical computations based on commonly estimated parameters, for the hypothetical case of a decision-maker of wealth 100,000 facing a one thousandth probability of a 50% loss. These numbers are of the order of magnitude of a house fire in the US per year.[15] The results are reported in Table 2.

In the case of hyperaverion to extreme risk, the effect ranges from 1.2 to almost 30 times the expected loss of the extreme risk (to be compared with a factor of 1.04 to 1.09 in the case of expected utility). The effect in the case of RDU is larger than in the case of CPT, due to the mitigating effect of the high marginal utility near the reference point. This figures confirm that for conventional values of parameters, the attitude toward extreme risk is of economic relevance. In particular, when faced with an unlikely risk of losing half her wealth with probability one thousandth, a Prelec CPT agent may accept to pay around three quarter of a percent of her wealth to get covered. For most extreme risk hyperaverse cases here, the agents would be willing to pay at least .20 % of their wealth not to bear the risk. It is noteworthy that despite the unboundedness of the derivative of the probability transformation near 0, the effect of hyperaversion may not show massively, especially for Kahneman and Tversky's specification: the figures obtained in the penultimate case are very similar to those obtained under expected utility, suggesting that for parameter values close enough the effect of hyperaversion could only appear for minuscule probabilities. The last estimation also shows that with credible parameter values, insensitivity can be obtained straightforwardly: in this case, where γ is only minimally lower that α, z is lower than the expected loss, which results in not subscribing an actuarially fair insurance. This tends to confirm that under plausible and not too unlikely values, the insensitivity formally described in section 4 can be present. Moreover, it gives weight to our conclusion that despite the intuition based on a very distorted probability transformation near 0, both hyperaversion and insensitivity can arise, and this for economically relevant circumstances.

These results also show how critical the distribution of parameters is: we used here only average estimated values, but given that the attitude to extreme risk will result from the relative values of the parameters (γ and α for Kahneman and Tversky's specification), their joint distribution is critical to the share of individuals that would be hyperaverse or insensitive. This, in particular, applies to catastrophic risk insurance decisions. To explore this question, we consider a random population of individuals characterized by a Kahneman and Tversky's specification, with parameters a and γ following two independent uniform distributions in $[\underline{\alpha}, \bar{\alpha}]$ and $[\underline{\gamma}, \bar{\gamma}]$ and compute their willingness to pay not to bear the risk. We do so in the context of an insurance decision against catastrophes (as in Table 2) but also for an investment decision, as in section 5, with the same ϵ, W and L and setting $i = 0.02$. Note that in the case of the decision of an investment, z is given by $\Delta_i W$. We simulate two different distributions for α and γ: first $\alpha \in [0.6, 1]$ and $\gamma \in [0.5, 0.9]$, and then $\alpha \in [0.5, 9]$ and $\gamma \in [0.5, 9]$. The numerical results are

[15] Every year, around 0.3% of households report fire damages for an average loss of around 20,000 USD, according to the National Fire Protection Association, for the 2009-2013 period.

Type of model	Specification	Parameters	WTP
Expected Utility			
Power function	$u(x) = x^{\alpha}$	$\alpha = 0.88$	$z/L = 0.10\%$ $\zeta(\epsilon) = 1.04$
Power function	$u(x) = x^{\alpha}$	$\alpha = 0.71$	$z/L = 0.11\%$ $\zeta(\epsilon) = 1.09$
Rank-Dependent EU			
Prelec 98	$\mu(p) = exp(-\beta(-ln(p))^{\sigma})$ $u(x) = x^{\alpha}$	$\sigma = 0.65$, $\beta = 1.0467$ $\alpha = 0.88$	$z/L = 1.12\%$ $\zeta(\epsilon) = 12.12$
KT 92	$\mu(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{\frac{1}{\gamma}}}$ $u(x) = x^{\alpha}$	$\gamma = 0.61$ $\alpha = 0.88$	$z/L = 0.24\%$ $\zeta(\epsilon) = 24.28$
Prelec 98	$\mu(p) = exp(-\beta(-ln(p))^{\sigma})$ $u(x) = x^{\alpha}$	$\sigma = 0.65$, $\beta = 1.0467$ $\alpha = 0.71$	$z/L = 1.23\%$ $\zeta(\epsilon) = 12.78$
KT 92	$\mu(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{\frac{1}{\gamma}}}$ $u(x) = x^{\alpha}$	$\gamma = 0.61$ $\alpha = 0.71$	$z/L = 2.26\%$ $\zeta(\epsilon) = 25.59$
Cumulative Prospect Theory			
Prelec 98	$\mu(p) = exp(-\beta(-ln(p))^{\sigma})$ $u_{-}(x) = -\lambda(-x)^{\alpha}$	$\sigma = 0.65$, $\beta = 1.0467$ $\alpha = 0.88$	$z/L = 1.53\%$ $\zeta(\epsilon) = 15.34$
KT 92	$\mu(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{\frac{1}{\gamma}}}$ $u_{-}(x) = -\lambda(-x)^{\alpha}$	$\gamma = 0.69$ $\alpha = 0.88$	$z/L = 0.44\%$ $\zeta(\epsilon) = 4.38$
Prelec 98	$\mu(p) = exp(-\beta(-ln(p))^{\sigma})$ $u_{-}(x) = -\lambda(-x)^{\alpha}$	$\sigma = 0.65$, $\beta = 1.0467$ $\alpha = 0.71$	$z/L = 0.56\%$ $\zeta(\epsilon) = 5.64$
KT 92	$\mu(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{\frac{1}{\gamma}}}$ $u_{-}(x) = -\lambda(-x)^{\alpha}$	$\gamma = 0.69$ $\alpha = 0.71$	$z/L = 0.12\%$ $\zeta(\epsilon) = 1.20$
KT 92 (Insensitivity)	$\mu(p) = \frac{p^{\gamma}}{(p^{\gamma}+(1-p)^{\gamma})^{\frac{1}{\gamma}}}$ $u_{-}(x) = -\lambda(-x)^{\alpha}$	$\gamma = 0.71$ $\alpha = 0.69$	$z/L = 0.008\%$ $\zeta(\epsilon) = 0.81$

We use varying α for the power function, broadly corresponding to parameters estimated in the literature. For probability transformation functions, we use Kahneman and Tversky's (1992) estimates for both gains ($\gamma = .69$) and losses ($\gamma = .61$) and Wakker's (2010) for Prelec's. For the last case (insensitivity), we chose a plausible pair of parameter values, that is not too far from the usually estimated average, with $\alpha < \gamma$.

Table 2: *Measurable effects of extreme-risk attitudes for RDU and CPT for* $\epsilon = .001$ *and* $L = 50,000$ *and* $W = 100,000$.

displayed in Table 3.[16]

[16]It is easy to show that the share of the population that is insensitive to extreme risk at 0 is given by:

$$Prob(\alpha < \gamma) = 1 - \int_{\underline{\gamma}}^{\bar{\alpha}} f_a(x)F_{\gamma}(x)dx = 1 - \int_{\underline{\gamma}}^{\bar{\alpha}} \frac{x - \underline{\gamma}}{(\bar{\alpha} - \underline{\alpha})(\bar{\gamma} - \underline{\gamma})}dx = 1 - \int_{\underline{\alpha}}^{\bar{\gamma}} \frac{x - \underline{\gamma}}{(\bar{\alpha} - \underline{\alpha})(\bar{\gamma} - \underline{\gamma})}dx - 1 + F_{\alpha}(\bar{\gamma})$$

Type of Setting	z/W (average)	Quartiles for z/W (%)	Share of $z \leq \epsilon L$	Share of $z > \epsilon L$
$\alpha \in [.6, 1]$, $\gamma \in [.5, .9]$				
Insurance	0.21 %	0.04, 0.11, 0.28	0.287	0.713
Investment	0.81 %	0.28, 0.57, 1.18	0	1
$\alpha \in [.5, .9]$, $\gamma \in [.5, .9]$				
Insurance	0.11 %	0.01, 0.05, 0.14	0.50	0.50
Investment	0.70 %	0.24, 0.49, 1.02	0	1

For direct comparability with the insurance case, the same γ parameter is used for gains and losses in the case of the investment setting. The loss aversion parameter λ is set to 2.

Table 3: *Hyperaversion and insensitivity for Kahnemann and Tversky's CPT specification in an insurance and an investment setting,* $\epsilon = .001$ *and* $L = 50,000$ *and* $W = 100,000$.

The figures obtained highlight several phenomena: first, the aversion to extreme risk is much stronger (and of a level likely to be of economic importance) for the investment setting than for the insurance one. Almost one point of percentage of spread would be required, on average, for agents to invest in the risky asset rather than in the safe one, to compensate for an expected loss of only half a thousandth of the initial wealth. Moreover, all simulated agents exhibit some form of hyperaversion to extreme risk in the investment setting. The picture is more nuanced for the insurance setting: while the average z is relatively high (.21 % of the wealth to be compared to a expected loss of .05 %) in the first case, the population is clearly heterogeneous with more than a fourth of individuals (first case) and half individuals (second case) who are willing to pay less that the expected loss to protect against the extreme risk. This simple simulation illustrates that some stylized facts found in the literature are compatible with the predictions based on Kahneman and Tversky's CPT specifications: first, a large share of individuals would not insure against extreme risk while the rest would be willing to pay quite a high premium; second, all agents are averse to extreme risk in the investment decision setting and require a large spread to engage in the extreme risk investment.

The spread due to the extreme risk is not negligible, and the numbers are of the relevant order of magnitude to explain a reluctance to take chances by individuals in even remotely risky assets. A probability of one thousandth to lose half one's wealth seems, in some cases, to be enough to explain the requirement of a few percentage points more of interests/return (in the absence of any other risk). This may strengthen the account for the equity premium puzzle based on extreme risk: it may be valid not because catastrophic events are more likely than usually thought (Rietz, 1988; Barro, 2006; Barro and Ursua, 2008a; Nakamura, Steinsson, Barro, and Ursúa, 2013), but because of individuals' hyperaversion to extreme risk. Beliefs

Which eventually gives:

$$Pr(\alpha < \gamma) = \frac{(\underline{\gamma} - \underline{\alpha})^2}{2(\bar{\alpha} - \underline{\alpha})(\bar{\gamma} - \underline{\gamma})} + \frac{E(\gamma) - \underline{\alpha}}{\bar{\alpha} - \underline{\alpha}}$$

in the likelihood of a macroeconomic disaster need not be of the percentage magnitude since a few per thousand may be enough with strongly overweighted probability weights. Using the distribution of the first case in Table 3 but setting ϵ to .005, we obtain an average z for the investment decision of 2.32 % of the total wealth (median: 1.85), a figure of the order of magnitude of the usually calculated equity premium.

7 Conclusion

This paper presents a simple way to approach attitudes toward extreme risk. In doing so, it shows that the spectrum of attitudes toward such risks can be rich, but also that the effect of extreme risks may be a few order of magnitudes more important than the expected loss. Under rank-dependent expected utility with conventionally estimated parameters, the most plausible attitude seems to be *hyperaversion to extreme risks*, that is a set of preferences for which, when the probability of the negative consequence approaches zero, the subjective cost of the risk tends to be infinitely larger than the corresponding expected loss. Cumulative prospect theory retains some features of rank-dependent expected utility but implies polarized attitudes toward extreme risks: in addition to hyperaversion, another solution of the model is insensitivity to extreme risk. This polarization fits nicely some of the results put forth in the empirical literature. A critical role is played by the marginal utility near the reference point that in some specifications more than counter-balance the marginal probability sensitivity near 0. It explains why agents, despite an unbounded marginal probability sensitivity, can exhibit insensitivity to extreme risk. A noteworthy point is that this insensitivity can only hold at the reference point, as in a typical insurance decision, and the same agent would exhibit hyperaversion to extreme risk everywhere else (as in the case of an investment decision). In sum, cumulative prospect theory predicts that individuals will in general be extreme risk hyperaverse, but at the reference point where a substantial share may act as if they simply discarded low probabilities.

These theoretical developments suggest at least three lines of research. The first one is that too little is empirically or experimentally known based on the probability transformation in the neighborhood of certainty. The results obtained here critically depend on the extrapolation of specifications whose aim was to account for intermediate probabilities rather than extremely small or large ones: the lower probability usually studied is 0.01. To be fair, some studies (Kunreuther and Slovic, 1978; Schade, Kunreuther, and Kaas, 2004) have dealt with extremely unlikely events but they were not aiming at measuring the behavior of the probability transformation at the boundaries, nor the marginal utility near 0. Although presenting some specific challenges in setting experimental conditions with extreme risks, measuring this effect and by extension the probability transformation for extreme probabilities seems a necessary step toward a better and more empirically grounded understanding of attitudes toward extreme risks.

Second, and provided that empirical assessments give some validity to existing specifications in the neighborhood of certainty, it would suggest a renewed approach to the study of low-probability high-consequence behavior. A prediction of the model is that individual could well neglect some extreme risk, because the cost to protect oneself applies at the reference point, yet they are, in general, extreme risk hyperaverse. A focus on reference points may in particular be critical in empirical studies, it may help rationalize apparently inconsistent behaviors. The predictions may be also used as a policy tool to induce more insurance coverage against natural disasters for instance. Indeed, using framing or the combination of the purchase of an insurance with some preexisting cost or gain (that would shift the level at which the cost is borne), authorities may be able to overcome individuals' insensitivity to extreme risk.

Third, if the general shape of the probability transformation function would be confirmed for very low and very high probabilities, it calls for theoretical developments based on hyperaversion to extreme risks in several economic contexts, such as finance, health, environment, etc. that may help explain a number of empirical "puzzles" found in the literature. Such puzzles include the "equity premium puzzle" as we argued, but also estimation of the value of statistical life (Bleichrodt and Eeckhoudt, 2006), the reluctance to rely on nuclear energy or, more generally, the reluctance to the introduction of new technologies with unlikely catastrophic consequences.

References

Abdellaoui, M., O. L'Haridon, and H. Zank (2010): "Separating curvature and elevation: A parametric probability weighting function," *Journal of Risk and Uncertainty*, 41, 39–65.

Aumann, R. J., and M. Kurz (1977): "Power and taxes," *Econometrica*, pp. 1137–1161.

Baillon, A., H. Bleichrodt, and V. Spinu (2017): "Searching for the reference point," Erasmus University of Rotterdam, mimeo.

Barro, R. (2006): "Rare disasters and asset markets in the twentieth century," *The Quarterly Journal of Economics*, 121(3), 823–866.

Barro, R., and J. Ursua (2008a): "Consumption disasters in the twentieth century," *The American Economic Review*, 98(2), 58–63.

——— (2008b): "Macroeconomic Crises since 1870," *Brookings Papers on Economic Activity*, pp. 255–335.

Barseghyan, L., F. Molinari, T. O'Donoghue, and J. C. Teitelbaum (2013): "The nature of risk preferences: Evidence from insurance choices," *American Economic Review*, 103(6), 2499–2529.

Bleichrodt, H., and L. Eeckhoudt (2006): "Willingness to pay for reductions in health risks when probabilities are distorted," *Health Economics*, 15(2), 211–214.

Botzen, W. J., and J. C. van den Bergh (2012): "Risk attitudes to low-probability climate change risks: WTP for flood insurance," *Journal of Economic Behavior & Organization*, 82(1), 151–166.

Botzen, W. W., H. Kunreuther, and E. Michel-Kerjan (2015): "Divergence between individual perceptions and objective indicators of tail risks: Evidence from floodplain residents in New York City," *Judgment and Decision Making*, 10(4), 365.

Camerer, C. F., and H. Kunreuther (1989): "Decision processes for low probability events: Policy implications," *Journal of Policy Analysis and Management*, 8(4), 565–592.

Chew, S. (1983): "A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox," *Econometrica*, 51, 1065–1092.

Diecidue, E., and P. P. Wakker (2001): "On the intuition of rank-dependent utility," *Journal of Risk and Uncertainty*, 23(3), 281–298.

Eeckhoudt, L., C. Schieber, and T. Schneider (2000): "Risk aversion and the external cost of a nuclear accident," *Journal of environmental Management*, 58(2), 109–117.

Erev, I., I. Glozman, and R. Hertwig (2008): "What impacts the impact of rare events," *Journal of Risk and Uncertainty*, 36(2), 153–177.

Fahri, E., and X. Gabaix (2016): "Rare disasters and exchange rates," *Quarterly Journal of Economics*, 131(1–52).

Farhi, E., and X. Gabaix (2016): "Rare disasters and exchange rate," *The Quarterly Journal of Economics*, 131(1), 1–52.

Foncel, J., and N. Treich (2005): "Fear of ruin," *Journal of Risk and Uncertainty*, 31(3), 289–300.

Gabaix, X. (2012): "Variable rare disasters: An exactly solved framework for ten puzzles in Macro-Finance," *Quarterly Journal of Economics*, 127(645–700).

Gonzalez, R., and G. Wu (1999): "On the shape of the probability weighting function," *Cognitive Psychology*, 38(1), 129–166.

Handa, J. (1977): "Risk, probabilities, and a new theory of cardinal utility," *The Journal of Political Economy*, pp. 97–122.

Hogarth, R. M., and H. Kunreuther (1989): "Risk, ambiguity, and insurance," *Journal of risk and uncertainty*, 2(1), 5–35.

Kahneman, D., and A. Tversky (1979): "Prospect theory: An analysis of decision under risk," *Econometrica*, 47, 263–291.

Koszegi, B., and M. Rabin (2006): "A model of reference-dependent preferences," *Quarterly Journal of Economics*, 71(4), 1133–1166.

Krawczyk, M. W., S. T. Trautmann, and G. van de Kuilen (2016): "Catastrophic risk: Social influences on insurance decisions," *Theory and Decision*, pp. 1–18.

Kunreuther, H. (1996): "Mitigating disaster losses through insurance," *Journal of risk and Uncertainty*, 12(2-3), 171–187.

Kunreuther, H., E. Michel-Kerjan, N. Doherty, M. Grace, R. Klein, and M. Pauly (2009): *At War with the Weather: Managing Large-Scale Risks in a New Era of Catastrophes.* MIT Press, Cambridge.

Kunreuther, H., and P. Slovic (1978): "Economics, psychology, and protective behavior," *The American Economic Review*, 68(2), 64–69.

McClelland, G. H., W. D. Schulze, and D. L. Coursey (1993): "Insurance for low-probability hazards: A bimodal response to unlikely events," *Journal of Risk and Uncertainty*, 7(1), 95–116.

Mehra, R., and E. Prescott (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

Nakamura, E., J. Steinsson, R. Barro, and J. Ursúa (2013): "Crises and recoveries in an empirical model of consumption disasters," *American Economic Journal: Macroeconomics*, 5(3), 35–74.

Prelec, D. (1998): "The probability weighting function," *Econometrica*, 66(1), 497–527.

Quiggin, J. (1982): "A theory of anticipated utility," *Journal of Economic Behavior and Organization*, 3, 324–343.

Rheinberger, C. M., and N. Treich (2016): "Attitudes Toward Catastrophe," *Environmental and Resource Economics*, pp. 1–28.

Rieger, M. O., and M. Wang (2006): "Cumulative prospect theory and the St. Petersburg paradox," *Economic Theory*, 28(3), 665–679.

Rietz, T. A. (1988): "The equity risk premium a solution," *Journal of Monetary Economics*, 22(1), 117–131.

Schade, C., H. Kunreuther, and K. P. Kaas (2004): "Probability neglect and concern in insurance decisions with low probabilities and high stakes," *Wharton Risk Management and Decision Processes Center Working Paper*, pp. 04–07.

Schmidt, U. (2016): "Insurance demand under prospect theory: A graphical analysis," *Journal of Risk and Insurance*, 83(1), 77–89.

Starmer, C. (2000): "Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk," *Journal of Economic Literature*, 38, 332–382.

Tversky, A., and D. Kahneman (1992): "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

Viscusi, W. K., and W. N. Evans (1990): "Utility functions that depend on health status: Estimates and economic implications," *The American Economic Review*, pp. 353–374.

Wakker, P. P. (2010): *Prospect theory: For risk and ambiguity.* Cambridge University Press.

Appendix

Proof of Proposition 1

It suffices to apply Proposition 2 to $\tau = Id$.

Proof of Proposition 2

Set $t = \tau(\epsilon)$ and $y(\tau(\epsilon)) = z(\epsilon)$, that is, y associates the weight to the willingness to pay, we have $\lim_{x\to 0}\tau(x) = 0$ by assumption. By definition of z and assumption on the sum of weights:

$$u\big(W - z(\epsilon)\big) = \tau(\epsilon)u(W-L) + \big(1-\tau(\epsilon)\big)u(W)$$

With a simple change of variables:

$$u\big(W - y(t)\big) = tu(W-L) + \big(1-t\big)u(W)$$

Differentiating gives:

$$y'(t) = \frac{(u(W-L) - u(W))}{u'(W)}$$

Moreover, the Taylor development of $y(t)$ is given near 0 by:

$$z(t) = ty'(0) + \circ(t^2)$$

Hence:

$$y(t) \simeq t\frac{u(W-L) - u(W)}{u'(W)}$$

Or:

$$z(\epsilon) \simeq \tau(\epsilon)\frac{u(W-L) - u(W)}{u'(W)}$$

Proof of Corollary 1

z is such that for all ϵ:

$$\mu(1-\epsilon)u(W) + (1-\mu(1-\epsilon))u(W-L) = u(W - z(\epsilon))$$

That is, differentiating wrt to ϵ:

$$z'(\epsilon) = \mu'(1-\epsilon)\frac{u(W) - u(W-L)}{\mu(1-\epsilon)u'(W) + (1-\mu(1-\epsilon)u'(W-L)}$$

If μ' is defined at 1:

$$z'(0) = \mu'(1)\frac{u(W) - u(W-L)}{u'(W)}$$

The first order Taylor expansion of z at 0 gives:

$$z(\epsilon) = \epsilon\mu'(1)\frac{u(W) - u(W-L)}{u'(W)} + \circ(\epsilon)$$

Proof of Proposition 3

z being bounded, monotonically increasing and of value 0 at 0, the fundamental theorem of calculus gives:

$$z(\epsilon) = \int_0^\epsilon z'(t)dt$$

For $t \in (0, \epsilon]$, we have from the proof of Proposition 2:

$$z'(t) = \mu'(1-t)\frac{u(W) - u(W-L)}{\mu(1-t)u'(W) + (1-\mu(1-t))u'(W-L)}$$

And

$$z'(t) > \mu'(1-t)\frac{u(W) - u(W-L)}{\max\big(u'(W), u'(W-L)\big)}$$

So that:

$$z(\epsilon) = \int_0^\epsilon z'(t)dt > \frac{u(W) - u(W-L)}{\max\big(u'(W), u'(W-L)\big)}\int_0^\epsilon \mu'(1-t)dt$$

Finally, μ' is defined and continuous on $(0, \epsilon]$, set $\psi(\epsilon) = \min_{x\in(0,\epsilon]} \mu'(1-x)$ (which is $\mu'(1-\epsilon)$ if $\mu'' > 0$ near 1). Then we have, for any ϵ small enough:

$$z(\epsilon) > \epsilon\psi(\epsilon)\frac{u(W) - u(W-L)}{\max\big(u'(W), u'(W-L)\big)}$$

It is clear that $\lim_{\epsilon\to 0} \psi(\epsilon) = \lim_{p\to 1} \mu'(p) = +\infty$.

Proof of Proposition 4

We have:

$$v(z(\epsilon)) = \mu(\epsilon)v(L)$$

Differentiating:

$$z'(\epsilon)v'(z(\epsilon)) = \mu'(\epsilon)v(L)$$

Hence:

$$z'(\epsilon) = \frac{\mu'(\epsilon)v(L)}{v'(z(\epsilon))}$$

And so:

$$z'(0) = \frac{\mu'(0)v(L)}{v'(0)}$$

The now usual Taylor expansion near 1 gives:

$$z(\epsilon) = \epsilon\frac{\mu'(0)v(L)}{v'(0)} + \circ(\epsilon) =$$

By normalization of u to v, we have $v(L) = \frac{u(L)}{u(L)} = 1$, hence:

$$z(\epsilon) \simeq \epsilon\frac{\mu'(0)}{v'(0)}$$

Proof of Proposition 5

$z(\epsilon)$ is given by:

$$v(z(\epsilon)) = \mu(\epsilon)v(L) = \mu(\epsilon)$$

v is a one-to-one mapping from $\mathbb{R}^+$ to itself, monotonically increasing and concave, hence v^{-1} is properly defined on $\mathbb{R}^+$, where it is monotonically increasing and convex. So we have for any ϵ:

$$z(\epsilon) = v^{-1}(\mu(\epsilon))$$

Deriving:

$$z'(\epsilon) = (v^{-1} \circ \mu)'(\epsilon)$$

By assumption $(v^{-1} \circ \mu)'$ converges at 0, then we have by Taylor approximation:

$$z(\epsilon) \simeq \epsilon z'(0) = \epsilon(v^{-1} \circ \mu)'(0)$$

Proof of Proposition 6

By the proof of Proposition 6, we have for any $\epsilon > 0$:

$$z'(\epsilon) = (v^{-1} \circ \mu)'(\epsilon)$$

It is hence immediate that $\lim_{p\to 0} \frac{z(p)}{p} = \lim_{p\to 0} z'(p) = \lim_{p\to 0}(v^{-1} \circ \mu)'(p) = +\infty$. Hence for any $M > 0$, for ϵ small enough, we have $z'(\epsilon) > M$. Moreover, developing:

$$z(\epsilon) = \int_0^\epsilon z'(t)dt = \int_0^\epsilon (v^{-1} \circ \mu)'(\epsilon)dt$$

v^{-1} is increasing, as the inverse function of an increasing one, μ is increasing, hence $v^{-1} \circ \mu$ is increasing and its derivative is positive. By assumption, $\lim_{p\to 0}(v^{-1}\circ\mu)'(p) = +\infty$ and $(v^{-1}\circ\mu)'$ continuous on $(0, 1)$. Eventually, for any $M > 0$, exists $\bar{\epsilon} > 0$ such that for all $\epsilon < \bar{\epsilon}$, it holds that $(v^{-1} \circ \mu)'(\epsilon) > M$. Said differently, for any $M > 0$, exists $\bar{\epsilon} > 0$ such that for $\epsilon < \bar{\epsilon}$:

$$z(\epsilon) = \int_0^\epsilon z'(t)dt > \int_0^\epsilon M dt = M\epsilon$$

Proof of Proposition 7

The first part of the Proposition, $\zeta_0 = 0$, is the same as Corollary 2. We hence focus on $\delta \neq 0$. First, for $\delta > 0$ and $\delta < L$, and given that we are interested in the behavior of z for ϵ approaching 0 (so that $\delta > z(\epsilon)$), we have:

$$v(\delta - z(\epsilon)) = -\lambda\mu(\epsilon)v(-\delta + L) + \mu(1 - \epsilon)v(\delta)$$

Differentiating:

$$z'(\epsilon)v'(\delta - z(\epsilon)) = \lambda\mu'(\epsilon)v(-\delta + L) + \mu'(1 - \epsilon)v(\delta)$$

Hence:

$$z'(\epsilon) = \frac{\lambda\mu'(\epsilon)v(-\delta + L) + \mu'(1 - \epsilon)v(\delta)}{v'(\delta + z(\epsilon))}$$

And so:

$$\lim_{\epsilon\to 0} z'(\epsilon) = \frac{\lambda v(-\delta + L) \lim_{\epsilon\to 0} \mu'(\epsilon) + v(\delta) \lim_{\epsilon\to 0} \mu'(1 - \epsilon)}{v'(\delta)}$$

Hence:

$$\lim_{\epsilon\to 0} z'(\epsilon) > \frac{\lambda v(-\delta + L) \lim_{\epsilon\to 0} \mu'(\epsilon)}{v'(\delta)}$$

$\lambda v(-\delta + L)$ and $v'(\delta)$ are bounded, so the result ensues.

The argument is similar for $\delta \geq L$ (although not of much economic relevance). For $\delta < 0$, we have:

$$-\lambda v(\delta + z(\epsilon)) = -\lambda\mu(\epsilon)v(\delta + L) - \lambda(1 - \mu(\epsilon))v(\delta)$$

That gives:

$$v(\delta + z(\epsilon)) = \mu(\epsilon)v(\delta + L) + (1 - \mu(\epsilon)v)(\delta)$$

Differentiating wrt ϵ:

$$z'(\epsilon)v'(\delta + z(\epsilon)) = \mu'(\epsilon)v(\delta + L) - \mu'(\epsilon)v(\delta)$$

Hence:

$$z'(\epsilon) = \mu'(\epsilon)\frac{v(\delta + L) - v(\delta)}{v'(\delta + z(\epsilon))}$$

And so:

$$\lim_{\epsilon\to 0} z'(\epsilon) = \frac{v(\delta + L) - v(\delta)}{v'(\delta)} \lim_{\epsilon\to 0} \mu'(\epsilon)$$

Since $v(\delta + L) - v(\delta) > 0$ and $v'(\delta) > 0$ and $v'(\delta) \neq \infty$, the result ensues.

Calculus

RDU specifications

Here we just calculate the limits of the derivatives of μ at 1 for the various specifications of μ.

Kahneman and Tversky (1992)

$$\mu(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}}$$

Then:

$$\mu'(p) = \frac{\gamma p^{\gamma-1}[p^\gamma + (1-p)^\gamma]^{\frac{1}{\gamma}} - \frac{1}{\gamma}[\gamma p^{\gamma-1} - \gamma(1-p)^{\gamma-1}]^{\frac{1-\gamma}{\gamma}}}{(p^\gamma + (1-p)^\gamma)^{\frac{2}{\gamma}}}$$

It follows that near 1:

$$\mu'(p) \sim \gamma - \frac{1}{\gamma}[\gamma - \gamma(1-p)^{\gamma-1}]^{\frac{1-\gamma}{\gamma}}$$

For $\gamma < 1$, we have $\lim_{p\to 1^-}(1-p)^{\gamma-1} = +\infty$ and then:

$$\lim_{p\to 1^-} \mu(p) = +\infty$$

For $\gamma = 1$, $\mu(p) = p$ and $\mu'(p) = 1$.

Prelec (1998)

The 1998 Prelec's specification is given by:

$$\mu(p) = e^{-\beta(-\ln(p))^\alpha} \tag{6}$$

With $\alpha, \beta \in \mathbb{R}^{++}$. Tedious calculations lead to, for any $p \in (0, 1)$:

$$\mu'(p) = \beta\alpha\frac{(-\ln(p))^{\alpha-1}}{p}e^{-\beta(-\ln(p))^\alpha} \tag{7}$$

The limit at 1 is immediately given by $\lim_{p\to 1^-} \mu'(p) = \lim_{p\to 1^-} \beta\alpha(-\ln(p))^{\alpha-1} = +\infty$ for $\beta > 0$ and $\alpha < 1$.

Wu and Gonzales (1999) The specification is given by:

$$\mu(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}$$

It gives:

$$\mu'(p) = \frac{\delta\gamma p^{\gamma-1}(\delta p^\gamma + (1-p)^\gamma) - \delta p^\gamma(\delta\gamma p^{\gamma-1} - \gamma(1-p)^{\gamma-1})}{(\delta p^\gamma + (1-p)^\gamma)^2}$$

Near $p = 1$ we have:

$$\mu'(p) \sim \gamma(1-p)^{\gamma-1}$$

Hence, for $\gamma < 1$, we have $\lim_{p\to 1} \mu'(p) = +\infty$.

CPT specifications

We take u as the power function: $u(x) = x^\alpha$. And $v(x) = \frac{u(x)}{u(L)} = \frac{x^\alpha}{L^\alpha}$. We then have: $v^{-1}(y) = Ly^{\frac{1}{\alpha}}$

Kahneman and Tversky (1992)

Then, for Kahneman and Tversky's (1992) specification:

$$(v^{-1} \circ \mu)(p) = L\Big(\frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}}\Big)^{\frac{1}{\alpha}} = L\frac{p^{\frac{\gamma}{\alpha}}}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\alpha\gamma}}}$$

After differentiation and simple yet tedious computation, we obtain:

$$(v^{-1} \circ \mu)'(p) = \frac{L}{\alpha}\frac{1}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\alpha\gamma}}}\Big[\gamma p^{\frac{\gamma-\alpha}{\alpha}} - p^{\frac{\gamma}{\alpha}}.\frac{p^{\gamma-1} + (1-p)^{\gamma-1}}{p^\gamma + (1-p)^\gamma}\Big]$$

Near 0, we have $(v^{-1} \circ \mu)'(p) \sim \frac{L\gamma}{\alpha} p^{\frac{\gamma-\alpha}{\alpha}}$

Prelec (1998)

For Prelec (1998), we also have:

$$v^{-1}(x) = Lx^{\frac{1}{\alpha}}$$

And so:

$$(v^{-1} \circ \mu)(p) = L\Big(e^{-\beta(-ln(p))^\sigma}\Big)^{\frac{1}{\alpha}} = Le^{-\frac{\beta}{\alpha}(-ln(p))^\sigma}$$

Hence:

$$(v^{-1} \circ \mu)'(p) = \frac{\beta\sigma}{\alpha}L\frac{e^{-\frac{\beta}{\alpha}(-ln(p))^\sigma}}{p(-ln(p))^{1-\sigma}}$$

Now when p approaches 0, $x = -ln(p)$ approaches $+\infty$. We hence have:

$$\lim_{p\to 0}(v^{-1} \circ \mu)(p) = \lim_{x\to+\infty}\Big[\frac{\beta\sigma L}{\alpha}.\frac{1}{e^{-x}}.x^{\sigma-1}.e^{-\frac{\beta}{\alpha}x^\sigma}\Big]$$

That is:

$$\lim_{p\to 0}(v^{-1} \circ \mu)(p) = \frac{\beta\sigma L}{\alpha}\lim_{x\to+\infty}\Big[x^{\sigma-1}.e^{x(1-\frac{\beta}{\alpha}x^{\sigma-1})}\Big]$$

More over, when x large enough, $e^{x(1-\frac{\beta}{\alpha}x^{\sigma-1})} \sim e^x$. So we have:

$$\lim_{p\to 0}(v^{-1} \circ \mu)(p) = \frac{\beta\sigma L}{\alpha}\lim_{x\to+\infty}\Big[x^\sigma.\frac{e^x}{x}\Big] = +\infty$$

Wu and Gonzales (1999)

For Wu and Gonzales' (1999) specification:

$$(v^{-1} \circ \mu)(p) = L\Big(\frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}\Big)^{\frac{1}{\alpha}} = L\frac{p^{\frac{\gamma}{\delta}}}{\Big(\delta p^\gamma + (1-p)^\gamma\Big)^{\frac{1}{\alpha}}}$$

Its derivative is given, after rearrangement, by:

$$(v^{-1} \circ \mu)'(p) = \frac{\gamma}{\delta} L \frac{p^{\frac{\gamma-\alpha}{\alpha}}}{(\delta p^\gamma + (1-p)^\gamma)^{\frac{1}{\alpha}}}\Big[1 - \frac{(\delta p^{\gamma-1} - (1-p)^{\gamma-1})p}{\delta p^\gamma + (1-p)^\gamma}\Big]$$

Near 0, we have $\frac{1}{(\delta p^\gamma + (1-p)^\gamma)^{\frac{1}{\alpha}}} \sim 1$ and $\delta p^\gamma + (1-p)^\gamma \sim 1$ and
$1 - (\delta p^{\gamma-1} - (1-p)^{\gamma-1})p = 1 - \delta p^\gamma - p(1-p)^\gamma \sim 1$, so that, near 0: $(v^{-1} \circ \mu)'(p) \sim \frac{\gamma L}{\alpha} p^{\frac{\gamma-\alpha}{\alpha}}$

Rieger and Wang (2006)

The probability transformation function is given by: $\mu(p) = \frac{3-3b}{a^2-a+1}(p^3 - (a+1)p^2 + ap) + p$. Denote for simplicity $c = \frac{3-3b}{a^2-a+1}$. Then:

$$(v^{-1} \circ \mu)(p) = L\Big(cp^3 - c(a+1)p^2 + (ca+1)p\Big)^{\frac{1}{\alpha}}$$

Differentiating gives:

$$(v^{-1} \circ \mu)'(p) = \frac{1}{\alpha} L\Big[3cp^2 - 2c(a+1)p + ca + 1\Big]\Big(cp^3 - c(a+1)p^2 + (ca+1)p\Big)^{\frac{1-\alpha}{\alpha}}$$

Quite straightforwardly, we have:

$$\lim_{p \to 0}(v^{-1} \circ \mu)'(p) = \frac{1}{\alpha} L(ca+1)(ca+1)^{\frac{1-\alpha}{\alpha}} \lim_{p \to 0}\Big[p^{\frac{1-\alpha}{\alpha}}\Big] = 0$$

for any $\alpha < 1$.

PUBLICATION 7: Magda Osman, Fabrice Le Lec et Bjoern Meder (2013), “Making decision in uncertain times – What cognitive scientists can say or learn from financial crises”, ***Trends in Cognitive Science***, 17(6):257-260.

Cell PRESS

Decision making in uncertain times: what can cognitive and decision sciences say about or learn from economic crises?

Björn Meder[1], Fabrice Le Lec[2], and Magda Osman[3]

[1] Max Planck Institute for Human Development, Center for Adaptive Behavior and Cognition, Lentzeallee 94, 14195 Berlin, Germany
[2] Institut Catholique de Lille, Faculté Libre de Sciences Economiques et de Gestion – LEM CNRS UMR 8179, 41 rue du Port, 59 016 Lille Cedex, France
[3] Queen Mary, University of London, Experimental Biology and Psychology Centre, London, E1 4NS, UK

> 'The essence of the situation is action according to opinion, of greater or less foundation and value, neither entire ignorance nor complete and perfect information, but partial knowledge. If we are to understand the workings of the economic system we must examine the meaning and significance of uncertainty; and to this end some inquiry into the nature and function of knowledge itself is necessary'
>
> Frank H. Knight, *Risk, Uncertainty, and Profit*, 1921

Economic crises bring to the fore deep issues for the economic profession and their models. Given that cognitive science shares with economics many theoretical frameworks and research tools designed to understand decision-making behavior, should economists be the only ones re-examining their conceptual ideas and empirical methods? We argue that economic crises demonstrate different forms of uncertainty, which remind cognitive scientists of a pervasive problem: how best to conceptualize and study decision making under uncertainty.

The challenge: uncertainty in various (dis)guises

Economic crises illustrate various types of real-world decision making under uncertainty within dynamic environments. These decisions involve dependencies in time and interdependencies amongst multiple agents [1] For instance, investors need to decide whether to provide loans to governments and banks without knowing how markets will develop and what policy makers will decide to do. Politicians weigh up the decision to bail out fragile banks and countries, while at the same time trying to appease the interests of their electorate. Here we can see that uncertainty is an inherent feature of the decision environment and of the agent (e.g., limitations in knowledge and information-processing capacities, conflicting goals). Uncertainty permeates all aspects of real-world decision problems, from constructing the action and outcome space to inferring the probabilities and values of outcomes and predicting the behavior of others. The question is, how can we best conceptualize decision making under uncertainty in all these various (dis)guises? More to the point, how can we characterize the many forms of uncertainty with which people have to cope in the real world?

Taking stock of the canonical framework for decision making

The canonical approach for conceptualizing decision making builds on the idea that possible states of the world can be associated with subjective probabilities and values (Box 1). In this view, all forms of decision making conform to two fundamental principles: (i) a trade-off between outcome probabilities and values, used to derive the expected utility of alternative actions; and (ii) decision makers act *as if* they maximize subjective expected utility (SEU) [2]. Although empirical research has revealed several departures from SEU theory, enriched variants try to take into account the peculiarities of human decision making, while preserving the core principle of utility maximization [3].

So, to what extent can this approach be used for understanding decision making in real-world contexts? Let us take the problem of investors deciding on whether to buy bonds from a struggling eurozone country. SEU would propose that investors consider the probability and value of future events, such as the risk of default. However, we already face a stumbling block: where do the probabilities come from? Unpacking this question involves turning to an economically informed distinction between different types of decision situations based on the agent's sources of knowledge regarding outcomes and probabilities. Knight [4] distinguished between: (i) *a priori* probabilities, which can be logically deduced, as in games of chance; (ii) statistical probabilities, derived from data; and (iii) estimates, arising from situations in which 'there is no valid basis of any kind for classifying instances' ([4], p. 225). Here, then, decision making under risk refers to situations in which probabilities are known (or knowable), whereas situations of uncertainty are characterized as cases where probabilities are neither logically deducible nor can be inferred from data. For instance, investors cannot refer to data to assign probabilities and value estimates to the consequences for the eurozone if a member defaults; therefore, decisions will only ever be based on (Knightian) estimates.

Corresponding author: Osman, M. (m.osman@qmul.ac.uk).

Trends in Cognitive Sciences June 2013, Vol. 17, No. 6

Box 1. The canonical framework for decision-making research

Rational choice theory. The canonical framework of decision making is based on two assumptions. First, the agent can order all possible situations according to her preferences; second, she always acts in accordance with them. Under some mild assumptions, this is equivalent to maximizing expected utility. In practice, however, specifications are required about what matters in such preferences, such as monetary or social welfare (e.g., Mother Theresa can be conceptualized as a perfectly rational agent by assuming that her utility function is based on the interests of the poor and sick).

Subjective expected utility (SEU) theory. The goal of SEU theory [2] is to give content to such preferences in the case of uncertainty. Agents are assumed to have preferences between actions, $a \in A$, from which the decision maker can choose. These preferences depend on possible states of the world, $s \in S$, which are beyond the agent's control, and 'consequences' or outcomes that she will eventually face. Savage showed that, under some arguably reasonable conditions, agents would choose acts as if they ascribed probabilities to states of the world and utilities u to consequences and maximized the corresponding expected utility, given by:

$$SEU(a) = \sum_{s \in S} p(s)u(a(s)) \quad [1]$$

Fundamental to this framework is the result that act a is preferred to act a' iff and only if the expected utility of a is larger than that of a'; that is, $a > a'$ iff $\mathrm{SEU}(a) > \mathrm{SEU}(a')$.

Expected utility and subjective expected utility. SEU can be seen as a generalization of expected utility theory as formalized by von Neumann and Morgenstern in 1944. This is decision making under risk, because the objective probabilities are known. Formally and operationally, the results have similar implications (i.e., the maximization of expected utility). However, at the conceptual level, Savage's theory is often interpreted as the origin of the applicability of expected utility theory for decision making under uncertainty, because it derives the existence of subjective probabilities from conditions on an agent's preferences.

Generalization of (subjective) expected utility theory. The descriptive accuracy of expected utility theory has been challenged by several empirical studies (e.g., the Allais and the Ellsberg paradox). The framework has been adapted to account for these findings through 'generalized' expected utility, which assumes the maximization of uncertainty-weighted expected utility (e.g., with a transformation of probabilities) and the separate representation of uncertainty and consequences of actions through probabilities and utility [3].

Although the distinction between risk and uncertainty is intuitively plausible, defenders of SEU have dismissed the risk/uncertainty distinction, arguing that the canonical framework assumes that decision makers act 'as if they assigned numerical probabilities to every conceivable event' ([5], p. 282). Thus, the claim is that people act rationally, given their subjective – not necessarily veridical – beliefs, with subjective probabilities and utility functions serving as building blocks for modeling decision making.

Others take the Knightian distinction between risk and uncertainty as a starting point for considering alternative ways to conceptualize decision making. They start from the view that many real-world problems are ill-structured and not easily formalized and that humans are cognitively constrained in their ability to process the informational complexities that arise (i.e., real-world agents are boundedly rational) [6]. As a consequence, heuristics and approximate strategies are used in decision making under uncertainty [7,8]. For instance, when dealing with dynamic decision problems and the need to achieve long-term goals, an aspiration level-based strategy may be used that does not require precise quantitative knowledge of the decision environment [9].

Rethinking decision making under uncertainty

Although there is considerable dispute about both the general usefulness of the risk versus uncertainty distinction and the ways by which decision making is modeled, these differences are not necessarily reflected in the empirical tools. Typically, researchers use decontextualized situations with well-defined probabilities and outcomes, such as lotteries and experimental games. Presenting participants with the probabilities and payoffs enables researchers to control the epistemological states of the decision maker, which in turn allows her to use SEU-like models as the normative benchmark. However, this approach limits the insights that can be gained for understanding decision making under (Knightian) uncertainty. If probabilities and values are given, there is no possibility of examining important predecisional processes. For instance, when it comes to applying SEU-like decision strategies, researchers cannot explore how people come up with probability estimates and outcome values. Alternatively, if we assume that people might use strategies other than SEU-like ones, there is no way of examining which pieces of information (other than probabilities and outcome values) people use to reach a decision. The upshot here is that the empirical tools are often too constrained to examine whether decision making under uncertainty is qualitatively different from decision making under risk. The implication is that, if there is a qualitative difference, there is a fundamental limit to any generalization of current models of decision making beyond the laboratory to real-world decision-making situations.

Although these problems are often recognized, empirically studying decision making in uncertain environments that approximate those in the real world is anything but trivial. However, some steps have been taken in this direction; for instance, by using dynamic decision-making tasks that set up microworlds, with various interdependencies of decisions in time and between multiple variables or agents [1,9]. Moreover, in economics too there is a growing trend to move away from traditional risk-based paradigms (i.e., lottery-type tasks) by employing a richer combination of tools [10,11] or by conducting field studies to examine decision making under out-of-the-laboratory conditions [12].

A major stumbling block for broadening the empirical scope is the lack of a clear framework for conceptualizing uncertainty in all its various forms. Knight's uncertainty category is essentially a negatively defined concept; namely, the absence of an objective basis for inferring probabilities. This is a helpful starting point, but refinement is needed. Some researchers have recently expanded on Knight's formulation and proposed different levels of uncertainty that consider the relationships that uncertainty

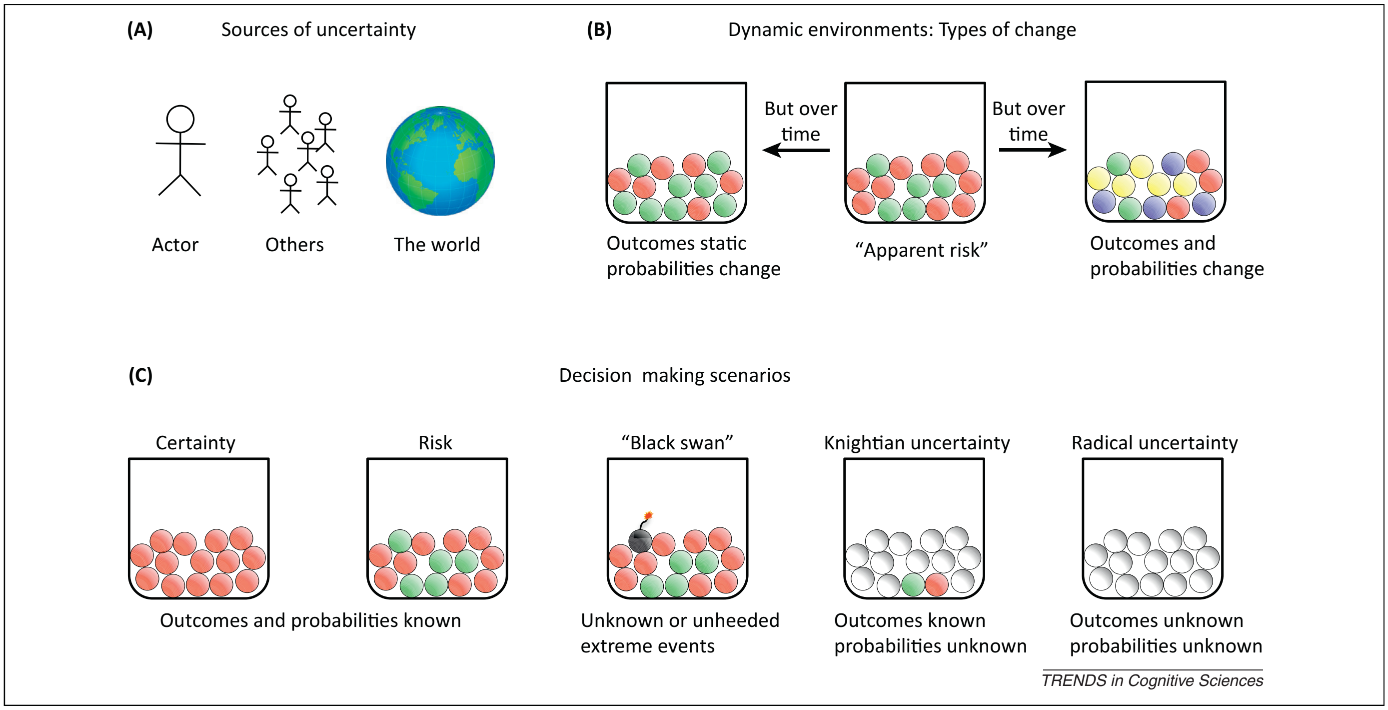


Figure 1. Uncertainty in its various guises. Illustrating sources of uncertainty and situations of decision making under uncertainty, using an urn model. (**A**) Uncertainty can reside in the mind of the boundedly rational agent. Uncertainty can also result from the decisions of and influences from other agents and from genuine randomness in the external environment (i.e., the data-generating process). (**B**) Examples of dynamic environments that involve changes in the decision-making situation over time. Left: The proportion of balls changes in unpredictable (or unknown) ways over time; therefore, probability estimates obtained at t_1 are of little use at t_2. Right: The outcomes themselves change over time, requiring a reformulation of the decision situation. (**C**) Examples of decision-making scenarios. From left to right: In situations of certainty and risk, the outcomes and their probabilities are known. In a 'black swan' situation, the urn contains a rare but highly consequential event (a 'bomb' or, in the case of a positive event, a 'diamond') that is either unknown to the decision maker or ignored in the representation of the decision situation. In a situation of Knightian uncertainty, the outcomes are known but not their probabilities. The right-most example is a situation of radical uncertainty, in which both the outcomes and their probabilities are unknown.

has to risk (e.g., whether we can reduce uncertainty to risk by sufficient amounts of data or whether even an infinite amount of data would be insufficient, because the data-generating process changes in unpredictable ways) [13]. Other authors have discussed variants of uncertainty from the perspective of inductive inference. They elaborate on problems arising from a misspecification of the hypothesis space (i.e., when the model used to derive predictions does not match the structure of the decision environment); this highlights breakdowns when applying models for situations of risk ('small worlds') to situations of uncertainty ('large worlds') [8,14].

We argue that real-world problems are a useful basis for characterizing variants of uncertainty and the types of uncertain environments with which decision makers (and cognitive systems in general) have to cope (Figure 1). For instance, economic crises illustrate uncertainty about the underlying dynamics of the conditions under which the decisions are being made, uncertainty in the feedback from decisions, uncertainty from interpreting the decisions and actions of multiple agents, and uncertainty in resolving conflicts between competing goals [1,9]. One may dispute whether the ultimate goal for theoretical and empirical research is to explain how decisions are made in complex real-world situations, where all of these uncertainties prevail, or whether the goal is to pinpoint characteristics of environmental structures to explain adaptive behavior and cognition. In any case, a first and necessary step is to identify types of uncertainty that can guide and expand theoretical and empirical practices.

Concluding remarks: coping theoretically and empirically with uncertainty

Our starting point was the claim that real-world problems like economic crises highlight the potential limitations in the way decision-making behavior is usually conceptualized in both economics and the cognitive sciences, particularly with respect to the many forms of uncertainty that people face outside the laboratory. In our view, the major challenge for developing a more comprehensive theory of decision making is the lack of a classification system that captures key elements of uncertainty and uncertain environments. If serious attempts are made to extend Knight's original formulation of uncertainty and develop a taxonomy of uncertainty to which researchers could adhere, perhaps this may shift the focus away from questions concerning the forms of rationality (or optimality) that decision-making behavior takes and onto questions about how best to conceptualize uncertainty in its many forms. This will not only provide a better foundation for modeling and studying decision making, but also set the stage for developing ways to aid decision makers when faced with real-world uncertainty [Haldane, A. (2012) *The dog and the frisbee*. Speech given at the Federal Reserve Bank of Kansas City's 36th economic policy symposium 'The Changing Policy Landscape', Jackson Hole, Wyoming. (http://www.bankofengland.co.uk/publications/Pages/speeches/2012/596.aspx)].

Acknowledgments

The authors thank Wayne Gray, Nils Stieglitz, and the ABC research group for helpful comments on earlier versions of this paper. Use of an urn model for illustrating different uncertainties (Figure 1) was inspired by stimulating discussions with Amit Kotyal, Mirta Galesic, Nadine Fleischhut, and Niklas Keller. B.M. was supported by a Visiting Scholar Award from the British Academy and Grant ME 3717/2 from the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program 'New Frameworks of Rationality' (SPP 1516).

References

1 Osman, M. (2010) Controlling uncertainty: a review of human behavior in complex dynamic environments. *Psychol. Bull.* 136, 65–86
2 Savage, L. (1954) *The Foundations of Statistics*, Wiley
3 Starmer, C. (2000) Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *J. Econ. Lit.* 38, 332–382
4 Knight, F.H. (1921/2006) *Risk, Uncertainty and Profit*, Houghton Mifflin
5 Friedman, M. (1962/1976) *Price Theory: A Provisional Text*, Aldine
6 Simon, H. (1991) Invariants of human behavior. *Annu. Rev. Psychol.* 41, 1–19
7 Gigerenzer, G. and Gaissmaier, W. (2011) Heuristic decision making. *Annu. Rev. Psychol.* 62, 451–482
8 Brighton, H. and Gigerenzer, G. (2012) Are rational actor models "rational" outside small worlds? In *Evolution and Rationality: Decisions, Co-operation, and Strategic Behavior* (Okasha, S. and Binmore, K., eds), pp. 84–109, Cambridge University Press
9 Selten, R. *et al.* (2012) Dealing with dynamic decision problems when knowledge of the environment is limited: an approach based on goal systems. *J. Behav. Decis. Making* 25, 443–457
10 Janssen, M.A. *et al.* (2010) Lab experiments for the study of social-ecological systems. *Science* 328, 613–617
11 Abdellaoui, M. *et al.* (2011) The rich domain of uncertainty: source functions and their experimental implementation. *Am. Econ. Rev.* 101, 695–723
12 Levitt, S.D. and List, J.A. (2009) Field experiments in economics: the past, the present, and the future. *Eur. Econ. Rev.* 53, 1–18
13 Lo, A.W. and Mueller, M.T. (2010) WARNING: physics envy may be hazardous to your wealth. *J. Investm. Manag.* 8, 13–63
14 Taleb, N. (2007) *The Black Swan: The Impact of the Highly Improbable*, Penguin

http://dx.doi.org/10.1016/j.tics.2013.04.008 Trends in Cognitive Sciences, June 2013, Vol. 17, No. 6

PUBLICATION 8: Benoît Tarroux et Fabrice Le Lec (à paraître), " On Attitudes to Choice: Some Experimental Evidence on Choice Aversion", ***Journal of the European Economic Association***.

ON ATTITUDES TO CHOICE: SOME EXPERIMENTAL EVIDENCE ON CHOICE AVERSION

Fabrice Le Lec
Catholic University of Lille,
Anthropo-Lab

Benoît Tarroux
Univ Lyon, Université Lumière Lyon 2,
GATE UMR 5824

The editor in charge of this paper was Juuso Valimaki.

E-mail: fabrice.lelec@univ-catholille.fr (Le Lec); tarroux@gate.cnrs.fr (Tarroux)

1. Introduction

Individuals routinely make decisions when faced with a large number of options. Consumers can choose from extensive product ranges, investors are offered very many opportunities and citizens in democracies have many political alternatives in frequent elections. This raises the question of how desirable are large choice sets. The standard view in Economics is that more choice generally brings about higher welfare (Arrow, 1995; Kreps, 1979): the more options there are available, the more likely an economic agent is to find the item that suits her needs or preferences. There are however two potential limitations to this view. First, as advocated by various philosophers such as Mill (1991[1859]), Marx (Marx and Engels, 1947[1845]) and Hayek (1960), freedom of choice may have some value in itself: individuals may attach intrinsic value to the freedom of choice, and so appreciate more choice *independently* of its (positive or negative) consequences for the quality of the subsequent decision (Sugden, 2003; Sen, 1988). Second, making choices from a large set of options may involve some cognitive, psychological or emotional drawbacks, such as decision costs, regret, temptation and self-control, or the fear of making bad decisions (Loewenstein, 1999).

There is only little empirical work on whether individuals appreciate having greater choice, and how they trade off the benefits and disadvantages of choice. This paper aims to examine whether the choice set over which individuals make decisions matters to them. All else equal, do individuals prefer larger or smaller choice sets? Or are they simply indifferent? And if we do find some specific attitude to choice, what underlying phenomenon can explain it? We answer these questions by running an experiment that elicits *individuals' preferences over choice sets* (or *menus*). The experimental design consists of a two-stage decision problem. In the first stage, subjects are asked to provide a monetary equivalent for various choice sets, including singletons. In the second stage, they have to choose one item from one of the choice sets that they "bought" and consume it in the lab. The goods in the choice sets are access to a given website for a compulsory 30 minutes of on-line presence at the end of the experiment.

The monetary valuations of choice sets and the difference in valuation between a choice set and its subsets provide information about the subjects' choice attitudes. For example, the comparison of the value of {a,b,c} to the maximum valuations of {a}, {b} and {c}, or {a,b}, {a,c} and {b,c}, reveals individual preferences over choice. If the individual values the larger set more than any of its components, she has a preference for choice; on the contrary, if the larger set is valued less than the maximum value of its components then she is choice-averse. We specifically focus on two measures of choice attitudes: first, the difference in valuations given to a choice set and its preferred element, and then the change in set valuation when a suboptimal item is added. The difference in valuations provides another indication of choice attitudes, and allows us to ask what underlying phenomenon may drive them.

Our main result is that subjects are choice-averse, with a negative value of more choice in both our measures. Choice sets are first valued less than their preferred element. Second, adding a suboptimal item reduces the value of the choice set, and all the more so when the suboptimal items are of low value. At the individual level, our data show that a majority of subjects are choice-averse, while only a third have a (mild) preference for larger sets. Despite this general negative attitude toward choice, subjects do have a preference over the quality of suboptimal options: the higher-valued are suboptimal items, the more attractive the choice set. In short, individuals are rather choice-averse, but still care about the quality of *all* available items. These results are robust to various controls and auxiliary explanations, such as subject misunderstanding or, more critically, the statistical bias induced by noise in subject valuations.

This paper is related to various fields of Economic literature. It first aims to provide an empirical foundation for theories of preferences over menus, as initiated by Kreps (1979), and the measurement of freedom of choice (Sen, 1988: see Foster, 2011, and Gravel, 2009, for literature reviews). Our main contribution in this respect is probably the lack of support for the premise that the size of the choice set is an important component of well-being, as suggested by a taste for freedom and autonomy (Pattanaik and Xu, 1990; Jones and Sugden, 1982) or preferences for flexibility (Kreps, 1979). Moreover, our results cannot be fully explained by the recent developments in the theory of preference over menus that incorporate cognitive and psychological phenomena, such as decision costs (Ergin and Sarver, 2010; Ortoleva, 2013), regret (Sarver, 2008), temptation and the disutility of self-control (Gul and Pesendorfer, 2001; Dekel et al., 2009), indecisiveness (Danan et al., 2012; Pejsachowicz and Toussaert, 2017) or aversion to responsibility (Dwenger et al., 2018). In particular, the preference for the quality of suboptimal options exhibited by individuals appears to be at odds with most of these models.

Our experimental results can be explained and interpreted in two different ways. The first is to reconsider how to model preferences over menus and the measurement of choice freedom by incorporating internal conflict between multiple ‘selves’ or ‘systems’ (Shefrin and Thaler, 1988; Bernheim and Rangel, 2004; Fudenberg and Levine, 2006; Brocas and Carillo, 2008; Sugden, 2007). Our results can here be rationalised by subjects’ fear of making decisions in the future that are not those that they would prefer today, due to their own bounded rationality or the possibility of an undesirable change in their preferences. The second explanation is that subjects can become confused and rely on cognitive short-cuts to assess choice sets. This heuristic evaluation leads them to consider the choice set as a whole and not to focus on its final consequences, i.e. the value of the option they will end up choosing.

Our findings relate to the empirical literature highlighting that large choice sets can lead to decision avoidance or deferral and damage decision quality as well as post-decision satisfaction (Iyengar and Lepper, 2000; Iyengar and Kamenica, 2010; Reutskaja et al., 2011; Dean et al., 2016).[1] Our paper differs from this existing work as it elicits individuals’ preferences for choice *before* they make a decision. To our knowledge, the only papers that are similar to ours in that they study pre-decision preferences over choice sets are Salgado (2006) and Toussaert (2018). The

1. This is generically referred to as choice overload; the robustness of this phenomenon is still disputed (Scheibehenne et al., 2010; Chernev et al., 2015).

first experimentally considers whether and why subjects are willing to reduce the size of their choice sets, and estimates the role of the anticipation of information and cognitive overload. Toussaert (2018) examines the effect of temptation on individual preferences over menus. She provides experimental evidence that individuals restrict their own choices as they attempt to avoid the costs of the self-control that is needed to prevent them from succumbing to temptation. This work is different from ours in many respects. While the Salgado (2006) experiment involves large and complex choice sets of up to 50 lotteries, ours concentrates on relatively small and simple sets of alternatives. And, contrary to Toussaert (2018), we do not focus on temptation and show that choice aversion goes beyond this type of explanation. Our experiment moreover allows us to test more diverse models of preferences over menus, such as anticipated regret or decision costs.

The remainder of the article is organized as follows. Section 2 provides the theoretical framework and motivations for various attitudes to choice. The next section describes our experimental design and testable hypotheses, and Section 4 sets out the results. Section 5 then discusses a number of potential explanations, and the last section concludes.

2. Theoretical Background

This section considers how different choice sets can be ranked. We assume that individuals are endowed with a preference relation $\succsim$ over choice sets, with its asymmetric and symmetric components denoted $\succ$ and $\sim$ respectively. The preference over singletons produces a ranking of options: x is preferred to y if, and only if, $\{x\} \succ \{y\}$. We also denote $\max(X)$, the set of preferred options in X: formally $\max(X) = \{x \in X : \{x\} \succsim \{y\}, \forall y \in X\}$. An element of $\max(X)$ is denoted x^*.

2.1. Broad attitudes toward choice

We focus on how preferences over choice sets are affected by set size and the quality of their suboptimal options. As we will argue at the end of this section, these two attitudes allow us to test the empirical relevance of various models of preferences over sets that have been proposed in the literature.

2.1.1. Attitudes towards the size of the choice set. Three attitudes toward choice set size can be distinguished: choice neutrality, preference for choice and choice aversion. The following definition formally states these attitudes:[2]

DEFINITION 1 (Attitudes toward choice-set size). *For all pairs of (non-empty and disjoint) choice sets X and X':*[3]

- *Choice neutrality:* $X \succsim X' \Rightarrow X \sim X \cup X'$.

2. These attitudes are constructed using the axiom in Dekel et al. (2009) of Positive Set Betweenness, an axiom itself based on that of Set Betweenness by Gul and Pesendorfer (2001).

3. For the sake of simplicity, all the definitions are given for strict preferences. Focusing on strict rankings is also useful for clearly distinguishing choice neutrality from the other two attitudes, especially in an empirical context. Weak versions of our definitions are straightforward, e.g. in the case of a preference for choice: for all X, X', $X \succ X' \Rightarrow X \succsim X'$ and $\exists X$ and X' s.t. $X \succ X'$.

• *Preference for choice:* $X \cup X' \succ X$.

• *Aversion to choice:* $X \succsim X' \Rightarrow X \succ X \cup X'$.

The preferences of choice-neutral individuals are not affected by the size of the choice set. In other words, enlarging the choice set is not preferred (nor dominated) if it does not allow a better item to be obtained. An individual is indifferent between a set and its preferred element. Under choice neutrality preferences over sets reduce to preferences over best items. On the contrary, with a preference for choice the mere presence of options is a source of satisfaction. An individual with choice preference will perceive the addition of option(s) as desirable, even when these new options are suboptimal. This has two noteworthy and easily-observable implications: $X \cup \{y\}$ is preferred to X, even if the best item in X is clearly preferred to y, and X is preferred to $\{x^*\}$. Last, choice aversion means that a set is preferred to a larger set as long as it is preferred to the additional subset. The intuition is that choice is an annoyance. Choice aversion implies that the best item of a set is preferred to the set. However, in contrast to the other two attitudes, the effect of adding a suboptimal option to a set cannot be formulated directly on the basis of preferences over items.[4] We hence consider a stronger version of choice aversion for two reasons: the first is related to symmetry with the other attitudes in the empirical tests, and the second is to allow for a tractable

4. To see this, consider three items such that $\{x\} \succ \{y\} \succ \{z\}$. Choice aversion implies $\{y\} \succ \{y,z\}$. By transitivity, $\{x\} \succ \{y,z\}$, which implies, by choice aversion, $\{x\} \succ \{x,y,z\}$. Choice aversion does not rule out that $\{y\} \succ \{x,z\}$, hence the relation between $\{x,z\}$ and $\{x,y,z\}$ is left undetermined, even though $\{x\} \succ \{y\}$.

definition of the importance of the quality of suboptimal options in preferences over sets. Weak choice-aversion appears in Definition 1 and the strong version is defined as follows:

DEFINITION 2 (Strong choice-aversion). *For all X and $y \notin X$, $\{x^*\} \succ \{y\} \Rightarrow X \succ X \cup \{y\}$.*

The interpretation of the strong version of choice aversion is more straightforward with respect to the best items in a set: as long as the additional items are not preferred to the best option in the set, choice reduces well-being. This definition ensures the comparability of the three attitudes, and we retain it for this reason in most of our analyses of the experimental data.

2.1.2. Attitudes towards the quality of suboptimal options. Another important characteristic of choice sets that may affect preferences is their composition. We focus here on the subjective quality of suboptimal options:

DEFINITION 3 (Attitudes toward the quality of suboptimal options). *For all choice sets X and for any elements x and y, such that $x, y \notin X$, $\{x\} \succ \{y\}$ and $x \notin \max(X \cup \{x\})$, then:*

- *Indifference to the quality of suboptimal options: $X \cup \{x\} \sim X \cup \{y\}$.*
- *Preference for the quality of suboptimal options: $X \cup \{x\} \succ X \cup \{y\}$.*
- *Aversion to the quality of suboptimal options: $X \cup \{y\} \succ X \cup \{x\}$.*

An individual is indifferent to the quality of suboptimal options if the replacement of a good, but suboptimal, option by a worse one does not affect the value of the choice set. With a preference for the quality of suboptimal options the individual prefers to have good options rather than poor ones, even if they are suboptimal. One the contrary, with aversion to the quality of suboptimal options the presence of valuable but suboptimal options reduces well-being. In other words, it is easier to choose between valuable and low-value items rather than between two items of value.

2.2. Theories of preference over choice sets

Many models of preferences over choice sets have been proposed in the literature. One of the advantages of our definition of attitudes toward choice-set size and the quality of suboptimal options is that the models can conveniently be classified regarding their predictions with respect to these two attitudes. Table 1 sets out the relationship between the most-prominent models of preferences over choice sets and these two attitudes. As there is very little overlap between the predictions, these provide clear-cut tests for these models. The formal derivations of these predictions appears in Appendix B.

Choice neutrality corresponds to the standard Economic view that choice only has instrumental value: i.e. it only matters in that it allows the decision-maker to choose better options. The indirect-utility criterion, as proposed by Arrow (1995), states that the value of a set amounts to the value of its preferred item(s). Although this instrumental approach may ignore important considerations, it provides a convenient

TABLE 1. The theoretical structure by choice attitude and the quality of suboptimal options

Models	Choice-set size	Quality of suboptimal options
Indirect utility	Indiff.	Indiff.
Preference for flexibility	Pref.	Pref.
Cardinality	Strict pref.	Indiff.
Search costs	Weak avers.	Indiff.
Endogenous decision costs	Weak avers.	Non-monotonic
Anticipated regret	Strong avers.	Avers.
Temptation and self-control	Weak avers.	Indiff. for non-tempting opt. Avers. for tempting opt.
Fear of making a bad decision	Weak avers.	Prob. distr. of errors

theoretical benchmark to which other elements can be added. One refinement of this instrumental approach is the preference for flexibility advocated by Kreps (1979), who assumes that future preferences are uncertain. When individuals are uncertain about their future tastes, it is rational to prefer larger sets, since more options imply more opportunities for the final choice, which is better according to future preferences.

In addition to instrumental motivations, there are a number of benefits or costs of choice that can play a role in individual attitudes to choice and the quality of suboptimal options. The mere presence of apparently irrelevant alternatives can improve individual satisfaction, when there is an intrinsic utility of decision. More choice allows individuals to express their own tastes and identity, and to exert their autonomy (Sen, 1988; Sugden, 2003). Following this justification, Pattanaik and Xu (1990) developed the cardinality approach, in which the number of options is used to compare choice sets. However, this purely quantitative approach is itself open to the criticism that it ignores the quality of the available items. As argued by Jones and

Sugden (1982), Puppe (1998) and Puppe and Xu (2010), this latter characteristic may justify the inclusion of individual preferences over items into the ranking of choice sets.

However, choice may not be as innocuous as assumed in the above theories, as it may produce cognitive and emotional costs, as well as other negative outcomes that lead to choice-aversion. One potential source of the latter is decision costs, i.e. the disutility of the time and effort needed to identify the best option within a set (Ortoleva, 2013; Ergin and Sarver, 2010). In its simplest version, the cost of a decision is proportional to the number of elements in the choice set, so that the decision cost refers to the effort of gathering information about the options (*search-costs* in Ortoleva's terminology). The negative effect of adding suboptimal options here should be the same for all options, regardless of their quality. More generally, a rational individual confronted with a choice set should trade off the cost of thinking and the (expected) gain in utility from the final decision. This implies choice-aversion: all else equal, adding a sub-optimal option implies greater decision costs and no expected benefit. Attitudes to the quality of suboptimal options are not monotonic: on the one hand, very low-valued and very high-valued sub-optimal options have little effect, as eliminating the former requires little thought about choice, while the latter generate little expected gain. On the other hand, intermediate-valued options are associated with higher costs of thinking about choice.

Another cause of choice-aversion is that individuals anticipate *psychological* or *emotional* costs associated with decisions. One of these is regret (Loomes and Sugden, 1982; Roese, 1997). Sarver (2008) applies the anticipation of regret to preferences over menus, where the individual is uncertain about her preferences over available options and thus cannot rule out that option y will be *ex post* better than x, even if x provides more satisfaction *in expectation* than y. When evaluating choice sets, she compares the available options based on their expected satisfaction, net of regret. The value of a set is given by the maximum expected level of net utility of the choice set. This model makes two predictions for the ranking of choice sets: first, individuals with regret-driven preferences are strongly averse to choice, as more (*ex ante* suboptimal) options imply more regret possibilities; second, the relative value of the suboptimal options affect set value as the higher the value of the suboptimal item, the greater the expected regret it may generate.

Temptation may also yield a preference for smaller choice sets. Individuals may prefer to commit to a certain option and restrict their opportunities, if these opportunities potentially mean the future choice of some tempting yet undesirable option as viewed from today. Gul and Pesendorfer (2001) propose a variation that takes into account the subjective cost of resisting tempting options, rather than the expected loss from succumbing. They refer to this as the cost of self-control (see also Dekel et al., 2009). In both cases, a decision-maker who correctly anticipates the disutility from temptation exhibits weak choice aversion.[5] The predictions are more ambiguous with respect to the quality of suboptimal options. Any clear

5. More precisely, Gul and Pesendorfer (2001) assume the set betweenness axiom, which requires that $X \succsim X' \Rightarrow X \succsim X \cup X' \succsim X'$.

attitude would require a correlation between the temptation that items generate and subjective evaluations: a negative correlation produces aversion to the quality of suboptimal options, and a positive one a preference for quality. As there are reasonable arguments in both directions, attitudes towards the quality of suboptimal options remain ambiguous.

Another explanation for choice-aversion is the *fear of making a bad decision*. When evaluating a choice set, an individual knows that she is more or less likely to make a bad decision in the future: i.e. she may not choose the option that she initially perceived as the best. There are a number of reasons: individuals may make simple mistakes given that they ‘tremble’ (they suffer from a lack of concentration), or may anticipate that moods, emotions or contextual elements will affect their decisions (Sugden, 2007; Shefrin and Thaler, 1988; Bernheim and Rangel, 2004; Fudenberg and Levine, 2006; Brocas and Carillo, 2008; Ozdenoren et al., 2012). Individuals who are scared of making bad decisions will naturally be weakly choice-averse. The implications for attitudes toward the quality of suboptimal options are less clear, and depend on the probability of (wrongly) selecting the suboptimal option and the difference in subjective value from the optimal option. These push in opposite directions: a bad option is less likely to be picked, but involves a large relative loss, while good suboptimal options are more likely to be picked but involve a smaller relative loss. Overall, attitudes to choice and the quality of suboptimal options yield two measures that can easily be related to theoretical models of preferences over choice sets. More generally, the relationship of the monetary valuation of choice sets to the components of the set allows us to infer these two attitudes and to test the empirical relevance of the theories.

3. Experimental Design

3.1. Experimental Protocol

The experiment has two stages: the first measures individual choice attitudes through the monetary valuation of various choice sets, in the spirit of *Becker-DeGroot-Marschak* (1964), while the second implements the first-stage decision. At the very start of the second stage, one of the valuation tasks participants that carried out in the first stage and a "price" are picked randomly to determine the choice set the subject faces. Only one choice set (round) is randomly drawn. If the price drawn is below the monetary valuation given by the subject, then she has to choose one item from those in the set and consume the selected item in the lab for 30 minutes. If the price is above her valuation, her payoff rises by the price amount, but she has to remain in the lab until the end of the second stage (30 minutes), i.e. in front of her (unusable) computer. The stages of the experiment are summarized in Table 1.

The valuation of choice sets is elicited a little differently from the standard BDM, using the *iterative Multiple Price List* (iMPL) method proposed by Andersen et al. (2006) that is known to produce more robust and consistent results. Participants are repeatedly asked to compare two "options": on the one hand, no consumption good in the second stage but a monetary amount of €m; on the other hand a choice set of

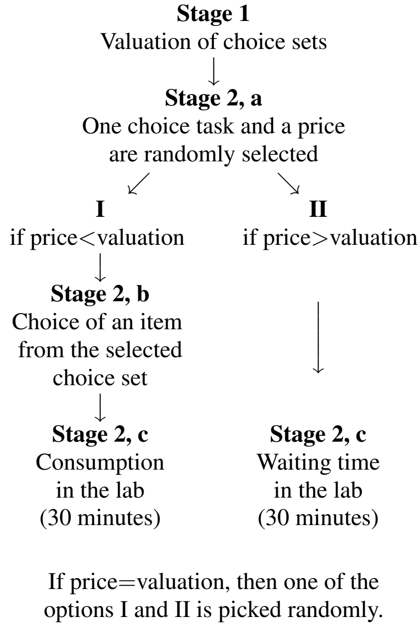


FIGURE 1. The experimental stages

consumption goods in the second stage. The amount m varies systematically to yield an estimate of subjects' indifference points. For each value of m, subjects can either have a strict preference for one option or be indifferent. In a first array, the monetary amounts vary from 0 to 8, in steps of €1. If the subject does not indicate indifference and instead switches directly from the consumption good to money between m and $m+1$, she fills out an additional array where the valuations vary from m to $m+1$ in steps of €0.10. The aim here is to obtain a fairly-precise approximation (i.e. at the €0.10 level) of the monetary valuation of the choice set.

The goods in the choice sets are websites that subjects can access during the 30-minute second stage. There are only four of these, referring to three TV channels and one newspaper, all of which are well-known in France and so expected to be familiar to subjects: TF1 (the most popular general private TV channel in France), M6 (a private TV channel, which is more youth-oriented than TF1), Arte (a publicly-funded TV channel, with a strong focus on Culture and other ‘highbrow’ content) and *Le Monde* (France’s leading mainstream quality newspaper). Subjects are required to value *all of the possible choice sets* formed by combinations of these four items, including singletons (a total of 15 sets), as well as access to a given website during the consumption period not presented as a set (4). The use of two frames for the individual items (as singletons, and then when not presented as a set as items) allows us to test for any unintended effects of the choice-set valuation framing in the case of singletons. In particular, we want to make sure that no negative reaction was triggered by the rather puzzling formulation of *“You can choose from among the following options:*” when only one option was proposed. This also allows us to have less-noisy estimates of the valuation of access to single websites.

In the second stage of the experiment, if the participant had valued the selected choice set higher than the random price drawn, she has to choose one website from this set (Stage 2,b), which is the only one she can access in Stage 2,c. The computers will prevent them from accessing any other website and subjects are not allowed to use any documents (books, magazines or others) or means of communication. When the participant chooses money rather than a choice set, she has to stay in the lab and remain in front of a blank computer screen for the 30-minute period. This rather unconventional procedure aims to provide incentives that are strong enough for subjects to value the choice sets carefully. More importantly, having subjects consume the good in the lab allows clear control of the circumstances of consumption and the choice set that subjects face. Providing goods that individuals can take home would instead create interference between choices in the lab and at home: an individual may always choose good B for the sake of choice, as they already have good A at home. Likewise, a choice-loving individual would favor money over the restricted choice set proposed in our experiment, as when the experiment is over she can "buy" herself more choice outside with her earnings.

3.2. Measurements and hypotheses

The monetary valuation yields a measure of the attractiveness of a given choice set, prior to any actual decision of which website to consult within the choice set. Let V_x be the monetary valuation of item x, V_X that of the choice set X and V_{x^*} that of the preferred item(s) in X. A first measure of choice attitudes is the valuation gap between a choice set and its preferred element: $V_X - V_{x^*}$. This is positive for choice-preference, zero for choice-neutrality, and negative for both weak and strong choice-aversion. A second measure of choice attitudes is the change in choice-set value from adding an item: $V_{X\cup\{y\}} - V_X$. The relevant test refers to situations where a suboptimal item is added, i.e. $V_y < V_{x^*}$. The difference between $V_{X\cup\{y\}}$ and V_X is zero for choice-neutrality and positive for choice-preference. Under the strong version of choice-aversion (Def. 2), $V_{X\cup\{y\}}$ is always lower than V_X, while under the weak version (Def. 1) this is only the case when y is of lower value than the original set X (with two or more items). Regarding attitudes to the quality of suboptimal options, the change in $\left(V_{X\cup\{y\}} - V_X\right)$

with respect to V_y and in $\left(V_X - V_{x^*}\right)$ with respect to V_x (with $x \in X \setminus \{x^*\}$) tells us about subjects' attitudes toward the quality of suboptimal options. These implications are summarized in Table 2.

TABLE 2. Predictions

Attitude towards choice-set size	Comparison between X and x^* with x^* s.t. $V_{x^*} = \max_{x \in X} V_x$	Comparison between $X \cup \{y\}$ and X with $y \notin X$ and $V_{x^*} > V_y$
Choice-Neutrality	$V_X = V_{x^*}$	$V_{X \cup \{y\}} = V_X$
Choice-Preference	$V_X > V_{x^*}$	$V_{X \cup \{y\}} > V_X$
Strong Choice-Aversion	$V_X < V_{x^*}$	$V_{X \cup \{y\}} < V_X$
Weak Choice-Aversion	$V_X < V_{x^*}$	$V_{X \cup \{y\}} < V_X$ iff $V_X > V_y$
Attitude towards the quality of suboptimal options	Change in $V_X - V_{x^*}$ with V_y	Change in $V_{X \cup \{y\}} - V_X$ with V_y
Indifference	No change	
Preference	Increasing in V_y	
Aversion	Decreasing in V_y	

3.3. Procedures

The order in which choice sets are presented in Stage 1 is reshuffled for every session so as to produce a balanced design. Four control questions are asked after the instructions were read out by the experimenter (and read by the participants). Given the particularity of our experimental design, the questions focus on the subjects' correct understanding of the meaning of a choice set. In particular, they test whether participants understand the procedure at the beginning of Stage 2: i.e. that the final choice will be theirs, and not someone else's or the result of a random device, and that during Stage 2 they will only be able to consult one website. Participants were paid a show-up fee of €11 (about $12), plus the earnings corresponding to the monetary-valuation procedures. The show-up fee is higher than usual as the subjects had to stay for half-an-hour during the second stage of the experiment, regardless of their first-stage decisions. The sessions took place at the laboratory of the University of Rennes-1 (LABEX-EM) with 72 subjects, almost all of whom were undergraduate students, majoring in various subjects. They earned on average around €14.70 (about $16), with the sessions lasting 60 to 70 minutes. All of the experiments were computerized with Z-tree (Fischbacher, 2007).

4. Results

4.1. Overview

The average monetary valuation of choice sets was of €2.20,[6] with considerable variance (s.d. €1.83). Only three subjects were only motivated by money (they preferred €0.10 to any item or choice set),[7] and slightly over 10% accepted to pay more than €4 on average. To obtain subjects' preferences over items and the value of

6. We estimate this figure as the average of the two monetary amounts that subjects switch between when they do not pick 'indifference'. A valuation of €0.05 means that the subject chose website access when the amount offered was €0, but preferred €0.10 to the website.

7. In the remainder of the paper, all analyses are based on the sample of individuals who are not only motivated by money, i.e. those who valued a website or choice set at over €0.05 on at least one occasion. Those solely motivated by money are not useful for our purpose here, although including them does not affect our analytical results.

the preferred item within a set, we can use: (a) the values given for the single item (i.e. the value of "having direct access to *x*"), (b) those for the singleton set (i.e. the value of "having the choice between the following option(s): *x*"), or (c) the average of these two. Table 3 displays the mean valuations for the different websites. The different measures are quite similar, with the ranking being consistent across columns.

TABLE 3. These figures are valuations of individual website access (in €)

	Single item	Singleton	Average
LeMonde	2.61	2.48	2.55
TF1	1.89	1.69	1.79
M6	2.04	1.97	2.01
Arte	2.21	2.07	2.14

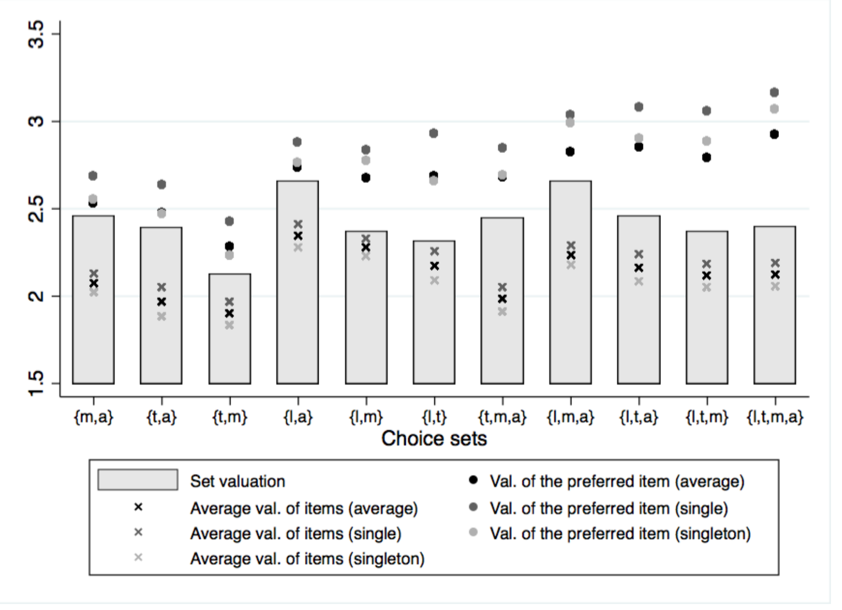
Average valuations across individuals, based on the value of the single item ("Direct access to TF1"), the singleton ("{TF1}"), and the average of the two.

4.2. Evidence of choice-aversion

4.2.1. Aggregate analysis. To elicit subjects' attitudes to choice, we first compare the value of a choice set to that of its best element.

RESULT 1. *At the aggregate level, the value of a choice set is lower than that of its preferred element. Moreover, the value of a choice set is larger than the average value of its components.*

Using the maximum of all valuations of single items within the choice set, the difference is -44.9 cents for an average set valuation of €2.20: subjects value the choice set on average 20% less than the preferred item. When we identify the best item based on the value of singletons or the average of values for single items and singletons, the results are similar, even though choice-aversion is lower at respectively



Note: This figure shows the average valuation of sets, that of the preferred item and that of all the items, using the three methods to identify preference over items. 'l' = lemonde.fr; 't' = tf1.fr; 'm' = M6.fr; 'a' = arte.tv.

FIGURE 2. Set valuations, valuations of the preferred item and the average value of items

-25.7 and -32.5 cents. Moreover, the values of sets are higher than the average values of their items. This difference ranges from 23 cents (using valuations of single items to measure preferences over items) to 37 cents (using valuations of singletons). In other words, the value of a choice set seems to lie between the value of its best option and the average value of all its options.

These results also hold when considering each choice set separately. Figure 2 shows the average value of each choice set, the average value of its preferred item, and the average value of all its items. Tables 4 and 5 present the exact numbers and the results of the statistical tests. First, monetary valuations for sets appear systematically lower than the valuation of the corresponding preferred item. As shown in Table 4, the difference in the distributions is in general confirmed by the two-sided signed-rank

TABLE 4. Difference in valuations between choice sets and the preferred item (in €)

	Difference between V_X and V_{x^*} Estimate of V_{x^*} based on:		
Choice set	Single items	Singletons	Average
{M6,Arte}	-0.23**	-0.32***	-0.07
{TF1,Arte}	-0.25**	-0.08	-0.09
{TF1,M6}	-0.30**	-0.11	-0.16
{LeM.,Arte}	-0.22	-0.11	-0.08
{LeM.,M6}	-0.47**	-0.40***	-0.31**
{LeM.,TF1}	-0.62***	-0.35***	-0.37***
{TF1,M6,Arte}	-0.40***	-0.24	-0.23
{LeM.,M6,Arte}	-0.38*	-0.33*	-0.16
{LeM.,TF1,Arte}	-0.62***	-0.44***	-0.40**
{LeM.,TF1,M6}	-0.69***	-0.52***	-0.42***
{LeM.,TF1,M6,Arte}	-0.77***	-0.67***	-0.53***

Notes:

a. These are the average differences in monetary valuation between X and the preferred element for the three methods of eliciting preference over items.

b. Level of statistical significance. * = significant at the 0.10 level, ** = 0.05 level, *** = 0.01 level for the Wilcoxon signed-rank test.

TABLE 5. Difference in valuations between choice sets and the average value of items (€)

	Difference between V_X and $(1/|X|)\sum_{x\in X} V_x$ Estimates of V_x based on:		
Choice set	Single items	Singletons	Average
{M6,Arte}	0.33*	0.43***	0.38**
{TF1,Arte}	0.34*	0.5***	0.42**
{TF1,M6}	0.16*	0.29*	0.22
{LeM.,Arte}	0.25**	0.38*	0.31**
{LeM.,M6}	0.039	0.14	0.089
{LeM.,TF1}	0.059	0.23***	0.14**
{TF1,M6,Arte}	0.4**	0.53***	0.47***
{LeM.,M6,Arte}	0.37**	0.48***	0.43***
{LeM.,TF1,Arte}	0.22**	0.37***	0.3***
{LeM.,TF1,M6}	0.18*	0.32***	0.25**
{LeM.,TF1,M6,Arte}	0.21	0.34**	0.27**
All sets	0.22***	0.383***	0.303***

Notes:

a. These are the average differences in monetary valuation between X and the average valuation of the items within X for the three methods of eliciting preference over items.

b. Level of statistical significance. * = significant at the 0.10 level, ** = 0.05 level, *** = 0.01 level for the Wilcoxon signed-rank test.

paired Wilcoxon tests. In addition the differences reported here increase in absolute value with the number of items, rising to between 53 and 77 cents for the set composed of the four items, i.e. over one quarter of the valuation of the choice set. Second, the value of each choice set is larger than the average value of its components, confirming that individuals are sensitive to the quality of the best option. Last, Appendix A presents some evidence that these two phenomena also hold at a more disaggregated level (subject$\times$choice set). Overall, our first measure of choice attitudes consistently suggests choice-aversion.

A similar picture results from our second measure of choice attitudes, namely that there is a change in choice-set valuations, when a suboptimal option is added:

RESULT 2. *At the aggregate level, adding a suboptimal item reduces the value of a choice set.*

Table 6 presents the results for the addition of a non-preferred item, that is the average difference between the value of a set X and the value of the same set plus some suboptimal item, i.e. $X \cup \{y\}$. On average, for the three methods of measuring preferences over items, adding a suboptimal item reduces the value of a set by slightly over 10 cents, which is 4.5% of the average set value. When the average difference is calculated over all possible comparisons for an individual, the Wilcoxon signed-rank tests yield statistically-significant results (at the 5% or 10% levels). Comparing singletons with pairs, and triples with quadruples leads to the same conclusion, although statistical significance is weaker. No difference is found when comparing pairs with triples.

We now ask whether the strong version of choice-aversion is found in our data, i.e. how the (negative) difference between $V_{X\cup\{y\}}$ and V_X varies with the sign of $V_X - V_y$.

RESULT 3. *At the aggregate level, adding a suboptimal item reduces the value of a set if the new item is valued less than the choice set; it increases its value if the new item is valued more than the original set.*

This can clearly be seen in the last three rows of Table 6. The average difference between $V_{X\cup\{y\}}$ and V_X is significantly negative when y is valued lower than X (between -0.33 and -0.52). On the contrary, $X\cup\{y\}$ is on average significantly higher-valued than X when y has a higher valuation than X, even though y is valued less than x^*. This suggests that subjects' preferences over sets are more in line with the weak than the strong version of choice-aversion.

TABLE 6. Differences in valuations when adding a suboptimal item (in €)

	Average diff. between $V_{X\cup\{y\}}$ and V_X with $V_{x^*} > V_y$					
	Estimates of V_x based on:					
	Single items		Singletons		Average	
	Average by subj.	Pooled data	Average by subj.	Pooled data	Average by subj.	Pooled data
All comparisons	-0.106*	-0.096	-0.107**	-0.112	-0.123**	-0.124
Singletons vs. pairs	-0.196	-0.154	-0.164	-0.195	-0.217***	-0.225
Pairs vs. triples	0.025	0.001	0.001	0.028	0.025	0.028
Triples vs. quadruples	-0.131	-0.135	-0.171	-0.168	-0.142	-0.138
if $V_y < V_X$	-0.312***	-0.260	-.352***	-0.301	-0.334***	-0.310
if $V_y = V_X$	0.682*	0.512	0.992***	0.757	0.786***	0.583
if $V_y > V_X$	1.027***	0.869	1.130***	1.030	0.987***	1.036

Notes:

a. This table shows the average differences in monetary valuations between $X\cup\{y\}$ and X for the three methods of eliciting preferences over items. We present statistics considering all observations (difference between two sets$\times$subject) as independent ("Pooled data") and averaged by subject ("Average by subj.").

b. Levels of statistical significance. * = significant at the 0.10 level, ** = 0.05 level, *** = 0.01 level for the Wilcoxon signed-rank test for the symmetry at 0, *only carried out for the individual average values.*

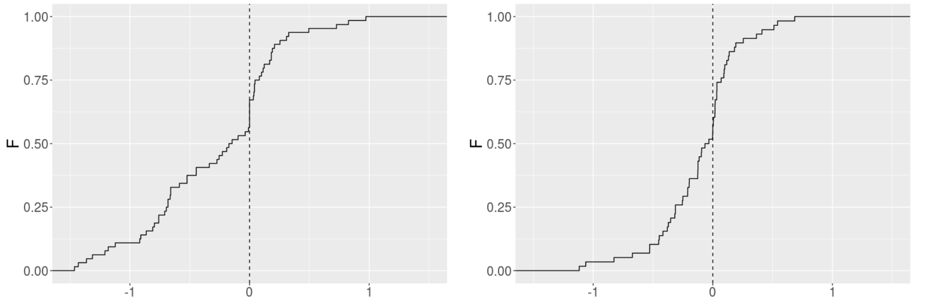
An additional test, which combines elements from both our choice-attitude measures, provides additional support for the importance of choice-aversion. This consists in comparing the value of choice sets of three items or more to the maximum value of the possible pair subsets. This difference is negative, with an average value of 60 cents. A series of Wilcoxon signed-rank tests on the five choice sets that have three or more items produce significant results in all cases ($p < .01$). Overall, subjects value choice sets less when a suboptimal option is added. This provides strong support for individuals being, on average, choice-averse.

4.2.2. Heterogeneity among subjects. We now ask whether this strong average effect conceals some heterogeneity in individual choice attitudes.

RESULT 4. *Even though we find individual heterogeneity, most subjects are choice-averse.*

Figure 3 shows the cumulative distribution functions of the difference averaged by subjects (a) between V_X and V_{x^*}, and (b) between $V_{X \cup \{y\}}$ and V_X. The majority of subjects are choice-averse, even though a significant proportion of subjects value larger choice sets more. According to this measure, 50% of subjects are on average choice-averse, and 25-30% have a preference for choice. The remaining subjects have a broadly neutral view of choice.[8]

8. Figure A.3 (throughout the paper the prefix A. refers to material in the Appendix) lists the average values of sets, their preferred component, and the average of all their items, all at the individual level, using the average value for single items and singletons to identify preferences over items. In addition, as shown in Figure A.4, a majority of subjects (about 60%) are found to value sets more than the average valuation of their items, while for about 10% of subjects these two values are the same. Last, 30% of subjects value choice sets less than the average items in them.



(a) The difference between choice sets and their preferred item ($V_X - V_{x^*}$)

(b) Adding an item ($V_{X\cup\{y\}} - V_X$)

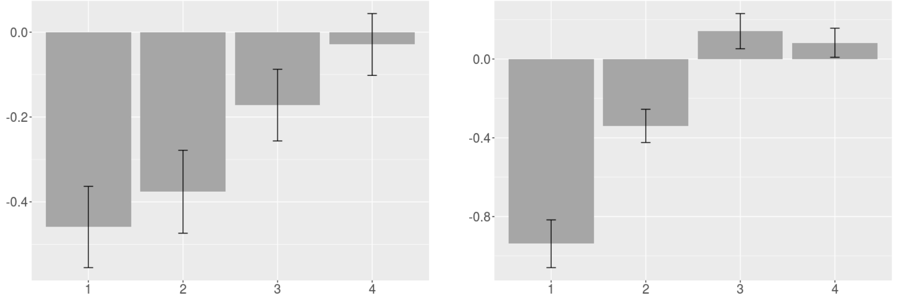
FIGURE 3. *CDFs of the differences between sets and relevant subsets, averaged by individual*

4.3. Evidence of preferences over the quality of suboptimal options

Our next findings concern the importance subjects attach to the quality of suboptimal options:

RESULT 5. *Subjects exhibit a preference for the quality of suboptimal options.*

Figures 4 and 5 show the effect of adding a suboptimal option, depending on the relative value of this option. Panel (a) of Figure 4 corresponds to the quartiles of the normalized differences between the average value of suboptimal options, denoted $V_{x^{-*}}$, and the value of the preferred item, that is, $V_{x^{-*}}/V_{x^*}$. Panel (b) shows the quartiles of the difference between the average value of suboptimal options and the value of the preferred element, $V_{x^{-*}} - V_{x^*}$. Clearly, the smaller is the relative value of the suboptimal options, the lower the value of the set. For example, the average difference is slightly over -80 cents for the first quartile of the difference between $V_{x^{-*}}$ and V_{x^*} (strong preference for the preferred item to others). On the contrary, when the preferred item is slightly preferred to the other items in the choice set, then V_X is close to V_{x^*}: the difference is close to zero when $V_{x^{-*}} - V_{x^*}$ is in the top quartile.



(a) Quartiles of the normalized average value (b) Quartiles of the average diff. in values

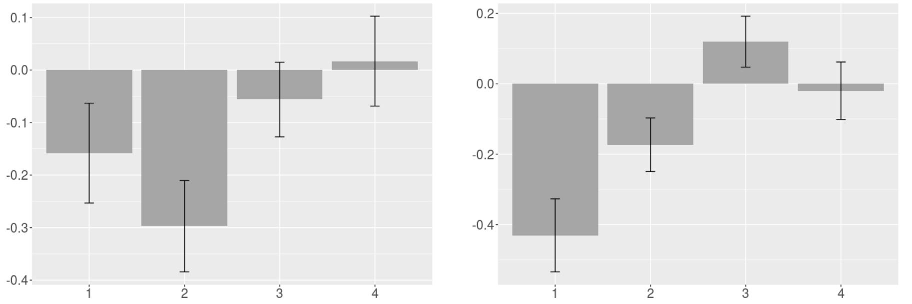
These figures show the difference between V_X and V_{x^*} according to the relative values of the suboptimal options. In Panels (a) and (b), the x-axis refers to the quartiles of $V_{x^{-*}}/V_x^*$ and $V_{x^{-*}} - V_x^*$ respectively. The confidence intervals are based on pooled standard errors.

FIGURE 4. *The difference in valuations between choice sets and the preferred item, by quartiles of the average quality of suboptimal items.*

A similar conclusion results when comparing V_X to $V_{X\cup\{y\}}$, where y is a suboptimal option. The effect of the quality of y, as measured by V_y, is depicted in Figure 5, again by quartiles, either of V_y/V_{x^*} or $V_{x^*} - V_y$. Overall, the data suggest a roughly monotonic effect of the quality of choice: the worse the suboptimal option, the less the choice set is valued. For instance, if x^* is strongly preferred to y according to the difference between V_y and V_{x^*}, $V_{X\cup y}$ is lower than V_X by €0.4, while there is no difference when y is close to x^*.

4.4. Robustness

We evaluate the robustness of our results by addressing two potential sources of mismeasurement in our experiment. The first relies on possible misunderstandings by the experimental subjects of the consequences of their decisions. The second is



(a) Quartiles of the normalized value of y, V_y/V_{x^*}

(b) Quartiles of the diff. in value between y and x^*, $V_y - V_{x^*}$

These figures show the difference between $V_{X\cup y}$ and V_X, according to the value of y. Panels (a) and (b): The x-axis refers to the quartiles of V_y/V_x^* and $V_y - V_x^*$ respectively. The confidence intervals are based on pooled standard errors.

FIGURE 5. *The difference in valuations when adding a suboptimal item, by quartiles of the average quality of the suboptimal options.*

related to the effect of noise in valuations that biases some of our comparisons. Last, we check our two main results via an econometric test that aims to take these possible confounding effects into account.

4.4.1. Subjects' misunderstanding of the experimental task. A possible artifactual explanation of our results could be subjects' confusion, in particular regarding what a choice set really means in Stage 2. It is fairly unusual to have to evaluate sets (and not single options): subjects may have thought that they themselves would not choose within the set in the second stage, but that some random mechanism (and not themselves) would pick one item from it, especially since these types of procedures (with the extensive use of lotteries and random devices in experiments) are very common. We do not believe that this explanation is germane, for at least three reasons. First, the written and computerized instructions made the description of the choice set as clear as possible, as we anticipated that the subjects would not be familiar with these procedures.

Second, four control questions were asked at the beginning to make sure that the experimental procedure was clear to subjects. Question 4 in particular checked for this possible misunderstanding.[9] Using only data from those who correctly answered Question 4 does not substantially affect the results: there is an average difference of 43 cents between the preferred item and the whole set, all sets are associated with a negative gap in value using the preferred item, and there is no major change in terms of statistical significance (Table A.2). The same holds for the difference between $V_{X\cup\{y\}}$ and V_X (Table A.1). When analyzing the effect of the quality of choice, the results are similar to those in the full sample (Figures A.1 and A.2). Overall, the removal of subjects who did not correctly answer the fourth control question testing for the correct understanding of a choice set has little effect on our results.

Third, and perhaps more importantly, we can rely on the difference between the value of sets and the average value of their items. This difference should be zero for a risk-neutral subject who wrongly thinks that an item is randomly picked at the beginning of the second part of the experiment. Risk aversion should then make the value of sets lower than the average value of their items. However, we find that a large

9. The correct answer to the fourth question, among the five possible, was explicit "I will choose one website among Lemonde.fr, Arte.tv, and M6.fr, and I will have access to this website for 30 minutes and will earn zero (in addition to the show-up fee)." See the Experimental Instructions in Appendix C for further details.

percentage of subjects did not confuse the choice set with a lottery, so that subject confusion seems unlikely as an explanation of our findings. First, as shown in Table 5, the value of nearly all choice sets is significantly higher than the average value of their items. Second, Figure A.3 indicates that few people value, on average, choice sets lower than the average value of their components. More specifically, only four subjects always gave a lower valuation to sets than the average value of their items. Most subjects then value choice sets between the value of its preferred element and the average value of all the set's components. This suggests that, although falling short of valuing a set at the exact value of its elements (the instrumental choice-neutrality view) individuals still take into consideration that they will have the ability to choose.

4.4.2. Noise in valuations and econometric estimations. Noise in subject set valuations may affect our analyses. As formally shown in Appendix A.4, noise has three implications. First, the gap between V_X and V_{x^*} may be underestimated, so that we overestimate the extent of choice-aversion or wrongly conclude that the value of the preferred element is larger than that of the set.[10] Moreover, the statistical bias falls with the gap between the valuation of the preferred and suboptimal items, and rises with noise variance. Second, the change in the value of a set when adding a suboptimal

10. Consider an option x whose valuation is given by a random variable $Z_x = U_x + \mu_x$ where U_x is the "true" valuation and μ_x zero-mean noise. Suppose similarly that the valuation of a set is given by $Z_X = U_X + \mu_X$. As shown in the Appendix, $E(Z_X) = U_X$ and $E\left(\max_{x\in X} Z_x\right) > \max_{x\in X} U_x$. Thus, $E(Z_X) - E\left(\max_{x\in X} Z_x\right) > U_X - \max_{x\in X} U_x$: subjects may be choice-neutral while our estimates suggest choice-aversion.

item is biased in the other direction. As there is a non-zero probability that y be wrongly estimated as suboptimal, the expected value of $X \cup \{y\}$ is larger than that of X for a choice-neutral individual with noisy valuations. Third, noise in valuations should lead the gap between $V_{X\cup\{y\}}$ and V_X to be close to zero, if x^* is strongly preferred to y (the probability of an error is close to 0) and rising in the subjective value of y. Given the prevalent choice-aversion in both our tests, it is unlikely that this only reflects bias due to noisy valuations. Moreover, subjects exhibit preferences for the quality of suboptimal items, so that again noise alone cannot explain our results. Last, looking at actual choices from the Stage-2 choice set, there is perfect consistency between valuations and actual choices.[11]

In any case, testing the robustness of our results to these biases is essential to confirm the presence of choice-aversion. To see how these biases affect our findings for $V_X - V_{x^*}$, we first use an alternative method to identify the preferred item. We next run regressions of $V_X - V_{x^*}$ using this new measure, and regressions based on the average of valuations for single items and singletons. To obtain our alternative estimate of the value of the best item, we first rank the four items using the average value of the sets containing each item, and then for each set take its value when presented alone or

11. Of 16 cases where subjects chose from non-degenerate choice sets, they in all cases chose the option with the highest value in the valuation tasks. Given this subsample size, this is not conclusive but suggestive.

as a singleton as the value of the preferred item, once we have identified the best item. We denote $\underline{V}_{x^*}$ as the estimated value.[12]

We next estimate the following generic equation:

$$V_X - V_{x^*} = \alpha_1 + \alpha_2.NbItems_X + \sum_{x \in X \setminus x^*} \alpha_3^x.\frac{V_x}{V_x + V_{x^*}} + \sum_{x \in X} \alpha_4^x.Noise_x \qquad (1)$$

$$+ \alpha_5.Understanding + \alpha_6.Period + \varepsilon$$

where $Noise_x$ is a measure of the noise in valuation the subject gives to item x, which is the absolute difference between the individual valuations of the item presented as a single item and as a singleton; *Understanding* is a vector of the four control questions for the understanding of the experimental instructions (1 if the subject gave the right answer and 0 otherwise); and *Period* is the period of the valuation task among all tasks. The presence of the relative value of the suboptimal items measured by $V_x/(V_x + V_{x^*})$ allows us to control for a feature of the potential statistical biases, that they fall with the difference between the preferred item and the others.

Tables 7 and 8 show the estimation results of equation 1 using $\underline{V}_{x^*}$ and V_{x^*} as the value of the preferred item in X. Columns (1) to (4) show OLS estimates with standard errors clustered by subjects. The *noise* in the valuations of items only appears in specification (4) and *misunderstanding* in specification (3). Another way to control for

12. We also directly perform the tests in Section 4.2.1 with this downward-biased estimate. As expected, we find even more support for Result 1, while support for Result 2 is weaker. Overall, the qualitative findings seem robust. See Appendix A.2.3 for details.

TABLE 7. *Econometric analyses of $V_X - \underline{V}_{x^*}$.*

	Endogeneous variable: $V_X - \underline{V}_{x^*}$					
	OLS Clustered SE	OLS Clustered SE (w/o singl.)	OLS Clustered SE	OLS Clustered SE	OLS Fixed eff.	Max. Likel. Random eff.
	(1)	(2)	(3)	(4)	(5)	(6)
NbItems	-0.789***	-0.756***	-0.819***	-0.510**	-0.854**	-0.831***
	(.275)	(.212)	(.277)	(.209)	(.353)	(.098)
$\frac{V_x}{V_x+V_{x^*}}$ for item 2	1.670**	1.740*	1.764**	1.013**	1.932*	1.848***
	(.785)	(.987)	(.790)	(.514)	(1.010)	(.302)
$\frac{V_x}{V_x+V_{x^*}}$ for item 3	1.568**	1.496**	1.647**	1.069**	1.724*	1.664***
	(.730)	(.583)	(.735)	(.512)	(0.975)	(.382)
$\frac{V_x}{V_x+V_{x^*}}$ for item 4	2.110***	2.028***	2.180***	1.601***	2.154***	2.124***
	(.566)	(.478)	(.571)	(.441)	(.747)	(.621)
Noise x^*				-0.373***		
				(.086)		
Noise for item 2				0.030		
				(.057)		
Noise for item 3				-0.117		
				(091.)		
Noise for item 4				-0.190		
				(.147)		
Understanding	no	no	yes (non-sig but Q3)	no	no	no
Period	-0.029**	-0.029**	-0.031**	-0.022**	-0.024**	-0.026***
	(.012)	(.012)	(.012)	(0.010)	(.010)	(.009)
Constant	0.957***	0.848***	0.755	0.894***	0.899**	0.950***
	(.337)	(.316)	(.525)	(.316)	(.395)	(.191)
$\bar{R}^2$	0.068	0.080	0.082	0.173	0.232	
Cond. R^2						0.242
Log Lik.						-1799 (8df)
No. of obs.	1035	759	1035	1035	1035	1035
No. of clusters	69	69	69	69	69	69

Notes:

a. The *relative value of suboptimal items* is measured by $V_x/(V_x+V_{x^*})$: for item 2, x not being the preferred item within X but preferred to all of the other set items; the relative values for items 3 and 4 are defined analogously.

b. *Noise* is the absolute difference between the individual valuations of the item presented as a single item and as a singleton.

c. *Understanding* is a vector of control questions for understanding the experimental instructions (1 if the subject gave the right answer and 0 otherwise). Period is the period of the valuation task among all tasks.

d. Significance level: *, ** and *** significant at the .10, .05 and .01 levels respectively. OLS clustered standard error p-values come from the Wald statistic, and those in the random-effect model by comparing nested-model fit via Maximum Likelihood. The results (not reported, except for model 2) are similar for estimations without singletons.

the noise is to run fixed- (5) or random- (6) effect regressions, since we can reasonably assume that the variance in the noise is mostly an individual characteristic that will be picked up by the individual effects.

TABLE 8. *Econometric analyses of* $V_X - V_{x^*}$.

	Endogeneous variable: $V_X - V_{x^*}$					
	OLS Clustered SE	OLS Clustered SE (w/o singl.)	OLS Clustered SE	OLS Clustered SE	OLS Fixed eff.	Max. Likel. Random eff.
	(1)	(2)	(3)	(4)	(5)	(6)
NbItems	-0.567***	-0.427**	-0.576***	-0.513***	-.579***	-.572***
	(.160)	(.211)	(.159)	(.168)	(.171)	(.084)
$\frac{V_x}{V_x+V_{x^*}}$ for item 2	1.561***	1.771***	1.587***	1.590***	1.644***	1.601***
	(.422)	(.457)	(.416)	(.406)	(.461)	(.213)
$\frac{V_x}{V_x+V_{x^*}}$ for item 3	1.194***	0.913*	1.207***	1.312***	1.154***	1.169***
	(.423)	(.533)	(.418)	(.389)	(.422)	(.282)
$\frac{V_x}{V_x+V_{x^*}}$ for 4	0.928*	0.589	0.974**	0.672	1.000*	0.960**
	(.477)	(.592)	(.475)	(.507)	(.523)	(.470)
Noise x^*				-0.034		
				(.059)		
Noise for item 2				-0.073*		
				(.044)		
Noise for item 3				-0.114		
				(.071)		
Noise for item 4				0.035		
				(.0.086)		
Understanding	no	no	yes (non-sig)	no	no	no
Period	-0.016	-0.016	-0.018*	-0.015	-0.019**	-0.017**
	(.010)	(.011)	(.010)	(.010)	(.009)	(.008)
Constant	0.642***	0.243	0.316	0.600***		0.655***
	(.160)	(.400)	(.295)	(.182)		(.155)
$\bar{R}^2$	0.062	0.069	0.068	0.073	0.125	
Cond. R^2						0.132
Log Lik						-1608 (8df)
No. of obs.	1035	759	1035	1035	1035	1035
No. of clusters	69	69	69	69	69	69

Notes:

a. The *relative value of suboptimal items* is measured by $V_x/(V_x + V_{x^*})$: for item 2, x not being the preferred item within X but preferred to all of the other set items; the relative values for items 3 and 4 are defined analogously.

b. *Noise* is the absolute difference between the individual valuations of the item presented as a single item and as a singleton.

c. *Understanding* is a vector of control questions for understanding the experimental instructions (1 if the subject gave the right answer and 0 otherwise). Period is the period of the valuation task among all tasks.

d. Significance level: *, ** and *** significant at the .10, .05 and .01 levels respectively. OLS clustered standard error p-values come from the Wald statistic, and those in the random-effect model by comparing nested-model fit via Maximum Likelihood. The results (not reported, except for model 2) are similar for estimations without singletons.

The results in Tables 7 and 8 confirm our previous conclusion that adding suboptimal items reduces set value. First, for both measures of the preferred-item value and all regressions, the number of set items always attracts a significant negative estimated coefficient. Second, the robust effects of the ratios confirms that the quality of suboptimal options matters for the subjective valuations of choice sets: the higher-valued these options relative to the preferred item, the higher is choice-set value. Third, experimental understanding does not affect choices, so that misunderstandings are unlikely to explain our results. Finally, since our results are robust across econometric specifications, and the noise variables are in general not significantly different from zero, our main result cannot be explained by noisy valuations alone.

5. Discussion

Our experimental results indicate that individuals are (weakly) choice-averse, but are sensitive to the quality of suboptimal options. This immediately excludes some theoretical explanations of choice aversion, as can be seen in Table 1.[13] First, and straightforwardly, observed behavior is not consistent with theories where more opportunities are always desirable, such as social-choice theories based on the intrinsic value of choice and the preference for flexibility. Second, models of anticipated regret cannot account for our preference for the quality of suboptimal options. Third, some versions of decision costs cannot predict the positive effect of the value of suboptimal items on the attractiveness of a set under the reasonable assumption that decision costs are non-decreasing in the set size and increasing in the (expected) value of the suboptimal options. Moreover, additional tests on our data rule out explanations based on decision costs that rise with item dissimilarity: we find no evidence of an effect of

13. All additional empirical tests related to the theoretical models appear in Appendix B.

the diversity of options on the valuation of the choice set (see Appendix B.2.2). Fourth, theories based on commitment or subjective cost of self-control are not consistent with some of our results. In particular, the addition of suboptimal items reduces the value of the set, irrespective of how tempting the option is: even adding the "highbrow" option Arte.tv produces a lower set value.

Two broad types of explanation are compatible with our results. The first is the *fear of making a bad decision* (See also Sugden, 2007).[14] When assessing choice sets, individuals know that they have a positive probability of making a bad choice decision. The intuition is as follows: when evaluating choice sets, the individual knows her true preferences but also that her motivation, mood or preferences at the moment of choice may change in undesirable ways. This set-up is the opposite to the Krepsian model of a preference for flexibility, in which tastes are revealed when choosing within a set. The individual may then rationally restrict her choice set or value smaller sets, producing choice-aversion. However, the implications for preferences over the quality of suboptimal options are less straightforward, depending on the relationship between the (subjective) probability of a future mistake and the gap in subjective values between options. We show that a simple plausible specification of this subjective probability (a logistic model based on option values) can generate a preference for the quality of suboptimal options (see Appendix B.5). This latter specification yields additional predictions about the effect of adding suboptimal options, with the set

14. We here propose an informal discussion of this explanation: the formal model is in Appendix B.5.

value effect depending on the (relative) value of the option added. Added options are desirable if and only if they are valued more than the original set. The intuition is simple: adding a third option z to a set $\{x,y\}$, with x strongly preferred to y, reduces the probability that x but also y be chosen. If the value of the lower probability of x is outweighed by the lower probability of y, then adding a high-quality (but suboptimal) option increases the value of the set. The fear of making a bad decision hence implies weak choice-aversion (Definition 1) rather than its strong version (Definition 2). This is consistent with Result 3 and with all our qualitative results (although it is worth noting that it requires that individuals have relatively little confidence in the quality of their future decisions). However, two phenomena may reinforce this fear of a bad decision even when there is a smaller probability of a mistake. First, psychological factors, such as regret or the burden of responsibility, may render mistakes subjectively more costly than the gap in option values. Second, probabilities may not be treated linearly (Wakker, 2010), and in particular individuals can strongly overweight small probabilities.

The second plausible explanation of our results is that individuals rely on some *imperfect choice-set valuation heuristic* in which they do not consider the whole sequence of the experimental task (i.e. the whole process and its consequences), but rather use some approximate cognitive procedure. One such plausible cognitive shortcut is to assess the choice set as the weighted average of its components. Individuals may spontaneously only have imperfect knowledge of their own preferences and not wish to reflect immediately, making this cognitive procedure appealing. The implications of this type of explanation are in line with our finding of weak choice-aversion and an (apparent) preference for the quality of suboptimal options.

This explanation may be thought to be implausible, as the task is not very complex and the cognitive demands from evaluating the second step – the choice of an item within the set – are arguably only low. In other words, failing to evaluate the second stage of the task implies considerable myopia. Nevertheless, two arguments can support this interpretation. First, subjects' preferences over options may not be completely transparent or fixed. Rather than first ranking options and then assessing the value of the best one, broadly assessing the set as a whole as a weighted average of its components may produce a relatively good approximation of the set value. This argument is comparable to that in Kreps (1979), in that individuals do not really know what they will want in the future, in the sense that they do not know the probability distributions of their future tastes, or are not sufficiently sophisticated to take them into account. In this case, turning to a reasonably good, but suboptimal, valuation heuristic seems to make sense. Second, if there are costs of ranking options, it makes sense to assign a rough value to the whole set, and only bear the cognitive costs of decision if this choice set appears at the end of the experiment.

6. Conclusion

This paper has provided evidence that individuals are choice-averse, and, in seeming contradiction, have preferences over the quality of suboptimal options. The first result contradicts models of preferences over menus with an intrinsic value of choice or a preference for flexibility. The second result contradicts many models of cognitive and psychological costs (such as regret, temptation etc.) in menu evaluation. We believe that two explanations fit our data. The first is *the fear of making a bad decision*, where preferences for smaller choice sets can be seen as an optimal response by individuals who anticipate their own bounded rationality and tendency to make bad decisions in the future. The second relies on individuals using some *heuristic process to value choice sets* which only imperfectly reflects the consequences of the task. One variation of this second explanation may involve indeterminacy in individual preferences over options and a consideration of decision costs.

Our results pose numerous questions for subsequent research. The first is the need to test their generality, be it with other option types, e.g. consumption goods, lotteries or the social allocation of resources (Evren and Minardi, 2016), or under more general circumstances. Another avenue for research is the better understanding of the roots of this negative attitude toward choice, not only between the two explanations provided here, but also whether other factors may play a role in reinforcing or mitigating the corresponding processes. While the fear of making bad decisions is consistent with our data, it may be reinforced by psychological phenomena such as regret, if applied to wrong decisions, rather than to an *ex-ante* optimal decision that turned out to be suboptimal *ex-post*, due to randomness. Equally, the heuristic-valuation hypothesis may interact with both preference indeterminacy and decision costs. The relationship between our results and decision procrastination (Tversky and Shafir, 1992; Shin and Ariely, 2004) or responsibility avoidance (Dwenger et al., 2018) is largely unexplored.

It is in addition fair to say that our experimental protocol did not allow for much preference for flexibility. The basic assumption underlying the latter is that there is sufficient uncertainty over future tastes. It may indeed be that with a longer delay between decisions over menus and the final decisions, future "true" preferences may become more uncertain than in our almost simultaneous setting. A preference for flexibility may come about when the delay between the decision over menus and the decision within menus is longer. In the same way, our experiment focuses on simple choice situations with a limited number of familiar options. It is plausible that decision costs may have a larger impact in more complex situations than those observed here.

The examination of preferences over choice sets or menus overall remains an emerging area of experimental research and there is still a great deal to be learned. We believe that our experiment contributes to this research by establishing that the default attitude, in the simplest of conditions, seems to be choice-aversion. However, there are many aspects of the choice situation (delay between decisions, complexity and the type of consequences) that could substantially affect individual choice attitudes.

Last, even though our experiment shows that subjects are not willing to enlarge their *own* choice set, this does not necessarily mean that they are willing to let someone else – the government or firms – *interfere* with their choice opportunities. Individuals' evaluations of interference from others in their opportunities raises the question of freedom as independence, as defended by Berlin (1969) and MacCallum (1967).[15]

15. Some recent research has addressed issues similar to ours by evaluating individual attitudes toward delegating/keeping decision rights and authority (Ahlert and Crüger, 2004; Bartling et al., 2014; Falk and

Exploring the taste for freedom as independence is likely a promising avenue for future research.

References

Ahlert, Marlies and Arwed Crüger (2004). “Freedom to veto.” *Social Choice and Welfare*, 22, 7–16.

Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and E. Elisabeth Rutström (2006). “Elicitation using multiple price list formats.” *Experimental Economics*, 12(3), 365–366.

Arrow, K. J. (1995). “A note on freedom and flexibility.” In *Choice, Welfare and Development: A Festschrift in Honour of Amartya K. Sen*, pp. 7–15. Oxford University Press.

Bartling, Björn, Ernst Fehr, and Holger Herz (2014). “The intrinsic value of decision rights.” *Econometrica*, 82(6), 2005–2039.

Becker, G., M. DeGroot, and J. Marshak (1964). “Measuring utility by a single response sequential method.” *Behavioral Science*, 9, 226–236.

Berlin, I. (1969). *Four Essays on Liberty*. Oxford U.K.: Oxford University Press.

Bernheim, B. Douglas and A. Rangel (2004). “Addiction and cue-triggered decision processes.” *American Economic Review*, 94(5), 1558–1590.

Kosfeld, 2006; Fehr et al., 2013; Kataria et al., 2014; Neri and Rommeswinkel, 2014; Owens et al., 2014; Ferreira et al., 2017).

Brocas, Isabelle and Juan D. Carillo (2008). “The brain as a hierarchical organization.” *American Economic Review*, 98(4), 1312–1346.

Chernev, Alexander, Ulf Böckenholt, and Joseph Goodman (2015). “Choice overload: A conceptual review and meta-analysis.” *Journal of Consumer Psychology*, 25(2), 333–358.

Danan, Eric, Ani Guerdjikova, and Alexander Zimper (2012). “Indecisiveness aversion and preference for commitment.” *Theory and Decision*, 72(1), 1–13.

Dean, M., Ö. Kibris, and Y. Masatlioglu (2016). “Limited attention and status quo bias.” *Journal of Economic Theory*, forthcoming.

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini (2009). “Temptation-driven preferences.” *Review of Economic Studies*, 76, 937–971.

Dwenger, Nadja, Dorothea Kuebler, and Georg Weizsaecker (2018). “Flipping a coin: Evidence from university applications.” *Journal of Public Economics*, 167, 240–250.

Ergin, Haluk and Todd Sarver (2010). “A unique costly contemplation representation.” *Econometrica*, 78, 1285–1339.

Evren, Özgür and Stefania Minardi (2016). “Warm-Glow Giving and Freedom to be Selfish.” *Economic Journal*, Forthcoming.

Falk, Armin and Michael Kosfeld (2006). “The hidden costs of control.” *American Economic Review*, 96(5), 1611–1630.

Fehr, Ernst, Holger Herz, and Tom Wilkening (2013). “The Lure of Authority: Motivation and Incentive Effects of Power.” *American Economic Review*, 103(4), 1325–1359.

Ferreira, João V., Nobuyuki Hanaki, and Benoît Tarroux (2017). “On the Roots of the Intrinsic Value of Decision Rights: Evidence from France and Japan.” Working Paper CREM 2017-11.

Fischbacher, Urs (2007). “z-Tree: Zurich toolbox for ready-made economic experiments.” *Experimental Economics*, 10(2), 171–178.

Foster, James E. (2011). “Freedom, Opportunity and Wellbeing.” In *Handbook of Social Choice and Welfare*, vol. 2, edited by K. J. Arrow, A. K. Sen, and K. Suzumura, chap. 24, pp. 687–728. Elsevier.

Fudenberg, Drew and David K. Levine (2006). “A dual-self model of impulse control.” *American Economic Review*, 96(1449-1476).

Gravel, Nicolas (2009). “Freedom.” In *Handbook of Economics and Ethics*, edited by J. Peil and I. Van Staveren, pp. 166–174. Edward Edgar Publishing.

Gul, Faruk and Wolfgang Pesendorfer (2001). “Temptation and self-control.” *Econometrica*, 69(6), 1403–1435.

Hayek, Friedrich (1960). *The Constitution of Liberty*. Routledge, London.

Iyengar, Sheena S. and Emir Kamenica (2010). “Choice proliferation, simplicity seeking, and asset allocation.” *Journal of Public Economics*, 94(7-8), 530–539.

Iyengar, Sheena S. and Mark R. Lepper (2000). “When choice is demotivating: can one desire too much of a good thing?” *Journal of Personality and Social Psychology*, 79(6), 995–1006.

Jones, Peter and Robert Sugden (1982). “Evaluating choices.” *International Journal of Law and Economics*, 2(1), 47–75.

Kataria, Mitesh, M Vittoria Levati, and Matthias Uhl (2014). “Paternalism with hindsight: do protégés react consequentialistically to paternalism?” *Social Choice and Welfare*, 43(3), 731–746.

Kreps, David M. (1979). “A representation theorem for ’preference for flexibility’.” *Econometrica*, 47(3), 565–577.

Loewenstein, George (1999). “Is more choice always better?” *Social Security Brief: National Academy of Social Insurance*, 7.

Loomes, Graham and Robert Sugden (1982). “Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty.” *Economic Journal*, 92(369), 805–824.

MacCallum, Gerald C. (1967). “Negative and positive freedom.” *The Philosophical Review*, 76.

Marx, Karl and Friedrich Engels (1947[1845]). *The German Ideology*. New York: International Publishers.

Mill, John Stuart (1991[1859]). *On Liberty*. Oxford: Oxford University Press.

Neri, Claudia and Henrick Rommeswinkel (2014). “Freedom, power and interference: An experiment on decision rights.” Mimeo.

Ortoleva, Pietro (2013). “The Price of Flexibility: Towards a Theory of Thinking Aversion.” *Journal of Economic Theory*, 148(3), 903–934.

Owens, Dave, Zachary Grossman, and Ryan Fackler (2014). “The control premium: A preference for payoff autonomy.” *American Economic Journal: Microeconomics*, Forthcoming.

Ozdenoren, Emre, Stephen W. Salant, and Dan Silverman (2012). “Willpower and the optimal control of visceral urges.” *Journal of the European Economic Association*, 10(2), 342–368.

Pattanaik, Prasanta K. and Yongsheng Xu (1990). “On ranking opportunity sets in terms of freedom of choice.” *Recherches Economiques de Louvain*, 56(3-4), 383–390.

Pejsachowicz, Leonardo and Severine Toussaert (2017). “Choice deferral, indecisiveness and preference for flexibility.” *Journal of Economic Theory*, 170, 417–425.

Reutskaja, Elena, Rosemarie Nagel, Colin F. Camerer, and Antonio Rangel (2011). “Search dynamics in consumer choice under time pressure: An eye-tracking study.” *American Economic Review*, 101, 906–926.

Roese, Neal J. (1997). “Counterfactual thinking.” *Psychological Bulletin*, 121(1), 133–148.

Salgado, Maria (2006). “Choosing to have less choice.” Working Paper Fondazione Eni Enrico Mattei 37-2006.

Sarver, Todd (2008). “Anticipating regret: why fewer options may be better.” *Econometrica*, 76(2), 263–305.

Scheibehenne, Benjamin, Rainer Greifeneder, and Peter M. Todd (2010). “Can there ever be too many options? A Meta-analytic Review of Choice Overload.” *Journal of Consumer Research*, 37(3), 409–425.

Sen, Amartya K. (1988). "Freedom of Choice: concept and content." *European Economic Review*, 32(2-3), 269–94.

Shefrin, Hersh M. and Richard H. Thaler (1988). "The behavioral life-cycle hypothesis." *Economic Inquiry*, 26, 609–643.

Shin, Jiwoong and Dan Ariely (2004). "Keeping Doors Open: The Effect of Unavailability on Incentives to Keep Options Viable." *Management Science*, 50(5), 575–586.

Sugden, Robert (2003). "Opportunity as a Space for Individuality: Its Value and the Impossibility of Measuring it." *Ethics*, 113, 783–809.

Sugden, Robert (2007). "The value of opportunities over time when preferences are unstable." *Social Choice and Welfare*, 29, 665–682.

Toussaert, Séverine (2018). "Eliciting temptation and self-control through menu choices: a lab experiment." *Econometrica*, 86(3), 859–889.

Tversky, Amos and Eldar Shafir (1992). "Choice under conflict: the dynamics of deferred decision." *Psychological Science*, 3(6), 358–361.

Wakker, Peter P (2010). *Prospect theory: For risk and ambiguity*. Cambridge University Press.

DOCUMENT DE TRAVAIL 2: Fabrice Le Lec, Marianne Lumeau et Benoît Tarroux (2016), “Preference change under choice overload”, version actualisée de “Choice or information overload?”, CREM Working paper 2016-07.

Preferences change under choice overload

Fabrice Le Lec[1], Marianne Lumeau[2] and Benoît Tarroux [3]

April 2019

Abstract

This paper aims to test how the profusion of choice directly affects the nature of individuals' final decisions and their preferences over options. To do so, we run an experiment where subjects have to choose between familiar (*i.e.*, easy, salient and relatively safe) and unfamiliar options under different choice contexts (Large or Small choice sets). The way the frequency with which familiar items are preferred to others varies with the number of options is a test for the *choice* overload effect in final decisions. Our experimental results show that subjects choose familiar items more frequently in larger choice sets, suggesting indeed choice overload has an effect on the nature of the final decision made. We also experimentally test whether this effect can be explained by the complexity or information abundance that naturally rise in larger choice sets and conclude that choice overload cannot be reduced to an effect of information abundance.

Keywords: choice overload; over-choice; preference; decision; bounded rationality.
JEL Code: C91, D01, D12.

Aknowledgements: The authors are grateful to Elven Priour for programming the experiment. The authors also thank Dorothea Kuebler, Marie Claire Villeval and seminar participants at the CREM (Rennes), the CEPN (Paris 13), the GATE (Lyon), the ASFEE conference (Besancon, France), the ACEI conference (Montreal), the ESA European Meeting (Prague) and the OSE-PSE meeting for helpful comments. Financial support from the MSHB and the LabEx ICCA is gratefully acknowledged. This paper is part of the research project ValFree (*The Value of Choice*, Grant No. ANR-16-CE41-0002-01) of the French National Agency for Research, whose financial support is gratefully acknowledged.

[1] Catholic University of Lille, Anthropo-Lab, Ethics EA 7446, 60-bis rue du Port, 59000 Lille, France, fabrice.lelec@univ-catholille.fr

[2] University of Angers, GRANEM and LabEx ICCA, 13 allée F. Mitterrand, BP 13633, 49000 Angers, France, marianne.lumeau@univ-angers.fr

[3] Univ Lyon, Université Lumière Lyon 2, GATE UMR 5824, 69130 Ecully, France, tarroux@gate.cnrs.fr, phone number: +33472866025 (*Corresponding author*).

1 Introduction

The proliferation of opportunities is a striking feature of modern economies. More than ever, consumers and investors face a huge number of options. How this choice proliferation affects people's decision-making and well-being is then a critical question. The standard microeconomic approach is that the more choice the agents have the better their decisions: agents provided with more options are more likely to find one that corresponds to their preferences. However, a number of contributions have shown that more choice may produce negative outcomes. First, with respect to *satisfaction*, there is growing evidence that the individual satisfaction from choosing from a larger menu can be lower than that from a smaller menu (Chernev, 2003). As argued by Reutskaja and Hogarth (2009), this seems to concern both post-decision satisfaction ("*outcome satisfaction*") and the satisfaction from the act of choosing ("*process satisfaction*"). Recent neurological evidence (Reutskaja et al., 2018) seems to confirm that the subjective value of a set falls in the number of options, once the latter exceeds some threshold. The second negative empirical effect of overchoice is of individual *decision-avoidance* or deferral (Iyengar and Lepper, 2000; Shin and Ariely, 2004; Tversky and Shafir, 1992; Boatwright and Nunes, 2001; Bertrand et al., 2010; Dean et al., 2017; Shah and Wolford, 2007; Beshears et al., 2013): when individuals face too-large choice sets , they put off the apparently difficult task of making a choice.[1] Numerous theoretical explanations have been proposed for these two dimensions of *choice overload*: regret (whether anticipated or not) (Irons and Hepburn, 2007; Sarver, 2008; Buturak and Evren, 2017), limited attention (Dean et al., 2017), the information borne by the composition of the choice set (Kamenica, 2008) and the complexity of the task (Gerasimou, 2017).

Our main contribution here is to analyze the effect of choice-set size on another dimension of decision-making: agents' preferences and final decisions. The consequences of the number of options on satisfaction and decision deferral do not imply that the decisions made under larger choice differ from those with smaller choice sets: lower satisfaction or a tendency to avoid decision-making can come about with the same option being chosen under the different conditions. In the language of economics, the preferences that emerge from a given choice condition are not necessarily affected

[1] Neither of these two phenomena is unambiguous (see the meta-analyses in Scheibehenne et al. (2010a) and Chernev et al. (2015), who come to rather opposite conclusions).

by the size of the choice set, even though this might delay decision-making. We here investigate whether overchoice affects individuals' preferences over options. When a decision-maker faces a large choice set, does she decide differently - or construct different preferences - than when facing a smaller choice set? Does she choose a different option as a function of choice-set size? If the size of the choice set does affect preferences, we should observe preference-reversals, or inconsistencies with respect to the standard revealed-preference properties. Inconsistent behavior has been amply-documented via *preference reversals* or *menu effects* (Huber et al., 1982; Simonson, 1989). However, little, if anything, is known about the effect of choice-set size on preference changes.

Generally speaking, a decision-maker appears inconsistent - or choice-overloaded - if her preference ranking over two items changes with the choice set size: an item is preferred to another one in a small choice situation and the reverse in a large choice situation. Following this general definition, we experimentally investigate whether increasing the number of available options produces inconsistent preferences for two types of options: familiar and unfamilar ones. Familar options are *easy, salient and relatively-safe* options that the individuals have already experienced and consumed. The rationale for the use of these options is that individuals have clear views and beliefs about the satisfaction they can expect from their consumption (*safe*), without having to exert much cognitive effort (*easy*), and their recognition by decision-makers is likely to make them worthy of attention (*salient*). Our measure of the effect of overchoice on preferences is hence based on the relative choice of these familiar options over other alternatives, and the overall inconsistencies in elicited preferences that choice overload may generate.[2]

The second contribution of this paper is to investigate the source of this effect, and in particular whether it stems from choice per se (pure choice overload) or the complexity and abundance of the information to be analyzed (cognitive or information overload), which naturally rises as the choice set expands. A great deal of empirical work has investigated the role of complexity and information abundance (Fasolo et al., 2009; Beshears et al., 2013; Persson, 2017), but the two (choice and information overload) may have different psychological origins, and whether choice

[2]Differently speaking, our definition of choice-overload might be interpreted in terms of exploring the choice set and one's preferences. A decision-maker is overloaded if he is more likely to neglect unfamiliar alternatives and to be reluctant to change his own preferences or tastes, that is, to explore new options, as the choice set size is expanded.

overload purely results from information (or complexity) overload remains an open question. It may be that individuals decide differently as the choice set expands because of attention filters, for instance, whether or not the situation has become more complex or the informational burden heavier.

To establish the impact of choice overload on elicited preferences and investigate the source of this effect, we run a lab experiment. The experimental method is particularly suitable for analyzing individual decision-making in contexts that differ in terms of choice and information sets. Our experiment places subjects in a relatively simple choice condition, to which subjects are accustomed. It proceeds as follows. Subjects are first presented with a choice set of experience goods that they may consume in their everyday life (websites).[3] They then have the opportunity to collect information about each available option. After the collection of information, subjects have to rank the goods according to their preferences using an incentive-compatible mechanism based on a 'real consumption' period. Our experimental manipulation consists in varying, between subjects, the size of the choice set (*Small*, with four options, or *Large*, with eight options) and the number of pieces of information subjects can collect (*No Info*, *Low Info* or *High Info*). All choice sets are constructed so that there is little variation, if any, in the proportion of familiar and unfamiliar options at the individual level. Varying the level of information is thought of both as a robustness check and as a test of the hypothesis that the effects of choice overload purely reflect cognitive or information overload (as suggested by Scheibehenne et al. 2010b and Fasolo et al. 2009).

Our results show a clear effect of choice-set size: independently of the level of information, subjects choose familiar products more frequently in large rather than small choice sets. This effect appears large and robust enough to suggest that choice overload affects dimensions other than satisfaction and decision deferral (Chernev, 2003; Iyengar and Lepper, 2000): here the final decision and the nature of individuals revealed-preference orderings. Showing that the type of options chosen by individuals differs with the size of the choice set, our results are related to Iyengar and Kamenica (2010).[4] Howerver we obtain this effect under an ordinary decision with

[3] Some recent work in experimental economics use 'real' goods to study decision-making, like snack items (Reutskaja et al., 2011) or access to websites (Le Lec and Tarroux, 2019).

[4] Other research, mostly based on the idea of a status-quo bias, can also be interpreted as examining this question (Samuelson and Zeckhauser, 1988; Kempf and Ruenzi, 2006; Dean et al., 2017). But we rather consider, in line with

common options (websites) whereas they provide evidence that individuals tend to choose simpler, or easier to understand, options as the choice set expands with types of options in which individuals have very little experience (lotteries and investment plans). We also show that the whole preference is affected, and not just the chosen option. The second result of this paper is that we observe only a limited effect of the quantity of information: familiar goods are chosen only mildly more often in the high- rather than low-information treatment. This suggests that the effect of choice overload on preferences cannot be reduced to only cognitive or information overload.

The remainder of the paper is organized as follows. Section 2 develops our measure of choice overload based on preference reversals, and Section 3 describes the experiment. Section 4 presents our main experimental findings, while Section 5 discusses these findings and concludes.

2 A measure of choice overload based on preference reversals

This section sets out a theoretical framework allowing for choice-overload effects on preferences. Let $\mathcal{X}$ be a (finite) set of alternatives. This set includes two kinds of items: *familiar* and *unfamiliar* objects. Denote by $\mathcal{F}$ and $\mathcal{N}$ the disjoint sets of familiar and unfamiliar items respectively, and $\mathcal{X} = \mathcal{F} \cup \mathcal{N}$. An item is said to be familiar if the consumer has already experienced it. As a consequence, she has accurate beliefs about the satisfaction from its consumption, so that this option is relatively safe. This belief results without much effort in terms of cognition, information-gathering or processing, and is hence an easy option. Last, this can be thought of as salient, as it is recognized spontaneously by subjects, and so immediately attracts attention.

The measure of choice overload is based on the relative attraction of the familiar and unfamiliar items (*i.e.* the change in preferences) as the choice set expands. We consider a decision maker who can face two choice sets, $C^- \in \mathcal{X}$ and $C^+ \in \mathcal{X}$ such that C^- is a strict subset of C^+. We denote $C^- = F^- \cup N^-$ and $C^+ = F^+ \cup N^+$, where F^- and F^+ are the subsets of familiar items within C^- and C^+ ($F^- \subset F^+ \in \mathcal{F}$), and N^- and N^+ are the subsets of unfamiliar items in C^- and C^+ ($N^- \subset N^+ \in \mathcal{N}$).

Dean et al.'s own interpretation, their results as additional evidence that individuals avoid choice by retaining the default/status-quo option, rather than changing their preferences.

The preference ordering over two (or more) items can be choice-set dependent to account for the effect of choice overload. Denote by $\succsim_C$ the preference ordering over alternatives in set C, and $\succ_C$ and $\sim_C$ its symmetric and asymmetric parts. For a set C and two (familiar or unfamiliar) items a and b in C, $a \succsim_C b$ means "a is weakly preferred to b when the consumer faces C". We focus on the profile of decision-maker preferences $(\succsim_{C^-}, \succsim_{C^+})$.

In the absence of a choice-overload effect on preferences, the decision-maker's preferences will satisfy the usual consistency requirement:

Property 1 (Consistency) *Consider two items a and b with $a, b \in C^- \subset C^+$. The decision maker's preferences are such that: $a \succ_{C^-} b \Leftrightarrow a \succ_{C^+} b$.*[5]

A rational decision-maker who prefers a to b in the small set will also prefer a to b in the large set, and vice versa. This holds irrespective of the nature of the items (familiar or unfamiliar). In other words, preferences over a and b are not choice-set dependent. This can be seen as a generalization of the preference ordering of the usual revealed-preference axioms.

A choice-overload effect on preferences implies by definition that the decision-maker's behavior is sometimes inconsistent across choice sets. To capture this idea, Dean et al. (2017) propose the axiom of contraction to two categories of options, the status quo and the other options. In our context, this axiom is adapted to mean that preferring an unfamiliar to a familiar item in a large set implies the same preference in a smaller set. Equivalently, if the decision-maker (weakly) prefers a familiar to an unfamiliar item in a small set, then she will have the same preference in a large set. The formal version of this property can be written as follows:

Property 2 (Contraction) *Consider two items $n \in N^- \subset N^+$ and $f \in F^- \subset F^+$. The decision-maker's preferences are such that: $n \succ_{C^+} f \Rightarrow n \succ_{C^-} f$.*

It is clear that the Consistency axiom implies Contraction but the contrary does not hold. It thus may be the case that an unfamiliar item is preferred to a familiar one in a small set, but not in a large set. This reflects a change in preferences with choice-set size.

[5] This property is easily generalized by considering the case where a and b belong to $\widetilde{C} \cap \widehat{C}$, where the former set is neither a subset nor a superset of the latter. However, we think that the text version is more intuitive, given our experimental design where subjects face one of two choice sets, one being a subset of the other.

To distinguish preference changes due to choice overload from other types of changes linked to well-documented menu effects (such as the decoy or compromise effects, for instance), we require, even for the choice-overloaded decision-maker, some consistency within the categories of items:

Property 3 (Within-Category Consistency) *Consider four items n', n'', f' and f'' such that $n', n'' \in N^- \subset C^- \subset C^+$ and $f', f'' \in F^- \subset C^- \subset C^+$. The decision-maker's preferences are such that: $n' \succ_{C^-} n'' \Leftrightarrow n' \succ_{C^+} n''$ and $f' \succ_{C^-} f'' \Leftrightarrow f' \succ_{C^+} f''$.*[6]

This property means that consistency holds, but only within a given category (familiar or unfamiliar items). Clearly, consistency implies within-category consistency.

These simple properties allow us distinguish between two types of decision-makers: rational and overloaded.

Definition 1 (Rational decision-maker)

A decision-maker is said to be rational if her preference orderings satisfy Consistency.

Definition 2 (Overloaded decision-maker)

A decision-maker is said to be prone to be choice overload if her preference orderings satisfy Contraction and Within-category consistency, but not Consistency.

Rational decision-makers (*i.e.*, when Consistency is satisfied) have preferences over options in a given choice set that are the stable restriction of a general preference ordering. Overloaded decision-makers (*i.e.* when Contraction and Within-Category Consistency hold, but not Consistency) can have preference reversals of the type $n \succ_{C^-} f$ and $f \succ_{C^+} n$: the reversal can only be in the sense of unfamiliar items being less well-ranked than familiar items in the large set. With overloaded decision-makers there can be some inconsistency (preferring f to n in large sets but the opposite in small sets), although we still require some general level of consistency. We consider any inconsistencies that are different from those described above as not reflecting choice overload.

[6]Again, this property can easily be generalized by considering the case where a and b belong to $\widetilde{C} \cap \widehat{C}$, where the former set is neither a subset nor a superset of the latter.

The behavioral implications are as follows. From these definitions, choice-overloaded individuals should rank familiar items better in larger than smaller choice sets. We test this hypothesis in lab experiments, and evaluate the probability that a familiar be preferred to an unfamiliar item as a function of choice-set size. A positive correlation would reveal that a significant percentage of subjects are choice-overloaded.

3 Experimental design and procedures

3.1 Experimental design

The experiment consists of six treatments in which subjects choose and consume an experience good in the lab, after obtaining pieces of information about the alternatives. The goods proposed are content websites.[7] For a participant, the experimental stake is the consumption of the preferred website for half an hour during the consumption period. We chose this type of good for a number of reasons: first, the typical subject is very experienced in online-content consumption decisions, as for most of them this is a daily activity;[8] second, the consumption of these goods is easy to implement and control in the lab. Moreover, it is easy to pick choice sets so that some options are very likely familiar to most if not all subjects.

(0)		(1)		(2)		(3)		(4)
Select a type of content	⇒	Choice set (websites)	⇒	Information (or not)	⇒	Rank items	⇒	Consumption (30 minutes)

Table 1: The stages of the experiment

We wish to test for choice overload, and see whether this is related to the level of information provided in the decision situation or whether it is rather a *pure choice-overload effect*. To do so, we divide the experiment into five stages, as summarized in Table 1.[9] We manipulate two of the experimental variables: the size of the choice set subjects face (Stage 1) and the number of pieces of information to be collected by the subjects (Stage 2). Table 2 summarizes the parameters in our different treatments.

[7]These websites offer their own content, which excludes e-commerce websites and social networks.

[8]We checked this in the post-experiment socio-demographic questionnaire: our subjects spend an average of 1 hour 49 minutes on the web per day.

[9]The instructions (translated from the French) can be found in the Appendix.

During the *preliminary stage* (0), subjects indicate the types of content in which they are interested. The available types were general news, sports, culture and pop culture, economics, games, movies and TV series, and cooking. In addition to making sure that subjects were familiar with some of the websites in the choice set, this stage also guarantees a sufficient level of subject motivation when evaluating the options.

In *Stage 1*, subjects are presented with the choice set they will face for the whole experiment. The options are labeled with their web addresses (URL), (*e.g.*, www.nytimes.com) to provide a minimal level of information. The number of alternatives differs across treatments (Table 2): four options in the *Small choice set* (C^-) and eight options in the *Large choice set* (C^+). We chose these numbers as they are relatively small, so that the situation is not *a priori* over-complicated. It has been established in cognitive psychology that the number of "objects" typical subjects can cognitively handle in their working memory is below seven (Farrington, 2011), so we expect overload to start at seven options. To consider the choice of familiar versus unfamiliar items, all choice sets are designed to have, as far as possible, similar proportions of familiar and unfamiliar websites at the individual level: each choice set was composed of fixed shares of "blockbuster", *i.e.* very famous websites, mildly famous ones and niche ones. We categorize websites according to the number of 'likes' on the local version of Facebook. The websites with a high number of 'likes' (blockbuster websites) are more likely to be familiar to subjects, while those with far fewer 'likes' (such as blogs) are likely unknown.[10] The subjects' familiarity with websites was also measured *ex post* at the individual level via a questionnaire. Subjects also say *ex post* whether they had heard of each website before the experiment in order to control for external information.[11]

In *Stage 2*, subjects collect pieces of information. There are three types of information for each website: a neutral description, a popularity index (the number of 'likes' on facebook.fr) and a user

[10]The average number of 'likes' of the most popular websites in a choice set is around 346,000 (C^+) and 313,000 (C^-), with a certain heterogeneity from 117,000 (Economics) to 700,000 (News). The least popular website has around 3,000 (C^+) and 5,000 (C^-) 'likes'.

[11]Constructing choice sets *a priori* with blockbuster/reasonably famous/niche criteria – complemented by *ex post* questions about familiarity with options – appeared more appropriate than straight *ex ante* questions about the subjects' familiarity with websites, which may lead to less incentives for truthful revelation due to chained decisions, and place subjects in a much larger choice-set context. The results in sub-section 4.1 provide descriptive statistics showing that most subjects knew at least some of these websites but very rarely all of them.

	Small Choice Set (C^-)	Large Choice Set (C^+)
No Information (I^0) (0% of info.)	C^-/I^0 4 options 0 info. piece	C^+/I^0 8 options 0 info. piece
Low Information (I^-) (25% of info.)	C^-/I^- 4 options 3 info. pieces	C^+/I^- 8 options 6 info. pieces
High Information (I^+) (50% of info.)	C^-/I^+ 4 options 6 info. pieces	C^+/I^+ 8 options 12 info. pieces

Table 2: Treatment matrix

comment.[12] The total number of available pieces of information depends on the size of the choice set (there are 12 pieces of information for four alternatives in C^-, and 24 pieces of information for eight alternatives in C^+), but subjects' information access is constrained according to the treatment. There are three information types (Table 2). In *Low Information* treatments (I^-), subjects select and consult 25% of the available pieces of information, whereas the analogous figure in the *High Information* treatment (I^+) is 50%. In both treatments, subjects have to consult the required number of pieces of information to guarantee their exposure to a given amount of information, but are free to choose which pieces of information they consult. In the *No Information* treatment (I^0), the second stage is not implemented and the only information is the web address of the choices. Strictly speaking, I^0 is more a minimal-information than a no-information treatment: the website URL may be informative about the content, and subjects may come to the lab knowing some of the websites (without necessarily having visited them). As noted above, subjects were asked at the end of the experiment whether they had previously heard of each website, in order to measure this initial information. From a methodological point of view, I^0 mostly serves as a control treatment to check that subjects are not motivated by other aspects of the experimental situation, for instance, curiosity about new websites linked to the specificity of the lab context. More generally, these different information treatments aimed to test the robustness of choice overload, and in particular to see if choice overload can be reduced to information overload.

[12]The comments were drafted as follows: 'I recommend this website/I do not recommend this website' plus an explanation. To control for the nature of the recommendations, 50% of the websites in a category (blockbuster, reasonably famous and niche) had positive comments and 50% negative comments. Moreover, the positive and negative recommendations for a given website were reversed across sessions to produce a balanced design.

The experimental variation in information reflects a *constant relative number of pieces of information*: subjects consult the same number of pieces of information *per website*. For a given choice-set size, the amount of information in I^+ is obviously greater than that in I^-. However, as choice-set size changes, it is less clear what more information means. Consider, for example, the C^- (four options) and C^+ (eight options) treatments: giving 6 pieces of information in C^-/I^+ and 12 in C^+/I^+ guarantees constant information per option, but there is arguably more information to be processed in the second case. It could hence be that any difference between C^-/I^+ and C^+/I^+ reflects the absolute level of information, and choice overload may in fact be information overload. Two features of our design allow us to test (and possibly rule out) this possible confusion: first, C^+/I^- can be compared to C^-/I^+, for which the *absolute level of information* is constant (at 6), while choice-set size varies; second, if the absolute level of information matters, we should observe differences between C^-/I^- and C^-/I^+ on the one hand and C^+/I^- and C^+/I^+ on the other. As a result, it seems that maintaining the relative level of information, rather than the absolute level, is a better way of identifying choice and information overload.

During *Stage 3*, subjects have to rank the different alternatives according to their preferences. To induce the revelation of true preferences, we design a real-consumption incentive mechanism: Subjects know that they will have to spend half an hour in the lab with access to one website as their only source of entertainment. This website is chosen randomly, but reflects their ranking: The probability of having the first-ranked website is higher than that of the second-ranked website, and so on. Table 3 presents the probability that a website be drawn by the computer program depending on the subject's ranking. This procedure ensures that revealing the true ranking is a stochastically-dominant strategy, and hence the procedure is incentive-compatible. Alternative measures of individual preferences are more problematic regarding our research question. In particular, asking subjects a series of binary questions has major drawbacks, as it may destroy the possible choice-set size effect. The situation under consideration would no longer be a relatively large choice set but simply two items to be compared.

The *last stage (Stage 4)* is the consumption stage. Subjects are informed of the website selected and then consult it, and no other website, for 30 minutes. The Internet access of the subject's

Rank	Small choice set (C^-)	Large choice set (C^+)
1	50%	45%
2	30%	25%
3	15%	15%
4	5%	7%
5	-	5%
6	-	2%
7	-	1%
8	-	0%

Table 3: The probability of drawing options depending on their rank

computer was locked to this particular website and subjects were not authorized to use their phone or any other type of documents. This 'real' consumption stage was introduced to reinforce the incentives associated with the ranking task, and to make subjects carefully consider the options and information available. This procedure is similar to that used in recent work in experimental economics (Reutskaja et al., 2011; Le Lec and Tarroux, 2019).

3.2 Procedures

The experiment was carried out at the Center for Research in Economics and Management (CREM), University Rennes 1, France. The experiment was computerized using the z-Tree program (Fischbacher, 2007) and consisted of 21 sessions, as summarized in Table 4. Given the specific nature of the experiment, we used a between-subject design. In total, 341 subjects (46.6% female) were recruited from a population of undergraduate students in a variety of majors.

The payoff was a fixed show-up fee of 12 Euros. The show-up fee was unusually high to account for the fact that the main experiment does not offer subjects the possibility to earn any additional money: incentive-compatibility is ensured by the fact that the subjects' choices will be implemented in the lab, where subjects will spend 30 minutes with an option. However, subjects could in addition earn some money in a side experiment, which took place at the end of the consumption stage, by completing a Holt and Laury (2002) task. We added this side experiment in order to measure subjects' attitudes towards risk, which may play a role in the probability of choosing a familiar (rather safe) option over a more risky unfamiliar one. The average total payoff was 15 Euros. After

Treatment	Choice-set size (C)	Information-set size (I)	Number of sessions	Number of subjects
C^+/I^+	+	+	5	66
C^+/I^-	+	-	4	60
C^+/I^0	+	0	2	47
C^-/I^+	-	+	4	60
C^-/I^-	-	-	4	60
C^-/I^0	-	0	2	48

Table 4: Characteristics of treatments

the main and side experiments, subjects completed a post-experimental questionnaire on their socio-demographic situation and their familiarity with the websites in the choice set presented (whether they had visited them before the experiment) and their previous knowledge of them (whether they had heard of them before the experiment). On average, the sessions lasted 90 minutes, including the initial instructions, lab consumption, the side experiment, the post-experimental questionnaire and subject payment.

4 Results

We first present a general description of subjects' initial familiarity with the websites they face, before considering their preferences for familiar items conditional on the size of the choice set, as well as a possible effect of the quantity of information.

4.1 Subjects' familiarity with items

Before looking at the effect of the context of choice (choice-set size and information) on preferences for familiar items, we control for subjects' initial familiarity with the websites that they face. As noted above, we asked subjects to report, for each website, whether they were familiar with it, that is whether they had already visited the website before the experiment.

On average, subjects were familiar with 31.9% of the websites, so that 68.1% of the websites were unfamiliar. Figure 1 presents the mean proportion of familiar items, which is fairly stable across treatments. The usual pairwise tests do not suggest any significant difference between the treatments in this respect, so that our *a priori* selection of the choice sets (based on the

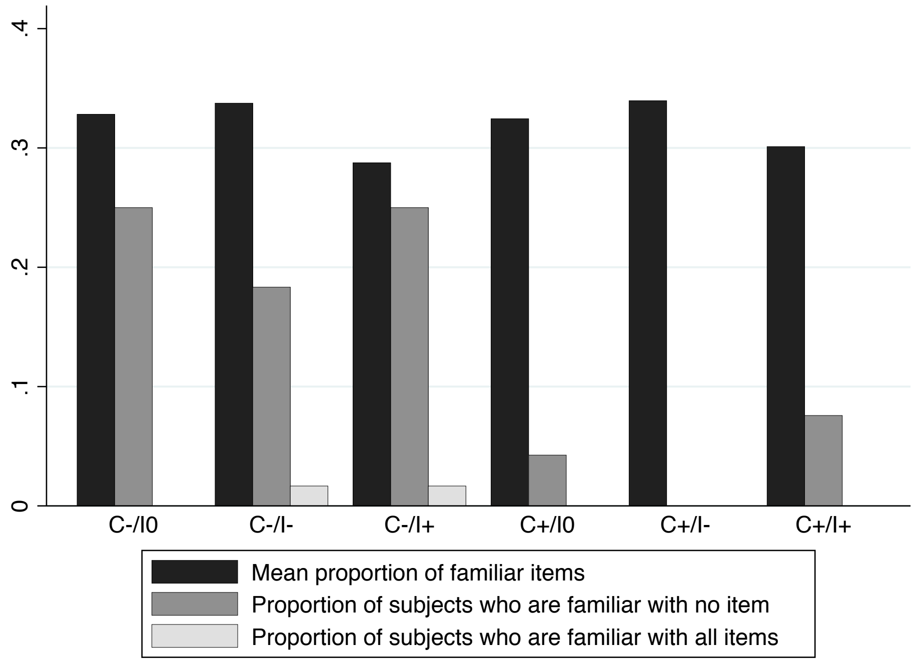


Figure 1: The difference in item familiarity between treatments

blockbuster/middle/niche criteria) does not seem to produce any major sampling problems. Mechanically, the proportion of subjects who are familiar with no items is significantly larger in the treatments with smaller rather than larger choice sets. Last, the number of subjects who are familiar with all of the items in the choice set is only very small.

4.2 Subjects' preferences for familiar items

We now consider how the choice context (choice-set size and information) affects the frequency with which familiar are preferred to unfamiliar items. We use the rank data to determine whether one option is preferred to another. We first rely on the simple proportion of this preference for familiar items, and then analyze the full rankings econometrically via a rank-ordered logit estimation.

4.2.1 Descriptive measures

Our analysis of whether a familiar option is preferred to another only considers the comparisons where both are present. For instance, for an individual who reports being familiar with one option of the four available, we have three valid comparisons. Subjects who are familiar with all of the

Choice set	No info (I^0)	Low info (I^-)	High info (I^+)	All
Small choice set (C^-)	88.6	65.1	69.3	73.3
	(131; 36)	(169; 48)	(153; 44)	(453; 128)
Large choice set (C^+)	93.2	72.6	82.4	81.6
(unrestricted)	(538; 45)	(741; 60)	(733; 61)	(2012; 166)
Large choice set (C^+)	93.6	72.0	75.6	79.0
(restricted)	(125; 37)	(168; 49)	(184; 54)	(477; 140)
All (unrestricted)	92.2	71.2	80.1	80.1
	(669; 81)	(910; 108)	(886; 105)	(2465; 294)
All (restricted)	91.0	68.6	72.7	76.2
	(256; 73)	(337; 97)	(337; 98)	(930; 268)

Note: The sample size appears in parentheses: the first figure is the number of valid comparisons and the second the number of subjects.

Table 5: The average frequencies with which a familiar is preferred to an unfamiliar item (in %)

options or none of them are automatically dropped from this analysis. Table 5 shows the frequencies with which a familiar is preferred to an unfamiliar option. We consider two sets of frequencies in the large choice-set condition: the proportion of cases where the familiar item is preferred over all possible comparisons (unrestricted), and the same proportion but restricted to the subset of options that are common to the Large and Small choice-set conditions. For the analysis we focus on the unrestricted case, as this provides more observations.

These descriptive measures of preferences for familiar options indicate that, on average, subjects prefer familiar to unfamiliar items. However, these preferences depend on information and choice-set size. For our main treatment (the effect of the choice set), the probability of choosing a familiar item rises with choice-set size, and this holds in each information condition. Pooling all the information treatments, this difference in proportions is significant with a cluster-adjusted χ^2 test (Donner and Klar, 1994) that takes into account the non-independence of preferences for familiar items at the individual level ($\chi^2 = 7.3$, $p = 0.007$). This, in line with our hypothesis, is consistent with choice overload. It moreover seems that the difference in the probability of choosing a familiar item rises with information: while the difference is 4.6 ppc in the absence of information, it is 7.5 and 13.1 in I^- and I^+ respectively. These differences represent 20 to 40 percent of the possible changes (given that, in the majority of cases, familiar items are preferred), with an average of 31% of possible

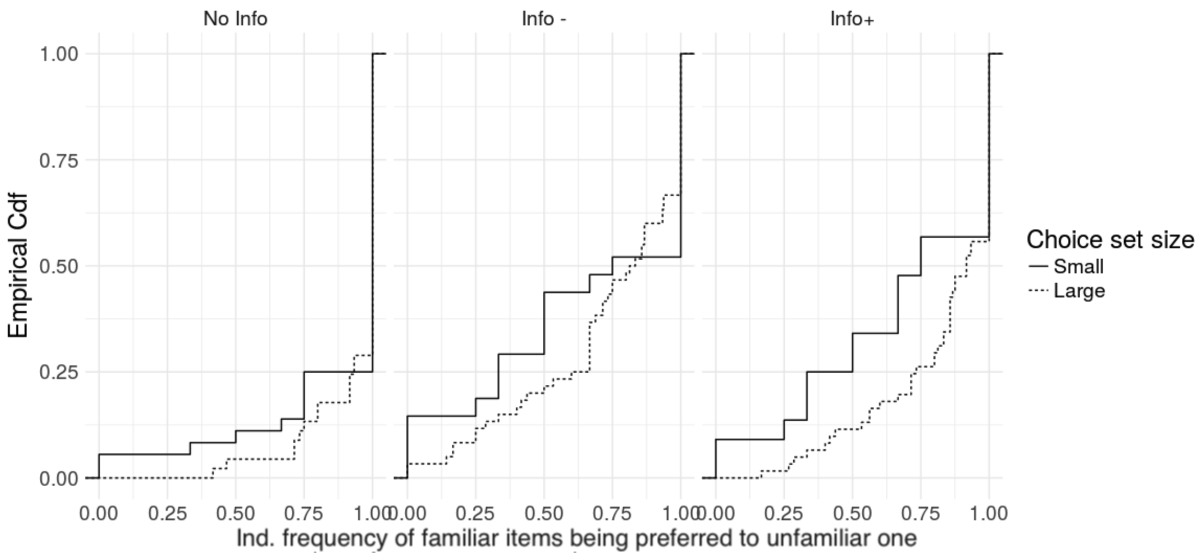


Figure 2: The empirical cumulative distribution functions of the individual frequencies of a familiar being preferred to an unfamiliar option.

changes in favor of familiar items.

We also consider the frequencies with which subjects prefer familiar to unfamiliar options. The corresponding empirical cumulative distribution functions are depicted in Figure 2. The conclusions drawn from the pooled frequencies seem to be valid at the individual level, where for all levels of information the distributions of the individual frequencies in which a familiar is preferred to an unfamiliar option in C^- (almost fully) dominate the distributions for (unrestricted) C^+.

Regarding information, Table 5 indicates that, in the absence of information, subjects overwhelmingly prefer familiar to unfamiliar items (in around 90% of cases). Table 5 also indicates that providing some information rather than none seems to strongly affect the probability of choosing a familiar option: the difference in the probability of choosing a familiar item is lower in the I^+ and I^- treatments compared to I^0, and strongly significantly so when using the cluster-adjusted χ^2 mentioned above ($\chi^2 = 22.4$, $p < 0.001$).[13] In addition, the probability of choosing a familiar item is minimal for I^-, and may rise from I^- to I^+ (or at least remain constant), the difference being significant in a cluster-adjusted χ^2 test ($\chi^2 = 4.1$, $p = 0.043$).

[13] The difference between I^0 and I^- may be also used to test how useful the information provided is for the subjects. As we observe that subjects are less willing to choose familiar over unfamiliar items when information is provided, this information is not totally irrelevant.

Carrying out our analysis on the restricted choice set (for which the number of observations, and hence the power, is smaller), we see no major differences from the unrestricted set. It is however worth noting that the statistical significance levels of similar tests change when testing the differences in treatments, and always in the direction of lower significance. This is consistent with the difference between C^+ and C^- not stemming from the difference in composition (of the small choice set and its complement in the large choice set).

To summarize, these descriptive measures are consistent with choice overload, as familiar items are better ranked in the large choice-set treatment. The effect of information is more ambiguous: providing some information increases the probability of choosing an unfamiliar option, but the quantity of information seems to have a slightly negative effect on this probability.

4.2.2 Econometric analysis

We consider these effects in depth via a rank-ordered logistic model (Hausman and Ruud, 1987). Let i be an individual and ℓ the total number of individuals. Each option is referred to as j, and J_i is the total number of options that individual i faces. For each individual i and each option j in the choice set of i, U_{ij} is the latent score of j for i, and i prefers j to k if $U_{ij} > U_{ik}$. The latent score corresponds to the value that the subject attaches to option i. U_{ij} can be decomposed into the sum of deterministic and stochastic terms: $U_{ij} = \upsilon_{ij} + \epsilon_{ij}$. As is usual in rank-ordered logit models, the ϵ_{ij}'s are assumed to be iid with an extreme-value distribution and the deterministic term depends linearly on various variables of interest or controls (see below). The likelihood of obtaining a given rank order for individual i is given by: $L_i = \Pi_{j=1}^{J_i} \frac{e^{\upsilon_{ij}}}{\sum_{k=1}^{J_i} \delta_{ijk} e^{\upsilon_{ik}}}$, with δ_{ijk} being the indicator function for the rank of j being lower than or equal to the rank of k for individual i. The log-likelihood for the sample of ℓ individuals is then given by:

$$\log L = \sum_{i=1}^{\ell} \sum_{j=1}^{J_i} \upsilon_{ij} - \sum_{i=1}^{\ell} \sum_{j=1}^{J_i} \log \big(\sum_{k=1}^{J_i} \delta_{ijk} e^{\upsilon_{ik}} \big)$$

We wish to estimate the impact of the choice context on the effect of the familiarity of the option on the latent score U_{ij}, and hence specify the deterministic term of the score in the following way:

$$\upsilon_{ij} = \delta f_{ij} + f_{ij}[\gamma_1 t_{ij}^1 + \gamma_2 t_{ij}^2 + ... + \gamma_K t_{ij}^K] + \alpha x_{ij}$$

where f_{ij} is a dummy variable for the familiarity of option j to i, t_{ij} the variables of interest (treatments, risk attitude, gender etc.), and x_{ij} a set of option-based controls. The objective is to estimate the interaction of the variables of interest (treatments) with the familiarity of the option.

The estimation results for a number of these specifications appear in Table 6. In all the specifications, *IsFamiliar* and *IsHeardOf* are two dummy variables for the individual reporting having visited or heard of the website.[14] In the first set of specifications (columns 1 and 2), C^+ indicates the unrestricted Large Choice-set treatment, *Info* is a dummy for information being provided in the treatment (I^- or I^+), and I^+ indicates the corresponding treatment. In column 2, individual controls are added as independent variables. These include the result of the Holt and Laury measure, gender, and the type of content chosen by the individual.

The estimation results from this first set of specifications in Table 6 are in line with the simple descriptive statistics in Table 5. First, a large choice set (C^+) is always associated with a stronger preference for familiar options (i.e. it attracts a positive coefficient) and significantly so, suggesting choice overload. Second, we confirm that providing some information reduces the propensity to choose a familiar item but, above some threshold, more information increases the probability that subjects prefer familiar items.[15]

The second set of specifications (columns 3 and 4 of Table 6) tests whether our choice-overload effect reflects a combination of choice and information overload. The careful reading of Table 5 reveals that most of the effect comes about when there is both more information and a larger choice set, and in particular when moving from I^-/C^- to I^+/C^+. We thus add an interaction

[14] When subjects were asked to report whether they had already visited the website before they took part in the experiment, they also said whether they had heard of the website. In the rank-ordered logistic model, this variable *IsHeardOf* corresponds to the set of option-based controls.

[15] We obtain similar results by working on ranks or by focusing on the preferred item in the choice set. However, these measures are less relevant considering our data.

	Endogenous variable: latent score for options						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Coef./(se)	Coef./(se)	Coef./(se)	Coef./(se)	Coef./(se)	Coef./(se)	Coef./(se)
IsHeardOf	.616***	.617***	.631***	.616***	.604***	.608***	.589***
	(.078)	(.078)	(.078)	(.078)	(.096)	(.078)	(.078)
IsFamiliar	2.151***	2.220***	2.240***	2.276***	1.405***	1.593***	2.349***
	(.212)	(.389)	(.233)	(.414)	(.165)	(.359)	(.238)
IsFam. ×							
C^-	(ref)	(ref)	(ref)	(ref)	(ref)	(ref)	(ref)
C^+	.673***	.655***	.536*	.565**	.754***	.605**	.360$^+$
	(.164)	(.171)	(.209)	(.218)	(.196)	(.208)	(.212)
Info	-1.572***	-1.702***	-1.565***	-1.696***			-1.490***
	(.214)	(.224)	(.213)	(.228)			(.216)
I^+	.576***	.587***	.332	.438			
	(.167)	(.176)	(.287)	(.304)			
$C^+ \times I^+$			.356	.218			
			(.340)	(.364)			
RelativeInfo					-.040	-.263*	
					(.033)	(.121)	
TotalInfo					-.086	-.009	
					(.112)	(.035)	
Controls	No	Yes	No	Yes	No	Yes	No
N	341	341	341	341	339	339	215
LogLik.	-2133	-2107	-2132	-2107	-2151	-2124	-1320
χ^2 (df)	87 (5)	138 (14)	88 (6)	139 (15)	38 (5)	92 (14)	58 (4)
p-value	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$

Notes: This Table reports the regression results for rank-ordered logit models. IsHeardOf = 1 if the subject has heard of the website, =0 otherwise. IsFamiliar =1 if the subject is familiar with the website, =0 otherwise. $C^- = 1$ if small choice set (4 items), =0 otherwise. $C^+ = 1$ if large choice set (8 items), =0 otherwise. $Info = 1$ if information is provided, =0 otherwise. $I^+ = 1$ if high information situation, =0 otherwise. RelativeInfo = number of pieces of information/number of unfamiliar websites. TotalInfo = total number of pieces of information in the treatment. Significance levels: $^+ = p < .10$; $* = p < 0.5$; $** = p < 0.01$; $*** = p < 0.001$. To calculate the likelihood-ratio test (χ^2), the null model is that with the two variables IsHeardOf and IsFamiliar.

Table 6: Rank-ordered logit estimates

term $C^+ \times I^+$ to the first and second specifications. This has very little effect on the estimated coefficient on C^+, suggesting that the choice-set size effect is independent of information. The estimated coefficient on the interaction term is never significantly different from zero, so that our main effect is not driven by the combined effect of choice and information.

In the third set of specifications (columns 5 and 6 of Table 6), we add two new information variables: *RelativeInfo* is the ratio of the number of pieces of information to the number of unfamiliar websites,[16] and *TotalInfo* is the total number of pieces of information in the treatment. The results suggest that relative information (how well informed the subject is about a particular option) may matter, but that absolute information has no effect.

The specification in column 7 uses a subsample to test whether the choice-set size effect is robust to keeping the total amount of information constant. We do so by looking at the data from treatments with *the same absolute number of pieces of information but different choice sets*, that is conditions C^-/I^0, C^+/I^0, C^-/I^+ and C^+/I^-. In the first two conditions, there is no information so only the choice set changes, while in the last two conditions the number of pieces of information is fixed at six. We find that larger choice sets are associated with greater latent value for familiar items, suggesting again that pure choice overload is at work. The estimated effect is only weakly significant, which may partly reflect the lower statistical power in the smaller sample. It worth noting that the negative effect still holds even when subjects are not information overloaded, despite the larger choice set (at least for C^-/I^+ and C^+/I^-) corresponding to subjects being less well-informed about unfamiliar options.

Regarding the controls, their inclusion has very little effect on the estimated coefficients on our variables of interest in all specifications. There are content categories with preferences for less-familiar websites ("Cooking", in particular). Women significantly value more-familiar choices, while the Holt and Laury risk-aversion measures attract a positive coefficient (more preference for familiar options), which is not however significant.

In summary, these rank-order logit estimations confirm the conclusions from the descriptive statistics. Subjects have a (relative) preference for familiar options as the choice set expands, and the availability of information reduces preferences for familiar options although the quantity of information matters only little.

[16] As a result, two observations were not used, as the subjects were familiar with all of the websites.

5 Concluding remarks

The previous literature on choice overload has shown that choice proliferation reduces (*ex ante* and *ex post*) satisfaction and may produce decision avoidance or deferral. We here focus on another dimension of choice sets: their effect on individual preferences and final decisions. To do so, we define a new experimental measure of choice overload as the relative preference for familiar over unfamiliar items, or in other words the general preference for safe, easy and salient options. Our experimental results show a clear effect of choice-set size: individuals choose familiar products more frequently in larger choice sets. This suggests that a variation in the number of alternatives leads to preference reversals, or inconsistences with respect to standard revealed-preference properties, and has an effect on the individuals final decision made. This constitutes another demonstration of choice overload.

Our second contribution is that this choice-overload effect on preferences does not seem to reflect complexity or information abundance: whatever the size of the choice set, providing individuals with more information reduces or does not change the preference for familiar items, so that the effect of choice overload on preference rankings cannot be reduced, or even directly explained, by some form of information overload. In contrast to Scheibehenne et al. (2010b) and Fasolo et al. (2009), these experimental results suggest that the effects of choice overload do not just show cognitive or information overload.

In this case, how can the impact of choice overload on the preference rankings be explained? Why does choice-set expansion produce preference reversals? We suggest three broad explanations. First, a number of asymmetric-attention models have been proposed to account for decision inconsistencies based on attention filters, (Lleras et al., 2017; Masatlioglu et al., 2012; Dean et al., 2017). As these generally assume that decision-makers first define a consideration set (via an attention filter) and then decide optimally within this set, they are not in general well-suited to explain more general changes in preferences.[17] However, a non-dichotomous extension of these models may be compatible

[17] Unless we assume that the decision-maker sequentially uses consideration sets. For instance, she first orders the options in her primary consideration set, then defines a secondary consideration set among those not yet ranked, etc. Likewise, saliency theory (Bordalo et al., 2013; Dertwinkel-Kalt et al., 2017) is not a likely candidate to explain instability in preference rankings, as the structural composition of choice sets is always the same, which should lead the salience of the various dimensions to be unaffected.

with more general changes in preferences. Decision-makers may pay more attention to familiar than unfamiliar options, leading to a less-balanced effort to update beliefs are about unfamiliar options, more subjective uncertainty about these options, and in the end less attraction to them. Alternatively, confusion may also be at play. Individuals may try to consider all of the options when establishing their subjective ranking, but may not be efficient in the sense that the more options they consider, the less clear their beliefs about them. This confusion may come from having too many mental elements to weigh at the same time, and may yield an asymmetric effect in terms of subjective beliefs about individual options: while familiar items are represented clearly, effortlessly and immediately, unfamiliar items may be more affected by confusion, leading to stronger feelings of subjective uncertainty about them. Last, decision heuristics can explain why choice proliferation affects preferences. Individuals have been found to rely on decision short-cuts and heuristics in a number of experimental situations (Kahneman, 2003). It is plausible that as the decision task increases in complexity, the heuristics (or the decision processes) used to rank alternatives change and focus more on simplicity.

What are the implications of choice proliferation leading to familiarity prevalence? Given that our results are obtained under relatively simple conditions – individuals in the experiment face a maximum of only eight alternatives – extrapolating to everyday external conditions, where economic agents face hundreds of alternatives, would suggest a considerable effect on decisions. This bias towards familiar products could lead to inertia in consumption: if consumers stick to familiar or habitual products, they may be reluctant to consider and experience new-comers or alternative options. In a long-term perspective, this may lead to more fixed individual consumption routines, as choice sets seem to be expanding endlessly in contemporary economies.

In the realm of industrial organization, choice overload effect may partly explain why markets are highly concentrated on some 'superstar products' (Rosen, 1981). In almost all countries, market regulation favors more competition between firms to ensure lower prices, but also more varieties and choice in markets. With more varieties, pure choice overload may simply lead consumers to limit their consumption to a small number of popular or habitual products, a phenomenon that may in the end, counter-intuitively, favor larger players.

Our study also echoes the debate over the importance of choice architecture in public policy (Sunstein and Thaler, 2003; Leonard, 2008). As empirical research indicates (see Cronqvist and Thaler, 2004; Bhargava and Loewenstein, 2015; Bhargava et al., 2015), the complexity of decisions makes options difficult to compare and can produce non-optimal choice or a reliance on *default* options or *status quo*s. Our results also suggests that a critical aspect of a successful choice architecture is choice-set size. In addition to the simplification of the information available, the format of information presentation and the existence of default options, maintaining small choice sets may be critical for individuals to make well-considered, if not optimal, decisions. As our results show that the most influential feature of choice situations is choice-set size, it might be preferable to offer a smaller number of options rather than change other features of the decision.

References

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., and Zinman, J. (2010). What's advertising content worth? Evidence from a consumer credit marketing field experiment. *Quarterly Journal of Economics*, 125(1):263–306.

Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2013). Simplification and saving. *Journal of Economic Behavior & Organization*, 95:130–145.

Bhargava, S. and Loewenstein, G. (2015). Behavioral economics and Public Policies 102: Beyond nudging. *American Economic Review: Papers and Proceedings*, 105(5):396–401.

Bhargava, S., Loewenstein, G., and Sydnor, J. (2015). Do individuals make sensible health insurance decisions? Evidence from a menu with dominated options. NBER Working Paper No. 21160.

Boatwright, P. and Nunes, J. C. (2001). Reducing assortment: An attribute-based approach. *Journal of Marketing*, 65(3):50–63.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843.

Buturak, G. and Evren, Ö. (2017). Choice overload and asymmetric regret. *Theoretical Economics*, 12(3):1029–1056.

Chernev, A. (2003). When more is less and less is more: The role of ideal point availability and assortment in consumer choice. *Journal of consumer Research*, 30(2):170–183.

Chernev, A., Böckenholt, U., and Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, 25(2):333–358.

Cronqvist, H. and Thaler, R. H. (2004). Design choice in privatized social-security systems: Learning from the Swedish experience. *American Economic Review*, 94(2):424–428.

Dean, M., Kibris, Ö., and Masatlioglu, Y. (2017). Limited attention and status quo bias. *Journal of Economic Theory*, (Forthcoming).

Dertwinkel-Kalt, M., Köhler, K., Lange, M. R., and Wenzel, T. (2017). Demand shifts due to salience effects: Experimental evidence. *Journal of the European Economic Association*, 15(3):626–653.

Donner, A. and Klar, N. (1994). Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *American Journal of Epidemiology*, 140(3):279–289.

Farrington, J. (2011). Seven plus or minus two. *Performance Improvement Quarterly*, 23(4):113–116.

Fasolo, B., Hertwig, R., Huber, M., and Ludwig, M. (2009). Size, entropy, and density: What is the difference that makes the difference between small and large real-world assortments? *Psychology & Marketing*, 26(3):254–279.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Gerasimou, G. (2017). Indecisiveness, undesirability and overload revealed through rational choice deferral. *Economic Journal*, 128(614):2450–2479.

Hausman, J. A. and Ruud, P. A. (1987). Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics*, 34(1):83–104.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Huber, J., Payne, J. W., and Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9(1):90–98.

Irons, B. and Hepburn, C. (2007). Regret theory and the tyranny of choice. *Economic Record*, 83(261):191–203.

Iyengar, S. and Kamenica, E. (2010). Choice proliferation, simplicity seeking, and asset allocation. *Journal of Public Economics*, 94(7-8):530–539.

Iyengar, S. and Lepper, M. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6):995–1006.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5):1449–1475.

Kamenica, E. (2008). Contextual inference in markets: On the informational content of product lines. *American Economic Review*, 98(5):2127–2149.

Kempf, A. and Ruenzi, S. (2006). Status quo bias and the number of alternatives: An empirical illustration. *Journal of Behavioral Finance*, 7(4):204–213.

Le Lec, F. and Tarroux, B. (2019). On attitude to choice : Some experimental evidence of choice aversion. *Journal of the European Economic Association*, Forthcoming.

Leonard, T. C. (2008). Richard H. Thaler, Cass R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness. *Constitutional Political Economy*, 19(4):356–360.

Lleras, J. S., Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2017). When more is less: Limited consideration. *Journal of Economic Theory*, 170:70–85.

Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2012). Revealed attention. *American Economic Review*, 102(5):2183.

Persson, P. (2017). Attention manipulation and information overload. Technical report, National Bureau of Economic Research.

Reutskaja, E. and Hogarth, R. M. (2009). Satisfaction in choice as a function of the number of alternatives: When goods satiate. *Psychology & Marketing*, 26(3):197–203.

Reutskaja, E., Lindner, A., Nagel, R., Andersen, R. A., and Camerer, C. F. (2018). Choice overload reduces neural signatures of choice set value in dorsal striatum and anterior cingulate cortex. *Nature Human Behaviour*, 2:925–935.

Reutskaja, E., Nagel, R., Camerer, C. F., and Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review*, 101(2):906–926.

Rosen, S. (1981). The Economics of Superstars. *American Economic Review*, 71(5):845–58.

Samuelson, W. and Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1:7–59.

Sarver, T. (2008). Anticipating regret: Why fewer options may be better. *Econometrica*, 76(2):263–305.

Scheibehenne, B., Greifeneder, R., and Todd, P. (2010a). Can there ever be too many options? a meta-analytic review of choice overload. *Journal of Consumer Research*, 37:409–425.

Scheibehenne, B., Greifeneder, R., and Todd, P. (2010b). Less may be more when choosing is difficult: choice complexity and too much choice. *Acta Psychologica*, 133:45–50.

Shah, A. M. and Wolford, G. (2007). Buying behavior as a function of parametric variation of number of choices. *Psychological Science*, 18(5):369–370.

Shin, J. and Ariely, D. (2004). Keeping doors open: The effect of unavailability on incentives to keep options viable. *Management Science*, 50(5):575–586.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 16(2):158–174.

Sunstein, C. R. and Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, 70(4):1159–1202.

Tversky, A. and Shafir, E. (1992). Choice under conflict: The dynamics of deferred decision. *Psychological Science*, 3(6):358–361.

Appendix suggested for online publication

A Instructions (translated from the French)

Instructions

You are going to participate in two experiments. You will be informed about everything you need to know to take part in these. You can earn a monetary amount during these experiments. The amount of money you can earn depends on your decisions. It is therefore important to make them carefully.

During this study, anonymity is guaranteed, i.e. your decisions cannot be linked to you. At the end of the experiment, you will confidentially receive the monetary amount that you earned during the experiment in the form of a check. You have already earned an initial payment of €12 at the beginning of the study. All the earnings we mention in these instructions will be added to this initial payment of €12.

Please note that communication is strictly forbidden during the entire experiment. It is also forbidden to use documents and materials (such as mobile phones) other than those made available to you at the beginning of the experiment. Communicating or playing around with the computer leads to exclusion from the study. We also inform you that you may only use the functions on the computer that are required to complete the study. In the case of the non-respect of any of these rules, you will be excluded from the experiment and all earnings will be lost.

If you have any questions, please raise your hand. We will then come and answer your questions at your place.

Procedure for the first experiment

Experiment 1 consists of two parts. In part 1, you will answer questions about various websites. In part 2, you will have access to one website for 30 minutes. The website to which you will have access will depend on the decisions you made in part 1. It is then important to make these carefully.

1. Preliminary questions

Before these two parts, you have to rank different content categories of websites (news, cooking, sport etc.). Your ranking will determine which kind of websites you have access to in part 2. More precisely, you will have access to one of websites in the category you rank first.

The following screen displays the website categories that you have to rank between 1 (the most-preferred category) and 7 (the least-preferred one):

Catégorie de sites (par ordre alphabétique)	Votre Classement
Actualité	
Cinéma et séries	
Cuisine	
Culture	
Economie	
Jeux vidéo	
Sport	

You have to assign an integer between 1 and 7 to each category according to your preferences. Once your ranking is made, click on the OK button (bottom-right of the screen) to move on to part 1.

Note that you cannot give the same rank to different website categories. Thus you can click on the OK button only if each category has a different rank from the others.

2. <u>Procedure for part 1</u>

Part 1 of the experiment consists of four stages.

During the **first stage**, you will be informed about the name of the websites to which you can have access in part 2. There are [*4/8 depending on the treatment*] websites available.

Here is an example of the screen:

wikipedia.org	facebook.com
tumblr.com	metropole.rennes.fr
france.meteofrance.com	yahoo.com
univ-rennes1.fr	twitter.com

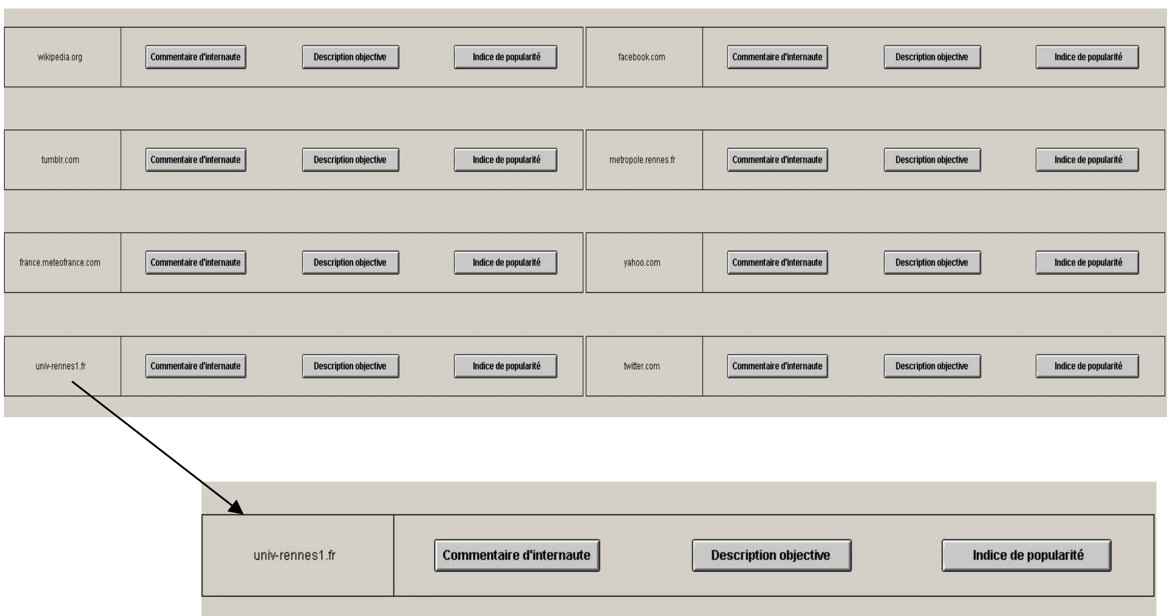
Once you have understood the websites available, click on the OK button.

During the **second stage**, you have the opportunity to be informed about the various websites. Three types of information are available:

- An *objective description*, which provides the basic characteristics of the website;
- An index of popularity, which is the number of times that people said they liked the website on facebook.com, i.e. the number of Facebook likes;
- A randomly-picked commentary from an internet user.

Note that you cannot read all of the pieces of information. More precisely, you have to read a total of [*0/3/6/12 depending on the treatment*] pieces of information on any of the websites you wish.

The screen is of the following type [*treatment C+/I+*]:



Clicking on an information button opens a popup window with the information. Once you have read the information, click on the OK button. The above screen will then appear again.

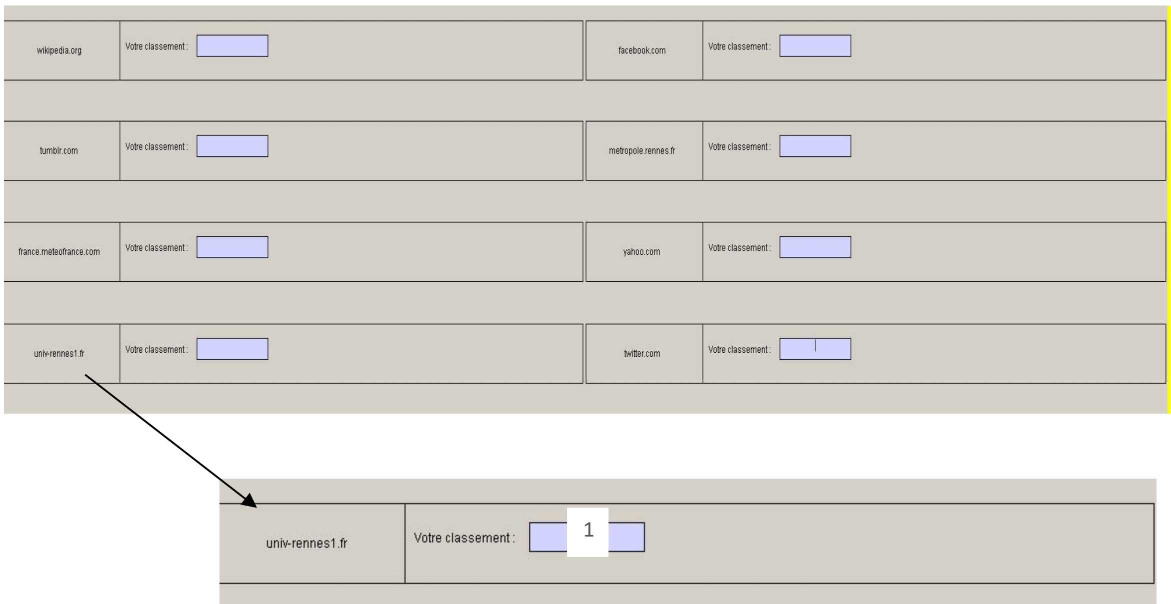
Since you have to read [*0/3/6/12 depending on the treatment*] pieces of information, you have to repeat this action [*0/3/6/12 depending on the treatment*] times. Note that you cannot read the same piece of information more than once. Paper and pen are at your disposal if you want to write down the information you read.

Once you have read the [*0/3/6/12 depending on the treatment*] pieces of information, the third stage starts automatically. However you have to wait until all the experimental participants have read all of the pieces of information.

During the **third stage**, you have to rank the websites between 1 (the most-preferred one) and [*4/8 depending on the treatment*] (the least-preferred one).

[The following screenshots and tables refer to the situation with eight options]

This task is carried out using the following screen:



Once you have made your ranking, click on the OK button to move on to the next stage. Note that you cannot give the same rank to different websites. Thus you can only click on the OK button if each website has a different rank to the others.

Note that your ranking determines the second part of the experiment, where you have access to a certain website. More precisely, one of the websites is randomly-picked with the probability that a website be chosen depending on its rank [*treatment C+/I+*]:

	Probability of being chosen
Website you ranked 1st	45%
Website you ranked 2nd	25%
Website you ranked 3rd	15%
Website you ranked 4th	7%
Website you ranked 5th	5%
Website you ranked 6th	2%
Website you ranked 7th	1%
Website you ranked 8th	0%

This means that the website you rank 1st has 45 chances in 100 to be chosen by the computer; the website you rank 2nd has 25 chances in 100 to be chosen; the website you rank 3rd has 15 chances in 100 to be chosen, etc. The **choice made by the computer will be implemented during the second part of the experiment**.

Ranking according to your preferences means that you are more likely to have access to a website you really want to visit than one for which you only have a weak preference. As such, it is in your own interest to rank the websites carefully and honestly.

For example, suppose that the website you rank in the 1^{st} position is univ-rennes1.fr. The probability that this website be selected by the computer is then 45% (see the table above). If so, you will have access to this website for 30 minutes during the second part of the experiment, and will not be able to visit other websites.

Note that you are informed about the selection of the website at the beginning of the second part of the experiment.

After this, a **fourth stage** starts. You have to indicate whether you previously knew of the website before the experiment and whether the available information was useful for ranking the websites. The first part of the experiment is now over.

3. Procedure for part 2

Once part 1 is finished, part 2 of the experiment starts. This part lasts 30 minutes and works as follows.

You are first informed about the randomly-selected website. In the previous example, you are informed that the website which was randomly drawn is: univ-rennes1.fr.

The selected choice is then implemented. For 30 minutes, you have access to a single website: that which was randomly drawn (in the example, univ-rennes1.fr). Your browsing is limited to this website. If you try to open (intentionally or not) another webpage, a no-entry message will appear. Clicking on the red cross at the top-right of the webpage allows you to close this. You will then return to the selected website. Please note that we cannot observe what you do on the website.

In you have any problems, do not hesitate to raise your hand. We will come and help you.

4. Evaluation of your satisfaction

Once the website is randomly selected, you will report your level of satisfaction with the website. You will answer a series of questions.

****

After answering these questions, the first experiment is finished. You will be given the instructions for experiment 2.

Summary of experiment 1

Preliminary questions: ordering website content

Part 1:

- Stage 1: presentation of available websites (belonging to the content categories you prefer).
- Stage 2: information about the websites.
- Stage 3: ranking the websites.
- Random draw/selection of a website according to the table below (you will be informed about the selected website at the beginning of Part 2).
- Stage 4: questions about your knowledge of the websites before the experiment and the use of the information.

Part 2:

- Information about the randomly-selected website
- 30 min browsing on the randomly-selected website

Evaluation of your satisfaction: questionnaire

Probabilities that a website be chosen

	Probability of being chosen
Website you ranked 1st	45%
Website you ranked 2nd	25%
Website you ranked 3rd	15%
Website you ranked 4th	7%
Website you ranked 5th	5%
Website you ranked 6th	2%
Website you ranked 7th	1%
Website you ranked 8th	0%

Understanding the questions

In order to check your understanding of these instructions, please answer these statements (TRUE or FALSE).

We will give you the right answer aloud when all the participants have answered them. If, at the end of the correction, you have questions, please raise your hand. We will then come and answer your questions at your place.

Q1. During experiment 1, I will have access to one website.

Q2. During experiment 1, I will have access to any website that I would like.

Q3. The website I will have access to depends on my decisions.

Q4. During part 1, I will answer questions about 8 websites with the same content.

Q5. The content of the website available in stage 1 depends on the ranking I give in the preliminary question.

Q6. During part 1, I have to read 12 information pieces per item.

Q7. During part 2, an item I ranked is proposed.

Q8. During part 2, I can have access to all the websites I ranked and the access time depends on my ranking.

Q9. I have more chances to browse the website I ranked first than that I ranked third.

Procedure for the second experiment

In this experiment, you can earn a certain amount of money, which adds to the initial endowment of €12. You have to make 10 decisions. Each decision consists in choosing between 2 options: option A and option B. For each option, you can earn some money/income with certain probabilities, as indicated on the screen that will face you.

The screen will indicate a list of 10 decisions. The table is of the following type:

Table: Choice of options

	Option A				Option B				Your choice	
Decision	Prob.	Earnings	Prob.	Earnings	Prob.	Earnings	Prob.	Earnings	A	B
1	10%	2	90%	1.6	10%	3.85	90%	0.1	○	○
2	20%	2	80%	1.6	20%	3.85	80%	0.1	○	○
3	30%	2	70%	1.6	30%	3.85	70%	0.1	○	○
4	40%	2	60%	1.6	40%	3.85	60%	0.1	○	○
5	50%	2	50%	1.6	50%	3.85	50%	0.1	○	○
6	60%	2	40%	1.6	60%	3.85	40%	0.1	○	○
7	70%	2	30%	1.6	70%	3.85	30%	0.1	○	○
8	80%	2	20%	1.6	80%	3.85	20%	0.1	○	○
9	90%	2	10%	1.6	90%	3.85	10%	0.1	○	○
10	100%	2	0%	1.6	100%	3.85	0%	0.1	○	○

Consider decision 4. Here option A gives you a 40% chance to earn €2 and 60% to earn €1.60. Option B gives you a chance of 40% to earn €3.85 and 60% to earn €0.10. You have to choose between options A and B.

For each decision, you give your choice between A and B by clicking on the option you prefer in the column on the right ("your choice").

Once you make your 10 decisions, the computer will randomly select one of them. Each decision has the same probability of being selected.

Next, the computer will randomly choose a number between 1 and 10, that determines the earnings associated with the option you have chosen.

You will be informed about your earnings and your total earnings.

Example.

Suppose that the computer selects the first line for the calculation of your earnings. For this line, option A gives you €2 if the randomly drawn number is 1 and €1.60 if it is between 2 and 10. Option B gives you €3.85 if the randomly drawn number is 1 and €0.10 if it is between 2 and 10.

If you have chosen A for this line and the computer randomly selects the number 1, then your earnings from experiment 2 are €2. The earnings are calculated in the same way for each of the decisions.

To sum up, you have to choose 10 times between option A and option B. When you have made all of the decisions, the computer randomly selects one of your decisions. Next, it randomly chooses a number, which determines your earnings.

The experiment is then finished.

****

If you have any questions, please raise your hand. We will come and answer your questions at your place.

PUBLICATION 9: Fabrice Le Lec, Marianne Lumeau et Benoît Tarroux (2017), "Le goût des autres", ***Revue Economique***, 68(5):773-784

LE GOÛT DES AUTRES

Fabrice Le Lec, Marianne Lumeau, Benoît Tarroux

Presses de Sciences Po (P.F.N.S.P.) | « Revue économique »

2017/5 Vol. 68 | pages 773 à 784

ISSN 0035-2764
ISBN 9782724635072

Le goût des autres

Fabrice Le Lec[*]
Marianne Lumeau[**]
Benoît Tarroux[***]

Cet article étudie pourquoi certains individus ont tendance à adopter des comportements de consommation conformistes. Plus précisément, notre attention se porte sur l'analyse des préférences intrinsèques pour la conformité : les préférences des individus auraient tendance à converger vers celles des autres, même dans des contextes où il n'existe pas d'incertitude quant à la qualité des biens et où les choix des individus ne peuvent être observés. Pour cela, nous avons conçu une expérience au cours de laquelle les sujets reportaient la satisfaction procurée par deux biens usuels testés en laboratoire, puis en donnaient une évaluation monétaire. Dans un premier traitement, les sujets étaient « isolés », alors que dans un deuxième traitement, une information sur le choix effectué par d'autres sujets leur était fournie juste avant les évaluations monétaires des biens. Nos résultats montrent que les sujets sont sensibles au choix des autres de manière asymétrique. Lorsque les sujets rapportent une plus grande satisfaction pour le bien également choisi par les autres, cette information n'affecte pas la valorisation monétaire des biens. En revanche, les sujets ayant des goûts différents de ceux des autres ont tendance à réduire la différence de valorisation entre le bien préféré et l'autre.

THE TASTE OF OTHERS

This paper studies why individuals adopt conformist consumption behavior. More precisely, we pay attention to intrinsic preference for conformity: individuals preferences tend to converge on the others' ones, even if there is no uncertainty about the quality of goods and individual choices can not be observed. For this purpose, we run an experiment, in which subjects report the satisfaction provided by two goods tested in the laboratory and then a monetary valuation. In a first treatment, subjects are "isolated," while, in a second one, an information about choice made by other subjects are provided before they give monetary valuations to both goods. Our results show that subjects are sensitive to the choice made by others in an asymmetric way. When subjects report a greater

* Université de Paris 1 et CES, Paris, France. *Correspondance :* Maison des sciences économiques, 106-112 boulevard de l'Hôpital, 75013 Paris, France. *Courriel :* fabrice.le-lec@univ-paris1.fr
** Université d'Angers, GRANEM, Angers, France et Labex ICCA, Paris, France. *Correspondance :* GRANEM et Université d'Angers, Faculté de droit, d'économie et de gestion, 13 allée François Mitterrand, 49036 Angers, France. *Courriel :* marianne.lumeau@univ-angers.fr
*** Université de Rennes 1 et CREM, Rennes, France. *Correspondance :* Faculté des sciences économiques, 7 place Hoche, 35000 Rennes, France. *Courriel :* benoit.tarroux@univ-rennes1.fr

Les auteurs remercient Elven Priour pour la programmation et son aide dans la conduite des sessions expérimentales. Ils remercient également Nicolas Jacquemet et deux rapporteurs anonymes qui ont permis par leurs commentaires et suggestions d'améliorer cette contribution, ainsi que les participants de la conférence de l'ASFEE à Paris et de la conférence internationale de l'ESA à Sydney pour leurs remarques. Cette recherche a bénéficié du soutien financier du Labex ICCA et de la MSH Bretagne.

Revue économique – vol. 68, n° 5, septembre 2017, p. 773-784

satisfaction for the good that is also chosen by others, this information does not affect the monetary valuations of goods. On the contrary, subjects differing from others in taste tend to reduce the gap between the valuation of the preferred good and the one of the other.

Classification *JEL* : C91, D12, D80.

On aime à imiter ; on imite souvent, même sans s'en apercevoir, et on néglige ses propres biens pour des biens étrangers, qui d'ordinaire ne conviennent pas.

La Rochefoucauld

INTRODUCTION

L'idée que le comportement d'un individu est influencé par celui des autres à travers des mécanismes d'imitation et de conformisme est (presque) aussi vieille que l'étude et l'observation des actions humaines. Des moutons de Panurge de Rabelais aux lois de l'imitation de Tarde, en passant par les innombrables dénonciations de la force aliénante du conformisme, nombre de penseurs classiques voient les actions humaines comme fortement influençables.

Cette tendance au comportement conformiste a été étudiée en détail par la psychologie sociale et le marketing. Parmi les nombreuses études en psychologie sociale, les plus connues sont probablement les expériences dites de « Asch » (Bond et Smith [1996]). En marketing, les influences interpersonnelles ont, en tant que déterminants des comportements de consommation, reçu une attention toute particulière (voir par exemple la revue de littérature de Lascu et Zinkhan [1999]). Récemment, le fort développement du bouche à oreille en ligne a relancé les travaux dans le domaine, montrant que les notes et les avis attribués par certains internautes sont très largement suivis par d'autres (pour une revue de la littérature, voir par exemple King, Racherla et Bush [2014]). En économie, l'étude empirique de l'influence des pairs est plus récente que dans les autres disciplines, mais néanmoins foisonnante, que ce soit en finance, en économie du travail, dans le domaine de la consommation ou encore des interactions stratégiques (Kandel et Lazear [1992] ; Falk et Ichino [2006] ; Salganik, Dodds et Watts [2006] ; Eriksson, Poulsen et Villeval [2009] ; Cai, Chen et Fang [2009] ; Zafar [2011]).

De ce point de vue, les comportements conformistes sont expliqués selon trois perspectives. La première est liée à *l'apprentissage social* : en situation d'incertitude, le comportement des autres agents véhicule une information et permet à l'agent de mettre à jour ses croyances. Ainsi, si une majorité d'individus choisissent une option donnée dans une situation similaire, ces derniers détiennent probablement une information que l'agent n'a pas quant à la pertinence de cette option (Cai, Chen et Fang [2009] ; Duffy, Hopkins et Kornienko [2015] ; Goeree et Yariv [2015] ; Zafar [2011]). L'influence des autres est alors instrumentale et ses relais sont l'information et les croyances. La seconde perspective est celle d'une préoccupation des agents pour l'*image sociale* qu'ils peuvent renvoyer : par respect des normes sociales ou par considération de statut, ce type de préoccupation peut engendrer des comportements conformistes (voir par exemple Engelmann et Friedrichsen [2013]). La troisième perspective est celle d'un *désir intrinsèque pour la conformité*, c'est-à-dire que les préférences des agents auraient tendance à converger vers celles des autres, même lorsqu'ils n'ont pas d'incertitude et/ou lorsque leurs actions ne sont pas observées. Dans la lignée des travaux sur la « construction des préférences » (voir par exemple Slovic [1995], Ariely, Loewenstein et Prelec [2003] et Isoni *et al.* [2016]), on peut en effet supposer que les agents utiliseraient les préférences des autres comme signal arbitraire pour construire leurs propres préférences.

L'objectif de cet article est d'étudier dans quelle mesure ce désir intrinsèque de conformité existe, ce qui, à notre connaissance, n'a jamais été fait. Pour isoler cet effet des deux premiers, nous avons défini un protocole expérimental dans lequel les individus acquièrent une bonne connaissance des biens par une expérience directe et où leur choix n'est pas publiquement observable par les autres. Dès lors, la comparaison des résultats d'un traitement où les sujets disposent d'information sur le choix d'autres individus à ceux d'un traitement où cette information n'est pas fournie offre un éclairage sur le rôle de cette variable dans la construction des préférences et sur l'existence de la convergence des préférences.

La démarche expérimentale s'avère ici particulièrement adaptée, car, contrairement aux études en conditions économiques réelles, le contrôle de l'environnement permet d'examiner la construction des préférences des individus en maîtrisant l'information qu'ils reçoivent et en garantissant que leur choix n'est pas observé. *A contrario*, les études en conditions économiques réelles (voir par exemple Chevalier et Mayzlin [2006] et Egebark et Ekstrom [2011]) ne permettent généralement pas d'identifier les mécanismes sous-jacents aux comportements conformistes car on ne connaît pas nécessairement la qualité de l'information dont disposent les agents avant d'observer le choix des autres (apprentissage social). Il est également difficile de déterminer dans quelle mesure le choix des individus est observé (image sociale), de même qu'il n'est pas toujours possible de distinguer un apparent comportement conformiste du fait que les individus ont tout simplement des préférences similaires à celles des autres.

L'expérience mise en œuvre pour étudier la convergence des préférences comporte deux traitements : un traitement de contrôle où les individus ne connaissent pas le choix des autres (traitement 0) et un traitement où ils disposent de cette information (traitement 1). Dans le traitement 0, les individus étaient invités à goûter deux biens de consommation usuelle : deux sodas de type « cola ». Pour chacun des colas, ils devaient reporter la satisfaction retirée de la consommation des deux biens, puis les évaluer monétairement[1]. Le traitement 1 est identique en tout point au traitement 0, à l'exception du fait que les individus disposaient d'une information sur le choix effectué par d'autres sujets. Cette information leur était transmise après le report de satisfaction, mais avant l'évaluation monétaire des colas. Dès lors, la différence d'évaluations monétaires entre le traitement 0 et le traitement 1 est une mesure de la sensibilité

1. Ces deux mesures permettent de révéler les préférences des individus : si la satisfaction déclarée pour le cola *x* (ou son évaluation monétaire) est plus élevée que la satisfaction déclarée (ou l'évaluation monétaire) pour le cola *y*, alors le cola *x* est préféré au cola *y*. Les préférences sont donc définies en termes microéconomiques. Dans le cadre de l'expérience, elles ne peuvent être basées que sur le goût des sujets dans le traitement 0, mais peuvent inclure d'autres dimensions dans le traitement 1.

Revue économique – vol. 68, n° 5, septembre 2017, p. 773-784

des individus aux choix des autres : elle révèle la manière dont les préférences peuvent être affectées par le choix des autres.

Nos résultats expérimentaux indiquent que l'information sur le choix effectué par d'autres sujets affecte le comportement des sujets de manière asymétrique selon qu'ils aient des préférences initiales identiques aux autres (*i.e.*, avant d'avoir de l'information sur le choix effectué par d'autres sujets). Ainsi, le comportement des individus ayant des préférences initiales identiques à celles des autres n'est pas modifié, ce qui indique que cette information ne renforce pas leurs préférences. *A contrario*, les individus ayant des préférences initiales différentes de celles des autres ont tendance à réduire l'écart de valorisation entre le bien préféré et le second, tout en maintenant la même hiérarchie entre les deux. Cette atténuation de l'intensité des différences suggère une tendance à la convergence des préférences, uniquement chez les sujets ayant des préférences initiales différentes de celles des autres.

La section suivante détaille le protocole expérimental. La troisième section présente les résultats expérimentaux. Enfin, les conclusions sont exposées dans la quatrième section.

PROTOCOLE EXPÉRIMENTAL

Design de l'expérience

L'expérience comporte deux traitements : un traitement sans information (traitement 0) et un traitement où une information sur le choix des autres est fournie aux sujets (traitement 1).

Le déroulement du traitement 0 est résumé dans le tableau 1. Il est composé de trois étapes. Lors de la première étape, les sujets sont invités à expérimenter deux biens distincts. Les biens choisis sont deux colas de marques différentes (Breizh Cola et Cola Super U)[2]. Ce type de produit a l'avantage d'être facilement testable en laboratoire par les sujets. Par ailleurs, les deux biens testés sont relativement homogènes, ce qui permet de s'assurer d'un contrôle sur la construction hors laboratoire des préférences des sujets et donc de limiter l'influence sociale s'exerçant à l'extérieur sur cette construction[3].

Afin d'expérimenter les deux biens, les expérimentateurs apportent à chaque sujet deux gobelets annotés A et B, contenant chacun 4,5 cl de boisson. Tout au long de l'expérience, ces deux colas sont en effet appelés de manière générique A ou B, de sorte que les sujets ne peuvent être influencés par les marques (McClure *et al.* [2006]). Par ailleurs, afin d'éviter les effets d'ordres, le contenu des colas A et B varie selon la session expérimentale : Breizh Cola est tantôt appelé cola A, tantôt appelé cola B.

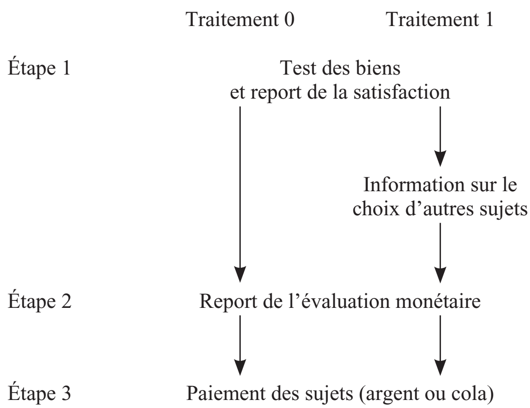
2. Ces deux marques ont été choisies du fait de leur relative faiblesse en part de marché, de sorte que relativement peu de sujets peuvent être familiers avec ces produits.

3. Les expériences préalablement menées sur des colas (voir par exemple McClure *et al.* [2006]) ont montré que, même en présence de biens homogènes de type cola, les sujets pouvaient apprécier différemment les deux colas.

Revue économique – vol. 68, n° 5, septembre 2017, p. 773-784

Après avoir goûté les deux colas, les sujets doivent reporter leur satisfaction sur une base déclarative. Pour chaque cola, les sujets doivent indiquer l'intensité de leur satisfaction sur une échelle de Likert, graduée de 0 (« pas satisfait du tout ») à 6 (« extrêmement satisfait »). Nous notons S_A la satisfaction déclarée pour le cola A et S_B la satisfaction déclarée pour le cola B. La comparaison de S_A et de S_B permet de disposer d'une mesure simple de la préférence initiale des sujets pour l'un ou l'autre des colas.

Tableau 1. *Déroulement de l'expérience*



Au cours de la deuxième étape, les sujets sont invités à donner une évaluation monétaire à chacun des colas à travers une liste à prix multiples (une version simplifiée d'« *Iterative multiple pricing list* » étudiée par Andersen *et al.* [2006]). Pour cela, ils doivent compléter le tableau 2 pour chacun des colas. La procédure est la suivante : à chaque ligne de ce tableau, chaque sujet doit indiquer s'il préfère recevoir une certaine somme d'argent (option I) ou s'il préfère recevoir deux litres du cola *x* (option II)[4]. On note V_A et V_B les valeurs estimées des colas A et B. Ces valeurs correspondent à la valeur monétaire minimale pour laquelle le sujet déclare préférer la somme d'argent aux deux litres de cola. La comparaison de V_A et V_B offre une seconde mesure des préférences des sujets. En plus de ces tâches de valorisation, les sujets font face à un choix binaire en indiquant s'ils préfèrent le cola A ou le cola B[5]. Afin d'inciter les sujets à révéler leur préférence, ils savent qu'une de leurs décisions (tâches de valorisation ou choix binaire) sera tirée au sort pour le paiement. Cette procédure permet de mesurer la valeur subjective attribuée aux biens tout en se fondant sur une suite de choix effectifs des sujets.

4. Il est précisé aux sujets que les deux litres de cola sont distribués sous forme de six bouteilles de 33 cl.

5. D'un point de vue expérimental, il est nécessaire de pouvoir collecter de l'information sur le choix des sujets ayant participé au traitement 0 pour pouvoir l'introduire dans le traitement 1 ; cette information est en effet transmise aux sujets participant au traitement 1 sous le label « Choix des autres ». La procédure est détaillée dans la suite de cette sous-section.

Tableau 2. *Liste à prix multiple*

Ligne	Option I	Option II	Votre choix ?	
			Option I	Option II
	2 litres du cola *x*	0 euro		
	2 litres du cola *x*	0,25 euro		
	2 litres du cola *x*	0,50 euro		
	2 litres du cola *x*	0,75 euro		
…				
	2 litres du cola *x*	3,25 euros		

La dernière étape est l'étape de paiement. L'ordinateur tire au sort une des tâches : soit l'évaluation monétaire de A, soit l'évaluation monétaire de B, soit le choix binaire entre A et B. Dans les deux premiers cas, l'ordinateur tire également au sort une ligne du tableau. Si, pour la ligne tirée au sort, le sujet a déclaré préférer la somme d'argent, son gain pour l'expérience sera cette somme d'argent. *A contrario*, s'il a déclaré préférer deux litres du cola *x*, son gain pour l'expérience sera deux litres du cola *x*. Si l'ordinateur tire au sort la tâche de choix binaire, le gain du sujet sera deux litres du cola qu'il a déclaré préférer.

Le traitement 1 est identique en tout point au traitement 0, à l'exception notable du fait qu'une information est fournie aux sujets sur le choix des autres. Plus précisément, ils sont informés qu'« une majorité de participants tirés au sort ont préféré le cola *x* »[6]. Le tableau 1 indique que cette information est fournie entre la première et la deuxième étape de l'expérience, c'est-à-dire entre le report de la satisfaction et l'évaluation monétaire des colas. Une telle procédure permet aux sujets de se former une représentation claire et non biaisée de leurs préférences sur les deux biens, car, lors du test des colas, leur perception n'est pas influencée par l'information sur le choix des autres. Dès lors, cette information ne devrait pas influer sur les niveaux de satisfaction. En revanche, les évaluations monétaires des biens, qui ont lieu après l'information sur le choix des autres, pourraient être affectées si les sujets sont sensibles au choix des autres[7].

6. Les sujets ayant participé à l'une des sessions du traitement 0 sont tirés au sort pour former des groupes de cinq sujets. Chaque groupe est ensuite apparié aléatoirement avec une session expérimentale du traitement 1. Dès lors, l'information sur le choix des autres correspond au cola que la majorité du groupe de référence a déclaré préférer. Cette procédure d'appariement a l'avantage de fournir la variance nécessaire à l'évaluation de l'effet (par le tirage au sort d'un groupe relativement restreint de sujets) tout en respectant l'absence de duperie des sujets (règle importante en économie expérimentale). Par ailleurs, cela permet de donner une information simple et facilement interprétable.

7. Il est toutefois possible que l'introduction d'une mesure de satisfaction avant la réception de l'information sur le choix des autres puisse atténuer la réaction des sujets à cette information par un biais de consistance. La mise en œuvre d'une telle procédure expérimentale paraît néanmoins préférable à une absence de contrôle sur les préférences des individus. Par ailleurs, un effet de demande pourrait également être possible, poussant les sujets à sur-réagir (dans un sens ou dans l'autre) à l'introduction d'une information sur le choix des autres. Pour limiter de tels effets, les étapes 1 et 2 de l'expérience sont espacées de dix minutes. Ce temps est consacré à la lecture des instructions relatives à l'étape 2 de l'expérience.

Déroulement de l'expérience

L'expérience a été programmée sous le logiciel z-Tree (Fischbacher [2007]) et réalisée au laboratoire d'économie expérimentale de l'Université de Rennes 1 (Labex-EM) en octobre 2014 et janvier 2015. Cent dix-huit étudiants de genre et d'horizons différents (45 % de femmes, 55 % d'étudiants en économie) ont été recrutés sur la base du volontariat pour participer à l'une des dix sessions de cette expérience.

Au début de l'expérience, les sujets sont informés qu'ils vont participer à une expérience de consommation réelle et qu'en plus d'un forfait de participation de 6 euros, ils peuvent gagner une certaine somme d'argent ou un bien de consommation[8]. Dans la mesure où l'expérience est simple à comprendre, les instructions sont disponibles au fur et à mesure de l'expérience[9]. À la fin de l'expérience, les gains sont attribués en privé, calculés selon la procédure décrite dans la sous-section précédente. Le tableau 3 indique que 45 participants ont obtenu un gain monétaire supplémentaire (en moyenne égal à 2,2 euros), 30 sont repartis avec deux litres de Breizh Cola et 43 avec deux litres de Cola Super U. En moyenne, les sessions durent trente minutes[10].

Tableau 3. *Gains de participants*

Nombre de sujets	T0	T1	Total
avec gain monétaire	23	22	45
avec gain Breizh	15	15	30
avec gain Super U	22	21	43
Total	60	58	118

RÉSULTATS EXPÉRIMENTAUX

Satisfaction

Le tableau 4 présente les préférences des sujets en termes de satisfaction déclarée lors de la première étape de l'expérience (après le test des colas). Nous considérerons, dans la suite de l'article, qu'un sujet a une préférence initiale pour x si celui-ci déclare une satisfaction plus élevée pour x que pour y. Le tableau 4 indique qu'en moyenne 45 % des sujets déclarent une satisfaction plus élevée

8. Les sujets sont préalablement informés de ces deux conditions ainsi que de la nature des biens (sous le terme générique de « sodas ») dans le courriel d'invitation à participer à cette expérience.

9. Pour l'étape 2 de l'expérience, des instructions papiers sont distribuées et lues aux sujets. Afin de s'assurer de leur compréhension, des exemples leur sont donnés. Ces instructions et les copies des écrans d'ordinateur sont disponibles sur demande auprès des auteurs.

10. Dans chaque session, les sujets ont préalablement participé à une autre expérience dont l'objet d'étude diffère de celui présenté dans cet article. Finalement, les deux expériences ont duré une heure.

pour Breizh Cola, 45 % des sujets déclarent préférer le Cola Super U et 10 % des sujets sont indifférents. Par ailleurs, des tests non paramétriques d'homogénéité des distributions (χ^2) indiquent qu'il n'existe pas de différence significative entre les traitements. Ce résultat est parfaitement cohérent avec le fait qu'à ce stade de l'expérience les traitements sont identiques et suggère donc une similarité des échantillons. Par ailleurs, comme indiqué dans la section précédente, les sujets ne connaissent pas la marque des colas[11], ce qui tend à montrer que, hors signaux marketing ou de marque, les deux produits sont, au plan agrégé, appréciés de manière similaire.

Tableau 4. *Préférence initiale des sujets en termes de satisfaction déclarée (nombre et proportion de sujets déclarant une satisfaction plus élevée pour le cola x)*

Préférence initiale	T0	T1	Tests du χ^2 (*p*-values)
Breizh	27 (45 %)	26 (44,83 %)	0,985
Super U	27 (45 %)	26 (44,83 %)	0,985
Indifférent	6 (10 %)	6 (10,34 %)	0,951

Évaluation monétaire

Le tableau 5 présente la manière dont les sujets valorisent les deux colas. Dans le traitement 0, 30 % des sujets attribuent une valeur monétaire supérieure au Breizh Cola, 28,33 % au Super U et 41,67 % des sujets reportent des évaluations monétaires identiques pour les deux colas. Dans le traitement 1, 29,31 % des sujets indiquent une valeur monétaire supérieure pour le Breizh Cola, 13,79 % pour le Cola Super U et 56,90 % une même valeur pour les deux boissons. Des tests non paramétriques de χ^2 révèlent des différences significatives à 10 % entre les traitements. Sans être conclusif, ce résultat suggère que l'information sur le choix d'autres sujets influence les évaluations monétaires des sujets. Toutefois, notons qu'au plan individuel il existe une forte cohérence entre les évaluations monétaires et les préférences initiales des sujets en termes de satisfaction, et ce, quel que soit le traitement : seul un sujet « renverse ses préférences » dans le traitement 0 contre deux dans le traitement 1 (la différence n'est pas significative)[12]. Si ce résultat s'explique immédiatement par le fait que les préférences sont fortement liées à la satisfaction, une explication complémentaire de ce phénomène serait l'existence d'un biais de consistance chez les sujets : étant réticents à se contredire, ils auraient tendance à respecter l'ordre de leur préférence initiale lors de l'évaluation monétaire des biens. Auquel cas, nous sous-estimerions le biais de la conformité, celui-ci étant contrebalancé par le biais de consistance.

11. Au cours de l'expérience, les colas étaient appelés cola A et cola B. Si on considère les préférences en termes de cola A et cola B (effet d'ordre par exemple), il n'existe également pas de différence significative entre les traitements.

12. De même, il n'existe pas de différence significative lorsque l'on considère les sujets qui reportent une valorisation monétaire similaire pour les deux colas alors qu'ils n'étaient pas indifférents en termes de satisfaction, ou inversement. Ces sujets représentent 35 % des sujets du traitement 0 et 46,5 % des sujets du traitement 1.

Tableau 5. *Évaluation monétaire des colas (nombre et proportion de sujets)*

	T0	T1	Tests du χ^2 (p-values)
$V_{Breizh} > V_{SuperU}$	18 (30,00 %)	17 (29,31 %)	0,935
$V_{Breizh} < V_{SuperU}$	17 (28,33 %)	8 (13,79 %)	0,053*
$V_{Breizh} = V_{SuperU}$	25 (41,67 %)	33 (56,90 %)	0,098*

Note : * $p < 0{,}1$; ** $p < 0{,}05$; *** $p < 0{,}01$.

Pour tester si, malgré un potentiel biais de consistance, le choix des autres influence les décisions des sujets, nous avons étudié la différence entre la valeur monétaire qu'un sujet *i* attribue au cola qu'il a initialement déclaré préférer et la valeur monétaire attribuée à l'autre cola : $\delta_{V_i} = V_{Colapréféré} - V_{Colamoinspréféré}$. Dans cette logique, les sujets indifférents entre les deux colas (en termes de satisfaction) ont été exclus de l'analyse (soit six sujets dans le traitement 0 et six sujets dans le traitement 1). L'intérêt de cette procédure réside dans la possibilité d'identifier finement l'effet de l'information sur le goût des autres, même en présence d'un biais de consistance. En effet, même si les sujets sont réticents à inverser la hiérarchie entre les deux produits, les différences de valorisation entre les deux produits peuvent être affectées par l'information sur le goût des autres.

En moyenne, δ_{V_i} est de 0,34 euro dans le traitement 0, 0,29 euro dans le traitement 1 lorsque la préférence initiale des sujets est identique à celle du choix des autres (soit 31 sujets) et 0,14 euro dans le traitement 1 lorsque la préférence initiale est différente de celle des autres (soit 21 sujets). Cela suggère que, dans le traitement 1, la moyenne des δ_{V_i} diffère selon que les sujets ont ou non une préférence initiale identique à celle des autres : alors que la différence moyenne d'évaluations monétaires reportées par les sujets du traitement 1 ayant la même préférence que les autres est relativement proche de celle des sujets du traitement 0, cette différence est deux fois plus faible pour les sujets du traitement 1 ayant des préférences différentes.

Le tableau 6 apporte une preuve formelle de ce résultat. Il présente les résultats d'estimations réalisées par la méthode des moindres carrés ordinaires où la variable dépendante est δ_{V_i}, la différence entre les évaluations monétaires du cola initialement préféré et de l'autre cola. Des variables indicatrices de traitement ont été introduites comme variables explicatives, avec, dans le traitement 1, une distinction entre les sujets ayant une préférence initiale identique ou différente à celle des autres. La variable de contrôle δ_{S_i} a également été introduite pour tenir compte de la différence de satisfaction entre les deux colas[13].

L'estimation 1 du tableau 6 indique que la différence des évaluations monétaires entre les deux colas n'est pas significativement différente entre les sujets du traitement 1 ayant la même préférence initiale que les autres et les sujets du traitement 0. En revanche, le coefficient de la variable *T1-Pref. différente* étant significatif et négatif, la différence des évaluations monétaires des sujets du traitement 1

13. Tous les résultats rapportés ci-dessous sont robustes à l'introduction de variables de contrôle, tant sociodémographiques, telles que le genre, l'âge, la discipline d'étude, etc., qu'internes à l'expérience, telles que l'ordre des colas, etc.

ayant une préférence différente de celle des autres s'avère significativement plus faible que celle des sujets du traitement 0. La deuxième spécification, colonne 2 du tableau 6, indique qu'il existe également une différence significative entre les sujets du traitement 1 n'ayant pas la même préférence initiale et ceux ayant une préférence identique. Ces résultats indiquent que l'information sur le choix des autres modifie le comportement des sujets n'ayant pas les mêmes préférences que les autres, mais n'a pas d'impact sur les sujets ayant des préférences identiques.

Une explication possible de ce phénomène pourrait être l'incertitude quant à la préférence pour l'un des produits : même si le sujet a individuellement testé les produits, il pourrait estimer sa préférence de manière floue ou incertaine (particulièrement face à deux biens relativement homogènes) ; dès lors, savoir que les autres ont une préférence différente de la sienne tendrait à le faire douter de ses préférences et réduirait donc la différence de valorisation monétaire. On pourrait alors expliquer cet effet par un mécanisme d'*apprentissage social* évoqué en introduction. Cette hypothèse est toutefois invalidée par la colonne 3 du tableau 6, qui suggère même un résultat contraire. En effet, le coefficient estimé, significatif et négatif, de la variable d'interaction δ_{S_i} × *T1-Pref. différente* indique que plus la différence de satisfaction entre les deux colas est élevée, moins la différence de valorisation des sujets ayant des préférences initiales différentes de celles des autres est élevée. Autrement dit, l'effet d'une préférence différente des autres est d'autant plus important que les préférences initiales du sujet sont claires. Ce résultat pourrait s'expliquer par la combinaison d'un désir intrinsèque de se conformer et d'un biais de consistance (ne pas contredire la hiérarchie entre produits initialement rapportée), conduisant les sujets à donner une évaluation monétaire identique aux deux biens même si l'un est (faiblement ou fortement) dominé. Ainsi, l'effet du choix des autres sur δ_{V_i} serait d'autant plus important que δ_{S_i} est élevé.

Tableau 6. *Estimations de la différence des évaluations monétaires (* δ_{V_i} *)*

	(1) T0 et T1	(2) T1	(3) T1
δ_{S_i}	0,1841***	0,1934***	0,1285***
	(0,0326)	(0,0384)	(0,0437)
T0	réf.		
T1-Pref. identique	0,0033	réf.	réf.
	(0,0924)		
T1-Pref. différente	– 0,1610**	– 0,1670**	– 0,2488
	(0,0844)	(0,0771)	(0,1729)
δ_{S_i} × T1-Pref. différente			– 0,2074**
			(0,0782)
Constante	– 0,0729	– 0,2543**	– 0,1111**
	(0,0929)	(0,0988)	(0,1078)
N	106	52	52
R2	0,2685	0,3587	0,4406

Note : OLS ; * $p < 0{,}1$; ** $p < 0{,}05$; *** $p < 0{,}01$.

Enfin, notons que le coefficient de la variable de contrôle δ_{S_i} est significatif et positif dans les colonnes 1 à 3 du tableau 6, indiquant une corrélation positive entre la différence des niveaux de satisfaction et la différence des évaluations monétaires. Cette corrélation positive peut être interprétée en termes de « consistance faible ». Par ailleurs, cela suggère que malgré son caractère déclaratif, la mesure de la satisfaction est une estimation raisonnable de la préférence des individus.

CONCLUSION

Cet article a présenté les résultats d'une expérience ayant pour but de mesurer l'impact du choix des autres sur les comportements de consommation des individus. Pour cela, nous avons comparé le choix de sujets ayant une information sur le choix d'autres sujets à celui réalisé par des sujets n'ayant pas cette information. Plusieurs enseignements peuvent être tirés de notre étude. D'une part, les préférences des individus semblent influencées par l'information sur le choix des autres ; d'autre part, cet effet apparaît asymétrique selon que les individus aient les mêmes goûts initiaux que les autres ou non. Pour les individus ayant des préférences initiales différentes de celles des autres, cette information les conduit à réduire leur différence de valorisation monétaire entre les deux biens, alors que le comportement des individus ayant des préférences initiales identiques de celles des autres n'est pas modifié par cette information. Ces résultats suggèrent que les préférences des individus ont tendance à converger (ou, du moins, à moins diverger) lorsque ces derniers disposent d'une information sur les préférences d'autres individus, même lorsqu'ils ont une bonne (sinon parfaite) connaissance de leurs goûts et que leur choix ne peut être observé par les autres. Dans la mesure où les travaux en marketing ont montré l'importance du contexte sur les choix des individus, il serait désormais intéressant de tester ces résultats dans différents contextes, par exemple en ayant recours à des biens dont les parts de marché sont élevées ou encore des biens de croyance. De même, l'expérience pourrait être répliquée sur une population plus large qu'une population estudiantine.

D'un point de vue économique, les résultats de cette expérience indiquent que, contrairement à ce que suppose la théorie microéconomique standard, les préférences des agents ne sont pas stables mais malléables. Elles sont déterminées par les goûts des sujets, mais peuvent également être influencées par d'autres facteurs, comme le choix des autres. Ce résultat s'inscrit dans la lignée des travaux mettant en exergue la « construction des préférences » (voir par exemple Slovic [1995], Ariely, Loewenstein et Prelec [2003] et Isoni *et al.* [2016])[14]. Les agents utiliseraient différents signaux en partie arbitraires (prix, marque, etc.) pour établir leurs préférences. Cette expérience tendrait à montrer que les préférences des autres agents pourraient ainsi jouer un rôle essentiel.

14. L'hypothèse de construction des préférences est distincte de la « découverte des préférences » (Plott [1996]). Dans la perspective de l'analyse des comportements conformistes, ce concept renvoie plutôt à l'apprentissage social.

RÉFÉRENCES BIBLIOGRAPHIQUES

Andersen S., Harrison G. W., Lau M. I. et Rutström E. E. [2006], « Elicitation Using Multiple Price List Formats », *Experimental Economics*, 9 (4), p. 393-405.

Ariely D., Loewenstein G. et Prelec D. [2003], « Coherent Arbitrariness: Stable Demand Curves without Stable Preferences », *Quarterly Journal of Economics*, 118 (1), p. 73-105.

Bond R. et Smith P. B. [1996], « Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1952b, 1956) Line Judgment Task », *Psychological Bulletin*, 119 (1), p. 111-137.

Cai H., Chen Y. et Fang H. [2009], « Observational Learning: Evidence from a Randomized Natural Field Experiment », *The American Economic Review*, 99 (3), p. 864-882.

Chevalier J. A. et Mayzlin D. [2006], « The Effect of Word of Mouth on Sales: Online Book Reviews », *Journal of Marketing Research*, 43 (3), p. 345-354.

Duffy J., Hopkins E. et Kornienko T. [2015], « Lone Wolf or Herd Animal? An Experiment on Choice of Information and Social Learning », *Miméo*.

Egebark J. et Ekström M. [2011], « Like What You Like or Like What Others Like? Conformity and Peer Effects on Facebook », *IFN Working Paper*, 886.

Engelmann D. et Friedrichsen J. [2013], « Who Cares for Social Image? Interactions between Intrinsic Motivation and Social Image Concerns », *CESifo Working Paper*, 4514.

Eriksson T., Poulsen A. et Villeval M. C. [2009], « Feedback and Incentives: Experimental Evidence », *Labour Economics*, 16 (6), p. 679-688.

Falk A. et Ichino A. [2006], « Clean Evidence on Peer Effects », *Journal of Labor Economics*, 24 (1), p. 39-57.

Fischbacher U. [2007], « z-Tree: Zurich Toolbox for Ready-Made Economic Experiments », *Experimental Economics*, 10 (2), p. 171-178.

Goeree J. et Yariv L. [2015], « Conformity in the Lab », *Journal of the Economic Science Association*, 1 (1), p. 15-28.

Isoni A., Brooks P., Loomes G. et Sugden R. [2016], « Do Markets Reveal Preferences or Shape Them? », *Journal of Economic Behavior and Organization*, 122 (C), p. 1-16.

Kandel E. et Lazear E. P. [1992], « Peer Pressure and Partnerships », *Journal of Political Economy*, 100 (4), p. 801-817.

King R. A., Racherla P. et Bush V. D. [2014], « What We Know and Don't Know about Online Word-of-Mouth: A Review and Synthesis of the Literature », *Journal of Interactive Marketing*, 28 (3), p. 167-183.

Lascu D.-N. et Zinkhan G. [1999], « Consumer Conformity: Review and Applications for Marketing Theory and Practice », *Journal of Marketing Theory and Practice*, 7 (3), p. 1-12.

McClure S. M., Li J., Tomlin D., Cypert K. S., Montague L. M. et Montague P. [2006], « Neural Correlates of Behavioral Preference for Culturally Familiar Drinks », *Neuron*, 44 (2), p. 379-387.

Plott C. R. [1996], « Rational Individual Behaviour in Markets and Social Choice Processes: The Discovered Preference Hypothesis », dans K. Arrow, E. Colombatto, M. Perleman et C. Schmidt (dir.), *Rational Foundations of Economic Behavior*, Londres, Macmillan, p. 225-250.

Salganik M. J., Dodds P. S. et Watts D. J. [2006], « Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market », *Science*, 311 (5762), p. 854-856.

Slovic P. [1995], « The Construction of Preference », *American Psychologist*, 50 (5), p. 364-371.

Zafar B. [2011], « An Experimental Investigation of Why Individuals Conform », *European Economic Review*, 55 (6), p. 774-798.

Revue économique – vol. 68, n° 5, septembre 2017, p. 773-784

PUBLICATION 10: Y. Loheac, H. Alia, C. Bazart, M. Bchir, S. Blondel, M. Bonescu, A. Bornier, J. Brouard, N. Chappe, F. Cochard, A. Flage, F. Galeotti, X. Hollandts, A. Hopfensitz, N. Jacquemet, F. Le Lec, M. Lefebvre, M. Leplat, C. Mantilla, G. Mateu, G. Peron, E. Peterle, E. Petit, E. Raiber, J. Rosaz, A. Rozan, J.-C. Tisserant, M.-C. Villeval, M. Willinger, A. Zylbersztejn et A. Sutan (2017), *Mise en place d'une expérience avec le grand public : entre recherche, vulgarisation et pédagogie*, **Revue Economique**, 68(5).

MISE EN PLACE D'UNE EXPÉRIENCE AVEC LE GRAND PUBLIC : ENTRE RECHERCHE, VULGARISATION ET PÉDAGOGIE

Youenn Lohéac, Alia Hayyan, Cécile Bazart, Mohamed Ali Bchir, Serge Blondel, Mihaela Bonescu, Alexandrine Bornier, Joëlle Brouard, Nathalie Chappe, François Cochard, Alexandre Flage, Fabio Galeotti, Xavier Hollandts, Nicolas Jacquemet, Fabrice Le Lec, Marianne Lefebvre, Mélody Leplat, Emmanuel Peterle, Emmanuel Petit, Anne Rozan, Marie Claire Villeval, Marc Willinger, Adam Zylbersztejn, Angela Sutan, Astrid Hopfensitz, Guillaume Peron, Cesar Mantilla, Guillermo Mateu, Julie Rosaz, Eva Raiber et Jean-Christian Tisserand

Presses de Sciences Po | « Revue économique »

2017/5 Vol. 68 | pages 941 à 953

ISSN 0035-2764
ISBN 9782724635072

Mise en place d'une expérience avec le grand public : entre recherche, vulgarisation et pédagogie

Youenn Lohéac*, Hayyan Alia, Cécile Bazart, Mohamed Ali Bchir, Serge Blondel, Mihaela Bonescu, Alexandrine Bornier, Joëlle Brouard, Nathalie Chappe, François Cochard, Alexandre Flage, Fabio Galeotti, Xavier Hollandts, Astrid Hopfensitz, Nicolas Jacquemet, Fabrice Le Lec, Marianne Lefebvre, Mélody Leplat, Cesar Mantilla, Guillermo Mateu, Guillaume Peron, Emmanuel Peterle, Emmanuel Petit, Eva Raiber, Julie Rosaz, Anne Rozan, Jean-Christian Tisserand, Marie Claire Villeval, Marc Willinger, Adam Zylbersztejn, Angela Sutan**

Nous présentons la mise en place d'une expérience lors d'un événement grand public national, de manière simultanée dans onze villes françaises, en septembre 2015. L'expérience a impliqué plus de 2 700 participants et a duré quatre heures ininterrompues. L'objectif de cet article est à la fois de fournir une feuille de route pour une éventuelle réplication et de penser à la manière dont la discipline peut être utilisée dans des terrains nouveaux (vulgarisation, pédagogie populaire, communication grand public).

METHODOLOGICAL CONSIDERATIONS ON IMPLEMENTING A PARTICIPATIVE EXPERIMENT

We present the implementation of an economic experiment conducted simultaneously in 11 French cities, with over 2700 participants, during four uninterrupted hours, during a popular-science event held in September 2015. Our goal is both to provide a roadmap for a possible replication and to discuss how the discipline can be used in new fields (science popularization, popular education, public communication).

Classification *JEL* : A11, C93.

* Brest Business School et CREM-CNRS (UMR 6211). *Correspondance :* Brest Business School, 2 avenue de Provence, 29200 Brest, France. *Courriel :* youenn.loheac@brest-bs.com

** Burgundi School of Business – Université de Bourgogne Franche-Comté, CEREN (EA 7477). *Correspondance :* Burgundy School of Business, 29 rue Sambin, 21000 Dijon, France. *Courriel :* angela.sutan@bsb-education.com

Institutions participantes : Université d'Angers, Université de Bourgogne Franche-Comté, ESC Dijon-CEREN, Brest Business School, Université de Bordeaux, Kedge Business School, Université de Montpellier, Toulouse School of Economics, Institute for Advanced Study in Toulouse, Université de Lyon, Paris School of Economics, ENGEES Strasbourg.

Les auteurs remercient la coordination nationale et les coordonnateurs locaux de la Nuit des chercheurs, particulièrement Lionel Maillot et Meriem Fresson, et sont reconnaissants du soutien financier à travers l'appel « Grande expérience participative ».

INTRODUCTION

Dans le cadre de la Nuit européenne des chercheurs 2015, nous avons conduit une expérience d'économie comportementale, de manière simultanée dans onze villes françaises. Dans cet article, nous présentons de manière synthétique et volontairement factuelle la mise en place de cette expérience, réalisée sous des objectifs et contraintes multiples, et qui est située à la frontière entre recherche scientifique, vulgarisation, pédagogie et action événementielle. L'expérience a impliqué plus de 2 700 participants et a duré quatre heures ininterrompues.

En économie comportementale, les méthodes de collecte des données se sont diversifiées pour répondre à des besoins nouveaux engendrés par la production de nouvelles théories ou pour tester ou renforcer la validité externe des expériences de laboratoire (expériences de terrain, essais randomisés à grande échelle, expériences en contexte « naturel », expériences de neuroéconomie, etc.). En parallèle, la discipline est aussi utilisée dans le domaine de la pédagogie (Eber et Willinger [2012]), et les expériences réalisées en classe sont nombreuses dans la transmission des concepts économiques. Dans cet article, nous présentons une extension du domaine des possibles dans deux dimensions. La première dimension est celle du terrain de collecte de données, en profitant d'un événement grand public, permettant un recueil massif sur une courte période. La seconde dimension est celle de la pédagogie. L'économie expérimentale est ici utilisée afin de communiquer sur les mécanismes économiques, non pas, comme c'est devenu classique, à destination des étudiants, mais à destination du grand public, et ce, afin d'échanger au sujet d'un thème contemporain. Cela permet de répondre aux faiblesses de diffusion et de sensibilisation des progrès de la science économique soulignées par le rapport Guesnerie [2008] sur l'enseignement des sciences économiques et sociales dans le secondaire.

Le projet a été mis en place en réponse à l'appel à projets lancé par les organisateurs de la Nuit européenne des chercheurs pour l'organisation d'une « Grande expérience participative » à conduire dans onze villes françaises prédéfinies. Cette Grande expérience participative devait être un événement de rencontre ludique entre les chercheurs et le public. Le projet devait donc s'inscrire dans un cahier des charges décrivant quatre objectifs et définissant pour nous, les chercheurs, autant de contraintes : 1) en analysant la façon dont les individus allaient adapter leurs stratégies empiriques en présence de variations de traitement dans le protocole, notre travail devait remplir un objectif scientifique et académique, à savoir apporter des informations sur la prise de décisions ; 2) cette expérience devait conduire à la diffusion de ces nouvelles connaissances à travers des publications spécialisées académiques et professionnelles en économie ; 3) le projet devait constituer un transfert original de connaissances de la discipline vers un public non spécialisé, et constituer une occasion pour le public d'apprendre et d'être sensibilisé à des enjeux économiques actuels ; 4) enfin, ce projet avait pour objectif une contribution méthodologique, c'est-à-dire la mise en œuvre de méthodes expérimentales valides sur un terrain « grand public », afin de partager ces méthodes et de faciliter l'investigation de nouveaux terrains.

La plupart des articles de méthodologie expérimentale en économie identifient trois catégories d'objectifs poursuivis par l'économie expérimentale (selon Roth [1986], [1988], [1995] : tester la théorie, rechercher des faits ou parler aux décideurs). Nous ajoutons, étant donné notre terrain, un objectif supplémentaire,

Revue économique – vol. 68, n° 5, septembre 2017, p. 941-953

celui de parler au « public », c'est-à-dire aux sujets expérimentaux, tout en respectant les canons de la discipline (contrôle expérimental, anonymat, incitations monétaires, etc.).

Mettre en place une expérience dans un contexte qui est à la fois *terrain* et *scène* implique la transmission d'une connaissance tacite au sens de Polanyi [1967], et donc, en parallèle avec la mise en place d'un protocole d'expérience, l'explication de ce protocole, des pratiques et compétences qui le composent et qui sont habituellement cachées à ceux qui ne sont pas experts. Il s'agissait ainsi de créer un cadre d'apprentissage expérientiel pour le public. Si l'on se réfère à la théorie de l'apprentissage de Kolb [1984], l'apprentissage expérientiel est efficace pour assurer cette transmission en cela que le spectateur devient aussi acteur de la prise de décision. Nous avons donc proposé une « approche expérientielle » au public, qui était en même temps notre sujet expérimental, et la rencontre a été basée sur trois principes : 1) l'implication du public dans l'échange avec le chercheur (par la participation à un jeu expérimental), 2) une coconstruction de la connaissance véhiculée dans les échanges (par l'explication *a posteriori* de la construction, de la méthode et des objectifs de l'expérience), et 3) un engagement fort du public dans cet apprentissage qu'il venait chercher dans un événement de vulgarisation scientifique (à travers la mise en place d'incitations réelles à la prise de décision, comme dans une véritable séance de laboratoire, et le choix d'un sujet d'expérience auquel le public était susceptible d'être sensible).

Nous devions par conséquent déterminer pour cette expérience un thème qui puisse concilier ces divers objectifs et répondre aux nombreuses contraintes. Si une démarche d'économie expérimentale peut en soi contribuer à la transmission d'un savoir, nous avons trouvé important de confronter les sujets à un thème pour lequel les besoins pédagogiques existent. Ainsi, c'est le thème de la tragédie des biens communs qui a été choisi pour cette expérience, afin de traiter d'un sujet d'intérêt général permettant un échange aisé avec les participants. Notre expérience a donc été construite de manière à répondre aux contraintes de la discipline et aux contraintes d'un exercice particulier. Le protocole expérimental utilisé lors de la grande expérience participative visait ainsi à reproduire, dans un contexte simplifié, la logique des décisions environnementales.

Cette expérience ainsi que les méthodes mises en œuvre sont décrites dans cet article de la façon suivante. La première section traite du contexte et des contraintes expérimentales propres au protocole. La deuxième explique le choix de la tragédie des communs. La troisième décrit la façon dont nous adaptons ce thème au terrain d'application. Le protocole expérimental et l'échantillon sont présentés dans la quatrième section. Enfin, nous concluons en tirant les enseignements de cette expérience de médiation scientifique recourant à l'économie expérimentale et en présentant des éléments de partage avec le public après l'expérience.

CONTEXTE ET CONTRAINTES EXPÉRIMENTALES

Cette expérience a été retenue lors d'un appel à projets dans le cadre de la dixième Nuit européenne des chercheurs[1], pour la France, intitulé « Grande expérience participative ». Cet événement est une occasion unique de médiation

1. Site Internet de la Nuit européenne des chercheurs : http://www.nuitdeschercheurs-france.eu/.

scientifique, moins institutionnelle que la Fête de la science, qui se déroule le soir, généralement dans un lieu emblématique. Elle permet au grand public de venir à la rencontre des chercheurs de leur ville et de les questionner sur leurs sujets, leurs pratiques, leur pensée, etc. Si dans les autres pays européens, cette soirée a lieu uniquement dans une ville qui change tous les ans, en France elle a lieu simultanément dans onze villes : Angers, Besançon, Bordeaux, Brest, Dijon, Le Mans, Limoges, Lyon, Metz, Paris et Toulouse.

Le cahier des charges de la Grande expérience participative indique que cette expérience doit pouvoir être conduite dans toutes les villes concernées, auprès d'un grand nombre de personnes en une seule soirée. Elle doit également être compréhensible par tous et donner lieu à des restitutions, à la suite des sessions expérimentales, d'une part, et après l'expérience, dans le cadre de communications scientifiques, expertes et de vulgarisation, d'autre part.

Les contraintes logistiques étaient les suivantes. Premièrement, il s'agissait de rassembler une équipe de scientifiques capables de conduire l'expérience dans onze villes. Le public souhaite, en effet, rencontrer de véritables chercheurs, qui parlent de leurs recherches. Il fallait donc que chacun d'entre nous maîtrise le domaine de l'expérience. Deuxièmement, il fallait mettre en place un protocole expérimental court, compréhensible par tous rapidement (la Nuit des chercheurs est un événement très grand public, qui accueille des visiteurs de tous âges) et qui ne demandait pas la mise en place de procédures informatiques compliquées (dans certaines villes, nous ne pouvions pas assurer une expérience à travers un programme informatique, du fait de l'affluence attendue et de la disponibilité des équipements, et pour ne pas limiter l'accès aux seuls détenteurs de smartphones, par exemple). Troisièmement, il était nécessaire que l'expérience permette un échange au sujet de son objet et de sa finalité. Les participants venaient surtout pour apprendre et pas uniquement pour jouer le rôle de cobayes. Quatrièmement, il nous fallait respecter les canons de la méthode expérimentale, afin que les productions issues de cette expérience puissent donner lieu à publication dans des revues scientifiques de référence. Enfin, il fallait mettre en place un protocole suffisamment flexible pour s'adapter à un nombre incertain et variable de participants par session. En effet, l'expérience allait avoir lieu dans des endroits très différents (grands amphithéâtres pouvant accueillir des centaines de personnes, petites salles avec un nombre fixe de personnes par session, places publiques, etc.). Nous devions donc prévoir une expérience qui pouvait être réalisée avec des sessions de tailles différentes. Le public de la Nuit des chercheurs visite les ateliers des chercheurs au gré de leurs envies. Il se pouvait ainsi qu'une session se déroulât avec quatre visiteurs seulement et une autre avec 400, en fonction de l'affluence.

Ainsi, la première étape a consisté à rassembler des chercheurs pouvant contribuer à la conception du protocole et assurer la mise en œuvre de l'expérience dans les différentes villes. La mobilisation a été réalisée à travers un appel au sein des chercheurs membres de l'Association française d'économie expérimentale (ASFEE). Les principes (et contraintes méthodologiques et matérielles) de l'expérience ont été ensuite présentés aux coordinateurs de la Nuit européenne des chercheurs de chaque ville.

Une partie de l'équipe rassemblée s'est ensuite chargée de préparer le protocole expérimental. Partant d'une idée de gestion de ressources naturelles communes, notre objectif était d'aboutir à un protocole simple et compréhensible par tous et permettant de produire une connaissance scientifique originale. Cette partie a donné lieu à un très grand nombre d'échanges et de discussions sur le type de jeu testé et sur les différents paramètres du protocole expérimental : les différents traitements, la définition de l'ensemble des décisions possibles pour les participants, le niveau d'information diffusé dans les différents traitements, le niveau de contextualisation et de réalisme, les incitations mises en œuvre. L'introduction d'une incitation réelle a fait l'objet de longues discussions. Nous avons opté pour des bons d'achats de 10 euros dont le rôle sera expliqué après la justification du choix du thème scientifique.

Nous avons ainsi abouti à un protocole expérimental dont les instructions tenaient sur une seule feuille et qui allaient être détaillées à l'oral par les expérimentateurs à l'aide d'une présentation PowerPoint. L'expérience a été conçue pour durer de 15 à 20 minutes et laisser une dizaine de minutes pour un échange avec les participants, sachant que la durée totale maximale d'une session fixée par les organisateurs de la Nuit des chercheurs était de 30 minutes.

LE CHOIX DU THÈME EXPÉRIMENTAL : LA TRAGÉDIE DES BIENS COMMUNS

La tragédie des biens communs est un des jeux « classiques » de l'économie expérimentale. L'exemple le plus standard de la tragédie des communs est l'épuisement des ressources naturelles communes du fait de leur surexploitation. Qu'il s'agisse de ressources halieutiques ou forestières, de minerais ou d'eau, une exploitation plus intense que la vitesse de reproduction de la ressource conduit au fil du temps à sa disparition. Nous avons choisi ce thème parce qu'il était familier à l'ensemble des chercheurs. De plus, de nombreuses expériences ont été conduites avec succès à ce sujet dans de nombreux contextes (du laboratoire à des terrains avec les personnes directement concernées, voir par exemple Cardenas, Janssen et Bousquet [2013] et Ostrom [2006]). Nous pouvions donc espérer que notre tâche allait être facilitée par une réception favorable du public. Nous avons aussi choisi ce thème parce qu'il constitue toujours un sujet d'actualité qui concerne les citoyens. Un élément déterminant de notre choix a aussi été le fait que, depuis Gordon [1954], le problème de la tragédie des biens communs est formalisé. Il résulte en effet de la volonté de chacun de maximiser son gain individuel au détriment de la recherche d'une solution socialement optimale. C'était donc un cadre idéal que nous pouvions exposer au public afin de lui montrer que les comportements font l'objet de modélisations et peuvent être anticipés de manière théorique en économie comportementale. De plus, ce problème est un sujet pris en charge par les décideurs publics ou privés (élus politiques, représentants de la société civile, etc.), qui cherchent à agir sur les variables susceptibles d'accroître la propension des individus à se diriger vers une autogouvernance durable (Ostrom [1999]).

Afin de conserver un protocole simple et qui parle au public de la Nuit européenne des chercheurs, nous nous sommes focalisés sur trois variables susceptibles d'influencer les décisions dans notre expérience. La première variable est le taux de renouvellement des ressources (faible ou fort ; cet élément est facilement intégrable par le grand public qui connaît dans la nature des ressources épuisables qui se renouvellent rapidement ou lentement). La deuxième variable est l'information dont disposent les individus sur les seuils de soutenabilité lors de l'exploitation de ces ressources (le public est sensibilisé par les décideurs et les médias sur les points de non-retour dans l'exploitation des forêts, par exemple). Enfin, la troisième variable est l'échelle de disponibilité de la ressource, c'est-à-dire son caractère local ou global (l'expérience allait avoir lieu dans onze villes françaises et le public en avait connaissance, ainsi la dimension locale et globale des décisions était un avantage à prendre en compte et aussi un élément facile à expliquer).

LE PROTOCOLE EXPÉRIMENTAL : ADAPTATION DU THÈME SCIENTIFIQUE AU TERRAIN D'APPLICATION

Dans le cadre des contraintes présentées précédemment, nous avons été conduits à faire des choix dans notre protocole et à hiérarchiser l'importance des canons d'économie expérimentale à respecter.

Une première adaptation vis-à-vis des règles habituelles de réalisation des expériences a consisté à contextualiser la situation de prise de décision. Après avoir réalisé plusieurs pilotes du protocole, nous avons conclu que ce format serait plus facile à transmettre au public. Ainsi, les participants se voient attribuer un rôle et sont exposés à une ressource naturelle au niveau local et au niveau national. Ils sont ainsi soit des bûcherons, soit des pêcheurs. S'ils sont des bûcherons, ils peuvent couper des arbres à la fois dans un bois auquel seuls les dix bûcherons de leur ville peuvent accéder à chaque génération et dans une forêt à laquelle les cent bûcherons du pays ont accès à chaque génération. S'ils sont des pêcheurs, ils peuvent pêcher des poissons à la fois dans un lac auquel seuls les dix pêcheurs de leur ville peuvent accéder à chaque génération et dans la mer à laquelle les cent pêcheurs du pays ont accès à chaque génération.

Dans le cadre de l'expérience, les participants étaient répartis aléatoirement entre quatre traitements et répondaient à deux questions (*cf.* tableau 1). Les traitements sont le croisement des informations sur le taux de reproduction de la ressource naturelle et sur le taux de soutenabilité de l'extraction de cette ressource. Chaque participant choisit dans le contexte auquel il est assigné le nombre d'unités (entre 0 et 5) de la ressource naturelle qu'il souhaite extraire au niveau local et au niveau national. Les fiches de décision des participants sont présentées en annexe.

Il est précisé que, à l'issue de l'expérience, les participants seront placés aléatoirement dans des générations de cent personnes au niveau national et de dix personnes au niveau local. Ils ne savent pas à quelle génération ils appartiennent, ni combien de générations il y a avant eux et après eux. La seule information est qu'il peut y avoir une génération avant eux et une génération après eux. Cela reproduit des situations réelles d'épuisement de ressources dans lesquelles les citoyens ont généralement connaissance du caractère épuisable de la ressource, mais ne savent pas exactement quel est le stock disponible, ni depuis combien de temps la ressource est exploitée, ni s'ils sont les premiers à l'exploiter (situation d'abondance) ou, au contraire, s'ils sont parmi les derniers à en bénéficier.

Revue économique – vol. 68, n° 5, septembre 2017, p. 941-953

Cette astuce du protocole nous permet aussi de faire face à la contrainte liée à la taille des sessions. Habituellement, en matière de jeux sur les prélèvements des ressources naturelles, les sessions sont conduites avec un nombre fixe d'individus. Notre variation nous permet de reconstituer des générations *a posteriori*, et donc de pouvoir conduire des sessions avec des nombres de participants différents à la fois entre les villes et d'une session à l'autre au sein d'une même ville.

Ce qui distingue les deux ressources naturelles, c'est leur niveau de reproduction entre deux générations : 5 % pour les arbres et 15 % pour les poissons. Cela mime la réalité (un arbre arrive à maturité plus difficilement qu'un poisson) et est facilement compréhensible par le public. Par ailleurs, quelle que soit la ressource, le stock initial est de 230 unités au niveau local (lac et bois) et de 2 300 unités au niveau national (mer et forêt). Cela nous permet de comparer les prélèvements dans chaque traitement et contexte de façon symétrique.

Un traitement informationnel est ajouté. En effet, dans certaines sessions, nous informons les participants sur le taux de soutenabilité des ressources naturelles, c'est-à-dire le niveau d'extraction permettant à la génération suivante de bénéficier du même stock de ressource que la génération actuelle. Que ce soit au niveau local ou au niveau national, ce taux est d'un arbre par bûcheron par génération et de trois poissons par pêcheur et par génération. Cet élément mime l'existence d'annonces publiques sur les points de non-retour.

Tableau 1. *Synthèse des traitements*

	Sans information	Avec information
Bûcherons / Arbres	(ASI) 10 bûcherons autour du bois 100 bûcherons aux abords de la forêt Taux de reproduction : 5 %	(AAI) 10 bûcherons autour du bois 100 bûcherons aux abords de la forêt Taux de reproduction : 5 % Niveau de soutenabilité : 1 arbre
Pêcheurs / Poissons	(PSI) 10 pêcheurs autour du lac 100 pêcheurs aux abords de la mer Taux de reproduction : 15 %	(PAI) 10 pêcheurs autour du lac 100 pêcheurs aux abords de la mer Taux de reproduction : 15 % Niveau de soutenabilité : 3 poissons

Chaque participant pouvait extraire entre 0 et 5 unités de ressource au niveau local et entre 0 et 5 unités de la même ressource au niveau national. Ils savaient qu'ils partageaient cette expérience avec d'autres participants présents dans d'autres villes de France au même moment (la liste des villes leur est communiquée). En termes d'incitation, chaque unité de ressource extraite donne le droit de participer à un tirage au sort lors duquel 300 bons d'achats de 10 euros sont à gagner.

Si l'incitation repose en partie sur du hasard, elle n'en est pas moins réelle et crédible (à en juger par les commentaires qui nous ont été faits par des participants). Dans ce cadre, nous indiquons que les participants sont ordonnés au sein de leur génération, que ce soit au niveau local ou au niveau national. Cela a une conséquence sur les unités de ressources éligibles au tirage au sort. En effet, s'il y a assez de poissons ou d'arbres pour sa génération, chaque participant reçoit autant de tickets de loterie qu'il a extrait de poissons ou d'arbres. En revanche, s'il n'y a pas assez de poissons et d'arbres pour toute sa génération, le rang du participant au sein de sa génération est pris en compte. Les premiers reçoivent autant de tickets de loterie qu'ils ont pris de poissons ou d'arbres, puis un participant reçoit moins de tickets de loterie que ce qu'il a pris en poissons ou en arbres, et les derniers ne reçoivent aucun ticket de loterie. Enfin, si la génération précédente n'a laissé aucun poisson ou arbre, alors les participants de la génération concernée ne reçoivent aucun ticket de loterie.

DÉROULEMENT DE L'EXPÉRIENCE ET PARTICIPANTS

Chaque session expérimentale s'est déroulée de la façon présentée dans le tableau 2, selon une procédure classique en laboratoire.

Tableau 2. *Procédures*

Étape	Activité
Étape 1	Accueil des participants avec distribution d'une fiche d'instructions et de réponse et de crayons
Étape 2	Lecture des instructions présentant le contexte et les rôles des participants
Étape 3	Explication du principe de génération
Étape 4	Description des stocks disponibles au démarrage de l'expérience
Étape 5	Description des décisions à prendre
Étape 6	Explication de la reproduction des ressources entre deux générations
Étape 7	Information sur le taux de soutenabilité, le cas échéant
Étape 8	Description et explication des incitations
Étape 9	Prise de décision par les participants, questionnaire postexpérimental et collecte des fiches
Étape 10	Information des participants sur le mode de publication des résultats
Étape 11	Échange avec les chercheurs présents au sujet de la méthode expérimentale, de l'expérience, de la tragédie des communs, de l'économie (sans présenter les résultats attendus du protocole)
Étape 12	Évacuation de la salle

Les participants ont été recrutés parmi les gens qui se rendaient à la Nuit européenne des chercheurs selon différentes modalités. Dans certaines villes, il était annoncé que les visiteurs pouvaient participer librement à des sessions d'expérience toutes les demi-heures. Lorsque l'expérience se déroulait dans un grand amphithéâtre, nous acceptions dans la salle tous les visiteurs intéressés. Si la salle avait une capacité limitée, les visiteurs étaient invités à revenir plus tard dans la soirée. Dans d'autres villes, un nombre limité de sessions avec un nombre fixe de places était prévu dans la soirée, à des heures précises. Les visiteurs s'inscrivaient sur ces créneaux-là.

2 813 personnes ont participé à cette expérience le 25 septembre 2015. 2 723 personnes ont répondu aux quatre questions essentielles pour l'analyse (extraction locale, extraction nationale, âge, sexe), soit 96,8 % des participants. La moyenne d'âge est de 28 ans (allant de 2 à 91 ans), avec des différences entre villes, et 55 % des participants sont des femmes. Le tableau 3 détaille le nombre de participants et d'observations par ville, ainsi que les âges moyens et la proportion de femmes. Globalement, la répartition entre les quatre traitements est équilibrée. Entre 597 et 784 personnes ont répondu dans chaque condition de contexte et d'information.

Tableau 3. *Effectifs et description des participants par ville*

	Part.	Obs.	Âge	Femmes	Obs. PSI	Obs. PAI	Obs. ASI	Obs. AAI
France	2 813	2 723	28,0	0,552	746	623	764	590
Angers	207	202	31,8	0,519	50	51	53	48
Besançon	224	221	28,2	0,583	52	53	62	54
Bordeaux	303	273	23,3	0,582	77	59	84	53
Brest	726	698	26,4	0,584	177	182	182	157
Dijon	487	478	29,5	0,546	142	97	142	97
Le Mans	123	122	35,1	0,475	51	10	53	8
Limoges	122	122	24,9	0,377	31	29	32	30
Lyon	168	165	28,1	0,563	54	31	48	32
Metz	129	128	30,0	0,625	23	41	23	41
Paris	126	117	31,1	0,461	39	21	37	20
Toulouse	198	197	26,8	0,568	50	49	48	50

Note : Part. : Participants ; Obs. : Observations ; PSI : Poissons sans information ; PAI : Poissons avec information ; ASI : Arbres sans information ; AAI : Arbres avec information.

L'analyse des données donnera lieu à une publication scientifique ultérieure, ainsi qu'à une diffusion sur le blog de la Grande expérience participative au sein du site Internet de la Nuit européenne des chercheurs.

CONCLUSION : UNE EXPÉRIENCE DE MÉDIATION SCIENTIFIQUE

La réalisation de cette expérience nous a permis, outre la partie académique classique, de nous plonger dans une expérience de vulgarisation scientifique. Cela permet de faire la transition de la *production* de données expérimentales à la *description* de ce processus. Cette transition implique la prise en compte de la critique de Collins [1985] qui précise qu'elle est possible uniquement lorsque, de manière explicite, la distinction peut être faite entre une « bonne » et une « mauvaise » expérience. Cette distinction doit être crédible et internalisée à la fois par les chercheurs de la discipline, dans la construction du protocole, mais aussi à l'extérieur de la discipline, par le public, qui doit comprendre nos choix. En économie expérimentale, plusieurs normes et pratiques unanimement acceptées par la communauté existent : la règle de non-tromperie des participants, l'existence des incitations, la connaissance commune des instructions, etc. Cela facilite la tâche en termes de partage des procédures mises en œuvre, mais nous oblige à faire des choix. La Grande expérience participative a été un événement unique, et l'exercice s'est situé à la croisée d'ambitions multiples (construire un protocole valide, récolter des données scientifiquement utilisables, vulgariser la méthode, vulgariser le thème, responsabiliser le public, etc.) et de contraintes logistiques, méthodologiques et institutionnelles. Nous avons donc entrepris d'expliquer comment nous les avons conciliées et avons fait des choix, en présentant, dans un premier temps, les principes de l'économie expérimentale et comportementale, et, dans un deuxième temps, comment nous avons construit le protocole pour pouvoir les respecter.

Au-delà de cet effort de vulgarisation, cette expérience nous a permis, d'une part, de nous interroger sur l'importance relative des principes de l'économie expérimentale : doit-on privilégier, si des choix sont à faire, la neutralité du contexte, les incitations monétaires, le contrôle, la non-tromperie des participants ? En effet, Guala [2005] précise que le savoir expérimental est construit, et qu'il est le fruit de règles et conventions qui doivent être reproduites, quel que soit le terrain de l'expérience. À ce titre, d'une part, Friedman et Sunder [1994] indiquent par exemple que l'économie expérimentale utilise des données créées dans des conditions contrôlées. Le contrôle constitue ainsi l'une des contraintes fortes à remplir lorsque le laboratoire est déplacé dans un événement de médiation scientifique. Avec un design astucieux, nous avons réussi à conserver tous ces principes dans notre protocole, en relâchant seulement la neutralité du contexte, et sans introduire un biais expérimental de demande (parce que la situation évoquée correspond à une situation très proche de la prise de décision réelle). D'autre part, ce protocole n'a pas été seulement porteur de contraintes, mais aussi d'opportunités. En effet, nous avons pu explorer un protocole générationnel qui n'était pas réalisable facilement en laboratoire, une situation d'incertitude sur la place de l'individu au sein d'une génération et de cette génération au fil du temps, ainsi que la distinction réelle par ville et par ressource locale-nationale. Enfin, nous avons pu collecter un nombre record de données (2 700 participants) en quelques heures.

De ce fait, la question de la sélection de notre échantillon s'est posée. D'après les statistiques de fréquentation de la Nuit des chercheurs, l'événement est réellement « grand public » : il n'y a pas de surreprésentation d'une tranche d'âge, d'un sexe, d'une catégorie socioprofessionnelle. Par ailleurs, notre échantillon est inégalement réparti en fonction des villes. Cela est dû à une organisation très différente de l'événement en fonction de la ville : si à Brest l'endroit accueille plus de 4 000 visiteurs dans la soirée, à Paris l'événement est volontairement conservé « petit » par les organisateurs.

Lors du débriefing avec le public, la mission pédagogique a consisté à introduire aussi le dilemme du bien commun, la tragédie des communs, en plus des éléments sur la méthodologie de l'économie expérimentale et, plus généralement, sur les sciences économiques (en répondant à des questions telles que « qu'est-ce que l'économie ? »). Cet objectif rejoint l'utilisation de l'économie expérimentale

Revue économique – vol. 68, n° 5, septembre 2017, p. 941-953

comme outil à visée pédagogique (Eber et Willinger [2012]). De façon générale, l'expérience est l'occasion de présenter le mode de penser de l'économiste à travers les concepts d'arbitrage, d'opportunité, d'optimisation, en faisant réfléchir les participants sur la façon dont ils ont agi (sans être normatif). L'approche par une expérience vécue permet également d'aborder les questions de reproductibilité des expériences, notamment à travers les contraintes de contrôle, ainsi que de comparer les différentes expériences portant sur le même sujet.

De plus, nous avons insisté sur le fait que, dans un contexte de raréfaction des ressources disponibles, l'examen des potentialités offertes par la mobilisation du « capital comportemental » des individus peut être un outil efficace pour améliorer la compréhension de la résistance des acteurs aux changements, notamment afin de favoriser la transition écologique. Nous avons introduit des exemples sur les leviers comportementaux susceptibles d'être activés afin de lever cette résistance et d'aboutir à des modifications comportementales durables. Pour cela, nous avons repris la définition d'utilisation et de préservation de l'environnement de Milfont et Duckitt ([2010], p. 81), selon laquelle « la préservation de l'environnement exprime la croyance générale que la priorité doit être donnée à la préservation de la nature et de la diversité des espèces naturelles dans leur état naturel originel, et à leur protection contre leur utilisation et leur altération par l'être humain », alors que l'utilisation de l'environnement exprime « la croyance générale qu'il est légitime, approprié et nécessaire que la nature et tous les phénomènes et espèces naturels soient utilisés et altérés pour répondre aux objectifs humains ». Cela nous a permis de discuter d'écocomportement avec le public, d'échanger au sujet des intérêts collectifs et personnels. Il s'agissait d'expliquer que l'individu va devoir effectuer une exploitation « raisonnable » de l'environnement, qui lui permettra non seulement de subvenir à ses besoins, mais qui pourra aussi permettre à des individus actuels ou futurs susceptibles de partager les mêmes ressources, de subvenir à leurs propres besoins. La pérennité des ressources doit ainsi être assurée à long terme. En surexploitant l'environnement, l'individu « non seulement diminue les possibilités offertes aux autres individus d'obtenir une part raisonnable de ressources, mais il impacte également la capacité de l'environnement à se régénérer sur le long terme » (Ajdukovic [2015]), et donc empiète sur la capacité des générations futures de prélever des ressources.

Nous avons enfin discuté avec les participants du dilemme de l'exploitation : alors que de nombreuses ressources sont communes, aucun utilisateur ne peut être privé de leur utilisation, même abusive (par exemple : la faune, les ressources en eau, l'air pur, etc.). Cela conduit à la « tragédie des biens communs » lorsque le comportement humain est entraîné par la maximisation des gains individuels et non par le désir de parvenir à une solution socialement optimale. Alors que les enquêtes confirment un désir largement répandu dans la population de prendre soin de telles ressources, des preuves empiriques de plus en plus alarmantes montrent leur surexploitation, pour certaines à un point de non-retour. En économie expérimentale, à travers des jeux incitatifs, nous avons expliqué pouvoir faire la différence entre ce que les gens déclarent et ce qu'ils font et ainsi éviter le biais déclaratif.

RÉFÉRENCES BIBLIOGRAPHIQUES

AJDUKOVIC I. [2015], *Appropriation et valorisation sociale de l'orientation pro-environnementale*, thèse de doctorat en psychologie, Université de Bourgogne, Dijon.

CARDENAS J., JANSSEN M. et BOUSQUET F. [2013], « Dynamics of Rules and Resources: Three New Field Experiments on Water, Forests and Fisheries », dans J. A. LIST et M. K. PRICE (dir.), *Handbook on Experimental Economics and the Environment*, Cheltenham, Edward Elgar, p. 319-345.

COLLINS H. M. [1985], *Changing Order*, Beverly Hills, Sage.

EBER N. et WILLINGER M. [2012], *L'économie expérimentale*, Paris, La Découverte.

FRIEDMAN D. et SUNDER S. [1994], *Experimental Methods: A Primer for Economists*, Cambridge, Cambridge University Press.

GORDON H. S. [1954], « The Economic Theory of a Common-Property Resource: The Fishery », *Journal of Political Economy*, 62 (2), p. 124-142.

GUALA F. [2005], *The Methodology of Experimental Economics*, Cambridge, Cambridge University Press.

GUESNERIE R. [2008], *Rapport au Ministre de l'Éducation nationale de la mission d'audit des manuels et programmes de sciences économiques et sociales du lycée*.

KOLB D. A. [1984], *Experiential Learning: Experience as the Source of Learning and Development*, New Jersey, Prentice-Hall.

MILFONT T. L. et DUCKITT J. [2010], « The Environmental Attitudes Inventory: A Valid and Reliable Measure to Assess the Structure of Environmental Attitudes », *Journal of Environmental Psychology*, 30 (1), p. 80-94.

OSTROM E. [1999], « Coping with Tragedies of the Commons », *Annual Review of Political Science*, 2 (1), p. 493-535.

OSTROM E. [2006], « The Value-Added of Laboratory Experiments for the Study of Institutions and Common-Pool Resources », *Journal of Economic Behavior & Organization*, 61 (2), p. 149-163.

POLANYI M. [1967], *The Tacit Dimension*, New York, Anchor.

ROTH A. E. [1986], « Laboratory Experimentation in Economics », *Economics and Philosophy*, 2 (2), p. 245-273.

ROTH A. E. [1988], « Laboratory Experimentation in Economics: A Methodological Overview », *The Economic Journal*, 98, p. 974-1031.

ROTH A. E. [1995], « Introduction to Experimental Economics », dans J. H. KAGEL et A. E. ROTH (dir.), *The Handbook of Experimental Economics*, Princeton, Princeton University Press, p. 3-109.

ANNEXE

EXEMPLE D'INSTRUCTIONS, TRAITEMENTS SANS INFORMATION

Poissons sans information - PSI

Vous êtes un pêcheur

Et vous pouvez pêcher dans deux lieux différents :

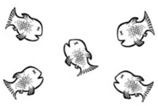
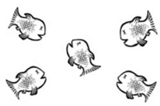
Un lac auquel seuls les 10 pêcheurs de votre ville ont accès, à chaque génération.

La mer à laquelle les 100 pêcheurs du pays ont accès, à chaque génération.

A chaque génération, un nombre fixe de pêcheurs ont accès au lac et à la mer. Il y a des générations avant vous et après vous. La génération à laquelle vous appartenez sera tirée au sort après la fin de la Grande Expérience.

Le stock initial de poissons est fixé au début de la Grande Expérience. A chaque génération, chaque pêcheur peut pêcher entre 0 et 5 poissons dans chacun des deux lieux. Les poissons restants vont se reproduire et leur population augmente de 15% pour la génération suivante.

ENTOUREZ et/ou COCHEZ LE NOMBRE DE POISSONS QUE VOUS SOUHAITEZ PÊCHER DANS LE LAC ainsi QUE DANS LA MER

LAC	MER
Stock initial : 230 poissons	Stock initial : 2300 poissons
Nombre de pêcheurs par génération : 10	Nombre de pêcheurs par génération : 100
Taux de reproduction des poissons par génération : 15%	Taux de reproduction des poissons par génération : 15%
	
0 1 2 3 4 5	0 1 2 3 4 5

Afin de recevoir vos gains en cas de succès à la loterie, merci de répondre aux questions suivantes:

Age :	
Sexe :	Féminin ☐ Masculin ☐
Téléphone ou adresse email :	
Souhaitez-vous recevoir des invitations à participer à d'autres expériences à l'avenir ?	Oui ☐ Non ☐

Arbres sans information - ASI

Vous êtes un bûcheron

Et vous pouvez couper des arbres dans deux lieux différents :

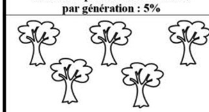
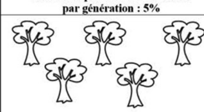
Un bois auquel seuls les 10 bûcherons de votre ville ont accès, à chaque génération.

La forêt à laquelle les 100 bûcherons du pays ont accès, à chaque génération.

A chaque génération, un nombre fixe de bûcherons a accès au bois et à la forêt. Il y a des générations avant vous et après vous. La génération à laquelle vous appartenez sera tirée au sort après la fin de la Grande Expérience.

Le stock initial d'arbres est fixé au début de la Grande Expérience. A chaque génération, chaque bûcheron peut couper entre 0 et 5 arbres dans chacun des deux lieux. Les arbres restants vont se reproduire et leur population augmente de 5% pour la génération suivante.

ENTOUREZ et/ou COCHEZ LE NOMBRE D'ARBRES QUE VOUS SOUHAITEZ COUPER DANS LE BOIS ainsi QUE DANS LA FORÊT

BOIS	FORÊT
Stock initial : 230 arbres	Stock initial : 2300 arbres
Nombre de bûcherons par génération : 10	Nombre de bûcherons par génération : 100
Taux de reproduction des arbres par génération : 5%	Taux de reproduction des arbres par génération : 5%
	
0 1 2 3 4 5	0 1 2 3 4 5

Afin de recevoir vos gains en cas de succès à la loterie, merci de répondre aux questions suivantes:

Age :	
Sexe :	Féminin ☐ Masculin ☐
Téléphone ou adresse email :	
Souhaitez-vous recevoir des invitations à participer à d'autres expériences à l'avenir ?	Oui ☐ Non ☐

PUBLIATION 11: Fabrice Le Lec, Theo Alexopoulos, Béatrice Boulu-Reshef, Marie-Pierre Fayant, Todd Lubbart, Franck Zenasni et Nicolas Jacquemet (2017), *The out-of-my-league effect*, comment on *Explaining financial and prosocial biases in favor of attractive people: Interdisciplinary perspectives from economics, social psychology, and evolutionary psychology* by Dario Maestripieri, Andrea Henry, Nora Nickels, **Brain and Behavioral Sciences**, 40.

Courtship behavior: "The Out-of-my-league effect"

Fabrice Le Lec, Théodore Alexopoulos, Béatrice Boulu-Reshef, Marie-Pierre Fayant, Franck Zenasni, Todd Lubart, Nicolas Jacquemet[1]

Comment on Maestripieri, Henry, Nickels: Explaining financial and prosocial biases in favor of attractive people: Interdisciplinary perspectives from economics, social psychology, and evolutionary psychology", forthcoming in Behavioral and Brain Sciences.

To explain financial and prosocial biases towards attractive adults, Maestripieri et al. defend a "strategic mating behavior" account. Their central argument relies on a causal relationship between viewing attractive individuals (A); a host of cognitive, emotional, and physiological changes (B); and financial/social generosity or other desirable behaviors (C). Yet, their reasoning is based on data from a collection of different experimental studies, and one cannot reliably determine how much (if any) of the effect of A (e.g., attractiveness) on C (e.g., financial decision) is actually explained by B (e.g., testosterone). Their review provides therefore no definitive evidence that mating motives or their proxies (e.g., physiological changes) are the actual causes of an attractiveness bias. There are in fact theoretical reasons to doubt the accuracy of a causal effect. This comment will focus on the idea that strategic mating behavior does not generally imply that favors should increase with attractiveness: a phenomenon we label the *out-of-my-league effect*.

The target article's argumentation is grounded on a mating model in which only the benefits of mating with attractive people are considered. This completely overlooks the effect of the probability of success in mating: A simple model of courtship behavior should take into account not only the benefit of mating with an attractive individual, but also the probability of doing so – itself determined by the mating opportunities of others. The potential "court maker," if motivated solely by mating per se as hypothesized, faces a trade-off between the benefit of mating and the probability of success: Whereas the former increases, the later realistically decreases with the attractiveness of the target. For a given attractiveness of the court maker, attractiveness-based matching implies that the probability of success decreases with the attractiveness of the potential mate – as the target's opportunities are likewise based on attractiveness. If the probability of success decreases more steeply than the benefit of mating increases given the potential mate's attractiveness, then the more attractive the potential mate, the lower the expected benefit of a match – that is, the benefit of a match weighted by its probability of occurrence. The out-of-my-league hypothesis states that one should not spend resources to court a very

[1] Le Lec, Boulu-Reshef: University Paris 1 Panthéon-Sorbonne (CES) ; Alexopoulos, Fayant, Zenasni, Lubart: University Paris Descartes (Psychology Institue); Jacquemet: Paris School of Economics & University Paris 1 Panthèon-Sorbonne.

attractive potential mate with a minimal probability of success but rather prefer a moderately attractive one with a reasonable chance of success. In a nutshell: To mate or not to mate is not the question, but rather with whom.

In fact, at the population level, "smart" courtship behavior is more likely to lead on average to a bias towards average-looking individuals: If the distribution of attractiveness is concentrated around its mean (such as in a normal distribution for example), individuals will most likely favor moderately rather than highly attractive mates. In terms of strategy, the average court-making agent is better off targeting individuals of intermediate attractiveness. Not only is it the rational strategy, but it is also the fittest one from an evolutionary standpoint: An individual systematically favoring much more attractive individuals than herself is less likely to mate, and this behavioral pattern is more likely to disappear from the population by evolutionary pressure. This explains the opportunity costs associated with trying to mate with very attractive individuals, which is largely excluded from Maestripieri et al.'s analysis.

In other words, there is no guarantee that the relationship between the level of effort by the courter will be monotonic with the attractiveness of the potential mate – quite the reverse may occur, because positive assortative matching implies that "birds of a feather flock together" (McPherson et al. 2001). The courtship explanation, implicitly based on an endogenous matching model, thus has implications at odds with the set of empirical facts it aims to explain. A complete model of endogenous mating is needed to understand the evolutionary explanation. Such a model has to account not only for the effect of attractiveness on individuals' decision making (as does the target article), but also for the general equilibrium implications of such mating behaviors that are distortions in the probability that mating is actually achieved given the relative attractiveness of the partners.

As mating motives do not necessarily explain the attractiveness bias, it appears premature to reject economic and social psychological explanations. Indeed, the main argument for favoring a mating-based over a stereotype-based account relies on a gender moderation of the attractiveness bias. However, such moderation can be easily explained, for example, from a social psychological perspective. The *opposite-sex* beauty premium effect could simply reflect stereotypic processes. Although the authors reject these based on the fact that an attractiveness bias occurs even when controlling for personality traits and independently from stereotype-induced expectations, one must distinguish the stereotype content from its accuracy and actualization in reality (Judd & Park 1993). Furthermore, as stereotypes operate most of the time on an unconscious level and their influence cannot be captured through explicit self-reports, the reviewed evidence is not a valid rebuttal of a stereotype-based explanation. Moreover, the same-sex negative bias could reflect self-threat because of comparisons and/or competition with attractive individuals. A possible self-threat regulation strategy (among others) relies on derogation and destructive behaviors towards attractive individuals. Such counterproductive responses to threatening comparisons occur routinely in the workplace (Lam et al. 2011). To protect their work environment from negative comparisons, individuals can even provide poor hiring recommendations (Garcia et al. 2010). Crucially, the attractiveness gender bias appears only for individuals who are sensitive to negative comparisons, whereas the rest show a gender-independent attractiveness bias (Agthe et al. 2014).

Although Maestripieri et al.'s "strategic mating behavior" account is scientifically attractive, we shed light on a theoretical argument that goes against their preferred explanation. As the out-of-my-league demonstration suggests: "One's man meat is another man's poison."

References

Agthe, M., Spörrle, M., Frey, D. & Maner, J. K. (2014). Looking up versus looking down: attractiveness-based organizational biases are moderated by social comparison direction. *Journal of Applied Social Psychology* 44:40-45.

Garcia, S. M., Song, H., & Tesser, A. (2010). Tainted recommendations: The social comparison bias. *Organizational Behavior and Human Decision Processes*, *113*(2), 97-101.

Judd, C. M, & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, *100*, 109-128.

Lam, C., K., Van der Vegt, G., S., Walter, F., and Huang, X. (2011). Harming high performers: a social comparison perspective on interpersonal harming in work teams. *Journal of Applied Psycholy.* 96, 588–601.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. Annual review of sociology, 415-444.

Muller, D., & Fayant, M.-P. (2010). On being exposed to superior others: Consequences of self-threatening upward social comparisons. *Social and Personality Psychology Compass, 4*(8), 621-634.

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 21, pp. 181-227). San Diego, CA: Academic Press. π