

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE AUGUSTIN COURNOT ED 221

BUREAU D'ÉCONOMIE THÉORIQUE ET APPLIQUÉE UMR 7522

THÈSE

pour l'obtention du titre de Docteur en Sciences Économiques

Présentée et soutenue le 15 novembre 2024 par

Anne-Gaëlle MALTESE

ESSAYS ON INDIVIDUAL, TEAM, AND ARTIFICIAL CREATIVITY: AN EXPERIMENTAL APPROACH

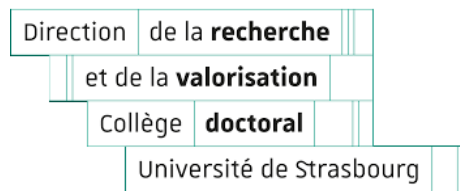
Préparée sous la direction des Professeurs Patrick LLERENA et Giuseppe ATTANASI

Membres du jury :

Marine AGOGUÉ	Professeure, HEC Montréal	Rapporteuse
Daniela GRIECO	Professeure, Université de Milan	Rapporteuse
Agnès FESTRÉ	Professeure, Université Côte d'Azur	Examinatrice
Robin COWAN	Professeur, Université de Strasbourg	Examineur
Patrick LLERENA	Professeur, Université de Strasbourg	Directeur
Giuseppe ATTANASI	Professeur, Université de la Sapienza à Rome	Directeur

À mes parents,
À Yann

L'Université de Strasbourg n'entend donner aucune approbation, ni improbation aux opinions émises dans cette thèse ; elles doivent être considérées comme propres à leur auteur.



Déclaration sur l'Honneur

Declaration of Honour

J'affirme être informée que le plagiat est une faute grave susceptible de mener à des sanctions administratives et disciplinaires pouvant aller jusqu'au renvoi de l'Université de Strasbourg et passible de poursuites devant les tribunaux de la République Française.

Je suis consciente que l'absence de citation claire et transparente d'une source empruntée à un tiers (texte, idée, raisonnement ou autre création) est constitutive de plagiat.

Au vu de ce qui précède, j'atteste sur l'honneur que le travail décrit dans mon manuscrit de thèse est un travail original et que je n'ai pas eu recours au plagiat ou à toute autre forme de fraude.

I affirm that I am aware that plagiarism is a serious misconduct that may lead to administrative and disciplinary sanctions up to dismissal from the University of Strasbourg and liable to prosecution in the courts of the French Republic.

I am aware that the absence of a clear and transparent citation of a source borrowed from a third party (text, idea, reasoning or other creation) is constitutive of plagiarism.

In view of the foregoing, I hereby certify that the work described in my thesis manuscript is original work and that I have not resorted to plagiarism or any other form of fraud.

Nom Prénom : Anne-Gaëlle MALTESE

Ecole doctorale : Augustin Cournot ED 221

Laboratoire : Bureau d'Économie Théorique et Appliquée UMR 7522

Remerciements

Mes remerciements vont en premier lieu à mes directeurs de thèse. À Patrick Llerena, qui m’a accompagnée, soutenue et encouragée depuis mon mémoire de Master. Je ne saurais pas comment traduire en mots toute la gratitude que j’éprouve suite à ces années de travail ensemble. Sa confiance et son ouverture m’ont toujours permise de m’exprimer librement et de me former en tant que chercheuse. Si l’on dit toujours que le choix d’un directeur est la décision la plus importante pour un thésard, je pense avoir pris la meilleure des décisions possibles. Pour votre temps, votre engagement, et pour toutes ces opportunités, merci Patrick !

Je remercie également Giuseppe Attanasi pour son accompagnement dans la réalisation de cette thèse. Ses conseils ont joué un rôle clé dans l’avancement de ce travail. Je lui suis reconnaissante pour la confiance qu’il m’a accordée et pour l’orientation qu’il a su me donner dans mes réflexions.

Je remercie vivement Marine Agogué, Robin Cowan, Agnès Festré et Daniela Grieco pour avoir accepté de compter parmi les membres de mon jury. Je suis honorée que ce travail de thèse puisse être évalué par des chercheurs et chercheuses ayant tant contribué à la recherche en innovation et créativité, ainsi qu’en économie expérimentale.

Je pense également à mes coauteurs qui ont collaboré aux travaux de cette thèse. Merci à Sara Gil Gallen, Pierre Pelletier et Rémy Guichardaz, collègues et aujourd’hui amis avec qui j’ai adoré travailler sur différents projets et, je l’espère, de nombreux travaux futurs. Leur soutien tout au long de cette thèse a été essentiel !

Je tiens aussi à remercier celles et ceux qui ont été à mes côtés au quotidien lors de ma thèse et depuis tant d’années. Merci à mes amis de toujours Romain et Lara, pour leur présence et leur humour à toute épreuve.

Merci au 126, mes collègues de bureau (le meilleur bureau). Si mes voisins ont été nombreux, je tiens à remercier plus particulièrement Romane et Jérôme pour leur bonne humeur au quotidien et pour tous ces moments et éclats de rire qui ont rendu ces journées de travail beaucoup moins longues (Romane, la plus stylée du BETA !)

(Jérôme, l'addict aux frites et le fou de l'interrupteur !). Merci à Ioannis pour nos discussions matinales passionnantes autour d'un café et ces quelques cours de grec (promis je continue mes efforts). Merci également à Benoît, Huy, Arnaud, Mali et Emmanouil.

Merci à tous les autres doctorants qui m'ont permis de créer des souvenirs inoubliables. Des soirées karaoké aux soirées raclette, des pauses déjeuners aux pauses cafés, des soirées Drag show au club couture, tous ces moments ont été incroyables ! Guillaume (pour ses rumeurs qui feraient pâlir n'importe quel journal à scandales), Agathe (ma compatriote du 68 qui nous manque beaucoup), Emilien (le meilleur trésorier des ACDD), Kenza (la plus gaffeuse mais c'est pour ça qu'on l'adore), Antoine (mon voisin d'amphi depuis mon premier jour à la FSEG) mais aussi Jean-Baptiste, Théo, Louis, Eva, Alexandre, Mathilde, Lucie, Julie, Sarah, Nicolas, Pauline, Vincent, Arnaud, Laura, Shengxi, Diletta, Kevin. Je pense aussi à Samuel qui a préféré voguer vers de nouveaux horizons mais qui a toujours soutenu mes projets "puzzles et fric". Enfin, merci à Chloé, tu me manques.

Si je dois ces belles années à mon laboratoire strasbourgeois et les personnes exceptionnelles que j'y ai rencontrées, je n'en oublie pas mon passage pour quelques mois à l'ETH Zürich au sein d'une équipe extraordinaire. Merci à Stefano Brusoni et Daniella Laureiro-Martinez de m'avoir reçue dans leur équipe. Merci à tous ceux avec qui j'ai pu échanger et discuter Axel, Sebastian, Ming, David, Jan, Pari, Anja, Tomoko, Ana et Zorica. J'ai une pensée toute particulière pour Silvia, Maria-Enrica et Oguzhan que j'ai le bonheur mais également la fierté de compter aujourd'hui parmi mes amis.

Un grand merci à ma famille, à ma mère Françoise et mon père Jean-Pierre. Vous êtes les meilleurs parents que j'aurais pu espérer avoir, sans qui je ne serais pas où je suis aujourd'hui. Toujours me permettre de rêver et m'encourager dans mes projets sont les plus belles choses que vous pouviez m'apporter. Merci également à mon frère Laurent pour son soutien au fil des années. Je vous aime plus que tout.

Finalement merci à Yann, mon compagnon, mon meilleur ami et mon meilleur soutien. Je n'aurai jamais pu réaliser tout cela sans toi, sans ta bienveillance et ton amour au quotidien. Je te remercie d'avoir supporté mes vacances perturbées, rendez-vous inopinés et surtout mes entraînements pour chaque présentation (oui, tu connais ma thèse par coeur) depuis 5 ans. Le meilleur reste encore à venir!

PS: j'ai une pensée émue pour la Anne-Gaëlle d'il y a quelques années qui pensait que cette épreuve allait être insurmontable. Tu l'as fait, je suis fière de toi !

Contents

General Introduction	14
Introduction générale	29
1 Rethinking Laboratory Experiments: The Case of Collective Experiments	45
1.1 Introduction	46
1.2 Historical and methodological insights	48
1.2.1 Development of the experimental methodology	48
1.2.2 The fundamental principles	55
1.2.3 The inherent role of the collective in experimental economics	58
1.2.4 Distinguishing factors from other disciplines	60
1.3 Rationale and Implications of Conducting Collective Experiments	63
1.3.1 Definition of collective experiments	64
1.3.2 Current limitations of the method	65
1.4 Necessity of removing intermediaries	67
1.4.1 From the subject to the group	67
1.4.2 The use of intermediaries	68
1.4.3 The resulting intermediary interpretative layer	69
1.4.4 The Collective Experiment Variables Matrix	70
1.5 Necessity for Innovative Tools and Technologies	75
1.5.1 Type of data	76
1.5.2 Required quantity of data	77
1.5.3 Enhancing understanding through technology	79
1.5.4 Integrating multimodal and multidisciplinary approaches	83
1.6 The <i>Social Interactions Lab</i>	84
1.6.1 The purpose	84
1.6.2 The research infrastructure	85
1.6.3 Implementation in this thesis	88
1.7 Conclusion	91
1.8 Appendix	93

2	The Many Faces of Creative Profiles: Exploring Task Openness	98
2.1	Introduction	99
2.2	Literature review	101
2.2.1	Creativity: definition and measures	101
2.2.2	Creative profiles	102
2.3	Experimental design	107
2.3.1	Timeline of the experiments	107
2.3.2	Experimental procedures	108
2.3.3	Hypotheses	110
2.4	Results	112
2.4.1	Creativity assessment	113
2.4.2	First Insights on Creative Performance	115
2.4.3	Regression analysis	119
2.5	Discussion and conclusion	128
2.6	Appendices	132
3	Diversity, Interactions, and Team Creativity: An Experimental Perspective	151
3.1	Introduction	152
3.2	Literature review	154
3.2.1	Individual and team creativity	154
3.2.2	Diversity	157
3.2.3	Social Interactions and Networks	159
3.3	Experimental design	161
3.3.1	A three-stage protocol	161
3.3.2	Experimental procedure	162
3.3.3	Hypotheses	165
3.4	Results	167
3.4.1	Creativity evaluation	168
3.4.2	Regression analysis	170
3.5	Discussion	176
3.6	Conclusion	179
3.7	Appendices	181
4	Does Creativity Thrive on Plot Twists? Exploring the Role of Surprise on Team Creativity	198
4.1	Introduction	199
4.2	Literature review	201
4.2.1	Team creativity	201
4.2.2	Apprehending the unexpected	205
4.3	Hypotheses	208
4.4	Method	210
4.4.1	Subjects and procedure	210
4.4.2	Tasks	211

4.4.3	Measures	214
4.5	Results	217
4.5.1	Descriptive statistics	217
4.5.2	Treatment testing	218
4.5.3	Regression Analysis	220
4.6	Discussion and conclusion	231
4.7	Appendices	236
5	Can AI Enhance its Creativity to Beat Humans?	248
5.1	Introduction	249
5.2	Literature review	252
5.2.1	Human and Artificial creativity	252
5.2.2	The specific case of Transformers	254
5.3	The experiment	256
5.3.1	Data collection and generation	256
5.3.2	Creativity assessment	258
5.3.3	Experimental procedure	260
5.4	Hypotheses	260
5.5	Methodology	263
5.5.1	Output Generation	263
5.5.2	Metrics	264
5.5.3	Analysis	266
5.6	Results	266
5.6.1	Text task	266
5.6.2	Alternative Uses task	269
5.6.3	Draw task	272
5.7	Discussion	274
5.8	Conclusion	278
5.9	Appendices	280
	General Conclusion	311
	Conclusion Générale	317
	Bibliography	323
	List of figures	356
	List of tables	358

General Introduction

Creativity, as a multifaceted and complex concept, plays a crucial role in driving innovation and progress. Given its importance, this thesis explores the determinants of creative performance by investigating various levels of observation, including individual, team, and artificial dimensions. Utilizing experimental methodology, this research seeks to uncover both intrinsic and environmental factors that influence creativity across these contexts, offering a deeper understanding of how creative potential can be fostered and harnessed. This introductory chapter sets out to outline the motivations behind this thesis, providing an overview of the history and definition of creativity while also detailing the structure and focus of the subsequent chapters.

To begin with, “*You can never have enough of [creativity] in most civilized societies*” [Glăveanu and Kaufman, 2019, p. 9]. Creativity drives the shaping and transformation of culture [Festinger, 1983] as it contributes to building, maintaining, and sometimes changing societies [Glăveanu, 2015]. From Nikola Tesla’s design of the modern alternating current system to Alexander Fleming’s “accidental” discovery of penicillin and, more recently, the adaptation of diving masks into ventilators during the COVID-19 crisis, the creative spark has played a pivotal role in advancing societies and well-being. These examples are just a few in a long list that underscores the central function of creativity: generating novel and useful ideas that might have tangible impacts on society.

While creativity may lead to disruptive innovations, it also serves more common purposes. Creativity is omnipresent in everyday life [Richards, 2007], urging us to investigate it in multiple domains and places, from art to science and schools to workplaces. As part of this dissertation, I draw my attention to creativity as one organisation’s core activity, underlining how indispensable creativity is for the economic sphere. Indeed, organisations operate in dynamic environments subject to

constant environmental, economic, and societal changes. To survive and thrive, they must continually seek new ideas that foster innovation [Amabile, 1988, Zhou and George, 2001]. Considered the first step toward innovation, creativity is a crucial competitive advantage for organisations [Amabile et al., 1996, Cook, 1998, Oldham and Cummings, 1996, Shalley, 1995]. As a result, creativity has been increasingly recognized as a valuable skill and asset that organisations should look for. This means that creativity should not only be conceived at the macro (societal) or meso-level (organisations). It also demands attention at the micro-level, focusing on the individuals or teams engaged in the creative process.

This leads us to question the way in which the economic discipline has addressed the issue of creativity. First and foremost, compared to the extensive literature on creativity across disciplines, mostly in psychology or management, only a few economic studies have been interested in creativity. Traditionally, the focus of economics has been placed on innovation more than creativity. The distinction between creativity and innovation has already been a subject of debate among scholars, with some of them advocating for their proper division, while others preferred to consider them as too strongly intertwined, limiting the possibility of separating them conceptually [Anderson et al., 2014]. As part of this thesis, I argue that it is crucial to distinguish creativity from innovation, correcting instances where these two terms have been mistakenly used as interchangeable terms. This distinction requires specific attention because the relationship between these two terms is not straightforward. If creativity is often considered the seed of innovation, it would be too simplistic to assume a sequential perspective on the relationship between creativity and innovation is sufficient [Amabile and Pratt, 2016, Sarooghi et al., 2015]. The road to implementing creative ideas is uncertain and far-reaching [Anderson et al., 2004], and creativity may occur intermittently, from the very first spark of an idea to its proper implementation. However, if we accept the need to focus on creativity and innovation separately, why has the economy favoured the latter? One key distinction between innovation and creativity lies in their tangible nature. Innovation is more readily quantifiable, often measured through metrics such as patents, while creativity—centred around the idea itself—is more abstract and less tangible. As economists tend to favour quantifiable data, creativity poses an additional challenge due to the absence of straightforward or concrete metrics for measurement.

However, it is essential to acknowledge that the current vision of creativity is relatively recent before examining more closely the conceptualisation of creativity,

its related concepts and the research questions that stem from them. Historically, creativity existed under alternative names - genius, talent, discovery, or imagination - and how it has been defined has changed throughout history and as society has progressed.

Historical perspective on creativity

The concept of creativity is broad and, as its history shows, tortuous. The term *creativity* itself, derived from the Latin *creatus* stemming from *creare* (to make, to produce), only appeared late in history. Although its use exploded during the 20th century, traces of the word creativity have been found throughout history. The word create appeared in the 14th century in Middle English when Geoffrey Chaucer wrote the *Parson's Tale*, a middle-age treatise on penance written in prose closing a collection of stories, where the term was associated with divine creation. Later, in *Métaphysique de l'art*, Mollière [1849] used the term creativity to refer again to creation as God's creation, never relating it to human creation. It is only in 1875 that Adolphus William Ward referred to creativity as a human feature, mentioning Shakespeare's "*poetic creativity*" [Weiner, 2000].

The relationship between religion and creation requires closer attention to understand the evolution of the concept of creativity. Ancient cultures did not conceive the act of creation as a truly human feature. Indeed, everything humans produced was only mere copies of Gods' creation. Creation only encompassed what was created ex-nihilo, which could only result from divine intervention. But did this view of creativity as solely God's work always prevail? It was acknowledged that humans could create, but only with divine intervention. In Asia, for Hindus, Confucianists, or Taoists, human "*creation was at most a kind of discovery or mimicry*" [Runco and Albert, 2010, p. 5]. In the Pre-Christian views of creativity, the concept of genius already emerged and was "*associated with mystical powers of protection and good fortune*" [Runco and Albert, 2010, p. 5]. Plato considered that poets could only write poetry through the influence of Muses. However, voices emerged on the possibility of human creation distanced from Gods' influence. Aristotle stated that creation came from the human mind within. In the Roman view, creation was, first and foremost, with the exception of childbirth, a male capacity that a man could pass to his children [Runco and Albert, 2010]. But these perspectives did not survive in the long term. Combined with our observations on Middle Ages conceptualization

of creation, there is no doubt that the strong and relentless influence of religious and political institutions, which also benefited from their growth in Western societies, limited the development of any questioning around the issue of thought, the spirit and, a fortiori, creativity [Lubart et al., 2015].

It was not until the Renaissance that more in-depth questioning of art, inspiration, or creativity developed. The mysticism behind creation was gradually evicted, and the idea that any creation was the manifestation of a superior entity could not remain. Though the Renaissance marked the rise of humanist thought and philosophy, with a conception of the world that is primarily human-centred, valuing intellect and individual achievement, creativity was still not defined as it is today. Creativity became the feature of “great men”, geniuses, humans endowed with a special gift. Thus, numerous authors have examined the characteristics of these creative geniuses. Among them, Duff [1767] differentiated genius from talent, the latter requiring a certain ability without any original thinking, which is the case for the former. But this change in the conception of creativity did not only occur due to philosophical considerations but also the historical and technological context - *“several key inventions, such as the printing press, led to an unprecedented ability to transmit ideas and gain new knowledge. It was also a time of invention and exploration (e.g., the discovery of the New World), of ingenuity and trade (anticipating the birth of capitalism), and one that encouraged individual thinking and hard work (through the Reformation)”*[Glăveanu and Kaufman, 2019, p. 13].

While this transition from divine to genius lived through the Renaissance and the Enlightenment, it is also interesting to look at the time lapse between the development of research and the spark of interest in creativity as a matter of research. In fact, several centuries passed from its first appearance in Middle English to its incorporation into the realm of research in the 19th century. It is rather trivial to recognize the role of mysticism and religious impregnation around creativity in its distancing from research [Sternberg and Lubart, 1996]. Thus, as soon as creativity was conceived as the feature of “great men”, the research on creative geniuses’ characteristics expanded, naturally leading thinkers to consider genetic and environmental factors at their source [Albert and Runco, 1999], negating the supernatural nature of creativity and grounding it in nature [Lubart et al., 2015] and bridging the concept of creativity to research concerns at the time. Two authors are often recognized as pivotal in developing creativity research: Charles Darwin and Francis Galton. While Darwinian theories contributed to the debate on individual characteristics of geniuses

based on the assumption of adaptation for survival, the work of Francis Galton marks the beginning of statistical studies on geniuses to discover the key determinants of their profiles. Ultimately, this eugenicist perspective on creativity, led by Galton, is considered one of the first attempts at an empirical study of creativity [Runco and Albert, 2010].

Years later, as soon as creativity had acquired its credentials, a large body of work emerged to make the concept a central research issue of the 20th century. A key element in this expansion was certainly the multidisciplinary dimension taken on by the concept. This multidisciplinary, including non-exhaustively philosophy, psychology, management and only recently economics, led to a series of seminal works that are still widely recognized and cited today. Among them, psychologist Joy Paul Guilford contributed to establishing creativity as a legitimate field of research, providing a framework to understand and measure creative thinking abilities. As the authors developed Creativity Tests, they also defined creativity's main components or criteria, encompassing, among others, fluency, flexibility, originality, and elaboration [Guilford, 1950, 1967, Wilson et al., 1954]. As a result, Guilford did not only link creativity to intelligence but a complex set of cognitive abilities where the creative process is influenced by the type of mental task observed and the type of thinking elicited, either divergent or convergent. Thus, much of the subsequent research on creativity in psychology resulted from Guilford's attempt to understand the creative process. The work of Ellis Paul Torrance was also crucial in the study and assessment of creativity, primarily through the development of the Torrance Tests of Creative Thinking (TTCT) that aimed to assess and nurture creative potential in individuals, particularly in educational settings. Similarly to Guilford, Torrance also introduced tasks eliciting different thinking processes, including the well-known Alternative Uses Task [Torrance, 1966], asking individuals to provide unusual uses of everyday objects. Even though Guilford's and Torrance's works have undergone several revisions, their influence can still be witnessed today.

Over time, the list of major works and the related theories underpinning social sciences research on creativity expanded. Different creativity theories have emerged since the 1950s as the interest in creativity rapidly grew. While this paragraph does not aim to provide an exhaustive list, it focuses on the two theories central to which the work to follow in this thesis will mainly relate. First, the *Componential theory*, mostly led by the seminal work of [Amabile, 1983, Amabile et al., 1996, Amabile and Pratt, 2016], is a comprehensive model that identifies creativity's social

and psychological components. This theory argues that creativity relies on three main components: task motivation, domain-relevant skills, and creativity-relevant processes. In addition, the authors acknowledge the crucial role of the environment surrounding individuals in their creative abilities, as there is a bidirectional relationship between individuals and their environment, where one influences the other conversely. Then, the *Interactionist theory*, a multilevel theoretical model of creativity, integrates the individual, the team, and the organisation involved in a complex set of interactions [Woodman et al., 1993]. In this theory, creativity can be explained by individual characteristics as well as contextual factors occurring at each level. Quite importantly, the shift from the individual to the team or organisational level is not realized by aggregating individual characteristics but by translating complex interactions and dynamics. The Interactionist theory underlines how complex creativity is, advocating for a holistic approach to understanding and fostering it within organisations. This thesis integrates these two theories to explore creative performance across different levels of observation, emphasizing the interactional nature of the creative process while also considering the complexity of the environment in which individuals or groups engage in this process.

Definition of creativity and related concepts

This complex history and the numerous emerging theories raise the question of whether a consensual definition of creativity could be found. Even though the conceptualization of surprise varies depending on the culture and era, researchers seem to have reached a consensus. In its simplest definition, creativity corresponds to generating novel and useful ideas [Amabile et al., 1996, Amabile and Pratt, 2016, Anderson et al., 2014, Eisenberger et al., 1999, Ochse, 1990, Oldham and Cummings, 1996, Sternberg and Lubart, 1995]. What is important to stress is the need for these two elements to work together. Indeed, a “good” creative idea must be both novel and useful [Rietzschel et al., 2010], knowing there exists a trade-off between these two elements, making the generation of creative ideas more challenging [Nijstad et al., 2010]. This naturally leads us to question what novelty is and what usefulness is.

On the one hand, the novelty of an idea, also referred to as uniqueness or originality [Amabile, 1983], varies according to the reference point. One idea may be new to an individual, a group of individuals or even an organisation, while another may be new to society as a whole. Novelty can also take different forms. An idea may

represent a deviation from existing ideas but also an overturning of existing concepts. On the other hand, usefulness is subject to more debate on its definition; sometimes associated with appropriateness or utility, sometimes with quality or effectiveness [Magni et al., 2023]. In any case, useful ideas must address the demands of the problem at hand while respecting the imposed constraints. This prevents creativity from falling into a trap where the creative solution would be completely detached from reality [Cropley and Cropley, 2010] for the sole purpose of finding something unusual.

If we define a creative idea as novel and useful, its validation requires some form of performance in these dimensions. The idea of creative performance is strongly related to the multifaceted nature of creativity. There is no absolute standard for assessing creative performance; a creative output is always evaluated relative to others or at least the evaluators' perception of it. This perception of creativity is highly influenced by the context in which the idea or output is produced. In fact, a creative assessment's result must be considered in its environment because it provides insight into the rationale behind this judgement [Plucker et al., 2004]. Thus, one needs to consider three components: the nature of the task, the specific criteria to evaluate, and the means of evaluation. First, regarding the nature of the task, at a more abstract level, we need to consider the thinking process elicited. One individual engaging in a creative process is engaged in a cognitive process. As such, the cognitive mechanisms behind generating novel and useful ideas might differ according to the type of task considered. The classical distinction made in the literature is the one between divergent and convergent thinking Guilford [1956], where the former corresponds to the generation of multiple ideas, while the latter derives a singular logical solution. As a result, tasks can be characterized by the type of thinking process they elicit. This differentiation between tasks can be translated into the openness of tasks, i.e., the importance of constraints imposed on the problem resolution. Beyond the usual dichotomy between open and closed tasks that respectively trigger a more divergent versus convergent thinking process, Attanasi et al. [2021] add a third type of task, namely open-with-constraints, that continues to elicit a divergent thinking process while reinforcing the constraints on the task resolution. Then, creative performance relies on the creativity criteria chosen to assess the performance. There exists an extensive array of possible criteria. As already mentioned, from the seminal work of Guilford [1967], authors have mainly assessed creative performance through originality (*how infrequent a particular solution is*), fluency (*how many ideas were*

generated to solve a specific problem), and flexibility (*how many themes cover the set of ideas generated*), or elaboration (*how detailed ideas are*). However, on a more practical note and more closely aligned with the definition of creativity, one could imagine tasks where the weight of originality is more important than usefulness, and vice versa: an artistic production versus a new product from an R&D department, for example. As such, the creativity criteria chosen would be assessed based on two dimensions: originality and feasibility. In any case, as a multifaceted concept, it is crucial to evaluate the creativity of an idea through various criteria that cover all the complexity of this concept. Finally, regarding the means of evaluation, we must consider two main distinctions: objective versus subjective measures and external judges versus peer judges. On the one hand, creativity criteria, depending on their definition, can be assessed by the subjective judgement of evaluators or objective measurements permitted by the quantitative analysis of ideas. In the context of this thesis, both types of measures will be used, combining the subjective assessment of the *consensual assessment technique* [Amabile, 1982] but also objective measures, partly computed using *Natural Language Processing* techniques. On the other hand, mostly for subjective measures, the creative assessment of outputs requires one or more judges to evaluate each criterion. As a result, the researcher faces two possibilities: creativity assessed by peer judges, i.e., the assessment via other subjects in the experiment, or external judges, i.e., individuals that are blind to the treatment. The second possibility is often considered the most reliable, which is why it has been chosen throughout this thesis.

However, besides the nature of the task, the chosen criteria, or the type of assessment, part of the creative context is also the chosen observation level. The literature recognises three main levels of observation: individual, team, and organisational creativity. Individual creativity is an important element of the Componential and Interactionist theories of creativity, where individual differences interplay with contextual factors. Even though studying individual creativity might seem limiting because it excludes higher levels of observation of creativity, individual creativity is the source of team and organisational creativity [Amabile, 1988], and stepping into the team and organisational creativity eluding this fundamental step would prevent researchers from grasping the full picture of the creative process. Studying individual creativity is crucial for identifying personal factors that influence creative performance, enabling organisations to better target the right individuals for specific tasks and, later on, to assess which individual profiles could complete each other once focusing

on a team level. Indeed, focusing solely on individual creativity may also not be sufficient. Creativity also needs to be studied from the perspective of teams, where teams are defined as “*interdependent collections of individuals who share responsibility for specified outcomes*” [Sundstrom et al., 1990, 120]. The importance of studying team creativity emerged since there is a growing tendency for many tasks to require collaboration within a team [Mathieu et al., 2017] due to the growing complexity of tasks and the expanding set of skills required to complete these tasks [Fiore et al., 2010, Kozlowski and Ilgen, 2006, Sundstrom and etc., 1998]. However, teams are not a simple aggregate of individuals but the complex interactions of these individuals and their specificities [Woodman et al., 1993]. Besides team members’ profiles, the study of team creativity requires including team characteristics, processes, and dynamics. As such, studying team creativity allows for the understanding of complex interactions between individuals involved in a creative process, reinforcing the social part of it. Finally, organisational creativity stems from individual and team creativity, though not all directly translate into organisational outcomes. As the upper level of observation, organisational creativity also strongly relies on contextual factors [Amabile et al., 1996, Andriopoulos, 2001, Chang and Chiang, 2008] but includes specific elements unique to the life of organisations. Regardless of the specific level of observation, it is not sufficient to study them separately but to consider them jointly, identifying their interrelations [Hargadon and Bechky, 2006, Oldham and Cummings, 1996, Shalley et al., 2009]. The cross-level analysis of creativity is then necessary to grasp the creative process and understand its intricacies. However, it is important to note that this thesis will primarily focus on the first two levels of creativity — individual and team — rather than the organisational level. This focus was chosen because individual and team creativity forms the foundation of organisational creativity, and the context in which our research was conducted did not involve organisational artefacts.

Finally, following on from the study of the determinants of creative performance, the question arises of the identity of the ideator. For a long time, creativity was seen as a human prerogative, recent advancements in Artificial Intelligence (AI), particularly following the launch of ChatGPT in 2022, have revived the debate on whether creativity is exclusively a human endeavour. Indeed, creativity is supposed to be part of what makes us human [Sawyer and Henriksen, 2024], but machines’ abilities to produce relevant creative outputs greatly challenges human minds. [Amabile, 2019] defines artificial creativity as “*the production of highly novel, yet appropriate, ideas,*

problem solutions, or other outputs by autonomous machines”(p. 3). What emerges from the study of artificial creativity is the tension surrounding the very ability of a machine to produce novel and useful ideas. For many individuals, AI should be used to handle our menial tasks, not be designed to take over our endeavours. In March 2024, Joanna Maciejewska’s message on the Platform X went viral. “*I want AI to do my laundry and dishes so that I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes*”. This plea to prevent the expansion of AI in the realm of creativity triggered debates on the possibility of artificial creativity. Specifically, this resistance to AI in creative fields not only stems from algorithmic aversion but also from the belief that creativity is central in our society and often considered the highest-level performance granted to humankind [Taylor, 1988]. In the end, when artificial creativity is examined, there is excitement and fear, and investigations are necessary to understand it and disentangle its implications and possible consequences. However, this paradigm shift does not diminish our interest in studying human creativity. Since AI lacks agency and requires initial human intervention to solve a creative problem. Especially highlighting the issue of problem-solving and problem-framing, and therefore the central place given to human-machine interactions.

Outline of the thesis

Although creativity has been studied in different research fields for several decades, many aspects remain under-explored. Before diving into the specific research questions addressed in each chapter of this thesis, it is important to present the broader research question this thesis aims to address. As underlined in the first paragraph of this introduction chapter, whether we focus on its individual, team, or artificial dimensions, creativity nurtures our ability to innovate and, thus, evolve. Consequently, there is an increasing need in our societies to explore how to enhance creative performance, ensuring that creativity—despite its inherent risks and uncertainty—can thrive and succeed. Different perspectives can be adopted to address the issue of creative performance, and it is crucial to emphasise this multi-dimensionality. Specifically, creative performance revolves around four main facets: (1) the identity of the creative agent performing a creative task, (2) the nature of the task evaluated, (3) the multiple creativity criteria chosen to assess performance, and (4) the intrinsic and environmental factors affecting it. As a result, although each chapter of this

thesis focuses on specific research questions, the underlying inquiry always relies on the multifaceted nature of creative performance.

Before diving into the outline of this thesis and each chapter's research questions and corresponding research strategy, it is important to present the chosen methodology and explain why this choice has been made. This thesis focuses on experimental methodology. This choice was not made at random. Creativity is a social process, with individuals and their interactions with the environment at its core. Thus, experimental economics offers a unique and promising framework for observing such a phenomenon, enabling researchers to observe complex mechanisms, processes, or dynamics in a controlled environment [Friedman and Sunder, 1994, Jacquemet and L'Haridon, 2018], regardless of the level of observation. While the current experimental methodology offers tools to study individual creativity (*Chapter 2 and 5*), one contribution of this work is also to investigate how the experimental methodology in economics can be adapted to the study of team processes (*Chapter 1, 3 and 4*).

Chapter 1 – *Rethinking laboratory experiments: The case of collective experiments*

This initial chapter discusses the advancement of the experimental methodology in economics with the promising implementation of collective experiments, defined as “*any experiment enabling the observation of collective processes through the social interactions of subjects without a systematic intermediary*”. While a tremendous number of protocols have been developed over the last 70 years, assuring that the experimental methodology in economics has become a stream of research in its own right, researchers focused primarily on the study of individuals. However, little attention has been paid to adapting this methodology to study collective processes and interactions between individuals. It must not be forgotten that economics relies on a solid ground of social interactions and interdependence. Therefore, diminishing the importance of studying collective dynamics, as has often been the case, is a mistake. Indeed, the experimental literature has explored situations of interdependence; however, few experimental protocols involve direct interactions between subjects with a specific focus on emerging collective processes and significantly influencing both individual and collective outcomes. This chapter proposes to reflect on this limitation and demonstrates how experimental economists should adapt their current methodological best practices to study collective processes. The chapter is structured around

four sections: a historical and methodological analysis of the experimental approach in economics, a justification of the need for collective experiments, the importance of eliminating intermediaries and using innovative tools to improve experimental protocols, and finally, an examination of the example of the *Social Interactions Lab* (SIL) for observing collective processes.

Chapter 2 – *The Many Faces of Creative Profiles: Exploring task openness*

The second chapter delves into the individual level of creativity. Creativity is a central element for the growth and evolution of our societies [Sawyer and Henriksen, 2024] as well as for organisations. Thus, understanding the key characteristics of creative profiles is even more important to identify individuals with a noticeable creative potential that may serve organisations and societies. Focusing on individual creativity is not something new in the literature. Creativity has long been envisaged as an individual and solitary endeavour, only achievable by the best humans. However, it has been recognized that creativity is not an innate quality of certain humans but a capacity that can be trained and improved. As such, it becomes natural to question what are some specific characteristics that explain why some individuals perform better than others. However, individual creativity is often analysed through intrinsic motivation, cognitive styles, or personality traits [Amabile and Pratt, 2016, Amabile, 1983]; less attention has been paid to knowledge and skills and how these elements are intertwined with other individual characteristics. In addition, due to the variety of tasks, one additional question is to what extent the degree of openness of creative tasks, as the importance of constraints on task resolution, influences individuals' creative performance. Building on the extensive literature on individual creativity, this chapter explores the determinants of creative performance, with a particular focus on how different creative profiles respond to varying degrees of task openness. In the end, what are the key characteristics of creative profiles? And, how do different degrees of openness of tasks moderate the relationships between profiles' characteristics and creative performance? To do so, a within-subjects experimental protocol has been implemented to investigate the relationship between individuals' profiles and their performance in a given task. The stemming results inform the effective development of diverse teams that can meet the multifaceted demands of organisations.

Chapter 3 – *Diversity, Interactions, and Team creativity: An experimental perspective*

Moving beyond the possibly limiting view of creativity as only an individual pursuit and recognizing its inherently collective nature, Chapter 3 naturally investigates team creativity and the determinants of team creative performance. If team creativity is a crucial element in the study of organisations [Amabile et al., 1996, Sarooghi et al., 2015, Chatzoglou and Chatzoudes, 2018, Rousseau et al., 2015], limited research has been pursued in economics on team creativity in an experimental context. This third chapter examines two main intrinsic factors of team creativity: team diversity and social networks. Especially, understanding how diversity and social networks jointly impact creative performance is primordial to grasp the complexity of social interactions in the team creative process. Even more so by studying the overlap of different diversity types [van Knippenberg and J. Hoeber, 2021]. In fact, to what extent does diversity in terms of team composition impact the team’s collective creative performance? And, how might individuals’ interactions, through pre-existing or in-situ networks, affect the team’s creative performance? Moreover, this chapter also aims to introduce, for the first time, a collective experiment in which intermediaries between subjects have been waived and where these subjects are involved in face-to-face interactions. As highlighted in Chapter 1, economics would benefit from new protocols that observe direct interactions between subjects without intermediaries, better reflecting the real creative process of teamwork. Of note, the experimental protocol used at the time is not unrelated to that used in Chapter 2 since the present chapter also includes the individual observation level, using team members’ individual creative performance as one predictor of teams’ creative performance. This allows us to conduct a cross-level analysis of creativity and bridge individual and team creative performance to comprehend the creative phenomenon fully. Knowing that, until now, the relationship between individual and collective creativity has not received enough attention [van Knippenberg and J. Hoeber, 2021]. Indeed, when studying cross-level performance, most studies only experimentally test team creativity, while questionnaires assess individual creativity. Finally, the results from Chapter 3 supplement the ones of Chapter 2 by testing team creative performance based on team members’ own performance and intrinsic factors to the team.

Chapter 4 – *Does Creativity Thrive on Plot Twists? Exploring the role of surprise on team creativity*

In contrast to Chapters 2 and 3, Chapter 4 does not only focus on the effect of intrinsic factors on creative performance at the individual or team level but also the effect of environmental factors, here surprise. Organisations operate in constantly changing environments prone to sudden and unexpected events that might affect their processes and performance. If creativity is one central element of organisations, looking at the relationship between creativity and surprise becomes essential. Especially since creativity, by nature, is an uncertain and risky undertaking, and this uncertainty is amplified by the presence of unexpected events that can alter the conditions under which novel and useful ideas emerge. In terms of literature, this chapter lies at the intersection between creativity, surprise and the schema theory to understand how team members experience and, possibly, overcome an unexpected event that affects the ongoing creative process. This chapter particularly questions how an unexpected event that takes the form of a change in constraints specification impacts a team's creative performance according to the degree of surprise introduced in a task. And how this relationship between surprise and the team's creativity is affected by other factors, such as the team's diversity, processes, and creative capacities. The stemming experimental protocol aims to answer these research questions by comparing teams experiencing surprise and those who do not and comparing different degrees of surprise, with low and high surprise. By doing so, this chapter sets the stage for a deeper understanding of the factors that determine the effect of unexpected changes on team creative performance, the role of team diversity, and the processes that facilitate or hinder a team's adaptability in the face of surprise.

Chapter 5 – *Can AI Enhance its Creativity to Beat Humans ?*

In line with the previous chapters, the fifth chapter of this thesis explores human creative performance while also considering the possibility of human creativity being challenged by artificial creativity. It emphasizes the importance of using multiple criteria to assess creative performance, aligning both forms of creativity. While creativity has been mainly studied as a human feature, the booming research on artificial creativity is an important avenue of research. The possibility of a machine outperforming a human in a (creative) task raises questions about substituting one for the other in carrying out that task. However, this requires a deep understanding

of the strengths and weaknesses of both agents, possibly advocating for complementarity rather than substitutability. This chapter specifically questions to what extent artificial creativity might challenge human creativity by comparing their performance in different types of tasks, varying in degree of openness, and according to the prompting strategy chosen to address the AI model. Gathering creative outputs from experimental protocols with human subjects and generating creative outputs through *ChatGPT* and *Dall-e*, an online experiment was designed for human evaluators to assess the creative performance of outputs produced by humans and AI. One important element of this chapter is the attention given to the specification of creative performance, as both objective and subjective measures of creativity have been used to assess whether artificial intelligence might truly challenge humans. To conclude, the results of Chapter 5 provide a detailed comparison between human and artificial creative performance, leading to discussions on the best way to envisage human and machine collaboration based on their strengths and possible complementarity.

Introduction générale

La créativité est un concept multidimensionnel et complexe, jouant un rôle crucial dans la stimulation de l'innovation et du progrès. Étant donné l'importance de la créativité au sein de nos sociétés, cette thèse explore les déterminants de la performance créative en étudiant différents niveaux d'observation, incluant les dimensions individuelle, d'équipe et artificielle. En s'appuyant sur une méthodologie expérimentale, cette recherche vise à identifier les facteurs intrinsèques et environnementaux qui influencent la créativité dans ces contextes, offrant ainsi une compréhension approfondie des moyens par lesquels le potentiel créatif peut être stimulé et exploité. Ce chapitre introductif a pour objectif de présenter les motivations de cette thèse, en offrant un aperçu de l'histoire et de la définition de la créativité, tout en détaillant la structure et l'orientation des chapitres à venir.

Pour commencer, “*You can never have enough of [creativity] in most civilized societies*” [Glăveanu and Kaufman, 2019, p. 9]. La créativité joue un rôle clé dans la formation et la transformation de la culture [Festinger, 1983], contribuant à la construction, au maintien, et parfois au changement des sociétés [Glăveanu, 2015]. Qu'il s'agisse de la conception du système moderne de courant alternatif par Nikola Tesla, de la découverte “accidentelle” de la pénicilline par Alexander Fleming, ou, plus récemment, de l'adaptation des masques de plongée en respirateurs durant la crise de la COVID-19, l'étincelle créative a toujours été centrale dans le progrès des sociétés et l'amélioration du bien-être. Ces exemples ne représentent qu'une petite partie d'une longue liste démontrant la fonction primordiale de la créativité : produire des idées nouvelles et utiles ayant des répercussions concrètes sur la société.

Bien que la créativité puisse mener à des innovations disruptives, elle sert également des objectifs plus ordinaires. Elle est omniprésente dans la vie quotidienne [Richards, 2007], ce qui justifie son étude dans une multitude de contextes, de l'art à la science, des écoles aux lieux de travail. Dans cette dissertation, je mets l'accent sur

la créativité en tant qu'activité centrale des organisations, soulignant à quel point elle est indispensable à la sphère économique. En effet, les organisations évoluent dans des environnements dynamiques, soumis à des changements constants, qu'ils soient environnementaux, économiques ou sociétaux. Pour survivre et prospérer, elles doivent sans cesse rechercher de nouvelles idées propices à l'innovation [Amabile, 1988, Zhou and George, 2001]. Considérée comme l'étape initiale de l'innovation, la créativité constitue un avantage concurrentiel essentiel pour les organisations [Amabile et al., 1996, Cook, 1998, Oldham and Cummings, 1996, Shalley, 1995]. En conséquence, la créativité est de plus en plus perçue comme une compétence précieuse, un atout que les organisations devraient activement rechercher. Cela signifie que la créativité ne doit pas être uniquement envisagée à l'échelle macro (sociétale) ou méso (organisationnelle), mais nécessite également une attention au niveau micro, en se concentrant sur les individus ou les équipes engagés dans le processus créatif.

Cela nous amène à examiner la manière dont l'économie en tant que discipline a abordé la question de la créativité. Comparée à l'abondante littérature sur la créativité dans d'autres disciplines, principalement en psychologie ou en gestion, peu d'études économiques se sont penchées sur ce sujet. Traditionnellement, l'économie s'est davantage intéressée à l'innovation qu'à la créativité. La distinction entre créativité et innovation a été longuement débattue, certains chercheurs plaidant pour une séparation stricte, tandis que d'autres estiment qu'elles sont trop étroitement liées pour être dissociées conceptuellement [Anderson et al., 2014]. Dans cette thèse, je défends l'importance de distinguer clairement la créativité de l'innovation, corrigeant les cas où ces deux termes ont été indûment utilisés de manière interchangeable. Cette distinction mérite une attention particulière, car la relation entre créativité et innovation est complexe. Si la créativité est souvent perçue comme le germe de l'innovation, il serait réducteur de supposer qu'une approche strictement séquentielle suffise à expliquer leur relation [Amabile and Pratt, 2016, Sarooghi et al., 2015]. Le chemin menant à la mise en œuvre d'idées créatives est incertain et souvent long [Anderson et al., 2004], la créativité pouvant survenir de façon intermittente, de la première étincelle d'une idée à sa concrétisation. Cependant, si l'on accepte la nécessité d'étudier séparément créativité et innovation, pourquoi l'économie a-t-elle privilégié cette dernière ? Une distinction clé réside dans la nature tangible de l'innovation. Cette dernière est plus facilement quantifiable, souvent mesurée par des indicateurs comme les brevets, tandis que la créativité — centrée sur l'idée elle-même — reste plus abstraite et difficile à mesurer. Les économistes, étant plus enclins à travailler

avec des données quantifiables, trouvent la créativité plus difficile à aborder en raison de l'absence de métriques claires pour la mesurer.

Toutefois, avant de se pencher de manière approfondie sur la conceptualisation de la créativité, ses concepts connexes et les questions de recherche qui en découlent, il est crucial de reconnaître que la vision actuelle de la créativité est relativement récente. Historiquement, la créativité existait sous d'autres appellations — génie, talent, découverte ou imagination — et sa définition a évolué au fil du temps à mesure que les sociétés se sont développées.

Perspective historique sur la créativité

Le concept de créativité est large et, comme son histoire le montre, tortueux. Le terme *créativité* lui-même, dérivé du latin *creatus* provenant de *creare* (faire, produire), n'est apparu que tard dans l'histoire. Bien que son utilisation ait explosé au cours du 20e siècle, des traces du mot créativité ont été trouvées à travers l'histoire. Le mot créer est apparu au 14e siècle dans l'anglais médiéval, lorsque Geoffrey Chaucer a écrit le *Parson's Tale*, un traité médiéval sur la pénitence écrit en prose, clôturant une collection d'histoires, où le terme était associé à la création divine. Plus tard, dans la *Métaphysique de l'art*, Mollière [1849] a utilisé le terme créativité pour se référer à nouveau à la création en tant que création divine, sans jamais la lier à la création humaine. Ce n'est qu'en 1875 qu'Adolphus William Ward a fait référence à la créativité comme une caractéristique humaine, mentionnant la “*poetic creativity*” [Weiner, 2000].

La relation entre la religion et la création nécessite une attention particulière pour comprendre l'évolution du concept de créativité. Les cultures anciennes ne concevaient pas l'acte de création comme une véritable caractéristique humaine. En effet, tout ce que les humains produisaient n'était que de simples copies de la création des Dieux. La création n'englobaient que ce qui était créé ex-nihilo, ce qui ne pouvait être que le résultat d'une intervention divine. Mais cette vision de la créativité comme étant exclusivement l'œuvre de Dieu a-t-elle toujours prévalu ? Il était reconnu que les humains pouvaient créer, mais seulement avec l'intervention divine. En Asie, pour les Hindous, les Confucianistes ou les Taoïstes, la création humaine “*was at most a kind of discovery or mimicry*” [Runco and Albert, 2010, p. 5]. Dans les conceptions pré-chrétiennes de la créativité, le concept de génie a déjà émergé et était “*associated with mystical powers of protection and good fortune*” [Runco and Albert, 2010, p. 5].

Platon considérait que les poètes ne pouvaient écrire de la poésie qu'à travers l'influence des Muses. Cependant, des voix se sont élevées pour affirmer la possibilité d'une création humaine éloignée de l'influence des Dieux. Aristote a déclaré que la création venait de l'esprit humain intérieur. Dans la vision romaine, la création était avant tout, à l'exception de l'accouchement, une capacité masculine qu'un homme pouvait transmettre à ses enfants [Runco and Albert, 2010]. Mais ces perspectives n'ont pas survécu à long terme. Combinée à nos observations sur la conceptualisation de la création au Moyen Âge, il ne fait aucun doute que l'influence forte et incessante des institutions religieuses et politiques, qui ont également bénéficié de leur essor dans les sociétés occidentales, a limité le développement de toute interrogation autour de la pensée, de l'esprit et, a fortiori, de la créativité [Lubart et al., 2015].

Ce n'est qu'à la Renaissance qu'un questionnement plus approfondi sur l'art, l'inspiration ou la créativité s'est développé. Le mysticisme derrière la création a été progressivement évacué, et l'idée que toute création était la manifestation d'une entité supérieure ne pouvait plus subsister. Bien que la Renaissance ait marqué l'essor de la pensée et de la philosophie humanistes, avec une conception du monde principalement centrée sur l'humain, valorisant l'intellect et les réalisations individuelles, la créativité n'était toujours pas définie comme elle l'est aujourd'hui. La créativité est devenue la caractéristique des "grands hommes", des génies, des humains dotés d'un don spécial. Ainsi, de nombreux auteurs ont examiné les caractéristiques de ces génies créatifs. Parmi eux, Duff [1767] a différencié le génie du talent, ce dernier nécessitant une certaine capacité sans aucune pensée originale, ce qui est le cas pour le premier. Mais ce changement dans la conception de la créativité ne s'est pas seulement produit en raison de considérations philosophiques, mais aussi du contexte historique et technologique - "*several key inventions, such as the printing press, led to an unprecedented ability to transmit ideas and gain new knowledge. It was also a time of invention and exploration (e.g., the discovery of the New World), of ingenuity and trade (anticipating the birth of capitalism), and one that encouraged individual thinking and hard work (through the Reformation)*" [Glăveanu and Kaufman, 2019, p. 13].

Bien que cette transition du divin au génie ait traversé la Renaissance et les Lumières, il est également intéressant de regarder le laps de temps entre le développement de la recherche et l'émergence de l'intérêt pour la créativité en tant que sujet de recherche. En effet, plusieurs siècles se sont écoulés entre sa première apparition dans l'anglais médiéval et son incorporation dans le domaine de la recherche au 19e

siècle. Il est plutôt trivial de reconnaître le rôle du mysticisme et de l'imprégnation religieuse autour de la créativité dans son éloignement de la recherche [Sternberg and Lubart, 1996]. Ainsi, dès que la créativité a été conçue comme une caractéristique des “grands hommes”, la recherche sur les caractéristiques des génies créatifs s’est étendue, conduisant naturellement les penseurs à considérer les facteurs génétiques et environnementaux à leur origine [Albert and Runco, 1999], niant la nature sur-naturelle de la créativité et l’ancrant dans la nature [Lubart et al., 2015], reliant ainsi le concept de créativité aux préoccupations de la recherche de l’époque. Deux auteurs sont souvent reconnus comme étant essentiels dans le développement de la recherche sur la créativité : Charles Darwin et Francis Galton. Alors que les théories darwiniennes ont contribué au débat sur les caractéristiques individuelles des génies en partant de l’hypothèse de l’adaptation pour la survie, le travail de Francis Galton marque le début des études statistiques sur les génies afin de découvrir les déterminants clés de leurs profils. En fin de compte, cette perspective eugéniste sur la créativité, dirigée par Galton, est considérée comme l’une des premières tentatives d’étude empirique de la créativité [Runco and Albert, 2010].

Des années plus tard, dès que la créativité a acquis ses lettres de noblesse, un vaste corpus de travaux a émergé pour faire du concept une question centrale de la recherche du 20e siècle. Un élément clé de cette expansion a certainement été la dimension multidisciplinaire adoptée par le concept. Cette multidisciplinarité, incluant non exhaustivement la philosophie, la psychologie, le management et, plus récemment, l’économie, a conduit à une série de travaux fondateurs encore largement reconnus et cités aujourd’hui. Parmi eux, le psychologue Joy Paul Guilford a contribué à établir la créativité comme un domaine de recherche légitime, fournissant un cadre pour comprendre et mesurer les capacités de pensée créative. Alors que les auteurs ont développé des tests de créativité, ils ont également défini les principaux composants ou critères de la créativité, englobant, entre autres, la fluidité, la flexibilité, l’originalité et l’élaboration [Guilford, 1950, 1967, Wilson et al., 1954]. En conséquence, Guilford n’a pas seulement lié la créativité à l’intelligence, mais à un ensemble complexe de capacités cognitives où le processus créatif est influencé par le type de tâche mentale observée et le type de pensée sollicitée, soit divergente, soit convergente. Ainsi, une grande partie des recherches ultérieures sur la créativité en psychologie découle de la tentative de Guilford de comprendre le processus créatif. Le travail d’Ellis Paul Torrance a également été crucial dans l’étude et l’évaluation de la créativité, principalement à travers le développement des *Torrance Tests of Crea-*

tive Thinking (TTCT) qui visaient à évaluer et à encourager le potentiel créatif des individus, notamment dans les contextes éducatifs. De manière similaire à Guilford, Torrance a également introduit des tâches sollicitant différents processus de pensée, y compris la célèbre tâche des *usages alternatifs* [Torrance, 1966], demandant aux individus de fournir des usages inhabituels d’objets du quotidien. Même si les travaux de Guilford et Torrance ont fait l’objet de plusieurs révisions, leur influence est encore perceptible aujourd’hui.

Au fil du temps, la liste des travaux majeurs et des théories connexes soutenant la recherche en sciences sociales sur la créativité s’est élargie. Différentes théories de la créativité ont émergé depuis les années 1950 à mesure que l’intérêt pour la créativité a rapidement grandi. Bien que ce paragraphe ne vise pas à fournir une liste exhaustive, il se concentre sur les deux théories centrales auxquelles se rapportera principalement le travail à suivre dans cette thèse. Tout d’abord, la *théorie componentielle*, principalement menée par le travail fondateur de [Amabile, 1983, Amabile et al., 1996, Amabile and Pratt, 2016], est un modèle complet qui identifie les composants sociaux et psychologiques de la créativité. Cette théorie soutient que la créativité repose sur trois principaux composants : la motivation pour la tâche, les compétences pertinentes au domaine et les processus pertinents pour la créativité. De plus, les auteurs reconnaissent le rôle crucial de l’environnement entourant les individus dans leurs capacités créatives, car il existe une relation bidirectionnelle entre les individus et leur environnement, où l’un influence l’autre réciproquement. Ensuite, la *théorie interactionniste*, un modèle théorique multiniveau de la créativité, intègre l’individu, l’équipe et l’organisation impliqués dans un ensemble complexe d’interactions [Woodman et al., 1993]. Dans cette théorie, la créativité peut être expliquée par des caractéristiques individuelles ainsi que par des facteurs contextuels se produisant à chaque niveau. Il est très important de noter que le passage de l’individu au niveau de l’équipe ou de l’organisation ne se fait pas par agrégation des caractéristiques individuelles, mais par la traduction des interactions et dynamiques complexes. La théorie interactionniste souligne la complexité de la créativité, plaidant pour une approche holistique de sa compréhension et de sa promotion au sein des organisations. Cette thèse intègre ces deux théories pour explorer la performance créative à différents niveaux d’observation, en mettant l’accent sur la nature interactionnelle du processus créatif tout en tenant compte de la complexité de l’environnement dans lequel les individus ou les groupes s’engagent dans ce processus.

Définition de la créativité et concepts liés

Cette histoire complexe et les nombreuses théories émergentes soulèvent la question de savoir s'il est possible de trouver une définition consensuelle de la créativité. Même si la conceptualisation de la surprise varie selon les cultures et les époques, les chercheurs semblent être parvenus à un consensus. Dans sa définition la plus simple, la créativité correspond à la génération d'idées nouvelles et utiles [Amabile et al., 1996, Amabile and Pratt, 2016, Anderson et al., 2014, Eisenberger et al., 1999, Ochse, 1990, Oldham and Cummings, 1996, Sternberg and Lubart, 1995]. Il est important de souligner que ces deux éléments doivent fonctionner ensemble. En effet, une "bonne" idée créative doit être à la fois nouvelle et utile [Rietzschel et al., 2010], sachant qu'il existe un compromis entre ces deux éléments, ce qui rend la génération d'idées créatives plus difficile [Nijstad et al., 2010]. Cela nous amène naturellement à nous interroger sur ce qu'est la nouveauté et ce qu'est l'utilité.

D'une part, la nouveauté d'une idée, également appelée unicité ou originalité [Amabile, 1983], varie selon le point de référence. Une idée peut être nouvelle pour un individu, un groupe d'individus ou même une organisation, tandis qu'une autre peut être nouvelle pour la société dans son ensemble. La nouveauté peut également prendre différentes formes. Une idée peut représenter une déviation par rapport aux idées existantes, mais aussi un renversement des concepts existants. D'autre part, l'utilité fait l'objet de davantage de débats sur sa définition ; parfois associée à l'adéquation ou à l'utilité, parfois à la qualité ou à l'efficacité [Magni et al., 2023]. Dans tous les cas, les idées utiles doivent répondre aux exigences du problème posé tout en respectant les contraintes imposées. Cela empêche la créativité de tomber dans le piège où la solution créative serait complètement détachée de la réalité [Cropley and Cropley, 2010], dans le seul but de trouver quelque chose d'inhabituel.

Si nous définissons une idée créative comme étant nouvelle et utile, sa validation nécessite une forme de performance dans ces dimensions. L'idée de performance créative est fortement liée à la nature multifacette de la créativité. Il n'existe pas de norme absolue pour évaluer la performance créative ; une production créative est toujours évaluée par rapport à d'autres ou du moins à la perception qu'en ont les évaluateurs. Cette perception de la créativité est fortement influencée par le contexte dans lequel l'idée ou la production est réalisée. En effet, le résultat d'une évaluation créative doit être considéré dans son environnement car il fournit un aperçu de la logique derrière ce jugement [Plucker et al., 2004]. Ainsi, il convient de considérer trois

composantes : la nature de la tâche, les critères spécifiques à évaluer et les moyens d'évaluation. Tout d'abord, concernant la nature de la tâche, à un niveau plus abstrait, nous devons prendre en compte le processus de réflexion suscité. Un individu engagé dans un processus créatif est engagé dans un processus cognitif. À ce titre, les mécanismes cognitifs derrière la génération d'idées nouvelles et utiles peuvent différer selon le type de tâche envisagé. La distinction classique faite dans la littérature est celle entre la pensée divergente et la pensée convergente Guilford [1956], où la première correspond à la génération de plusieurs idées, tandis que la seconde dérive d'une solution logique unique. En conséquence, les tâches peuvent être caractérisées par le type de processus de réflexion qu'elles suscitent. Cette différenciation entre les tâches peut se traduire par l'ouverture des tâches, c'est-à-dire l'importance des contraintes imposées à la résolution du problème. Au-delà de la dichotomie habituelle entre les tâches ouvertes et fermées, qui déclenchent respectivement un processus de pensée plus divergente ou plus convergente, Attanasi et al. [2021] ajoutent un troisième type de tâche, à savoir les tâches ouvertes-avec-contraintes, qui suscitent encore un processus de pensée divergente tout en renforçant les contraintes sur la résolution de la tâche. Ensuite, la performance créative repose sur les critères de créativité choisis pour évaluer la performance. Il existe une vaste gamme de critères possibles. Comme mentionné précédemment, à partir des travaux fondateurs de Guilford [1967], les auteurs ont principalement évalué la performance créative à travers l'originalité (*à quelle fréquence une solution particulière est-elle donnée ?*), la fluidité (*combien d'idées ont été générées pour résoudre un problème spécifique ?*), et la flexibilité (*combien de thèmes couvrent l'ensemble des idées générées ?*), ou l'élaboration (*à quel point les idées sont-elles détaillées ?*). Cependant, de manière plus pratique et plus proche de la définition de la créativité, on pourrait imaginer des tâches où l'importance de l'originalité est plus grande que celle de l'utilité, et vice versa : une production artistique par rapport à un nouveau produit issu d'un département RD, par exemple. En conséquence, les critères de créativité choisis seraient évalués sur deux dimensions : l'originalité et la faisabilité. Dans tous les cas, en tant que concept multifacette, il est crucial d'évaluer la créativité d'une idée à travers divers critères qui couvrent toute la complexité de ce concept. Enfin, en ce qui concerne les moyens d'évaluation, nous devons prendre en compte deux distinctions principales : les mesures objectives par rapport aux mesures subjectives et les juges externes par rapport aux juges pairs. D'une part, les critères de créativité, selon leur définition, peuvent être évalués par le jugement subjectif des évaluateurs ou par des

mesures objectives permises par l'analyse quantitative des idées. Dans le cadre de cette thèse, les deux types de mesures seront utilisés, combinant l'évaluation subjective de la *consensual assessment technique* [Amabile, 1982] mais aussi des mesures objectives, en partie calculées à l'aide de techniques de *Natural Language Processing*. D'autre part, principalement pour les mesures subjectives, l'évaluation créative des résultats nécessite qu'un ou plusieurs juges évaluent chaque critère. Par conséquent, le chercheur fait face à deux possibilités : la créativité évaluée par des juges pairs, c'est-à-dire l'évaluation via d'autres sujets dans l'expérience, ou des juges externes, c'est-à-dire des individus aveugles au traitement. La seconde possibilité est souvent considérée comme la plus fiable, c'est pourquoi elle a été choisie tout au long de cette thèse.

Cependant, en plus de la nature de la tâche, des critères choisis ou du type d'évaluation, une partie du contexte créatif est également le niveau d'observation choisi. La littérature reconnaît trois principaux niveaux d'observation : la créativité individuelle, la créativité en équipe et la créativité organisationnelle. La créativité individuelle est un élément important des théories composant et interactionniste de la créativité, où les différences individuelles interagissent avec des facteurs contextuels. Même si l'étude de la créativité individuelle peut sembler limitante parce qu'elle exclut les niveaux d'observation supérieurs de la créativité, la créativité individuelle est la source de la créativité en équipe et organisationnelle [Amabile, 1988], et s'engager dans la créativité en équipe et organisationnelle sans cette étape fondamentale empêcherait les chercheurs de saisir l'ensemble du processus créatif. L'étude de la créativité individuelle est cruciale pour identifier les facteurs personnels qui influencent la performance créative, permettant ainsi aux organisations de mieux cibler les bons individus pour des tâches spécifiques et, par la suite, d'évaluer quels profils individuels pourraient se compléter une fois concentrés à un niveau d'équipe. En effet, se concentrer uniquement sur la créativité individuelle peut également ne pas être suffisant. La créativité doit également être étudiée du point de vue des équipes, où les équipes sont définies comme "*interdependent collections of individuals who share responsibility for specified outcomes*" [Sundstrom et al., 1990, 120]. L'importance de l'étude de la créativité en équipe a émergé depuis qu'il y a une tendance croissante pour de nombreuses tâches à nécessiter une collaboration au sein d'une équipe [Mathieu et al., 2017] en raison de la complexité croissante des tâches et de l'ensemble élargi de compétences requises pour mener à bien ces tâches [Fiore et al., 2010, Kozlowski and Ilgen, 2006, Sundstrom and etc., 1998]. Cependant, les équipes ne sont pas un simple

agrégat d'individus mais les interactions complexes de ces individus et leurs spécificités [Woodman et al., 1993]. En plus des profils des membres de l'équipe, l'étude de la créativité en équipe nécessite d'inclure les caractéristiques, les processus et la dynamique de l'équipe. Ainsi, étudier la créativité en équipe permet de comprendre les interactions complexes entre les individus impliqués dans un processus créatif, renforçant ainsi la dimension sociale de celui-ci. Enfin, la créativité organisationnelle découle de la créativité individuelle et de la créativité en équipe, bien que toutes ne se traduisent pas directement en résultats organisationnels. En tant que niveau d'observation supérieur, la créativité organisationnelle dépend également fortement des facteurs contextuels [Amabile et al., 1996, Andriopoulos, 2001, Chang and Chiang, 2008] mais inclut des éléments spécifiques uniques à la vie des organisations. Quel que soit le niveau d'observation spécifique, il n'est pas suffisant de les étudier séparément mais de les considérer conjointement, en identifiant leurs interrelations [Hargadon and Bechky, 2006, Oldham and Cummings, 1996, Shalley et al., 2009]. L'analyse inter-niveaux de la créativité est donc nécessaire pour saisir le processus créatif et comprendre ses subtilités. Cependant, il est important de noter que cette thèse se concentrera principalement sur les deux premiers niveaux de créativité — individuel et en équipe — plutôt que sur le niveau organisationnel. Ce choix de concentration a été fait parce que la créativité individuelle et en équipe forme la base de la créativité organisationnelle, et le contexte dans lequel notre recherche a été menée n'impliquait pas d'artefacts organisationnels.

Enfin, suite à l'étude des déterminants de la performance créative, se pose la question de l'identité de l'idéateur. Pendant longtemps, la créativité a été considérée comme un privilège humain, les avancées récentes en intelligence artificielle (IA), en particulier suite au lancement de ChatGPT en 2022, ont ravivé le débat sur la question de savoir si la créativité est exclusivement une entreprise humaine. En effet, la créativité est censée faire partie de ce qui nous rend humains [Sawyer and Henriksen, 2024], mais les capacités des machines à produire des résultats créatifs pertinents défient considérablement l'esprit humain. [Amabile, 2019] définit la créativité artificielle comme “*the production of highly novel, yet appropriate, ideas, problem solutions, or other outputs by autonomous machines*”(p. 3). Ce qui émerge de l'étude de la créativité artificielle, c'est la tension entourant la capacité même d'une machine à produire des idées nouvelles et utiles. Pour de nombreuses personnes, l'IA devrait être utilisée pour gérer nos tâches ménagères, et non être conçue pour prendre en charge nos efforts. En mars 2024, le message de Joanna Maciejewska sur la plateforme *X* est de-

venu viral. “*I want AI to do my laundry and dishes so that I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes*”. Cet appel à empêcher l’expansion de l’IA dans le domaine de la créativité a déclenché des débats sur la possibilité de créativité artificielle. Plus précisément, cette résistance à l’IA dans les domaines créatifs découle non seulement d’une aversion algorithmique, mais aussi de la conviction que la créativité est centrale dans notre société et souvent considérée comme la performance de plus haut niveau accordée à l’humanité [Taylor, 1988]. En fin de compte, lorsque la créativité artificielle est examinée, il y a de l’excitation et de la peur, et des investigations sont nécessaires pour la comprendre et démêler ses implications et ses conséquences possibles. Cependant, ce changement de paradigme ne diminue pas notre intérêt à étudier la créativité humaine. Puisque l’IA manque d’agence et nécessite une intervention humaine initiale pour résoudre un problème créatif. Soulignant particulièrement la question de la résolution et du cadrage des problèmes, et donc la place centrale accordée aux interactions humain-machine.

Plan de la thèse

Bien que la créativité ait été étudiée dans divers domaines de recherche depuis plusieurs décennies, de nombreux aspects demeurent encore peu explorés. Avant d’aborder les questions de recherche spécifiques à chaque chapitre de cette thèse, il est essentiel de présenter la question de recherche générale que cette thèse vise à traiter. Comme souligné dans le premier paragraphe de ce chapitre d’introduction, qu’il s’agisse de ses dimensions individuelles, collectives ou artificielles, la créativité alimente notre capacité à innover et, par conséquent, à évoluer. Il existe donc un besoin croissant, au sein de nos sociétés, d’explorer comment améliorer la performance créative, afin de permettre à la créativité—malgré ses risques et incertitudes inhérents—de s’épanouir et de réussir. Diverses perspectives peuvent être adoptées pour examiner la question de la performance créative, et il est crucial de souligner cette multidimensionnalité. Plus précisément, la performance créative repose sur quatre facettes principales : (1) l’identité de l’agent créatif qui effectue une tâche créative, (2) la nature de la tâche évaluée, (3) les multiples critères de créativité choisis pour évaluer la performance, et (4) les facteurs intrinsèques et environnementaux qui influencent cette performance. Ainsi, bien que chaque chapitre de cette thèse se concentre sur des questions de recherche spécifiques, l’enquête sous-jacente est tou-

jours ancrée dans la nature multidimensionnelle de la performance créative.

Avant d’aborder la structure de cette thèse, les questions de recherche propres à chaque chapitre, ainsi que la stratégie de recherche adoptée, il est important de présenter la méthodologie choisie et d’expliquer les raisons de ce choix. Cette thèse s’appuie principalement sur la méthodologie expérimentale. Ce choix n’a pas été fait au hasard. La créativité est un processus social, au cœur duquel se trouvent les individus et leurs interactions avec l’environnement. Ainsi, l’économie expérimentale offre un cadre unique et prometteur pour observer un tel phénomène, permettant aux chercheurs d’étudier des mécanismes, des processus ou des dynamiques complexes dans un environnement contrôlé [Friedman and Sunder, 1994, Jacquemet and L’Haridon, 2018], et ce, quel que soit le niveau d’observation. Alors que la méthodologie expérimentale actuelle permet d’étudier la créativité individuelle (*Chapitres 2 et 5*), une des contributions de ce travail est d’examiner comment cette méthodologie peut être adaptée à l’étude des processus d’équipe (*Chapitres 1, 3 et 4*).

Chapitre 1 – *Rethinking laboratory experiments : The case of collective experiments*

Ce chapitre introductif discute de l’avancement de la méthodologie expérimentale en économie avec la mise en œuvre prometteuse d’expériences collectives, définies comme “*toute expérience permettant d’observer des processus collectifs à travers les interactions sociales des sujets sans intermédiaire systématique*”. Bien qu’un nombre énorme de protocoles aient été développés au cours des 70 dernières années, garantissant que la méthodologie expérimentale en économie est devenue un courant de recherche à part entière, les chercheurs se sont principalement concentrés sur l’étude des individus. Cependant, peu d’attention a été accordée à l’adaptation de cette méthodologie pour étudier les processus et les interactions collectives entre les individus. Il ne faut pas oublier que l’économie repose sur un socle solide d’interactions sociales et d’interdépendance. Par conséquent, diminuer l’importance de l’étude des dynamiques collectives, comme cela a souvent été le cas, est une erreur. En effet, la littérature expérimentale a exploré des situations d’interdépendance ; cependant, peu de protocoles expérimentaux impliquent des interactions directes entre sujets avec un accent spécifique sur les processus collectifs émergents et influençant significativement à la fois les résultats individuels et collectifs. Ce chapitre propose de réfléchir à cette limitation et démontre comment les économistes expérimentaux devraient adapter leurs meilleures pratiques méthodologiques actuelles pour étudier les

processus collectifs. Le chapitre est structuré autour de quatre sections : une analyse historique et méthodologique de l'approche expérimentale en économie, une justification du besoin d'expériences collectives, l'importance d'éliminer les intermédiaires et d'utiliser des outils innovants pour améliorer les protocoles expérimentaux, et enfin, un examen de l'exemple du *Social Interactions Lab* (SIL) pour observer les processus collectifs.

Chapitre 2 – *The Many Faces of Creative Profiles : Exploring task openness*

Le deuxième chapitre aborde le niveau individuel de la créativité. La créativité est un élément central pour la croissance et l'évolution de nos sociétés [Sawyer and Henriksen, 2024], ainsi que pour les organisations. Ainsi, comprendre les caractéristiques clés des profils créatifs est d'autant plus important pour identifier les individus ayant un potentiel créatif remarquable qui pourrait servir les organisations et les sociétés. Se concentrer sur la créativité individuelle n'est pas une nouveauté dans la littérature. La créativité a longtemps été envisagée comme un effort individuel et solitaire, uniquement réalisable par les meilleurs humains. Cependant, il a été reconnu que la créativité n'est pas une qualité innée de certains humains, mais une capacité qui peut être entraînée et améliorée. En tant que tel, il devient naturel de se demander quelles sont certaines caractéristiques spécifiques qui expliquent pourquoi certains individus performant mieux que d'autres. Cependant, la créativité individuelle est souvent analysée à travers la motivation intrinsèque, les styles cognitifs ou les traits de personnalité [Amabile and Pratt, 2016, Amabile, 1983] ; moins d'attention a été accordée aux connaissances et compétences et à la manière dont ces éléments sont entrelacés avec d'autres caractéristiques individuelles. De plus, en raison de la variété des tâches, une question supplémentaire est de savoir dans quelle mesure le degré d'ouverture des tâches créatives, ainsi que l'importance des contraintes sur la résolution des tâches, influencent la performance créative des individus. S'appuyant sur l'abondante littérature sur la créativité individuelle, ce chapitre explore les déterminants de la performance créative, en se concentrant particulièrement sur la manière dont différents profils créatifs réagissent à des degrés variés d'ouverture des tâches. En fin de compte, quelles sont les caractéristiques clés des profils créatifs ? Et, comment différents degrés d'ouverture des tâches modèrent-ils les relations entre les caractéristiques des profils et la performance créative ? Pour ce faire, un protocole expérimental intra-sujets a été mis en place pour examiner la relation entre les profils

des individus et leur performance dans une tâche donnée. Les résultats qui en découlent informent le développement efficace d'équipes diverses capables de répondre aux exigences multifacettes des organisations.

Chapitre 3 – *Diversity, Interactions, and Team creativity : An experimental perspective*

En dépassant la vision potentiellement limitante de la créativité comme une simple poursuite individuelle et en reconnaissant sa nature intrinsèquement collective, le Chapitre 3 explore naturellement la créativité d'équipe et les déterminants de la performance créative des équipes. Si la créativité d'équipe est un élément crucial dans l'étude des organisations [Amabile et al., 1996, Sarooghi et al., 2015, Chatzoglou and Chatzoudes, 2018, Rousseau et al., 2015], peu de recherches ont été menées en économie sur la créativité d'équipe dans un contexte expérimental. Ce troisième chapitre examine deux facteurs intrinsèques principaux de la créativité d'équipe : la diversité des équipes et les réseaux sociaux. En particulier, comprendre comment la diversité et les réseaux sociaux impactent conjointement la performance créative est primordial pour saisir la complexité des interactions sociales dans le processus créatif d'équipe. D'autant plus en étudiant le chevauchement des différents types de diversité [van Knippenberg and J. Hoeve, 2021]. En effet, dans quelle mesure la diversité en termes de composition d'équipe impacte-t-elle la performance créative collective de l'équipe ? Et comment les interactions des individus, à travers des réseaux préexistants ou in situ, affectent-elles la performance créative de l'équipe ? De plus, ce chapitre vise également à introduire, pour la première fois, une expérience collective dans laquelle les intermédiaires entre les sujets ont été supprimés et où ces sujets sont impliqués dans des interactions en face à face. Comme souligné dans le Chapitre 1, l'économie bénéficierait de nouveaux protocoles qui observent des interactions directes entre sujets sans intermédiaires, reflétant mieux le véritable processus créatif du travail d'équipe. À noter que le protocole expérimental utilisé à l'époque n'est pas sans rapport avec celui utilisé dans le Chapitre 2, puisque le présent chapitre inclut également le niveau d'observation individuel, utilisant la performance créative individuelle des membres de l'équipe comme un prédicteur de la performance créative des équipes. Cela nous permet de mener une analyse croisée de la créativité et de relier la performance créative individuelle et celle de l'équipe pour comprendre pleinement le phénomène créatif. Sachant que, jusqu'à présent, la relation entre la créativité individuelle et collective n'a pas reçu suffisamment d'attention

[van Knippenberg and J. Hoever, 2021]. En effet, lors de l'étude de la performance à plusieurs niveaux, la plupart des études ne testent expérimentalement que la créativité d'équipe, tandis que les questionnaires évaluent la créativité individuelle. Enfin, les résultats du Chapitre 3 complètent ceux du Chapitre 2 en testant la performance créative de l'équipe sur la base de la propre performance des membres de l'équipe et des facteurs intrinsèques à l'équipe.

Chapitre 4 – *Does Creativity Thrive on Plot Twists ? Exploring the role of surprise on team creativity*

Contrairement aux Chapitres 2 et 3, le Chapitre 4 ne se concentre pas seulement sur l'effet des facteurs intrinsèques sur la performance créative au niveau individuel ou d'équipe, mais aussi sur l'effet des facteurs environnementaux, ici la surprise. Les organisations opèrent dans des environnements en constante évolution, sujets à des événements soudains et inattendus qui pourraient affecter leurs processus et leur performance. Si la créativité est un élément central des organisations, examiner la relation entre la créativité et la surprise devient essentiel. D'autant plus que la créativité, par nature, est une entreprise incertaine et risquée, et cette incertitude est amplifiée par la présence d'événements inattendus qui peuvent modifier les conditions dans lesquelles émergent des idées nouvelles et utiles. En termes de littérature, ce chapitre se situe à l'intersection de la créativité, de la surprise et de la théorie des schémas pour comprendre comment les membres d'une équipe vivent et, éventuellement, surmontent un événement inattendu qui affecte le processus créatif en cours. Ce chapitre questionne particulièrement comment un événement inattendu prenant la forme d'un changement dans la spécification des contraintes impacte la performance créative d'une équipe selon le degré de surprise introduit dans une tâche. Et comment cette relation entre la surprise et la créativité de l'équipe est affectée par d'autres facteurs, tels que la diversité de l'équipe, ses processus et ses capacités créatives. Le protocole expérimental qui en découle vise à répondre à ces questions de recherche en comparant les équipes éprouvant la surprise et celles qui ne le font pas, et en comparant différents degrés de surprise, avec une surprise faible et une surprise élevée. Ce faisant, ce chapitre prépare le terrain pour une compréhension plus approfondie des facteurs qui déterminent l'effet des changements inattendus sur la performance créative de l'équipe, le rôle de la diversité de l'équipe, et les processus qui facilitent ou entravent l'adaptabilité d'une équipe face à la surprise.

Chapitre 5 – *Can AI Enhance its Creativity to Beat Humans ?*

Dans la continuité des chapitres précédents, le cinquième chapitre de cette thèse explore la performance créative des humains, tout en intégrant la possibilité que la créativité humaine soit remise en question par la créativité artificielle. Ce chapitre met également en lumière l'importance d'utiliser des critères multiples pour évaluer la performance créative, afin d'aligner ces deux formes de créativité. Si la créativité a principalement été étudiée comme une caractéristique humaine, la recherche florissante sur la créativité artificielle constitue une voie de recherche importante. La possibilité qu'une machine surpasse un humain dans une tâche (créative) soulève des questions sur la substitution de l'un à l'autre dans l'exécution de cette tâche. Cependant, cela nécessite une compréhension approfondie des forces et des faiblesses des deux agents, plaidant peut-être pour une complémentarité plutôt que pour une substituabilité. Ce chapitre interroge spécifiquement dans quelle mesure la créativité artificielle pourrait défier la créativité humaine en comparant leurs performances dans différents types de tâches, variant en degré d'ouverture, et selon la stratégie de demande choisie pour aborder le modèle d'IA. En rassemblant des productions créatives provenant de protocoles expérimentaux avec des sujets humains et en générant des productions créatives via *ChatGPT* et *Dall-e*, une expérience en ligne a été conçue pour que des évaluateurs humains évaluent la performance créative des productions réalisées par des humains et de l'IA. Un élément important de ce chapitre est l'attention accordée à la spécification de la performance créative, car des mesures à la fois objectives et subjectives de la créativité ont été utilisées pour évaluer si l'intelligence artificielle pourrait véritablement défier les humains. Pour conclure, les résultats du chapitre 5 fournissent une comparaison détaillée entre la performance créative humaine et artificielle, menant à des discussions sur la meilleure façon d'envisager la collaboration entre humains et machines en fonction de leurs forces et de leur possible complémentarité.

Chapter 1

Rethinking Laboratory Experiments: The Case of Collective Experiments

Summary of the chapter

This chapter aims to explore potential advancements in experimental economics by conducting *collective experiments*. Initially, economics was deemed unsuitable for experimentation due to the absence of necessary preconditions favourable to its development. The chapter highlights how economics evolved to adopt experiments as its theoretical and technical landscape matured. Despite these advancements, experimentalists did not invest fully in the study of collective processes. This chapter argues that the lack of *collective experiments* is rooted in the historical, methodological, and technological context of the discipline's practices and emphasizes the potential of removing intermediaries and leveraging innovative tools to enhance our understanding of collective processes. Finally, the chapter calls for adapting the research infrastructure itself, the experimental laboratory, to this new methodological approach.

Part of this chapter was published as:

Gil-Gallen, S., & Maltese, A. G. (2023). How can technologies help disclose new insights into collective behaviours?. *Journal of Behavioral Economics for Policy*, 7(2), 21-32.

1.1 Introduction

For a long time, economics was not considered an experimental discipline. Although some disciplines, such as biology or chemistry, are recognised as experimental, economics has only resorted to experiments in recent decades. In order to understand why economics was not considered suitable for experimentation, we need to consider the context and stage of the economic field at the time. Friedman and Sunder [1994] argue that a discipline’s context determines its ability to conduct experiments, rather than the discipline being inherently experimental or non-experimental, i.e., a discipline will only become experimental when the theories and the techniques in use allow it. Even *hard* sciences, traditionally characterized as experimental, initially engaged in such methodologies at different points in history. This was the case for physicians, who only implemented experimental protocols during the 16th century, or chemists and biologists, who resorted to experiments during the 19th century — several centuries after physics but only a few decades before psychology, one of the first experimental social sciences. Experimental economics followed this pattern, and its development over the last century is a consequence of the shift from the inherent non-experimental nature of the discipline to the gained legitimacy of running experiments in a controlled environment. In the same vein, Weisberg and Muldoon [2009] state that any research approach can be characterized through four different components: “(1) *the research questions being investigated*, (2) *the instruments and techniques used to gather data*, (3) *the methods used to analyse the data*, and (4) *the background theories used to interpret the data*” (p. 228). Given the importance of these four components in shaping a research approach, directing our attention toward them might reveal the reasons behind the delayed emergence of experimentation in economics. The intuition further developed in this chapter is that economics could finally resort to experimentation because the four elements constituting the experimental research approach aligned with the economic field at that time. In other words, research questions close to the interests of the economic field were considered, the instruments and techniques required to run experiments appeared and grew on a larger scale, the method used for data analysis suited the approach, and finally, the background theories aligned.

With the inherent non-experimental nature of economics being discarded, the

experimental field has seen a dramatic increase in interest and publications, and economists outside the realm of experimentation allocated credit to this new methodology. One of the most salient examples of the experimental methodology's quick development is from the work of Samuelson and Nordhaus, who, between two editions of the same book on the basic principles of economics, changed their tune on this issue. In 1985, the authors supported that “*economists cannot perform the controlled experiments of chemists or biologists because they cannot easily control other important factors. Like astronomers or meteorologists, they generally must be content largely to observe.*” [Samuelson and Nordhaus, 1985, p. 8]. Less than ten years later, the same authors considered experimental economics as an “*exciting new development*” [Samuelson and Nordhaus, 1992, p. 5].

Year after year, experimental economics has established itself as a methodology and field of research in its own right, enabling the study of individuals' behaviour, decisions, and actions. Nonetheless, although experimental economics has now become a significant part of the economic discipline and a cornerstone of microeconomics, the existing literature lacks a focus on adapting this methodology to study collective processes and interactions between subjects. While numerous experimental protocols aim to observe behaviours in situations of interdependence, such as public goods or dictator games, it is rare to find references that do include interactions between subjects, even more so when those interactions are not regulated by an interface, either physical or digital. In fact, while the study of individuals has facilitated the aggregation of the behaviours, decisions, and choices of economic agents, the group as a unit of analysis has received very little attention. Thus, the issue is the fundamental difference between examining aggregates of individuals and studying a proper entity such as a group. This lack of focus on the collective dimension is all the more damaging given that, as will be explained later in this chapter, (experimental) economics is naturally rooted in collective processes. Nonetheless, based on previous considerations, we assume that this gap in experimental economics was not intentional but rather due to the state of the prevailing approach towards conducting experiments at the time. Indeed, the context in which the first experimentalists evolved largely influenced their decisions as to the proper conduct of experiments in economics, leading them to formulate founding principles in specific directions, leaving little room for the development of a collective approach to experimentation. Today's literature is still based on these same principles, making it difficult to move away from them. This chapter reflects on these elements and demonstrates how experimentalists could

carry out the changes required to develop collective experiments, allowing a deeper understanding of collective processes.

This chapter is articulated into four distinct parts. Section 1.2 considers historical and methodological insights on the prevalent experimental method in economics. Then, Section 1.3 focuses on the rationale for conducting collective experiments. Section 1.4 and Section 1.5 respectively underline the need for removing intermediaries and using innovative tools and technologies to increase the potential of experimental protocols in the context of collective experiments. Consequently, Section 1.6 builds on the example of the *Social Interactions Lab* (SIL), an experimental laboratory designed to observe collective processes. Finally, Section 1.7 provides the concluding remarks and reflections on this chapter.

1.2 Historical and methodological insights

1.2.1 Development of the experimental methodology

As highlighted in the first lines of this chapter, beginning as a theoretical and observational discipline, economics has recently been portrayed as an experimental one. In its simplest definition, experimental economics can be described as a methodology driven by research questions in economics, with the aim of generating and collecting data in a controlled environment to observe subjects' decisions, actions, and behaviours. Over the last 70 years, this stream of research has gained recognition despite being relatively young compared to the extensive history of economic sciences. However, while the term *experimental economics* only appeared in 1967 in the work of Heinz Sauermann [Roth, 1995], determining the origin of the method is more complex. The aim of this section is not to provide a complete history of the experimental field but to shed light on certain key elements, especially in consideration of other factors discussed in this chapter.

Before going into more detail on these issues, it is important to note that this chapter is primarily interested in experiments in a controlled environment within an experimental laboratory. However, it should be noted that the development of the experimental methodology in economics has not only led to the establishment of laboratory protocols but also led to other branches of experimentation. These include field experiments (where observations are produced from “(i) a random allocation of individuals to the treatment, but (ii) in a ‘natural’ or ‘real’ environment” [Jacquemet

and L’Haridon, 2018, p. 83]), natural experiments (where “*the experimenter does not establish the perturbation but instead selects sites where the perturbation is already running or has run*” [Diamond, 1986, p. 12]), or even simulation experiments (where “*simulation [is] a numerical technique for conducting experiments with certain types of mathematical and logical models describing the behaviour of an economic system on a digital computer over extended periods of time*” [Naylor et al., 1967, p. 1316]). These other ramifications of the experimental method, not being included in a laboratory context, important though they are, will therefore not be a central element of the pages to come.

The origin of the field

If several authors have shown interest in the history of experimental economics [Buda, 2000, Cot and Ferey, 2016, Serra, 2012a], all seem to agree that the first formal protocol was developed by Vernon L. Smith in 1962, making him one of the founding fathers of the discipline. In reality, Smith [1962]’s famous protocol was a response to the experiment carried out a few years earlier by Chamberlin [1948]. As Smith explained later, he got his inspiration for this protocol when “[he] *was, wide awake at 3 a.m., thinking about Chamberlin’s “silly” experiment*” [Smith, 1991a, p. 155].

More specifically, E. H. Chamberlin conducted this “silly” market experiment with his students, wherein they engaged in pairs to discuss sale prices. After face-to-face negotiations, the buyers and sellers reported their results to the teacher. This preliminary research protocol aimed to extend Chamberlin’s theoretical work on how real-world markets deviate from the idealized concept of perfect competition [Chamberlin, 1946]. Beyond this contribution, Chamberlin emphasized the pedagogical value of such a method, allowing his students to play different roles in a market, and to compare the situation in the classroom with real-world markets. But, what truly intrigued him at the time was that “*It is a commonplace that, in its choice of method, economics is limited by the fact that resort cannot be had to the laboratory techniques of the natural sciences.*” [Chamberlin, 1948, p. 95]. Stressing the limitations of the experimental procedure then in place, Vernon L. Smith, who was also interested in the problem of imperfect information, chose to modify Chamberlin’s experiment by running it over several days and by introducing an auction market mechanism to organise the negotiations between buyers and sellers. This raises the question of the true origins of the discipline, or at least why Vernon L. Smith was and still is considered to be the founding father of experimentation in economics. As

a result, should the genesis of experimental economics be attributed to Smith? To Chamberlin?

Let's take Roth [1993]'s analogy on the discovery of America. According to him, the attribution of a discovery is often detached from its exact temporality but relates more to the moment when the discovery is integrated into common knowledge. Although America had been visited before Columbus, "*After Columbus, America was never lost again*" [Roth and Sotomayor, 1990, p. 170]. Serra [2022] takes over this analogy as the reason Vernon L. Smith was seen as the founder of experimental economics: by taking the methodology from being highly novel to a methodology in its own right.

In the end, although Vernon L. Smith is seen as the instigator of the first rigorous experimental protocol, numerous steps have enabled the field of experimental economics to emerge. As highlighted before, the development of the experimental methodology in economics was the result of concomitant individual and isolated initiatives that shaped a favorable environment where researchers benefited from the influence of others in other fields or geographical areas. In particular, two regions resorted as fertile ground for experimental economics to develop: the United States and Germany. Although these branches emerged at relatively the same time, they evolved separately, with a thin bridge between Reinhard Selten in Germany and Austin Hoggatt in the US, who discussed their work and experience. It is certainly for these reasons that the history of experimentation in economics may seem more tortuous. The list of possible antecedents of the discipline is long and creates disagreements around the first building block of experimental economics.

Looking back over the years, or even centuries, numerous episodes bear witness to the infancy of the methodology. The groundbreaking book of von Neumann et al. [1944], the appearance of business games in 1926 [Kaufmann et al., 1976], Wesley Clair Mitchell's advocacy on the necessary transformation of economics into an experimental discipline during the 37th Congress of the *American Economic Association* in 1924 [Mitchell, 1925], Bernoulli and the St Petersburg Paradox Roth [1993], or even David Hume's¹ intuition to study human minds following a rigorous methodology close to the one used in natural sciences [Moffatt et al., 2009, Sugden, 2006] are just a few examples of the first steps of experimentation in the economic discipline. Paradoxically, we could even trace the origin of experimental economics to the very prohibition laid down by Mill [1836], who denied the moral sciences the possibility

¹In *A Treatise of Human Nature* [Hume, 1739-40].

of resorting to experimentation, opening the door to discussion and possible emancipation from this statement.

The shaping of the field

From the seminal work of von Neumann and Morgenstern [1964] on the “*Theory of games and economic behaviour*,” a new impetus has emerged on the possibility of running experiments in economics. However, it would appear that the discipline’s beginnings were difficult before accelerating some twenty years later. As a result, different perspectives were adopted to understand the development of the experimental field, most of them choosing a linear historical approach relating pivotal events or phases. Among them, Serra [2012a] distinguished four consecutive major stages: (1) emergence (1945-1960), (2) commencement (1960-1970), (3) take-off (the 1980s), and (4) maturity (since the mid-1990s). Cot and Ferey [2016] identified three phases: (1) the early steps (1945-1970), (2) the foundations (1970-1990), and (3) the success (1990-2010). Yet, both of them similarly chronicle the same dynamics that served the expansion of the experimental field in economics.

While none of these approaches is incorrect, it also seems appropriate to put the evolution of experimental economics into perspective through the prism of Weisberg and Muldoon [2009]’s representation of a research approach. As a reminder, the authors present scientific approaches as “*narrow specifications of how an individual scientist or research group investigates the topic*” and distinguish “(1) *the research questions being investigated*, (2) *the instruments and techniques used to gather data*, (3) *the methods used to analyse the data*, and (4) *the background theories used to interpret the data*” (p. 228). Following the same logic, the central topic would be the individuals’ behaviours or decisions, expecting that this could be then divided into several branches for narrower interests.

The alignment of research questions and background theories

To understand the development of experimental economics, it is important to comprehend what research questions, but also which background theories, justified its expansion at the time and why it was not the case before. We must first consider the experimental discipline’s main objective in the social sciences beyond the economic question. The first aim of experimentation is to observe and understand the behaviour of individuals concerning their decisions and actions within a social framework. As a result, the advent of experimental economics represented a novel

approach to investigating agents' strategic behaviours from a fresh perspective that sparked considerable interest within the economic discipline.

In its early stages, experimental economics pursued three primary interests: (1) individual decision-making under certainty, uncertainty, or risky conditions, (2) decisions involving interactions among a limited number of agents that could be modelled, and (3) impersonal exchange mechanisms through market institutions [Serra, 2012a, Davis and Holt, 1993]. Regarding individual decision-making, the literature highlights seminal works such as Thurstone [1931]'s exploration of indifference curves, Mosteller and Nogee [1951]'s study of decisions in risky environments involving lotteries, and the presentation of Allais' Paradox at an international colloquium in Paris [Allais, 1953], which challenged the predictions of expected utility theory. In terms of decision interactions, notable experiments included the iconic prisoner's dilemma by Flood [1958], testing non-cooperative equilibria derived from John F. Nash's theoretical work, as well as studies on negotiation in duopoly and oligopoly by Sauermann and Selten [1959], Hoggatt [1959], and Siegel and Fouraker [1960]. Lastly, experimental studies on markets expanded significantly, starting with Chamberlin [1948]'s initial work and subsequent contributions by Vernon L. Smith.

Nevertheless, the development of these research themes is not coincidental and demonstrates a strong parallel with the state of microeconomics at that time. As stated by Guala [2010], "*with the slow exhaustion of general equilibrium theory [...], the turmoil in macroeconomics, and an increasing disillusionment about econometrics, the seventies created the conditions for the seeds of the 1940s and 50s to finally blossom. Experimental economists were in a position to take advantage of this situation.*" (p. 104). Thus, under the hegemony of standard microeconomic theory, the experimental methodology developed as an extension for the study of these research themes. Even today, we find traces of this co-evolution in the choice of many experimentalists to base their work on an initial theoretical model before conducting their experimental procedure. Thus, in a fairly straightforward manner, the issues addressed above can naturally be linked to considerations in the economic field, creating a bridge between experimentation and more conventional approaches to economics. At the time, experimental economics, by examining real-world phenomena and dynamics that the existing economic theories and models may not fully capture, was able to explore a broader range of research questions by providing an iterative and dynamic contribution to the discipline.

A final point that appears decisive in the development of experimental economics

is its integration into the mainstream. Even though the first experiments in economics were conducted by psychologists, for some researchers, “*experimental economics originated as an outgrowth of game theory*” [Innocenti, 2000, p. 2]. As a result, game theory and experimental economics intrinsically share common features when looking at their development. The most important common ground at the time was their heterodox orientation, explaining why their effective introduction in the economic field was postponed until the 1960s or 1970s [Innocenti, 2000]. In fact, it is crucial not to fall into the relatively classic trap of considering the birth of experimental economics as based solely upon a desire to validate or invalidate existing economic theories. In this sense, experimental economics benefited from its duality, between exploring unknown strategic behaviours of agents and testing already existing theories. However, for experimentation in economics to achieve its full potential, it was also necessary to move closer to more conventional concerns in the economic field.

The implementation and growth of instruments, techniques, and methods to gather and analyse data

If we reflect on John Stuart Mill’s warning on the use of experimentation in economics, we may well ask what enabled this reversal, including in the collection and analysis of data. Over time, economics became a science based on instruments [Morgan, 2023]. Indeed, it was when economists began to consider the need for *scientific rigour* in their discipline that the door was opened to the development of experimental methods [Cot and Ferey, 2016]. While we are aware that this growing interest was concomitant with the increasing weight given to econometrics and modelling methods, the experimental method, through the analytical tools available, could only emerge favourably. On the side of statistics, it’s important not to overlook the fact that, over the decades, the notion of “*experimetrics*” has become established, with a shift from more descriptive quantitative techniques to the advanced use of econometrics. Indeed, while Siegel’s early work focused on non-parametric statistical analysis, the development of the discipline showed a clear preference for the use of more complex techniques [Vallois and Jullien, 2018]. Furthermore, regarding modelling, Guala [2012] critically points out that “*From Mill to Marshall it was more or less taken for granted that economics was mainly concerned with the study of “real-world” markets, it was now possible to argue that economics was concerned with the study of whatever could be modelled by economic theory*” (p. 600) and to conclude

that “*the rise of modelling is probably the most relevant phenomenon for the birth of experimental economics*” (p. 600). The popularity of these analytical tools paved the way for experimental economics as a complementary method for testing theoretical projections. Experimental economics not only validates or invalidates these projections but also refines hypotheses and contributes to theoretical debates, especially when multiple theories are in opposition.

However, when considering the development of experimental economics, tools such as modelling and experimetrics are not the only aspects to examine. Long before the results are analysed, the question arises of how data are collected. If the protocol from Chamberlin [1948] did not seem to require much more investment in terms of material or tools, the rapid growth of the experimental field can be explained by the emergence of complementary tools and technologies facilitating the organisation of experiments. As envisioned by Morgenstern [1954], the use of computerised experiments has dramatically transformed economics and the prospect of developing an established experimental field. This revolution was instigated by some pioneer researchers whose initiatives have enabled the practice of computerised experiments to take root on a long-term basis. Historically, it is possible to identify the first computerised laboratory for controlled experimentation in economics, established by Austin Hoggatt in the 1960s. This specific shift from pen-and-paper experiments to human-to-computer interfaces is all the more important knowing that experimental economics could be then described as an instrument-based discipline where the infrastructure itself, the experimental laboratory, is regarded as a location for generating data, as well as the heart of experimental economics communities [Svorenčák, 2018].

As a central element in understanding experimentation in economics and its development, it is necessary to understand how the empirical contribution and the methodological contribution have co-evolved. However, growing interest in experimentation came with challenges. As it has been underlined by Roth [1987] on his own experience, “*When [he] began [his] own experimental work about a dozen years ago, it was most convenient to publish the results in journals of psychology and business*” (p. 1). Similarly, Reinhard Selten, admitted that he had turned away from experimentation for some time because of the suspicion that reigned around the seriousness of the experimental approach when he first started out [Selten, 2003]. Those breaks certainly influenced what Cot and Ferey [2016] highlighted as a case of *images of knowledge* preceding the *body of knowledge*, i.e., the development of a

field where the guiding principles or selectors emerged prior to the facts and open problems [Corry, 1989]. In other words, a world where experimentalists had to first ensure the scientific rigour of their protocols in order to defend the results emerging in the experimental literature. This point is all the more important if we consider that the need for rigour in research methods has led economists to shape a certain approach to experimentation, possibly hermetic to any incursion into the complex study of collective processes.

1.2.2 The fundamental principles

From testing to reshaping existing theories (*'Speaking to Theorists'*), by seeking to detect irregularities in individual behaviour (*'Searching for Facts'*), or by evaluating and contributing to the policy-making process (*'Whispering into the Ears of Princes'*), researchers have provided a wide range of roles to their experimental protocols [Roth, 1986]. Beyond the sole scientific purpose of experiments, this methodology has even taken on an additional role with experimentation as a tool to promote learning [Chamberlin, 1948, Villion, 2010]. Regardless of the role assigned to the experiments carried out by a researcher, the experimental methodology is based on a solid grounding of specific fundamental principles reflecting the willingness of scientific rigour mentioned earlier.

To start with, experimental economics revolves around a fairly simple triptych: environment, institutions, and behaviours [Smith, 1982]. The environment encompasses all subjects' characteristics, such as preferences, distribution of costs, or dotations. Institutions constitute the set of rules governing exchanges between subjects. Behaviours, as a function of the environment and institutions, are observed through the actions and decisions of subjects, along with their potential aggregation. These three components form the microeconomic system², which underpins the experimental method. As a result, experimental economics could be defined as studying changes of behaviours relying on the environment's modifications and institutions induced by artificial variations known as treatments.

Beyond those theoretical considerations, experimentalists have developed a number of rules and principles designed to structure the proper conduct of an experiment. While these principles are manifold, all of them converge towards one goal: *control*.

²In addition, Smith [1982] presents five precepts that shape the experimental methodology in economics: non-satiation, saliency, dominance, privacy, and parallelism.

Control ensures replicability³ and robustness of results. More specifically, control is defined as “*the capacity to manipulate laboratory conditions so that observed behaviour can be used to evaluate alternative theories and policies.*” [Davis and Holt, 1993, p. 15]. This is an essential cornerstone of the experimental methodology because it enables researchers to establish a clear causal relationship between variables. Researchers can effectively isolate the specific factor they wish to investigate while minimizing the impact of nonessential factors, and any observed effects can be attributed to the manipulated variable rather than being influenced by confounding factors.

Besides the aforementioned notion of replicability, other fundamental principles do shape the way experimentalists run their protocols; among them are monetary incentives, deception, and anonymity [Eber and Willinger, 2005]. To start with, it is mandatory to provide monetary compensation to subjects for their participation in an experimental session. A distinction must be made between two components of this payment: the fixed and the variable part. The first is unconditional and compensates the individual disutility generated by his or her participation in the experiment. It can also be perceived as a “moral debt” towards the experimentalist to ensure the subject’s engagement in tasks. The second is often dependent on the subject’s behaviour during the experiment, mainly performance-based. This part of the payment provides incentives for subjects to take the experiment seriously and make thoughtful choices closer to their real behaviour.

Then, economists almost unanimously agree that it is prohibited to lie to the subjects taking part in an experiment. The argument behind this ban is that a climate of suspicion might have a harmful effect on subjects’ behaviour if they were to discover that the experimenter had lied to them. The possibility of a lie would then bias all behaviour and distort the data collected. However, Hey [1998] made the following point: “*there is a world of difference between not telling subjects things and telling them the wrong things. The latter is deception, the former is not.*”(p. 397). As a result, the perception of deception and all the grey areas that can appear is heterogeneous [Charness et al., 2022]. It is, therefore, clear that when designing an experimental protocol, the researcher walks on a tightrope and, depending on the interpretation of each person, the researcher’s choices may be a genuine lie, a simple omission, or an ambiguous language.

³Replicability is defined as the ability given to other researchers to reproduce a specific experiment and find consistent results Davis and Holt [1993]. Understanding the experiment and how it was designed becomes crucial in certifying its proper construction and internal validity.

In addition, experimentalists rely on anonymity, which prevents subjects from identifying each other and allows experimentalists to limit social contagion problems or undesirable effects of possible interactions between them. The principle of anonymity can then be extended to a double-anonymity where the subject is also unaware of who the experimenter is. Whether it is single or double anonymity, the principle protects the researcher from losing control over his or her experimental setting.

Yet, another fundamental principle should also draw our attention. Decontextualization, or limited contextualization, allows experimentalists to observe decisions and behaviours true to the artificial variations induced by treatments instead of contextual elements that may pollute the results. In other words, subjects will act in the laboratory as they would act in the real world based on their own experiences and what they naively understand from the presented institutions, not based on their interpretation of a specific role or vision assigned to them by the experimental design, which would create a loss of control over the environment. Since the experimentalist will not be able to control how the context is perceived and envisaged by the different subjects in a way that is unique and specific to each, it is preferable to eliminate this context. A nuance can be added to this rule in the case of protocols with a decision support perspective, where context specification can be beneficial [Eber and Willinger, 2005].

The last two principles will be the main focus of the next sections. Indeed, both anonymity and decontextualization have major implications for the place of economics within experimental methods in social sciences and also for the investigation of collective processes in an experimental context. The reason why this focus is interesting lies in the idea that such founding principles of the discipline have formed a straitjacket around what can and can not be considered in an experimental approach. Our initial premise is the following: experimentalists might consider that Chamberlin [1948] underestimated the effect of the context, and possibly non-anonymity, in his classroom experiment as this could explain the discrepancy between Smith [1962] and his first results [Rullière, 2003]. In fact, when Vernon L. Smith modified Chamberlin's experiment due to biases introduced by the pre-existing relationships between the students, he stated good practices and first rules on how to run an experiment. However, could not the problem raised by Vernon L. Smith have been approached differently? An alternative approach might have been to understand and incorporate these disruptive elements into the experimental procedure. Therefore, the funda-

mental principles that did not really change since the early days paved the way for experimentation in economics but may have also created excessive path dependency for the lack of development of collective experiments.

1.2.3 The inherent role of the collective in experimental economics

Even though experimental economics is often referred to as an individual-focused field, it is rooted in collective processes. If this statement might seem surprising, we need to go back to the very definition of economics and the major role played by the collective component. In its most simple definition, economics is interested in the allocation of scarce resources in societies where needs are infinite. This allocation takes the form of exchanges between agents, where the exchanged element can take multiple forms encompassing both tangible and intangible goods. However, another way to consider those exchanges is to conceive it through the prism of interactions. In fact, the societies we live in are shaped by all the social interactions that occur in our daily lives. These interactions can take infinite forms, going from buying from the local grocery shop, taking a course at university, or working on a joint project in a team within an organisation. The study of these interactions is all the more important because they are based on the principle of the interdependence of agents. Interdependence is defined as “*the fact or condition of depending each upon the other; [a] mutual dependence*”⁴. This creates a bidirectional relationship between the social interactions of agents and their decisions and behaviours as they influence each other as they go along. Although an individual approach may be appropriate for the study of an economic phenomenon, the fact remains that this individual aspect is part of a larger, more complex, collective whole.

We can then ask ourselves how this collective aspect translates into the various areas of economics and, more specifically, into experimental economics. If experimental economics is often seen as the study of individuals’ decisions, behaviours, and actions, it would be incorrect to assume that the experimental literature completely overlooked all collective phenomena. In fact, research has been conducted on the outcomes of collective decisions, primarily focusing on the individual (or aggregated) behaviours and gains that result from such decisions. However, additional and deeper collective aspects of these interactions have often been excluded, with

⁴Definition from the Oxford English Dictionary (July 2023)

limited attention given to understanding the underlying mechanisms at play. Thus, the primary emphasis has been on analysing how individuals are affected by these interactions rather than delving into the intricate dynamics and processes driving these interactions.

Historically, experimental economics has been developed around three main interests: markets, individual decision-making under uncertainty, and interdependency situations [Davis and Holt, 1993, Serra, 2012a]. This led to the collision between two worlds, that of mathematicians and game theorists interested in negotiation, bargaining, or coalition building, and the field of experimentation. Among them were Kalisch et al. [1954], stating that *“because of the relatively undeveloped status of the theory, the authors feel that the use of an experimental approach is strongly indicated.”* (p. 302). Eminent names from the world of game theory, such as Thomas C. Schelling, Martin Shubik, or John F. Nash, have taken part in experiments to help bring these two disciplines closer together. Thus, when examining the main themes from the early days of the discipline, it is evident that experimentalists recognised the importance of interdependence among economic agents. The focus on interactions was present from the beginning. The primary difference between this early collective aspect and what will be discussed in this chapter’s following sections lies in the experiment’s purpose. While interactions were embedded in the experiments’ mechanisms or institutions, they did not critically examine the induced collective processes stemming from subjects’ interactions. As a result, it is crucial to explore the various types of collective processes encompassed within the existing experimental literature to understand how they have been studied (or not).

Based on our initial premise, social interactions greatly influence our behaviours and choices, making their study challenging due to the diverse range of individuals involved and the complex environmental factors at play. Given the intricate nature of social interactions and collective processes, the experimental methodology emerges as one of the most viable and advantageous means to investigate the fundamental mechanisms behind them. We then need to define the phenomena or types of interaction in question. In fact, competition, cooperation, coopetition, and collaboration are ways individuals or groups can interact and work towards the same goal or objective. However, they differ in the relationships and dynamics between the involved parties. First, competition refers to a situation where individuals or groups strive to outperform each other to achieve a goal or win a prize. Competition often involves a sense of rivalry and the desire to be the best or to come out on top.

In a different direction, cooperation refers to the act of working together towards a common goal or objective. It involves individuals or groups actively contributing to the achievement of the goal, but they may be working independently rather than coordinating their efforts. Then, coopetition is a concept that combines elements of cooperation and competition. It refers to a strategic approach where individuals simultaneously collaborate and compete with each other in order to achieve mutual benefits and maximize their own competitive advantage. Finally, collaboration involves more direct and active coordination between individuals or groups. It involves a higher level of interaction and communication, as well as a shared decision-making process. However, in the overwhelming majority of studies on interactions, the focus is on individual characteristics (e.g., altruism in a public good game) and/or the aggregated individualized outcome (e.g., individual gain in a dictator game compared to other subjects) by disregarding the social interaction process leading to that outcome, resulting in scarce references on those dynamics. In the end, we can observe that the current literature in experimental economics on possible collective aspects of behaviours is often limited to certain dynamics (competition, cooperation, or cooperation) but neglects others (such as collaboration). This is a crucial point, and Section 1.3 will explore how to address it.

1.2.4 Distinguishing factors from other disciplines

In order to understand experimental economics, we certainly need to dig into its development over time and the fundamental principles, but it might also be thought-provoking to compare it with other experimental disciplines that focus on similar phenomena, even though their principles or methodologies might differ. This is especially the case for psychology when we consider that the first experiments on economic phenomena were run by psychologists who specialized or were interested in decision-making or negotiation. In addition, it should not be forgotten that experimental economics has also given rise to behavioural economics, whose central principles are based more on the psychology of agents. This splintering of the discipline was not insignificant at the time, and the leading figures, Amos Tversky and Daniel Kahneman, made their differences clear with a return to the roots of experimentation anchored in the work of psychologists. This close relationship justifies comparing these two disciplines, including their similarities but also theoretical or methodological differences.

Focusing first on their shared characteristics, both psychology and economics,

as social sciences, are interested in individuals' decisions, actions, and behaviours. The strong connection between them, which also laid the groundwork for the emergence of experimental economics, mainly relies on their mutual perspective on the non-rationality of individuals. For this reason, Reinhard Selten, influenced by the work of Herbert Simon, underlines that “*the structure of boundedly rational economic behaviour cannot be invented in the armchair, it must be explored experimentally.*” [Selten, 2024].

If these two disciplines are similar in this respect, then we might wonder what has caused them to go off course and set themselves apart. While experimental economics has certainly found solid foundations in the early experimental work of psychologists, again, it has also naturally taken an interest in the work from the existing economic approaches and, more specifically, within game theory, as already described in Section 1.2, and the importance given to rules and institutions Smith [1962]. This can also be identified through one major concept already mentioned in the fundamental principle of experimental economics: contextualization.

As a starting point, while economics tries to suppress the context as much as possible, psychology attaches great value to it. Ariely and Norton [2007] defend that the main difference between economics and psychology, when it comes to experimental approaches, relies on the type of abstraction induced by the protocol. While bringing a phenomenon into the laboratory, economists aim to capture only essential elements and dispose of contextual factors as “*having a general normative theory allows for very general abstraction*” (p. 337). On the contrary, psychologists consider that all decisions are sensitive to the context and that bringing such specific context to the laboratory enables the researcher to observe behaviours as they might occur in the real world. This divergence in terms of abstraction justifies, to some extent, the theoretical and methodological differences in terms of incentives and deception [Ariely and Norton, 2007].

On the one hand, as was highlighted earlier, payment of the subjects taking part in an experiment is compulsory in economics. This has been justified by the willingness to trigger certain behaviours, the assurance of providing sufficient cognitive effort, or the perception of a moral debt towards the experimenter. However, there appears to be a sharp theoretical divide between economics and other social sciences when it comes to those financial incentives [Camerer and Hogarth, 1999]. Smith [1991b] attested that financial incentives “*are commonly absent in research of psychologists. This has made their work vulnerable to the criticism that the results are*

not meaningful” (p. 887). In fact, unlike economists, psychologists do not resort to financial incentives due to pragmatic, theoretical, and ethical considerations [Hertwig and Ortmann, 2003]. However, it is interesting to note that this distinction in the method of remunerating subjects has not always been such a clear-cut between economics and psychology. If we can trace the origins of experiments in economics back to the seminal work of psychologists, the same applies to remuneration practices. The first time subjects were paid in an experimental context was in the work of Sidney Siegel, then a psychologist, who was to be a major inspiration for the work of Vernon L. Smith, including on this point. Siegel’s work focused on the importance of real incentives in motivating subjects. Over the years, psychologists turned away from this practice, leaving the use of remuneration as a characteristic of economic experiments. Amos Tversky did not hesitate to rule and reiterate that paying subjects during an experiment was a bad practice [Serra, 2022].

On the other hand, when it comes to experimentation in psychology, there exists no ban on the use of deception in a protocol. In fact, psychology does not recognise any problem related to a possible negative effect on subjects’ behaviour, as is the case in economics, and can even consider it more a requirement than a choice [Kimmel, 1998]. This is strongly related to the value attributed to contextual factors and the need to bring subjects to a specific situation. Conversely, economists perceive deception as a threat since the protocol does not require it due to the introduced abstraction, and it might be the cause of detachment from the real-world situation.

In addition to the differences in contextualization, psychology seems to have given greater prominence to subjects interactions, as anonymity does not hold such a strong role of control in experiments. Indeed, psychologists and experimenters in psychology seem to consider anonymity as a research question in itself [Spears, 2021] more than an instrument for control. This explains why anonymity is generally not considered a fundamental principle for good practices in experimentation by psychologists, in contrast to its high importance in economics.

Beyond this fairly descriptive comparison between these two branches of experimentation in social sciences, there is a real interest in examining the similarities and differences between psychology and economics in setting up an experimental protocol. More particularly, as two disciplines in their own right, it does not seem that one is more valuable than the other in terms of the validity of its results.

The reasoning set out in this chapter highlights the idea that the methodological principles of a scientific approach can evolve over time as progress in data capture

and collection leads to more solid and reliable results. Although certain disciplines may have originated from others before fully diverging, as in the case of experimentation in economics, which partly stems from psychology. This does not prevent these disciplines from advancing by incorporating new shared insights. While a collective approach to experimental economics may seem to deviate from the established rules of the discipline, the broader experimental approach in the social sciences relies on various principles that, while sometimes discordant, are still valid. Understanding the practices of researchers who have already investigated the collective aspects of experimentation would contribute to implementing new practices in economic analysis.

1.3 Rationale and Implications of Conducting Collective Experiments

Social interactions greatly influence our behaviours and choices as much as behaviours and choices are influenced by social interactions. This makes the study of both even more challenging when researchers try to embrace the wide and diverse range of individuals involved and the complex environmental factors at play. In order to better understand these phenomena, one possible approach is experimentation and, more precisely, the use of collective experiments. As exposed in previous sections, laboratory experiments allow for observing behaviours in a highly controlled environment. The statement made in this chapter is that laboratory experiments do not allow researchers to study at their full potential the social interaction phenomenon due to the rigidity of established principles that would diminish the validity of their protocols. However, as a complementary method, field experiments offer the possibility to have a more natural representation of the phenomenon in terms of environment and sample. The distinction made between laboratory experiments and field experiments becomes relevant since what makes one of the strengths of field experiments, i.e., ecological validity⁵, represents the vulnerability of laboratory experiments, even more so when it comes to collective experiments. This section argues that this inherent limitation comes from the fundamental principles of anonymity and isolation of

⁵The ecological validity is defined as “*the relation between real-world phenomena and the investigation of these phenomena in experimental contexts*” [Schmuckler, 2001, p. 420] — the degree to which the results of a study reflect the natural, everyday environment in which the behaviour or phenomenon being studied would occur. Ecological validity is considered a subtype of external validity [Andrade, 2018].

subjects in experimental economics, where the solution to this issue is twofold: the removal of intermediaries and the reconsideration of data collection.

1.3.1 Definition of collective experiments

Yet, before examining the potentially limiting principles of the experimental method in economics, it is necessary to delve into the very definition of collective experiments. Collective experiments rely on two main necessary conditions. First, the capacity of such experiments to observe and capture collective processes. Second, the fact that subjects interact directly with each other without an imperative or systematic intermediary regulating their exchanges. We define an intermediary as *the deliberate virtual or physical separation of subjects, which creates a sense of distance and artificially restricts their potential interactions*, e.g. the computer interface widely used in the experimental literature. The use of the word systematic is crucial because it prevents researchers from using intermediaries unless this is required to address a specific research question. To provide an illustration, let's consider an experimental scenario where subjects engage in interactions via video calls. In this case, if the researcher employs this intermediary to replicate and simulate a real-world phenomenon rather than solely for the purpose of maintaining control, we can classify it as a collective experiment. In simpler terms, this second necessary condition emphasizes the significance of non-regulated interactions, recognizing the value of exchanges that are not impeded by excessive control. We will return later to the motivations behind this choice and the importance of such a change. As a result of the two conditions mentioned above, we define a collective experiment as *an experiment enabling the observation of collective processes by means of the social interactions of subjects without a systematic intermediary*.

It also becomes crucial to nuance the very definition of collective experiments to differentiate possible scenarios. This distinction is made according to the modalities of interactions and the nature of the outcome. When implementing a collective experiment, one researcher must recognise that the presence of a collective outcome alone does not determine the collective nature of the protocol. In fact, it is also defined by the effective interactions among subjects and how they result in specific decisions and behaviours. Our definition of a collective experiment should not be perceived as a straitjacket but more as a spectrum of possible experimental settings relying on the strength of the collective characteristics of the protocol. Figure 1.1 considers the aforementioned aspects of interaction regulation and experimental outcome types by

incorporating two fixed elements: the exclusion of intermediaries driven by control rather than experimental design and the constant focus on collective processes. This table serves as a helpful tool for researchers to determine whether their experimental protocol aligns with our definition and, if it does, where to situate it on the spectrum of collective experiments.

First, situated at the intersection of *unregulated interactions* and *collective outcome*, such experimental protocols represent the one end of the spectrum with the most advanced type of collective experiment. Then, loosening one constraint at a time, protocols situated at the intersection of *unregulated interactions* and *individual outcomes* or *regulated interactions* and *collective outcomes* both belong to our spectrum of what a collective experiment is. Finally, at the intersection of *regulated interactions* and *individual outcomes*, we consider collective experiments at the other end of the spectrum where the collective dimension is most diminished. In any case, no experimental protocol can be characterized as collective if no interactions exist between subjects, whether the outcome is individual or collective.

Figure 1.1: Categorization of Collective Experiments

		Outcome	
		Individual	Collective
Interactions	None	Non-collective experiment	Non-collective experiment
	Regulated		
	Unregulated		

Figure 1.1 is a 3x2 matrix categorizing experiments based on 'Interactions' (rows) and 'Outcome' (columns). The 'Interactions' axis has three levels: 'None', 'Regulated', and 'Unregulated'. The 'Outcome' axis has two levels: 'Individual' and 'Collective'. The top row ('None') shows 'Non-collective experiment' for both outcomes. The bottom two rows ('Regulated' and 'Unregulated') are shaded blue, indicating they are part of the 'Collective experiment' spectrum. A dashed box encloses the 'Regulated' and 'Unregulated' rows across both outcome columns. A white callout box with the text 'Collective experiment' is positioned over the intersection of 'Regulated' interactions and 'Individual' outcome.

1.3.2 Current limitations of the method

Experimental economics relies on solid and rooted principles that ensure the proper conduct of experiments and the replicability and robustness of results. However,

these principles can also impose some limitations when observing collective processes. Most social interactions are not anonymous if we detach ourselves from the experimental or even scientific approach. These social interactions occur between individuals who, most of the time, can communicate and identify with each other. In addition, there is no principle of systematic isolation of subjects or systematic intermediaries between them. Considering both the social nature of interactions and our definition of a collective experiment, we can then ask ourselves what limits the development of research questions related to this issue in the field of experimental economics. To dig deeper into the cause of such a theoretical and methodological barrier, we must address both the anonymity and isolation problems and the systematic intermediary principles. First, as presented earlier, *control* is the reason for implementing strong rules and principles in experimental economics. As experimentalists value *subject's anonymity* and *isolation*, subjects involved in the same experimental session should remain unable to identify one another. Although Eber and Willinger [2005] acknowledge that this requirement can be waived if the research question calls for it, e.g. with reputation effect games, the typical experimental method in economics still involves, in most cases, that the decisions are carried out through an interface or an intermediary leading to the *isolation* principle. Indeed, traditional experimental laboratories consist of individual computer stations delimited by partitions preventing subjects from interacting. The reason for this is that some experimentalists may see anonymity as a treatment procedure in itself, as a subject's knowledge of others could influence their decision-making and behaviour, avoiding contagion problems.

Owing to this condition of anonymity and the resulting isolation of subjects, Jacquemet and L'Haridon [2018] argue that the application of the following four basic principles ensures the proper conduct of the experimental method: (1) minimizing the probability that subjects know each other, (2) prohibiting communication, (3) isolating subjects, and (4) preventing identification. According to the authors, these good practices eliminate any possible external effects on the protocol and treatments, whether positively or negatively affecting the outcomes. However, readers might also consider a situation where the experimenter is willing to observe collective phenomena in a controlled environment by trying to stay as close as possible to real-life situations of social, often face-to-face, interactions, considering then the above principles as harming the proper conduct of the experimental protocol, mainly in terms of ecological validity.

Nonetheless, we need to specify the type of anonymity that this chapter focuses on. Anonymity is a multidimensional notion, and Burkell [2006] differentiates three types of it: identity protection, visual anonymity, and action anonymity. If identity protection and action anonymity are respectively defined as a “*form of withholding name or other identifying information*” (p. 200) and a situation where “*actions either cannot be seen, or cannot be individuated*” (p. 201), what interests us in this chapter is the visual anonymity that emerges in the context of interpersonal interactions when subjects are not able to visually identify each other. Indeed, if we aim to run collective experiments, strict rules on anonymity mechanically prevent researchers from implementing protocols focusing on collective processes with no systematic intermediaries between subjects. Consequently, while the subject’s anonymity and, by extension, the isolation of subjects are considered a thumb rule in experimental economics, this also represents the obstacle for experimentalists to run collective experiments as the second necessary condition would be violated. This brings experimentalists face-to-face with a research trap preventing them from developing collective experiments, as we defined earlier. As a result, experimentalists need to consider solutions to overcome the barriers to conducting collective experiments, with the first theoretical argument being the removal of systematic intermediaries between subjects and the second being the use of appropriate tools and technologies.

1.4 Necessity of removing intermediaries

As highlighted in the previous section, the rules put in place to establish the experimental method have also limited its development in terms of collective experiments. The following paragraphs will examine the transition from an individual unit of analysis to a collective unit of analysis, the existence of intermediaries, the stemming *intermediary interpretative layer*, and, finally, the classes of variables that then become central to the study of collective processes in an experimental context.

1.4.1 From the subject to the group

In the traditional experimental approach, experimentalists focus mainly on one specific unit of analysis: the *subject*. Experimental studies observe and capture the behaviour of subjects performing tasks in a controlled environment, and the results are obtained from the aggregation of subjects’ behaviour. The shift towards collective experiments implies introducing another unit of analysis, the *group*, and thus

the possible interactions between at least two subjects. It is important to note that incorporating a collective dimension to the analysis entails introducing a third entity, the *environment*⁶. On this last entity, if the *environment* is suppressed as much as possible in an individual setting where the individual behaviour is isolated or distanced from other subjects on purpose, observing group behaviours implies also observing the environment in which those groups evolve. These considerations bring us to reasoning on different scales and dimensions where the experimentalist needs to drive between individual and collective layers as well as the complex nature of the environment.

1.4.2 The use of intermediaries

When setting up an experimental protocol, the question of organising the layout of the experimental room is rarely raised. In fact, most laboratories adopt the general scheme of organising an experimental room by isolating subjects with partitions or cubicles. Then, we need to focus first on this physical aspect of intermediaries. As mentioned above, all this is done to ensure almost total control over what happens during the experiment, all the more so given the principle of anonymity [Serra, 2012b]. Looking back at the early days of running computerised experiments, this emerging form of intermediary was considered a standardization of the relationship between the experimenters and the subjects, as well as the relationships between subjects. From pen-and-paper economic experiments where everything was facilitated by humans to the increasing part of computerised protocols culminating in the 1990s, “*experimentalists’ control became inextricably linked to spatial arrangements (separation, communication, flow of information, randomization) of subjects in relation to each other and to experimentalists inside the laboratory*” [Svorenčík, 2023, p. 799]. However, what interests us in this chapter is to distinguish those two types of relationships and focus on the case of subject-to-subject interactions.

Until now, the prevailing view in the literature has been to recognise the major role played by the development of computer interfaces as intermediaries. However, it would be interesting to take the opposite view. An example often used is that of market experiments for which, as explained by Svorenčík [2023], “*Various auction formats were studied, including ones never conceived before, utility markets, and*

⁶Please note that the terminology *environment* in the context of this section does not correspond to the definition provided by Smith [1982], but rather to the idea of the framework in which interactions between subjects take place.

other institutions with alternative allocation algorithms requiring substantial computational capacities” (p. 769). The increasing implementation of computerised experiments, partly due to the simplified access to such technology over the years, allowed researchers to automate and access more complex protocols. However, the weight given to the now institutionalized rules surrounding their usage, including the principle of systematic intermediaries, has also limited the development of other types of experiments. Indeed, looking at the literature⁷, the very existence of intermediaries, guaranteeing control over both the subjects’ ability to communicate and their perception of the experimental environment, has restricted the development of collective experiments. By removing these systematic intermediaries between subjects, it would be possible to loosen the constraints surrounding the implementation of collective experiments.

1.4.3 The resulting intermediary interpretative layer

Beyond their physical form, we need to question the repercussions of intermediaries on the subjects’ behaviours. Building on the example of the computer interface, which is still the most used intermediary in experimental settings, one might question the significance of removing it when studying collective processes. While the literature recognises that visual anonymity can have certain benefits, such as greater willingness for self-revelation, there is nonetheless a number of factors that have a negative impact on interpersonal interactions, such as a decrease in interpersonal commitment, lower levels of helping behaviours, and higher levels of aggression [Burkell, 2006]. Beyond those detrimental factors, what particularly interests us is the subject’s perception of an interaction involving other subjects when an intermediary is introduced. As simple as it may seem, one will always wonder, even unconsciously, what is on the other side of the computer. Let’s envision a scenario where a subject engages in an experiment that involves interactions with other subjects through a computer interface. The subject knows that she or he is interacting with another human being. Having this interface between them creates an *intermediary interpretative layer*. In the case of interactions without intermediaries, subjects are able to gather information on their interlocutor from both direct and indirect sources of information also due to their physical or visual proximity. Then, when an intermediary is added, subjects face an information gap. One could argue that

⁷It was the case for Chamberlin [1948] who organised direct face-to-face interactions between his subjects, but very few other examples appeared over the years.

this would imply less information to process, which could be beneficial to let the subjects focus only on the task she or he is asked to complete, but it also represents an additional cost for the subject who needs to fill those gaps according to her or his beliefs and expectations. Hence, this intermediary interpretative layer gives rise to ambiguities or uncertainties that impose a cognitive burden on the subject. We can define this intermediary interpretative layer as *the implicit distance that exists between individuals when their interactions are regulated by any form of intermediary, resulting in an additional cognitive load for information detection and processing*. As a consequence, these uncertainties can potentially lead to misinterpretations and misleading cues or at least a distance to reality, ultimately altering the subjects' behaviour and adversely affecting the experimental outcomes. The additional effort required to bridge the information gap, shifting the subject's cognitive focus away from the task at hand, can potentially undermine the accuracy and reliability of the results.

One possible solution would be for the experimenter to provide pre-defined information to subjects in order to reduce their cognitive load. However, when considering the variables that are relevant to cognitive processes, there is no comprehensive or exhaustive list available that can substitute real-life interactions. It is crucial to acknowledge that certain vital variables that influence collective processes can not be effectively captured through an interface or a restricted list. Neglecting these variables could negatively affect the observation of collective processes. Paradoxically, the implementation of intermediaries that were considered as a means of control becomes a possible harmful barrier between subjects, distancing the behaviour of agents in the laboratory from their real behaviour. Once again, while this does not pose a problem in the case of individual behaviour, it seems clear that the transition to a collective level of observation requires the removal of this systematic intermediary principle in order to answer the key research questions about collective processes fully.

1.4.4 The Collective Experiment Variables Matrix

To investigate interactions between subjects in a controlled environment, one must consider the complex set of variables that characterize those interactions. This is all the more critical in face-to-face interactions, where intermediaries have been removed. While intermediaries are not prohibited, they can not represent a necessary condition for the implementation of a protocol involving interactions between sub-

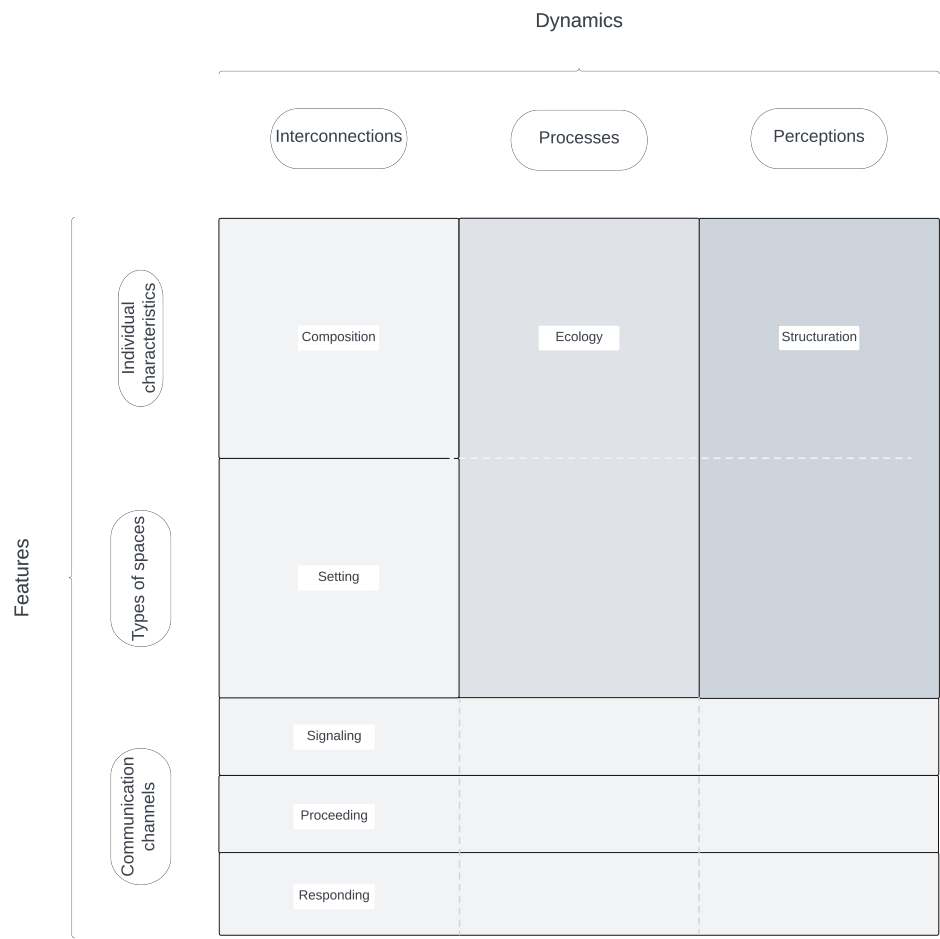
jects. A comprehensive and holistic perspective is essential to envision and control the environment where these interactions take place. The following paragraphs provide a guide to the relevant elements to consider when examining collective processes in a controlled environment. Figure 1.2 illustrates these elements in the form of a matrix. It should be noted that the following paragraphs describe the variables arising from subjects' interactions but do not consider extra-interaction variables, which can be categorized as contextual information encompassing the environment and institutions in the sense of Smith [1982].

First of all, there are two structuring dimensions of interactions between subjects in the laboratory: the *Features* and *Dynamics*. *Features* refer to the descriptive and static aspects of interactions, while *Dynamics* pertain to the changing and active aspects. Each of these structuring dimensions is organised around three components.

Features can be broken down into individual characteristics, types of spaces, and communication channels:

- **Individual characteristics:** In an experimental protocol, experimentalists are interested in studying subjects' decisions, actions, and behaviours. Thus, having a clear vision of their profiles and idiosyncratic characteristics is crucial, as this will play a key role in identifying the entities involved in specific interactions.
- **Types of spaces:** Space is a multifaceted concept with different types of spaces coexisting and possibly overlapping each other. If the laboratory itself might count as the physical or absolute space in which interactions take place, the structure or arrangement of such a space matters. In addition, beyond this physical aspect of spaces, one should also consider spaces in terms of their relational or social nature. As for Individual characteristics, this element becomes essential when putting in perspective with occurring social interactions.
- **Communication channels:** Defined as “*the transactional process in which people simultaneously create, interpret, and negotiate shared meaning through their interactions*” (p. 14), communication is one essential feature of a group [Adams and Galanes, 2011]. As a matter of fact, having a deep understanding of the specific communication channels allowing subjects to communicate becomes

Figure 1.2: Collective Experiment Variables Matrix



decisive in embracing the complexity of subjects' interactions occurring within the controlled environment of the laboratory.

Dynamics can be broken down into interconnections, processes, and perceptions:

- **Interconnections:** In its simplest definition, interconnections are all the relationships or links between entities. Those interconnections can take different forms, including physical or symbolic interconnections, but, in all cases, those interconnections forge a situation of interdependence between all involved entities.
- **Processes:** As a consequence of subjects' interactions, subjects in a collective experiment participate in different processes that can prove to be of different types, e.g., associative (e.g., cooperation) or dissociative (e.g., conflict) processes.
- **Perceptions:** These are the ways in which subjects might envision, understand, or interpret interactions and processes emerging in collective experiments.

Having introduced the components of the two axes from the matrix, we can now focus on the key classes of variables that emerge from their intersections. In the end, we account for seven elements that can be defined as follows:

- **Composition:** At the intersection of Interconnections (*Dynamics*) and Individual characteristics (*Features*), the composition is defined as a reflection on “*Which profiles compose the sets, groups, or teams of subjects ?*”. As collective experiments induce interactions, it is necessary to take into account individual profiles and their aggregation.
- **Setting:** At the intersection of Interconnections (*Dynamics*) and Types of spaces (*Features*), the setting interrogates “*How the different spaces are shaped?*”. The physical and symbolic configuration of spaces can either facilitate or limit social interactions, depending on their modalities. This element is very important since changes in terms of setting might affect, either positively or negatively, subjects' decisions, actions, or behaviours.

- **Signaling, Proceeding, and Responding:** At the intersection of the Communication channels (*Features*) and the three components of the axis regarding Dynamics (Interconnections, Processes, and Perceptions), those three elements are articulated around the concept of communication that encompasses both the content (the *What?*, as the shared elements) and the relations (the *How?*, as the expression of the perceived relationship) and is composed of two main aspects: the symbolic and the transactional. On the one hand, communication relies on symbols defined as all things that “*arbitrarily represent something else*” [Adams and Galanes, 2011, p. 51]. On the other hand, communication is a transactional process. All symbols sent, received, and interpreted by individuals enter a simultaneous and multidimensional process that allows those individuals to exchange and understand each other. Relying on the *interpretation layer* presented earlier, communication mediated by technology (or computer-mediated communication, CMC⁸) seems to generate a lack of *social presence* detrimental to group members to process the information as if it was shared in face-to-face (FtF) exchanges. If verbal communication can find substitutes, non-verbal communication faces heavy challenges when comparing CMC to FtF communication. Non-verbal communication relies on three principles: (1) non-verbal communication is ambiguous, (2) individuals never stop sending non-verbal signals, and (3) non-verbal communication, compared to verbal communication, is less subject to deliberate control and, thus, often appears more trustful. In the end, any communication relies on the simultaneous sub-processes that we resume in signaling (*Which (in)direct and (un)conscious signals are sent to other subjects?*), proceeding (*How are those signals processed and interpreted?*), and responding (*How do receivers react and answer to those signals?*). However, the boundaries between those three concomitant elements are blurred as they occur sequentially but in a rather short amount of time.
- **Ecology:** At the intersection of Processes (*Dynamics*) and of both Individual characteristics/Types of spaces (*Features*), the ecology interrogates “*How groups arrange themselves under specific conditions, and how those arrangements affect the emerging processes?*”. The term “arrangements” focuses here on two sides: the arrangements based on group composition and those based on the setting. Again, the processes are influenced by the interactions emerging

⁸This CMC exists through numerous types of mediation: emails, chat rooms, webinars, and so on.

during the collective experiment.

- **Structuration:** At the intersection of Perceptions (*Dynamics*) and all three Individual characteristics/Types of spaces/Communication patterns (*Features*), Structuration underlines the question, “*How are the sets, groups, or teams of subjects structured?*”. Structuration can take different forms, such as roles or norms, and will impact subjects’ behaviours as they play a role or are influenced by subjects’ perceptions.

Due to the complex nature of interactions, bidirectional relationships between all of our classes might exist. Therefore, if we are interested in the classes emerging from interactions between subjects, we must also consider that these may be strongly linked and mutually influence each other. Collective experiments thus provide a valuable way to observe these complex interrelationships and understand their dynamics. Beyond studying the classes themselves, understanding their relationships becomes a crucial area of research.

To conclude on the question of the *Collective Experiment Variables Matrix*, it is essential to stress the importance of understanding interactions between subjects using a complex and comprehensive approach. Experimentalists must consider the relevant variables to the study of collective decisions, actions, and behaviours emerging from the experimental context. As a consequence, Figure 1.2 aims to constitute a useful tool for researchers to identify those relevant variables. If this matrix does not aim to mention all possible variables, it allows researchers to identify the core classes of interest. Those intersections create a research agenda for experimentalists in economics willing to explore collective processes.

1.5 Necessity for Innovative Tools and Technologies

As extensively discussed in the previous part, the complexity of analysing interactions between subjects in collective settings arises from the numerous variables involved. When researchers aim to conduct collective experiments and incorporate relevant variables, they encounter challenges in data collection and analysis. The following paragraphs emphasize critical factors to consider when collecting data for observing social interactions and collective processes. These factors encompass the nature of

the collected data, the required quantity of them, and the tools employed for their collection in line with multimodal and multidisciplinary strategies.

1.5.1 Type of data

The primary consideration in studying groups of individuals revolves around the type of data that needs to be considered. We distinguish three pivotal ways to categorize types of data: their level of analysis, their temporality, and their degree of objectiveness.

First, data can be categorized relative to their level of analysis. In the context of a collective experiment, as has already been pointed out, the researcher must consider three distinct layers: the subject, the group, and the environment. Conducting collective experiments forces the experimentalist to go back and forth between these layers to understand the entire phenomenon. To start with, while the most intuitive way to move from the individual to the group level is to aggregate from the first to the second, there are also group-specific elements that require additional considerations. Indeed, some specific variables related to groups can not be derived solely from aggregating individual-level data. These variables are inherent to the existence of the group itself. Moreover, on the environment side, the researcher must turn to variables in the very space where the experiment is taking place and go beyond the variables specific to the subjects.

Afterward, in terms of temporality, data extends beyond the actual duration of the experiment. Both individual and group-related data emerge prior to the experiment itself. The first type of data that inherently exists beyond the experimental timeline pertains to the subjects' profiles. If experimenters are used to account for those, another element is crucial to observe. Indeed, the lifespan of the group may differ from the actual duration of the experiment, and this requires additional consideration and data collection. This matter may arise when the group already exists prior to the experiment. However, even in the case of a group formed specifically for the experiment, an *ad hoc* or temporary group, the subjects may have some level of pre-existing connections. The probability of the subjects being complete strangers to each other might not always be zero. If this is totally discarded when it comes to conventional experiments by introducing systematic intermediaries, this is not the case for collective experiments that require additional control. Therefore, it becomes essential to gather additional information on these pre-existing networks or social dynamics, e.g., as self-declared data in questionnaires, as they can significantly in-

fluence the processes, decisions, or actions that occur during an experiment.

Finally, data also rely on their degree of objectiveness. By objectiveness, one should define and understand whether the data has been collected based on subjects' self-assessments or from external and more objective sources. If both objective and subjective data sources are valid, this distinction implies differences in terms of methodologies, measures, and interpretations. On the one hand, objective data refers to information that is measurable, quantifiable, and based on observable facts or evidence. It is independent of individual experiences, perceptions, or interpretations. Subjective data, on the other hand, relates to those personal experiences, perceptions, or interpretations that may vary from person to person. Both objective and subjective data have their advantages and disadvantages. When it comes to objective data, measurements are precise and unbiased. However, certain aspects may not be observable and would necessitate the subjects themselves to provide the required information, e.g., mood or perception of conflicts. On the other hand, subjective data can be obtained directly from the subjects, but it introduces the possibility of multiple biases stemming from the individuals' own perspectives and interpretations. As a consequence, researchers should always be interested, if possible, in combining both data sources and rely on their triangulation to enhance the overall validity and reliability of findings. This is all the more important when interactions between subjects in a group and a specific environment come into play as additional processes emerge besides those involved in more individual protocols.

When considering these three perspectives altogether, the transition from an individual to a collective experiment offers a richer pool of data, albeit with the challenge of studying more intricate interactions and relationships.

1.5.2 Required quantity of data

Regarding experimental results, the researcher must consider the amount of data to be collected in order to obtain valid, significant, and robust results. It is crucial to consider both the set of variables and the sample size to reach this goal.

To start with, utilizing multilevel observations for collective and group experiments results mechanically in a higher volume of data collected due to the growing number of critical variables. Indeed, based on the *Collective Experiment Variables Matrix* presented earlier, the number of variables to consider once running a collective experiment significantly increases when we introduce interactions between subjects. Ultimately, this larger quantity of data, in both depth and breadth, neces-

sitates careful management and analysis strategies to extract valuable insights from this abundance of cross-level information.

In addition, we need to consider the number of observations. In order to define the optimal sample size, it is necessary to take into account the significance level of the power of the hypothesis test⁹, and the minimum detectable effect size [List et al., 2011]. Apart from those theoretical considerations, researchers face additional practical considerations, including the monetary and time costs induced by collecting experimental data. From a practical standpoint, conducting data collection in a collective experiment requires a larger sample size, as each group consists of multiple individuals. If a study requires n group observations and the research setting necessitates i subjects in each group, then the total number of subjects should be at least $n \times i$ subjects. Thus, a significant sample of subjects will have to exceed the existing experimental economics rules. While, traditionally, experimentalists have relied on rules of thumb, such as the $n \geq 30$ subjects per treatment, to determine appropriate sample sizes, they should also consider integrating power calculations in their methodology. Nevertheless, looking at five high-ranked economic journals between 2000 and 2018¹⁰, Grüner [2020] observed that the fraction of papers using power analysis is low¹¹. In addition to the complexity associated with assessing statistical power, researchers may opt to forgo this option due to limited resources or opportunities and instead rely on smaller samples of subjects. Recognizing this limitation, experimentalists should also take into account the type of protocol chosen, as certain designs inherently reduce statistical power [Vasilaky and Brock, 2020]. It would then be advisable to use alternative designs that are less demanding regarding statistical power. In any case, the scaling from individual to collective experiments creates an additional burden on researchers and their trade-off between statistical power and practical costs. It seems important to underline that those considerations in terms of quantity of data hold once we assume to run econometric analysis while other methods are also available to treat smaller experimental samples such as non-parametric tests. In fact, as has been underlined by Vallois and Jullien [2018],

⁹Power is “*the probability that the researcher will be able to reject the null hypothesis of no treatment effect*” [Gerber and Green, 2012, p. 93]. Then, a study is considered underpowered if too few subjects are recruited.

¹⁰Quarterly Journals of Economics, American Economic Review, Experimental Economics, Journal of Economic behaviour and organisation, and Environmental and Resource Economics

¹¹8.64% for Quarterly Journals of Economics, 5.38% for American Economic Review, 1.68% for Experimental Economics, 2.64% for Journal of Economic behaviour and organisation and 2.67% Environmental and Resource Economics between 2000 and 2018.

econometrics was never the only methodology chosen to analyse results. However, it has become more and more salient with the development of *econometrics*.

1.5.3 Enhancing understanding through technology

To gain a comprehensive understanding of new technologies and their implications in terms of collective experiments, we need to remember that one of the reasons why the use of collective experiments was hampered is the inadequacy of the techniques available at the time to provide the necessary control legitimizing their development. What truly is interesting and promising is to combine traditional experimental methods in the laboratory with external technologies and tools, also from different disciplines, to understand the complex phenomenon of collective processes better. In fact, these technologies might provide valuable insights into interactions between subjects and additional information on the two-way relationship between individuals or groups and their environment.

It is crucial to note that the tools and technologies selected in our argumentation focus on collective experiments and not individual protocols. This is why the range of available tools is not complete. Other technologies might be added if we consider both individual and collective perspectives. In addition, they are considered as novel as they are not widely or traditionally used in experimental economics.

This section will divide the relevant tools and technologies into categories tailored to specific interests, namely subjects' **(a) emotional states** and their **(b) motion**, emphasizing the critical elements and characteristics and comparing them. Additional information is summed up in the Appendix 1.8.

Emotion recognition

Emotions play a significant role in influencing both the physiological and psychological state of individuals, and the complex interplay between these factors makes emotion recognition a challenging task. Moreover, interactions between individuals naturally lead to communication, which can take the form of verbal or non-verbal signals. Incorporating both aspects is essential to comprehend fully any collective process and distinguish the different mechanisms behind them. In the case of collective experiments, both verbal and non-verbal communication signals can be used to analyse the subjects' emotions. The related tools can also be distinguished based

on their biometric or non-biometric nature¹². In characterizing emotions, we follow the circumplex model of Wundt [1948] and its two dimensions of arousal (active or passive) and valence (positive or negative). Thus, the following paragraphs will look at different sources for emotion recognition based on the comprehensive technological reviews of Dzedzickis et al. [2020], Egger et al. [2019], Pal et al. [2021], Reiss and Amft [2015], Saganowski et al. [2023], Schmidt et al. [2019]¹³. The Appendix 1.8 provides additional information on using the following tools to capture and analyse emotions in collective experiments. Finally, beyond other methodologies and techniques mentioned in the following paragraphs, when it comes to biometric tools, researchers in social sciences could benefit from some knowledge of biostatistics and the standard practices in the related fields to collect, analyse, and link such data to the emotional state at play.

Body language

Three main non-contact¹⁴ techniques in the context of emotion recognition through body language are Facial Expressions (FE), Body Posture (BP), and Gesture Analysis (GA) [Glowinski et al., 2011, Kleinsmith and Bianchi-Berthouze, 2013]. These physical external signals for valence and arousal, distinct from internal physiological signals discussed later, can be analysed using automated tools. Advancements in body language analysis technology enable faster and cost-effective tracking through video recordings and algorithms, replacing manual coding with more accurate and objective automated recognition of movements that do not require additional expertise from researchers except possible algorithmic competencies.

Brain activity

By examining brain activity, researchers gain insights into the underlying processes associated with emotional arousal and valence. The main technique for observing brain activity is electroencephalography (EEG), a contact sensor that allows researchers to capture the electrical activity of subjects' brains while facing different

¹²Experimentalists can use biometric methods and affective computing, defined as a “*research area that studies and develops systems to sense the emotional state of a user (using sensors) and process them using computer systems to recognise the emotions*” [Pal et al., 2021, p. 1] to observe and measure bodily expressions.

¹³To be noticed, all tools presented in this article are composed of non-invasive sensors, i.e., sensors without penetration into the body in order to obtain data or measurements.

¹⁴Contact sensors or tools refer to devices that require direct connection to the subject's body or skin to obtain data or measurements.

stimuli. The resulting waves and their frequencies allow researchers to assess subjects' EEG responses to experimental stimuli [see, e.g., Coricelli et al., 2019].

Skin conductance and temperature

Emotional arousal can also be detected by capturing variations in Skin Conductance (SC) and Temperature (SKT), two contact tools. First, SC¹⁵ can be defined as the “*continuous measurement of electrical parameters of human skin*” [Dzedzickis et al., 2020, p. 9]. The changes in subjects' sweating provide valuable insights into the temporal aspects and frequency of emotional responses. Then, SKT provides valuable information about the autonomic regulation of blood flow to the skin that leads to changes in skin temperature.

Heart rate variability

Heart Rate Variability (HVR) can be defined as the variation in time intervals between successive heartbeats. Here, we focus on two main contact tools to assess HVR [Egger et al., 2019, Sayed Ismail et al., 2022]. On the one hand, Electrocardiography (ECG) is based on the electrical activity of the heart, and on the other hand, Photoplethysmography (PPG or rPPG for remote application) is based on Blood Volume Pressure (BVP). Both represent two viable solutions to assess HRV and the subjects' emotional arousal.

Respiration rate

The Respiration Rate Analysis (RR) provides respiratory data through thoracic activity, shedding light on subjects' emotional states. RR allows researchers to observe subjects' emotional states, but many measurement techniques exist [Dzedzickis et al., 2020]. First, we can distinguish two main types of non-contact tools: the video-based detection of body movements and signals related to subjects' respiration (displacement of reference point) and on the other hand, the use of thermal cameras to detect temperature fluctuations. Second, as a contact sensor, RR can be provided by Respiratory Inductive Plethysmography, which measures the chest and abdominal wall movement.

¹⁵Also called Galvanic Skin Response (GSR) or Electrodermal Activity (EDA).

Verbal communication

By capturing and analysing the verbal exchanges among subjects, researchers can gain valuable insights into the form and content of subjects' interactions. This leads to a distinction between Speech Recognition (SR) and Voice Recognition (VR). The former is defined as a semantic analysis of exchanges, while the latter focuses on the acoustic aspect of these exchanges [Egger et al., 2019]. Either on the content or the form, it becomes necessary for researchers to gain knowledge of some analysis techniques, among which the quantitative approach of Natural Language Processing or the qualitative approach of Discourse Analysis through coding. From a practical point of view, the choice of microphone type is crucial, as it affects the capture of both voice lines and the surrounding environment. While an ambient microphone provides a means of control, it offers lower quality for analysing verbal interactions than individual microphones.

Motion detection

By allowing the motion of subjects into an experimental setting, the researcher introduces an additional dynamic dimension that needs to be controlled by an extensive knowledge of the subjects' positions and evolving trajectories. At the individual and collective levels, analysing the subjects' motion in the laboratory enables a deeper understanding of their impact on the experiment's outcome and how stimuli can also impact those shifts. While the use of human coding is possible, the use of technological tools allows for more accurate data collection. The Appendix 1.8 provides additional information on two main tools that can be profitable for researchers to understand subjects' motion: subject-tracking devices (non-contact) and inertia sensors (contact). If such tools make it possible to visualize and map interactions between subjects, they also make it possible to quantify these interactions, understand their dynamics, and identify patterns, roles, or other phenomena resulting from them, e.g., knowledge flows, social contagion, or preferential attachment. The resulting analysis requires additional expertise in social network analysis. Moreover, those technologies can seamlessly integrate with video and audio systems, enabling researchers to disentangle ambiguous situations. For instance, in scenarios where two subjects are in close proximity, the combination of tracking data with video footage and audio recordings can determine if they are facing each other, engaged in conversation, or simply back-to-back.

To conclude, in collective experiments involving the mobility of subjects, re-

searchers need to consider using suitable sensors that are minimally intrusive and allow for movements. *Wearable devices* – both for detecting subjects’ motion and, in principle, for emotion recognition – would allow researchers to observe phenomena in settings closer to real-life environments and to run experiments with stronger ecological validity. While static devices are commonly employed in traditional laboratory settings, there has been a recent and ongoing development of non-intrusive and wireless tools designed to facilitate accurate measurements when dealing with mobile subjects. These innovative tools aim to collect reliable and robust data in scenarios where subjects are in motion. In this section, certain tools are consciously discarded, such as Electromyography devices, that do not seem to be the most suitable for laboratory experiments in economics due to lower practicality and the existence of sufficient alternatives. This does not prevent experimental economists from using them, and readers are encouraged to question the implementation of these tools.

1.5.4 Integrating multimodal and multidisciplinary approaches

Multimodality

Beyond the previous considerations on the valuable tools and technologies that experimentalists should integrate into their protocols, one important thing is also to consider those tools and technologies as complementary elements. In this sense, the data collection needs to be multimodal. To fully understand the ongoing processes, experimentalists must gather, combine, and synchronize several tools. This multimodality of measurements ensures a more robust and trustful analysis of collective processes, enhancing scientific interpretations and contributions. By integrating a subjective (self-assessment) and objective (tools’ measurements) approach to their data collection and ensuring the combination of several data sources and technologies, researchers are able to create new research avenues for the current experimental practice in economics. The main issue arising from this remark is the high demanding problem of data synchronization. Depending on the tools and technologies, this requires either real-time or *a posteriori* synchronization. Furthermore, while it is possible to analyse each source individually for a smaller number of observations, the accumulation of data sources and the growing number of observations due to the introduction of groups as observation units mean that the results need to be more automated. Such expertise needs to be acquired by economists who are not used to these techniques. This naturally leads us to the following part on the increasing

necessity of multidisciplinary.

Multidisciplinary

The previous paragraphs highlighted the importance of combining, synchronizing, and automating experimental data sources and analyses with various technological tools in order to better understand the complex phenomenon of collective processes and to improve the control of direct social interactions as well as the collection of physiological information in a laboratory setting. Although some of these technologies are already widely used in other fields, such as psychology, interdisciplinary collaboration is crucial. However, beyond technical interdisciplinarity, a research project also stands to benefit from methodological interdisciplinarity, as analysing the collected data requires expertise that is not yet prevalent in economics but is well established in disciplines such as medicine and psychology. It becomes essential to consider existing knowledge from other fields regarding their theoretical models, methodologies, and analyses when using these tools alongside traditional experimental economics. Consequently, most references cited in this section are from engineering and computer science journals for the tools used and from medicine and psychology journals for data analysis, since technical and methodological knowledge is often necessary for both. Nonetheless, effective communication, including shared terminology, methodological criteria, interests, and openness to new approaches are essential for the success of multidisciplinary projects.

1.6 The *Social Interactions Lab*

1.6.1 The purpose

As mentioned earlier, the literature on experimental economics does not just consider traditional laboratory experiments. In fact, among the possible alternatives for running experiments, two opposite directions are offered to experimentalists: staying in a laboratory or going to the field. We can then differentiate two main types of experimental data: the *field*- and the *laboratory-happenstance* data Buda [2000]. Experimenters can then imagine a simplified scope of experimental facilities, from the most traditional experimental laboratory where subjects are isolated from each other to the pure field experiment within a real-life setting. Shifting from one end of the scope to the other, protocols face both positive and negative effects. The

main potential of field experiments is that they offer two factors that improve the external and ecological validity: the representativeness of the environment and the sampled population [List, 2007]. Therefore, experimentalists face a trade-off between control of the experimental setting and external factors (internal validity) and the proximity to a real-life setting (ecological validity). In between those two extremities, researchers can encounter *lab-in-the-field* setting integrating a naturalistic environment for the relevant population using the laboratory paradigm. However, other research infrastructures might emerge as another form of compromise between conventional laboratory and field experiments. This is the case of the *Social Interactions Lab* (SIL) that has been thought to offer a facility halfway between the two extremes of the scope described above. More importantly, this laboratory has been thought to provide an experimental setting allowing the study of collective behaviours and processes in a controlled environment based on the principle of non-systematic intermediaries between subjects and the possibility of face-to-face interactions.

1.6.2 The research infrastructure

Located in Strasbourg (France), the SIL is an experimental room of more than 100m² and designed to allow researchers interested in collective processes to carry out their experiments in a space different from the more conventional offer of experimental laboratories¹⁶. While in economics, the usual experimental rooms are composed of fixed stations separated by partitions in which each subject is installed totally isolated, the previous parts of this chapter argued that the study of collective processes requires another form of a layout where subjects can interact without any form of systematic intermediary. Another key point of the SIL's design is the modularity of the room. In fact, the SIL layout can be adjusted depending on the researchers' needs relative to furniture quantity and setting and/or delimited areas.

To fully describe this novel research infrastructure, we need to describe also the measuring tools available for experimentalists when running an experimental protocol at the SIL¹⁷. As was discussed earlier in Section 1.5, it is essential to have novel

¹⁶To be noted, there is also a control room attached to the creativity room that gathers all the measurement tools that can be used in the various experiments.

¹⁷All the measurement and data collection equipment used at the SIL was acquired through Noldus, a company specializing in the development of software and solutions for behavioural and neuroscience research. More specifically, the video systems and part of the audio system are integrated into the *Viso* solution for creating video and audio recordings to capture behaviours and interactions of your subjects, and the tracking system is integrated into the *Tracklab* solution for collecting spatial data and behavioural statistics on subjects' motion.

and appropriate measurement tools and technologies to capture and understand the intricacies of collective behaviours and processes. In the context of the SIL, three main tools are available: an audio system, a video system, and a tracking system. Figure 1.3 provides a visual representation of these tools and their setup within the SIL, as well as the modularity aspect. The subsequent paragraphs will draw upon this figure to provide further insights into each element and its functionality.

Audio system

The audio system is made up of two components: an *ambient audio system* and an *individual wearable audio system*. The purpose of the ambient system is to capture the sound atmosphere of the room using four portable microphones focusing on a particular area¹⁸. The individual wearable audio system is designed to capture the speech track of each subject and is in the form of a clip-on microphone. By combining these two sound-recording components, we ensure that we can analyse the audio data of any experiment at both the global and individual levels. This comprehensive audio system grants us control over subjects' direct communication signals and offers an enhanced understanding of their interactions throughout the experimental session.

Tracking system

The tracking system is designed to follow and capture the motion and trajectories of the subjects. Specifically, the room is equipped with beacons that capture the subjects' movements, while each subject wears a non-invasive and non-contact wireless transmitter. This device enables us to track the paths, relative distances, and speeds of the subjects. The primary objective is to analyse the individual and/or group dynamics that occur within the space during the experimental session. In essence, this tool also aims to mitigate the loss of control that may arise from the subjects' unrestricted movements in the room.

Video system

The video system comprises four *fixed corner cameras*, which provide wide-angle coverage of the room, and four *fixed zooming cameras* that allow us to zoom in on specific points of interest. This equipment empowers us to observe and analyse the

¹⁸Those areas can be modified due to the modularity of the room and can change according to the research purpose.

Figure 1.3: Measuring Tools Setting



occurrences during the experiment, encompassing subjects' movements, as well as indirect communication signals, conveyed through posture and gestures. This tool can also serve as a means of control for the audio and tracking systems installed in the room that might create ambiguities, e.g., ensuring that two subjects are facing each other or identifying each speaker correctly.

1.6.3 Implementation in this thesis

While this chapter presents a theoretical and methodological argument regarding the conduct of collective experiments in economics, part of this thesis will focus on applying these elements. This thesis aims to address the determinants of creative performance encompassing individual, team, and artificial creativity. Understandably, the following paragraphs focus on the chapters focusing on team creativity and so, introducing collective experiments. As the protocol in Chapter 2, “*The Many Faces of Creative Profiles: Exploring Task Openness*”, has neither subject interactions nor a collective outcome, it cannot be categorized as a collective experiment. The same applies to Chapter 5, “*Can AI Enhance its Creativity to Beat Humans?*”, which has the same characteristics. Nevertheless, Chapter 3 and Chapter 4, entitled “*Diversity, Interactions, and Team Creativity: An Experimental Perspective*” and “*Does Creativity Thrive on Plot Twists? Exploring the Role of Surprise on Team Creativity*” respectively, are characterized by experimental protocols with unregulated interactions and collective outcomes, enabling them to be positioned as collective experiments. In addition, those two chapters investigate research questions relevant to the study of collective processes that have already been presented in the *Collective Experiment Variables Matrix*. Notably, while the protocol of a collective experiment can be associated with nearly all matrix classes, this discussion will focus on the dominant classes in each chapter.

Diversity, Interactions, and Team Creativity: An Experimental Perspective

This chapter aims to contribute to the literature on team creativity and the intrinsic factors affecting team performance, i.e., diversity and social networks. As a result, it seeks to address mainly two classes within the *Collective Experiment Variables Matrix*. First, *composition* as the analysis integrates diversity measures to conclude the effect of diversity on team creative performance. Then, *setting* as the laboratory setting was voluntarily modified to observe different phases during the same experimental session.

Indeed, the experimental protocol was conceived as a three-stage protocol during which subjects came three times in the laboratory: (1) to answer a comprehensive questionnaire on their profile, (2) to perform individual creative tasks, and (3) to perform a collective creative task. Here, let's focus mainly on the third stage, which

was the first attempt at running a collective experiment as defined in the present chapter. This third stage has been articulated around four different phases encompassing (1) an individual phase, (2) a strategic phase, (3) a collective phase, and (4) a presentation phase. For the purpose of this chapter, it is worth taking a closer look at phases (2) and (3). On the one side, the strategic phase was based on unregulated interactions with individual outcome mechanisms as subjects exchanged information and ideas with one another but did not directly pursue a collective goal. The interactions consisted of exchanges of tokens as a form of payment for others' ideas. It was mentioned to them that those exchanges would help them gain information on others' ideas, knowing that one key element of their further evaluation as a group would rely on their distinctiveness. On the other side, the collective phase aimed at unregulated interactions with collective outcome mechanisms where subjects worked as a team to answer the problem and converged towards a common idea or project. During this phase, subjects sat around a table, discussed their ideas, and elaborated on new ones to settle on a final idea or project. For both phases, the interactions took place without a physical intermediary. In this way, the subjects could identify each other and exchange information without an additional filter.

The suppression of this *intermediary interpretative layer* forced us to include additional controls in our experiment with a specific questionnaire on pre-existing networks between subjects participating in the same experimental session. Using such questionnaires presents an interesting option for experimenters. It builds on established economic methodologies to gain control over their experiments, specifically by using widely used techniques in social network analysis to account for pre-existing relationships between subjects.

In the end, these two phases, yet different in terms of collective experiment conditions as presented in Section 1.3, provide us insights and learnings on running collective experiments. The main hands-on learning is that cross-level protocols are highly time-consuming and resource-demanding. We experienced a noticeable loss of subjects while running the entire protocol, as not all subjects from the first stage decided to follow the entire protocol. This leads us to the main limit of this research work: the number of observations as the unit of analysis was not the subject anymore but the group. Indeed, as the protocol was ambitious and we needed to follow each subject over all stages, we observed 35 teams (146 subjects). Even though we obtained robust results on the factors affecting team creativity, future protocols need to consider this demanding cross-level and multi-stage protocol factor. In addition,

we did not control for conversations between subjects while they were interacting. As a result, our analysis of the collective process is mainly based on their perceived experiences and self-declared impressions.

Does Creativity Thrive on Plot Twists? Exploring the Role of Surprise on Team Creativity

This chapter examines team creativity by exploring the impact of surprise, or unexpectedness, as environmental factors on team creative performance. This protocol primarily addressed two classes within the *Collective Experiment Variables Matrix*. First, *composition* as individual and team-level characteristics were integrated into our analysis. Then, *ecology* through the arrangement of teams that affect their processes, with a specific focus on the processes affected by surprise and the ones aiming to overcome it. As presented in this thesis's conclusion, we aim to improve Chapter 4 through the integration of speech data, enabling us to address *signaling*, *proceeding*, *responding*, and *structuration*. This will help refine our analysis of subjects' perception of surprise and its influence on their organization as a team.

We observed 122 teams (366 subjects) performing two tasks, one of which included our treatment of surprise. Each team was composed of three members who exchanged to develop a creative idea. Throughout the experiment, subjects interacted exclusively with other team members, with no intermediaries regulating their interactions. In other words, this protocol has been conceived to observe unregulated interactions between subjects within a team that had to produce a common collective outcome.

As a means of control, we introduced again questionnaires on pre-existing networks between subjects and implemented a recording of subjects' conversations while performing the tasks, which was lacking earlier.

As for the first protocol, this second protocol presents strengths and weaknesses. On the one hand, this second protocol gained statistical robustness due to a much higher number of observations. On the other hand, while the first protocol was demanding due to the multi-stage structure, this second protocol was more challenging in terms of the tools and technologies used. Besides the usual organisation related to an experimental protocol, as experimenters, we needed to master the use of microphones and develop skills related to such data collection. If the resulting data have not yet been integrated into the final results, their collection has illustrated the purpose of Section 1.5 on the crucial acquisition of complementary skills by economists

when dealing with more advanced technological tools.

1.7 Conclusion

From “*Chamberlin’s ‘silly’ experiment*” [Smith, 1991a, p. 155] to “*the protagonist of one of the most stunning methodological revolutions in the history of science.*” [Guala, 2010, p. 1], experimental economics has undeniably earned recognition within the field of economics. Despite its tumultuous beginnings, the discipline has firmly established itself and validated its methodology. It would be an oversimplification to state that experimental economics is only about individualistic behaviours. However, it is important to recognise that the analysis of collective processes is still limited in this area of research. Especially when we recognise the important role played by the collective in the economic discipline as a whole, but also in the beginnings of experimentation. The contribution of this chapter is twofold: first, it aims to understand the reasons behind the lack of exploration of protocols in the context of *collective experiments* throughout the history and development of the discipline. Second, it addresses the theoretical and methodological implications of implementing such protocols, demonstrating how this has been accomplished in this thesis and within a new research infrastructure, the *Social Interactions Lab*.

In examining the reasons for the scarcity of collective experiments in the literature, this chapter identifies that the fundamental principles that have been widely and rapidly adopted by experimentalists on the proper conduct of experiments did limit the development of such protocols. As was already presented in the historical part of this chapter, Vernon L. Smith brought changes to Chamberlin’s experiment due to the possible factors introducing bias in subjects’ behaviour, including pre-existing relationships between students in the classroom who were interacting without any form of intermediary. These biases polluted the experimental results and motivated Vernon L. Smith to introduce guiding rules. Among these rules, the anonymity and isolation of subjects constrained the development of protocols involving direct interactions between subjects. This also limited the creation of protocols designed to understand complex collective processes, which could ultimately lead to collective outcomes. However, Smith and other experimentalists could have envisioned and resolved the flaws of Chamberlin’s and other experiments from another perspective. One perspective could have been understanding these “polluting” elements and integrating them into the experimental procedure instead of trying to erase them.

Once experimentalists recognise the possibility of conducting *collective experiments*, they might believe that collective processes could be studied using existing tools and principles. This chapter has shown that this turns out to be more complex and that the main reason lies in the concept of the *intermediary interpretative layer*. This concept assumes that adding an intermediary between subjects in an experiment imposes a cognitive burden on them due to their lack of information. In addition, such an intermediary will limit the emerging social processes, distancing even more behaviours in the laboratory from behaviours outside the laboratory and reducing the ecological validity of such protocols. The resulting recommendation is, therefore, to dispense with intermediaries as soon as their use is not mandatory to reproduce a particular context and to integrate the relevant variables resulting from direct interactions as indicated in the *Collective Experiment Variables Matrix*.

Although it is conceivable that this withdrawal of intermediaries could cause a loss of control, it also seems possible to use tools and techniques that can be adapted to the environment's context. From questionnaires on pre-existing networks to more advanced tools and technologies for emotion and motion recognition, many paths are opened for researchers to explore the potential of collective experiments. While this represents a certain advantage, it is also essential to emphasize the challenges involved in strategies for both collecting and analysing data using a multimodal and multidisciplinary approach.

Finally, given the elements presented in this chapter, it appears necessary to advocate for adapting laboratories to this new type of protocol. If the *Social Interactions Lab* is one example of what might be the future of some experimental laboratories in economics, there is still a long way to go regarding the required managing and technical skills of these infrastructures and the legitimacy of their use.

To conclude this chapter, as was the case in the early days of experimentation in economics, I advocate for some guiding principles for conducting experiments. Let's be cautious; this does not mean that these recommendations should become a straitjacket for any researcher interested in implementing collective experiments but that they should serve as a basis for both theoretical and practical debate on implementing protocols aiming to observe behaviours, actions, and decisions necessitating interactions between subjects. As a result, the following recommendations aim to consider the challenges brought by collective experiments to provide valid and robust results and to contribute effectively to the literature.

- **Recommendation 1:** *The experimentalist must reconsider how subjects interact within the laboratory to closely mimic the interaction modalities elicited by the research questions and organise interactions in terms of both physical and symbolic configurations.*
- **Recommendation 2:** *The experimentalist must consider collecting more extensive information on subjects due to the increased number of variables affecting their interactions, including for pre-existing networks that account for prior relationships that strongly influence interactions.*
- **Recommendation 3:** *The experimentalist must integrate as much as possible a multimodal data collection to grasp all dimensions that emerge from interactions, encompassing motion, communication, and physiological elements.*
- **Recommendation 4:** *The experimentalist should envision multidisciplinary research projects to benefit from the expertise of other disciplines in data collection, methodologies, and analysis.*

While the above comments serve as initial suggestions for conducting collective experiments, I hope this chapter will spark further discussions on implementing such protocols and exploring potential advancements in experimental practice. It is evident that the collective dimension can no longer be overlooked. Experimentalists must engage with this issue to address the possible challenges highlighted in this work, and this chapter serves as a plea for such an approach.

1.8 Appendix

Wearable Devices for Motion and Emotion Recognition

Motion		Emotion recognition - Body language		
Tools	Subject-tracking devices	Inertial sensors (1)	Video-based facial expression recognition	Video-based Body Posture and Gestures recognition
Aim	Accurately and continuously monitor and capture the position, movement, or orientation of the subjects.	Capture accelerations or changes in the velocity of subjects to investigate movements, activity, and gestures.	Capture and define the emotional, communicative, and psychological aspects of facial expressions.	Capture and define the non-verbal movements and spatial interactions of individuals through the alignment of the body (posture) and expressive movements (gestures).
Focus features	Full body	Torso/Back, Hand/Wrist, Feet/Ankle	Head/Face	Full body
Possible human coding	Yes	Yes	Yes	Yes
Degree of unconsciousness for the subject	Low	Low	Low	Low
Garment-based devices	Headband (Kim et al., 2008)	Headband (Kim et al., 2008), Wristband (Gao et al., 2020), Shirt (Rosso et al., 2010; Pourbarnany et al., 2022)	Not applicable	Not applicable
Accessory-based devices	Badge (Lederman et al., 2018), Smartphones (Thorpe et al., 2019)	Watch (Maurer et al., 2006), Bracelet (Malhi et al., 2012)	Not applicable	Not applicable
Contact (2)	Non-contact	Both	Non-contact	Non-contact
External altering factors (3)	No specific factor	Vibration, shocks, or movement patterns.	Lighting, glasses, masks, cultural factors, occlusions.	Lighting, subjects' clothing, cluttered background, occlusions.
Two-dimensional emotion recognition	Not applicable	Not applicable	Valence, Arousal (Egger et al., 2019)	Valence, Arousal (Glowinski et al., 2011; Kleinsmith and Blanchi-Berthouze, 2013)

(Cont.)		Emotion recognition - Brain activity		Emotion recognition - Skin conductance and temperature		Emotion recognition - Heart Rate Variability	
Tools		Electroencephalography	Galvanic Skin Response devices	Skin Temperature Measurement devices	Electrocardiography (4)	Photoplethysmography	
Aim		Measure the electrical activity of the brain to gain insights into the underlying neural processes and their relationship to behavior and cognition.	Measure changes in skin conductance to provide valuable insights into the physiological aspects of emotional and psychological processes.	Measure the skin surface temperature to understand the impact of environmental and psychological factors on the body's thermal responses.	Measure the electrical activity of the heart to assess cardiac autonomic regulation and the effects of stimuli on the cardiovascular system.	Measure changes in blood volume and flow in the skin to understand the cardiovascular dynamics and their relationships with various physiological and psychological processes.	
Focus features		Head/Face	Torso/Back, Hand/Wrist, Feet/Ankle	Torso/Back, Hand/Wrist	Torso/Back	Head/Face, Hand/Wrist	
Possible human coding		No	No	No	No	No	
Degree of unconsciousness for the subject		High	High	High	High	High	
Garment-based devices		Cap (Vargas et al., 2021)	Scarf (Guo et al., 2016), Wristband (Gao et al., 2020)	Wristband (Gao et al., 2020)	Shirt (Rosso et al., 2010), Scarf (Guo et al., 2016)	Headband (Kim et al., 2008), Pocket (Teichmann et al., 2015)	
Accessory-based devices		Headpiece (Ahn et al., 2019), Glasses (Kosmyna et al., 2019)	Bracelet (Iadarola et al., 2021)	Bracelet (Malhi et al., 2012)	Bracelet (Malhi et al., 2012), Headpiece (Ahn et al., 2019)	Ring (Asada et al., 2003), Wristband (Yang et al., 2019)	
Contact		Contact	Contact	Contact	Contact	Both	
External altering factors		Muscle artifacts (eye blinks or head movements), environmental noise, subjects' hair, and scalp condition.	Cleanliness or natural dryness of the skin, temperature in the room, psychological factor prior the experiment.	Temperature in the room, type of clothing, subjects' age, gender, overall health.	Subjects' weight, alcohol consumption, environmental noise, or movements.	Lighting, movements, temperature.	
Two-dimensional emotion recognition		Valence; Arousal (Egger et al., 2019; Dzedzickis et al., 2020)	Arousal (Egger et al., 2019; Dzedzickis, 2020)	Arousal (Egger et al., 2019)	Valence; Arousal (Egger et al., 2019; Sayed Ismail et al., 2022)	Valence; Arousal (Egger et al., 2019; Sayed Ismail et al., 2022)	

(Cont.)	Emotion recognition - Respiration Rate			Emotion recognition - Verbal communication	
	Tools	Video-based respiratory movements recognition	Thermal cameras for respiratory movements recognition (5)	Respiratory inductive plethysmography	Audio-based devices for speech and voice recognition
Aim		Measure respiratory movements from external body movements or signals to identify changes in respiration rate, breathing depth, or breathing regularity associated with stress, anxiety, or emotional arousal.	Capture the thermal radiation emitted by objects and create images based on temperature variations to gain insights into respiratory function, respiratory disorders, and the effects of interventions on respiratory parameters.	Measure and analyze respiratory parameters and patterns to gain a comprehensive understanding of respiratory function and behavior.	Study verbal interactions, speech patterns, language processing, or communication dynamics by analyzing the content, quality, or characteristics of speech, including speech rate, tone, pitch, or speech errors.
Focus features	Full body	Full body	Full body	Torso/Back	Speech
Possible human coding	No	No	No	No	Yes
Degree of unconsciousness for the subject	High	High	High	High	Low
Garment-based devices	Not applicable	Not applicable	Not applicable	Shirt (Rosso <i>et al.</i> , 2010 ; Pourbarny <i>et al.</i> , 2022)	Shirt (Rosso <i>et al.</i> , 2010)
Accessory-based devices	Not applicable	Not applicable	Not applicable	Belt (Ramos-Garcia <i>et al.</i> , 2017)	Watch (Maurer <i>et al.</i> , 2006)
Contact	Non-contact	Non-contact	Non-contact	Non-contact	Non-contact
External altering factors	Lighting, clothing, respiratory conditions, motion artifacts.	Lighting, clothing, respiratory conditions, motion artifacts.	Ambient temperature, clothing, respiratory conditions, natural body heat variation.	Clothing, respiratory conditions, motion artifacts, environmental noise.	Background noises and recording environment, speech articulation and pronunciation.
Two-dimensional emotion recognition	Valence, Arousal (Egger <i>et al.</i> , 2019)	Valence, Arousal (Egger <i>et al.</i> , 2019)	Valence, Arousal (Egger <i>et al.</i> , 2019)	Valence, Arousal (Egger <i>et al.</i> , 2019)	Valence (6); Arousal (Egger <i>et al.</i> , 2019)

(Cont.)
(1) Inertia sensors can also be considered for body language measurements.
(2) Contact if a direct contact with the skin is necessary.
(3) External factors do not take into considerations the technological issues that can occur, e.g., poor cameras or microphones quality or placement, weak signals emission or reception, missing data, or sensors misplacement.
(4) ECG can also be considered for RR.
(5) Thermal cameras can also be considered for SKT measurements.
(6) The efficacy of valence measures is still in debate in the literature, see <i>Wagner et al. (2023)</i> .

Chapter 2

The Many Faces of Creative Profiles: Exploring Task Openness

This chapter was co-authored with

Sara GIL-GALLEN and Patrick LLERENA

Summary of the chapter

Given the rising complexity of work environments and the importance of multidisciplinary skills, understanding the diversity of creative profiles and how these profiles perform across different tasks is essential for fostering innovation in both individual and organizational contexts. Moreover, these characteristics may interact differently in the creative process depending on the tasks individuals face. We distinguish three degrees of openness to differentiate tasks, indicating which one is more divergent or convergent thinking dominant. Additionally, we explore subjects' profiles considering socio-demographics, social habits and attitudes, skills, and creative performance (self-perceived or cross-task). We conclude that the impact of subjects' characteristics on creativity strongly varies depending on both the openness of the task and the selected criteria for creativity.

2.1 Introduction

“As the source of anything new and effective that comes about through human efforts, creativity is also the driver for most societal growth, change, and discovery over time” [Sawyer and Henriksen, 2024, p. 3]. Without creativity, there would be no innovations [Amabile et al., 1996] which are vital to organisations [Woodman et al., 1993] and economic growth [Aghion and Howitt, 2008]. Moreover, knowing that collective activities are expanding [Mathieu et al., 2017] and that a significant percentage of tasks require multidisciplinary skills to complete because of their growing complexity [Fiore et al., 2010, Kozlowski and Ilgen, 2006, Sundstrom and etc., 1998], understanding the diversity of creative profiles is essential. In the end, *“firms can increase innovation by identifying potentially creative individuals and encouraging them to implement new ideas”* [Sarooghi et al., 2015, p. 13]. But what makes us creative as individuals?

Creativity has been at the heart of research in psychology and management for decades and, more recently, in economics. Regardless of the field of research, an essential element in the study of creativity is the idea that creativity can emerge from everyday acts, from the most insignificant to the major. While the creativity types are numerous, this work mainly refers to it as everyday creativity, which is defined as *“a phenomenon in which a person habitually responds to daily tasks in an original and meaningful way”* [Ilha Villanova and Pina e Cunha, 2021, p. 18]. Moreover, a single individual can not be expected to perform well in every task he or she may be presented with. This helps identify the factors influencing individuals’ creative performance in specific tasks, enabling them to focus on key areas they master and, if needed, develop other aspects to enhance creativity. Creativity is not an innate aptitude of individuals but a skill that can be trained and improved. In the end, detecting creative profiles allows for identifying profiles’ complementarity and beyond, individual creativity, it gives the possibility to set up effective work teams for which the diversity of profiles will make it possible to cover the multiple needs a specific task might require. As a result, this study aims to explore two main research questions: What are the key characteristics of the creative profiles that improve individuals’ performance? And how do different degrees of task openness moderate the relationships between profiles’ characteristics and creative performance?

Nevertheless, in the study of creativity, we can distinguish two central positions: the individualistic and the socio-cultural approach. The first aims to study single

individuals engaged in the creative process, while the second considers individuals within a social and cultural context. Since this study explores the specificities of creative profiles, it combines both approaches, focusing on the individuals and considering elements of their environment. To do so, our study focuses on creative profiles through different dimensions, including sociodemographics, social habits and attitudes, skills, and creative performance. On the other hand, while much of the literature focuses on the dichotomy between an open and a closed task, the fact remains that the degree of openness of a creative task is a continuum. Especially in real-life scenarios, an environment in line with an open task is rarely achieved. Generally, in the creative process, individuals face constraints such as time or economic sources. Thus, this study will focus on three types of tasks: open, open-with-constraints, and closed [Attanasi et al., 2021].

Our within-subject experimental protocol consists of three creative tasks varying in degree of openness. The first task is an open task that consists of the making of a drawing and that elicits a more divergent thinking process. The second task is the widely used Alternative Uses task, which can be defined as an open-with-constraints task that requires divergent thinking processes introducing some constraints, in this case, to find unusual uses for everyday objects with a certain degree of implementability. Finally, the third task elicits a more convergent thinking process since it is a closed task requiring the resolution of a tangram puzzle. In total, 135 subjects took part in this experiment.

The main conclusion of this study is that the factors influencing individual creativity vary depending on the openness of the task and, in one case, even on the specific creativity criterion applied. In fact, across tasks, our hypotheses did not reveal consistent patterns, despite being performed by the same individuals. This highlights the importance of distinguishing between task types, and the thinking processes they elicit.

The remainder of this article is structured as follows. First, Section 2.2 encompasses the key concept and literature on creative profiles and the type of tasks studied later. Then, Section 4.4 presents the experimental design of this study, and Section 4.5 the results stemming from our protocol. Finally, Section 2.5 presents the discussion and concluding remarks of this work.

2.2 Literature review

2.2.1 Creativity: definition and measures

For centuries, authors have questioned creativity. While some saw it as divine inspiration [Gaut, 2012, Stokes and Paul, 2016], others followed the Romantic view, seeing creative individuals as geniuses [Miller, 1996, Sawyer and Henriksen, 2012]. In both cases, creativity remains a mystery as it is envisioned as an inexplicable force. However, this elitist view of creativity has evolved, and creativity is now perceived as a more complex process where everyone can be involved and no longer as a privilege reserved for a few gifted individuals. Also, creativity is no longer confined to the scientific or artistic domains, but now extends to more ordinary aspects of individuals' life. As a result, the reasons to study creativity are numerous. According to Sawyer and Henriksen [2024], these reasons include the identification and realization of creative potential, such as the better response to challenges, the improvement as problem solvers, or even the contribution to mental health and education. Creativity is seen as a transformative force that benefits the entire society.

Creativity can be defined as “*the ability to come up with ideas or artefacts that are new, surprising and valuable*” [Boden, 2004, p. 1]. Beyond these conditions of novelty and usefulness, the study of creativity remains complex because of its multi-dimensional nature. Indeed, it would be shallow to label an output as merely creative or not. In fact, the creative assessment of outputs needs to be based on different elements, and the rest of this work will refer to them as creativity criteria. Throughout the literature on creativity, whether in psychology, management, or economics, various criteria have been used to determine the creative performance of outputs. The four main criteria to assess creativity are *originality* (how infrequent a particular solution is), *fluency* (how many ideas were generated to solve a specific problem), *flexibility* (how many themes cover the set of ideas generated), and *elaboration* (how detailed are the given ideas) [Cassotti et al., 2016, Guilford, 1967]. However, other criteria have emerged to evaluate an output, and the range of criteria that this work focuses on is given in Section 2.4.1.

In addition to the evaluation criteria, if the task context influences creative performance, the creative assessment also relies on the type of task considered. Attanasi et al. [2021] argue that creative tasks differ in degree of openness. They differentiate three types of tasks: a *closed* task, which is defined as finding the right and

often unique solution, an *open* task, which relies on an “outside the box” type of reasoning favouring the generation of multiple solutions, and an *open-with-constraints* task, which is similar to an open task but closer to reality, requiring to respect some constraints. Stemming from the degree of openness of a task, one can determine which thinking process is the most likely to be activated. Indeed, when facing a task, individuals do not always activate the same type of thinking process, and so, when analyzing their performance, we need to focus also on the task itself and the related thinking process. Following Guilford [1967], two main thinking processes can be elicited. On the one hand, the convergent thinking process, i.e., to derive a singular logical solution. On the other hand, the divergent thinking process, i.e., to generate multiple solutions. We can connect the concepts as a closed task relies more on convergent thinking, while open and open-with-constraints tasks rely more on divergent thinking.

Finally, the question remains on the creativity assessment and who will judge this performance. Focusing on the subjective assessment of creativity¹, the literature is concentrated on two predominant options: peer evaluations (participants from the same experiment) or external judges. If the first option allows for direct evaluation of outputs, it might also be biased, knowing that subjects are not blind to the experiment or the treatments. For this reason, the most relevant and widely chosen evaluation procedure is certainly the *consensual assessment technique* [Amabile, 1982], in which we consider subjects creative (or not) when all external judges independently agree upon their evaluations. Further details on the creative assessment procedure are given in Section 2.4.1.

2.2.2 Creative profiles

One of the key aspects to explore in creativity research is the profile of the idea generator. The creative process is a social process as it exists a bidirectional relationship between the individual and their environment. The creative performance of an individual is influenced by their idiosyncratic features as much as their past experiences. Creative performance relies on a complex set of individual specificities, so one should try to categorize them. The following paragraphs distinguish

¹The type of task also indicates if the creativity assessment is objective or subjective. Divergent thinking (open-with-constraints and open tasks) is measured subjectively because we foster the generation of multiple ideas as responses, which are heterogeneous among them. Closed tasks (convergent thinking) are usually judged according to objective rules by the experimenters or a computer. Because in them, subjects provide a unique solution to a problem.

the relationship between individual creative performance and sociodemographic, social habits and attitudes, skills, and creative performance. These elements were not selected randomly; instead, they represent various factors that can influence an individual's intrinsic disposition for creativity, as well as the effects of their environment. It is also important to distinguish between variables that are directly and indirectly related to creativity.

Sociodemographics

When we look at sociodemographic variables, the question of gender naturally arises. This element is not lacking in the creativity literature, as many authors have sought to understand whether gender influences creative performance. Although some references seem to indicate a significant effect of gender [Charness and Grieco, 2019, Bradler, 2015, Karwowski et al., 2016], the final conclusions on this relationship are not consistent [Hardy and Gibson, 2017]. In the end, it seems that the majority of studies conducted on this question conclude that there is no effect of the gender variable on creative performance [Baer and Kaufman, 2008].

The same logic can be applied to the age variable, as there seems to be no clear results on creative performance. For some studies, age has no significant impact on it [Eder and Sawyer, 2007]. For others, this relationship is more complex with a curvilinear age pattern [Massimiliano, 2015], and a possible effect of age on the rate at which older adults pursue divergent thinking process [Foos and Boone, 2008], or a simple decrease in cognitive abilities with age [Salthouse, 2009].

Additionally, sociodemographics also provide us with information on one individual's cultural background. Cabra and Guerrero [2022] argue that creativity must be taken in a certain context, including cultural and geographical. The diversity of results does not either allow us to come to a firm conclusion. One classical result is on the comparison between Easterners and Westerners; while there does not appear to be a clear difference in terms of creative performance between those populations, the creative performance is determined by social habits and leisure for Westerners and intrinsic motivations and self-perception for Easterners [Attanasi et al., 2019a, 2021]. In addition to that, if the geographical and cultural context matters when studying individuals' creative performance, it is also important to observe how multicultural influences might affect this performance. Leung et al. [2008] do observe a positive relationship between exposure to multicultural experiences and individuals' creative performance as it enhances creativity-supporting cognitive processes. Moreover, cre-

ative performance can also be influenced by individuals' educational background. The literature shows that a more highly educated environment enhances creative performance [Jankowska and Karwowski, 2019] as it offers greater intellectual challenges and opportunities arising from greater exposure to diverse ideas [Pang et al., 2020].

Moreover, there exists a positive relationship between creativity and education achievements [Pretz and Kaufman, 2017, Gajda, 2016]. In the end, sociodemographic variables do not seem to present a clear-cut relationship with creative performance, and we aim to explore this relationship in line with the openness of tasks to see if there are more distinct results on the relationship between sociodemographic variables and individual creative performance.

Social habits and attitudes

It is also important to investigate the relationship between social habits or social attitudes and creativity. For that purpose, we differentiate two main elements in the literature: openness to experience and risk. First, the literature on openness to experience presents a positive relationship with creative performance. This relationship is two-fold. On the one hand, openness to experience can be linked to curiosity or an inclination to discover new things that are “out of the ordinary” or conventional. Individuals considered open to experience are more flexible and inclined to experience new perspectives and feelings. This personality trait translates to a preference for exploration that would contribute to creativity, allowing individuals to access a broader range of possible ideas. Hardy et al. [2017] presents the concept of diversity curiosity as the search in breadth for unfamiliar topics. Also, this curiosity would contribute positively to performance by triggering individuals' intrinsic motivation to get involved in a creative process [Hennessey and Amabile, 1988]. Openness to experience, as opposed to satisfaction with the familiar, is a reason for the openness of thoughts and problem-solving strategies increases [Feist, 1998]. On the other hand, openness to experience can also be linked to the related personality trait of extroversion. Feist [1998] observed that, depending on the type of creativity considered, extroversion benefited scientists' creativity, whereas introversion favoured artistic creativity. For both approaches to openness to experience, this personality trait is a source of exposure to diverse and alternative perspectives, points of view, or ideas which fuel creativity. From curiosity, extroversion, or exposure, this allows us to conclude that social activities and the relationships they foster positively affect

individuals' creative performance [Attanasi et al., 2019a].

Furthermore, creativity is intrinsically related to risk. Creativity is an uncertain process, as engaging in a creative process puts you at risk of failing. The relationship between creativity and risk is all the more interesting because these two elements are essential in the innovation process [Koellinger, 2008], and individuals' engagement in a creative task depends on their propensity or willingness to take risks. Indeed, Mueller et al. [2012] underline that the relationship between creativity and risk propensity is complex as the authors found ambivalent feelings towards creativity. Similarly, Ma [2009] find that embracing risk might positively impact creativity. One interpretation is that risk-takers are more inclined to engage in challenging activities. In contrast, risk-averse individuals may limit themselves, leading to a focus on more conventional ideas and stifling their creativity. Risk-takers tend to place a higher value on creative ideas [Mittone et al., 2022], while the uncertainty associated with risk may hinder individuals' ability to recognize valuable creative ideas [Mueller et al., 2012]. Thus, regarding task openness, one might question if the role of risk differs according to the type of task. In fact, risk-averse individuals would demonstrate higher accuracy considering the satisfaction of the constraints imposed in a task than risk lovers [Attanasi et al., 2019b], favouring risk-averse individuals in closed tasks. However, the results in the literature on risk and creative performance are still scarce.

Skills

In addition, one needs to explore individuals' creative ability, defined as the “*skills or competencies relevant to creative performance*” [Choi, 2004, p. 189]. In reality, no one is creative by nature, but the creativity of individuals will vary according to the nature of the task as well as the specific domain it applies to [Baer, 2015]. “*After all, creative people are not creative in a general, universal way; they're creative in a specific sphere of activity, a particular domain*” [Sawyer and Henriksen, 2024, p. 64]. This is why it is important to look at the determinants of creative performance in terms of skills specific to the task at hand. Within this set of possible skills are the creativity-relevant skills such as particular cognitive styles or problem-solving skills [Amabile, 1988, Woodman et al., 1993]; as an example, MacKinnon [1966] argue that writers performed better thanks to verbal intelligence while architects are favoured by spatial intelligence. Again, we also know that skills are not innate to individuals and that they are acquired through learning. This point is even more relevant when

we consider the possibility of acquiring specific training in creativity. However, as highlighted by Baer [2015], “*the effects of creativity training are [also] very domain specific*” (p. 169). In the end, one can not just train to be creative but has to acquire skills and expertise in specific areas related to creativity and the specific task to solve.

Creative performance

Beyond an individual’s skills related to creativity, there is also the perceived or measured creative performance. First, on self-perceived or declared creativity, individuals’ belief in their abilities to perform influences how they behave [Bandura and Wessels, 1997]. “*People who feel more confident in their creative abilities would be more willing to engage in creative behaviour*” [Sawyer and Henriksen, 2024, p. 182] and this engagement contributes to their performance. However, the momentum when individuals are asked about their self-perceived performance is decisive. In fact, self-perceived performances are reliable when individuals evaluate them right after they have just completed a specific task, but when asked in a more general manner, creative performances often do not correspond to their actual performance [Pretz and McCollum, 2014]. Finally, self-perceived performance must be considered in relation to the specific tasks individuals are facing. Whether the task elicits a more divergent or more convergent thinking process, self-perceived creativity will be impacted differently [Furnham and Bachtiar, 2008, Reiter-Palmon et al., 2012]. This last point on divergent and convergent thinking processes can also be related to effective performance, as not all individuals would perform well in all types of tasks. Divergent thinking favours intuition, low evaluation, tolerance to ambiguity, and high ideation, while convergent thinking prevails with structured thinking, high evaluation, measuring, and intolerance to ambiguity [Da Costa et al., 2015, Guilford, 1967, Ma, 2009]. Therefore, examining performance across tasks is essential in identifying individual creative profiles. Even though these two thinking processes might be complementary, some individuals may excel in tasks involving more divergent thinking, others more convergent, and some may perform well in both or neither. Taken together, all these considerations on the close and complex links between an individual’s performance in a creative task and the many and varied characteristics of his or her profile lead us to ask the following research question. To what extent is the creative performance in a task influenced by individuals’ characteristics, and how does the degree of openness of the task moderate those relationships?

2.3 Experimental design

2.3.1 Timeline of the experiments

Our experiment is structured around two main phases. The first phase consists of the completion of a comprehensive questionnaire to define subjects' profiles, and the second phase is organized around three creative tasks that differ according to their degree of openness in a within-subjects design. Those two phases aim to understand the intricacies of creative profiles in order to determine which characteristics benefit certain types of tasks more or, conversely, would harm performance. The following paragraphs describe in more detail those two phases, the resulting experimental procedure, and the main propositions that we aim to test in Section 4.5.

The first phase of our protocol is an in-lab pre-questionnaire. This first comprehensive questionnaire provides us with a complete profile of each subject. In total, 66 questions, divided into four main topics² compose this questionnaire. The topics covered include subjects' sociodemographic information, social habits³, *Diversity of Life Experiences*⁴, and areas of expertise⁵. Then, the second phase is an in-lab pencil and paper within-subjects experiment articulated around three different tasks. These three distinct tasks differ in their degree of openness⁶ and are the following⁷.

The first task, the **Draw task**, is an open task that aims to evaluate subjects' creative performance based on a creative drawing of an alien animal Ward [1994]. As this task presents no specific constraints, it mainly elicits a divergent thinking process. Indeed, if selecting a more common idea requires less cognitive effort, the additional effort of exploring more unusual, more original ideas will lead to better creative performance. This can be referred to as the *path of least resistance* [Ward, 1994, Ward et al., 2004]. This is all the more important when we bear in mind that this type of behaviour tends to appear more in cases where fewer constraints apply. In the specific case of this task, the literature reports that as a result of having to imagine and create an alien animal, the individuals tended to attribute numerous terrestrial characteristics to their creations (bilateral symmetry, sensory

²The full questionnaire is available upon request.

³Adapted and translated from Attanasi et al. [2019a].

⁴Adapted and translated from Douthitt et al. [1999].

⁵As presented in Carson et al. [2005] and Dutcher and Rodet [2022].

⁶To avoid any order effect, some of the subjects completed the tasks from the least constraining to the most and the others from the most to the least.

⁷See Appendices 2.6 for more details on each task instructions.

organs, etc.) [Ward, 1994, Ward and Sifonis, 1997], having the effect of underlining a certain fixation effect on the core concept of animal. More creative individuals will then distinguish themselves by detaching themselves from the most familiar elements, although “*the ability to generate a creative idea begins with known concepts*” [Birdsell, 2019, p. 45].

Then, the **Alternative Uses Task** (from now on AUT), where subjects have to find unusual uses for everyday objects [Torrance, 1966, Guilford, 1967] has been chosen as our *open-with-constraints task*. In fact, the Alternative Uses task is one of the most widely used tasks for assessing an individual’s creative performance. The goal of this task is to evaluate individuals’ ability to generate a wide range of novel ideas, which elicits a divergent thinking process. But it also requires assessing the usefulness or appropriateness of these ideas, necessitating an ability to converge towards feasible solutions. Therefore, in this task, we elicit divergent thinking, but with some constraints; thus, it is *open-with-constraints task*.

Finally, the **Tangram task** where subjects have to solve a tangram puzzle in order to evaluate their performance in a *closed task*. This task, similar to the “Packing Quarters” task in Ariely et al. [2009], aims to measure individuals’ creative performance in the context of a real-effort task by assessing their lack of performance, low-performance, or high-performance. Through this task, we aim to assess individuals’ cognitive flexibility as their ability to produce alternative solutions to a problem and navigate between them to select the most fitted one [Spiro et al., 1988, Stevens, 2011]. This goes in line with the objective of closed tasks that elicit convergent thinking.

Ultimately, an additional questionnaire⁸ was given to subjects to analyze how they perceived their own creative performance. Subjects were asked about their general self-perceived creative performance. Additionally, they were asked (1) what difficulties they encountered during this task, (2) their self-perceived creative performance for each task, and (3) if they think they performed better than others on average.

2.3.2 Experimental procedures

A total of 135 subjects took part in those experimental sessions conducted in the *Social Interactions Lab* (SIL) at the University of Strasbourg. All experiments were conducted in French. The experimental session had two sets depending on the mo-

⁸See Appendices 5.9.

ment they took place. Set 1 refers to sessions in April 2021, and Set 2 refers to sessions in December 2021. There were 3 sessions for the pre-questionnaire divided by the three programs. While for the individual experiments, we have 13 sessions with in average 10 subjects (13, 14, 8, 9, 11, 12, 7, 8, 10, 6, 8, 14 and 15 subjects in each). Table 4.3 presents the descriptive statistics of our sample of student subjects:

Table 2.1: Summary of the Descriptive Statistics

Variables	N	mean	s.d	min	max
Gender	135	0.459	0.5	0	1
Age	135	23.274	4.077	19	45
French speaker	135	0.637	0.483	0	1
French nationality	135	0.585	0.495	0	1

Note: Gender is equal to 1 for women and 0 for men. French speaker is equal to 1 if Yes and 0 if No. French nationality is equal to 1 if Yes and 0 if No.

Regarding the reward system for this experiment, the subjects received a flat payment of 17 euros. This amount is decomposed into two components: 7€ dedicated to the completion of the individual questionnaire and 10€ dedicated to the completion of tasks. As mentioned earlier, subjects performed three different types of tasks (open, open-with-constraints, and closed). In the specific case of creativity, performance-based payments demonstrate different effects, positive and negative, on the subjects' performance [Charness and Grieco, 2019]. Amabile [1983], Amabile et al. [1996] refers to the possible negative effect of monetary incentives as the crowding-out effect negatively affects the intrinsic motivations of subjects, diminishing their creative performance. As those differences in terms of impact rely partly on the nature of the task and its degree of openness [McCullers, 1978, McGraw, 1978], the same payment scheme was applied to all tasks to maintain consistency across tasks and prevent any harmful effects. As a result, subjects were paid a fixed amount rather than a performance-based reward. Moreover, Attanasi et al. [2021] in their survey, all the contributions where performance-based and flat-payment were

introduced did not find a significant difference among them.

2.3.3 Hypotheses

Based on the established literature presented in Section 2.2 and the elements emerging from our experimental protocol, the following paragraphs present the main propositions of this study. Firstly, with regard to sociodemographic variables, we assume that the results currently available in the literature suggest a general effect on creative performance that is not significant [Baer and Kaufman, 2008, Eder and Sawyer, 2007]. Nonetheless, bearing in mind that this work aims to integrate a dimension of the degree of openness of the task studied, we may wonder about the possible moderating effect of eliciting different thinking processes. We formulate the following hypothesis :

Hypothesis 1: *Sociodemographic factors (age, gender, education, and culture) have no significant effect on individuals' creative performance regardless of task openness.*

Secondly, regarding social habits and attitudes, we break down our hypotheses into two parts. On the one hand, we assume that subjects who have been exposed to a variety of experiences will have an advantage in terms of creativity [Hardy and Gibson, 2017, Hey, 1998]. As presented previously in Section 2.2, an individual, through their social habits and attitudes, can show greater openness to experience, which leads them to display inquisitive and possibly extroverted personality traits [Feist, 1998]. Thus, they may have been exposed to different perspectives and ideas that nurture their creative potential. In our opinion, these different possible perspectives contribute more to a divergent thinking process, which aims for fruitful and *outside the box* ideation, more than a convergent thinking process. As this would more significantly improve performance in an open task than in a closed one, it also applies to open-with-constraints tasks, which are mainly geared toward fostering divergent thinking. Therefore, we formulate the following hypothesis:

Hypothesis 2a: *Individuals' exposure to diverse experiences positively impacts their performance in the context of more open-ended tasks (open and open-with-constraints), while this has no significant effect on closed tasks.*

On the other hand, an individual engaged in a creative activity is involved in a risky activity due to the nature of the creative process. While a creative process can

result in producing one or more ideas, it is also highly probable that it might be unsuccessful. As a result, individuals identified as risk-lovers may benefit from this in terms of their creative performance [Ma, 2009, Mueller et al., 2012]. Paradoxically, in such a risky activity, a greater propensity to take risks can lead to the exploration of more uncommon areas of knowledge by generating more novel ideas and divergent thinking processes. On the contrary, risk-averse individuals are expected to outperform on tasks requiring a convergent thinking process [Attanasi et al., 2019b]. Indeed, a closed task that elicits a convergent process will benefit from an attitude focused on the implementation of constraints and the feasibility of the idea. This leads to the formulation of the following hypothesis:

Hypothesis 2b: *Individuals' propensity to take risks positively impacts more open-ended tasks, while risk-averse individuals perform better in closed tasks.*

Afterwards, we assume that the subjects' creative performance can be determined by their skills. As a result, performing well in a creative task might be determined by the individual's expertise in areas closely related to creativity or directly connected to the task itself [Amabile and Gryskiewicz, 1989, Woodman et al., 1993]. In addition, if creativity is a muscle that anyone can train, we can assume a positive impact on subjects' performance when those subjects have already benefited from creativity training [Baer, 2015]. We formulate then the following hypotheses:

Hypothesis 3a: *Individuals' creative performance is positively impacted by their expertise in creativity-related areas, where this positive impact is reinforced when considering task-relevant expertise.*

Hypothesis 3b: *Individuals with creativity training perform better regardless of task openness.*

Finally, regarding creative performance, we distinguish between the analysis of self-perceived creative performance and the actual cross-performance across different tasks. On the one hand, we expect self-perceived creative performance to be a good predictor for subjects' performance [Bandura and Wessels, 1997]. However, we assume that creativity, as defined in the literature, is not perceived similarly by subjects. Based on the subjects' definition of creativity, provided in their first

questionnaire⁹, we assume a tendency of subjects to consider something creative more in line with the conditions of open tasks rather than closed tasks, inducing that self-perceived creativity is a better predictor in the case of open tasks than closed ones [Furnham and Bachtiar, 2008, Reiter-Palmon et al., 2012].

Hypothesis 4a: *Individual’s self-perceived creativity is a good predictor of their actual creative performance for more open tasks, less for closed tasks.*

On the other hand, there might be the case that some individuals can perform well in any creative task presented to them. However, as different degrees of openness of tasks do elicit different types of thinking processes, we expect some subjects to perform better in more open tasks or more closed tasks [Guilford, 1967]. Switching from one thinking process to another requires a certain degree of cognitive flexibility. Since the literature is scarce regarding the cross-performance of subjects in open, open-with-constraints, and closed tasks, we can not predict whether the correlation between performances in open-ended and closed tasks will be positive or negative. Because our experimental setting does not provide evidence that subjects possess a high degree of cognitive flexibility, we prefer to state the following hypothesis regarding the negative correlation:

Hypothesis 4b: *Individuals performing well in more open-ended tasks perform less in more closed tasks, and conversely.*

2.4 Results

In this section, we analyse the results emerging from our experimental protocol considering the three different tasks assigned to subjects. Section 2.4.1 presents the creativity measures stemming from our creativity assessment procedure¹⁰. Then, Section 2.4.2 provides the first insights into subjects’ creative performance across tasks. Section 4.5.3 finally presents our regression analysis for each task and the related significant characteristics of creative profiles.

⁹Their definitions mostly rely on terms such as “*imagination*” or “*novelty*”, but they also insist on the ability to produce ideas or “*Let oneself be guided without restraint or limitation*”.

¹⁰In Appendices 2.6, we present a Codebook with a detailed explanation of all the variables used for the analysis.

2.4.1 Creativity assessment

This section defines our dependent and independent variables related to creativity. First, in the *Tangram task*, we elicit the convergent thinking process given that the subjects overcome the task only by solving the tangram. In line with the literature, we distinguish subjects' performance according to the time spent to solve the task. The final categorical variable distinguishes subjects who failed the task (*failure*) and those who performed above the average time¹¹ (*low-performance*), and those who performed below the average time (*high-performance*) [Ariely et al., 2009].

Then, in the *AUT*, we are dealing with an open-with-constraints task, where divergent thinking with significant restrictions is elicited. Due to the sufficient openness of the task, a subjective assessment is implemented, and we asked an external jury, blind to the protocol and which characteristics are depicted in table 3.4, to evaluate subjects' creative outputs. Each evaluator has to determine the number of valid ideas for each of the five everyday objects presented to each participant and from those valid ideas, they determined the theme or category connected to each idea. Finally, they evaluated on a scale from 0 to 10 the level of elaboration of each set of ideas. Once we had the evaluators' scores, we could determine subjects' ideas *fluency* (number of ideas) and *flexibility* (number of themes) scores. Thus, we could compute two other measures of creativity: *Expansivity* (frequency)¹² and *Persistency* (Number of ideas/Number of themes)¹³ and *Elaboration* (the details introduced in the explanation of the unusual use). It should be noted that both Expansivity and Persistency measures provide a more nuanced analysis of the originality of subjects' ideas. The measures can be interpreted as the internal fixation effect for Persistency (frequency of ideas in the themes of the own participant) and the external fixation effect for Expansivity (compares the frequency of themes across subjects). We selected to describe subjects' creativity in Alternative Uses task through Expansivity and Persistency, knowing that their Cronbach's alpha values are acceptable and reliable [an alpha reliable from 0.6-0.7 and over 0.7 acceptable and reliable, see Ursachi

¹¹In the context of this experiment, the average resolution time was 406.477 seconds (6.77 minutes).

¹²Expansivity, as defined by Camarda et al. [2017], refers to “*the ability to provide solutions outside the fixation effect*” (p. 353) compared to all ideas generated by the entire sample. However, we did not compute the Expansivity score as in the study of Camarda et al. [2017]. Instead, we opted to consider all ideas without categorization, basing scores solely on average frequency.

¹³This Persistency score aims to evaluate subjects' own fixation effect by studying how diverse themes were across ideas.

et al., 2015]¹⁴. Indeed, the Cronbach’s alpha for Expansivity is 0.8493 and for Persistence, 0.8739.

Table 2.2: External Judges’ Profiles

<i>Judge</i>	<i>Set</i>	<i>Gender</i>	<i>Position</i>	<i>Domain</i>
1	1	M	PhD Student	Cliometrics and Growth theory
2	1	F	Associate Professor	International innovation management, Values-based innovation management
3	1	M	CEO	Research and development, Health
4	1	F	PhD Student	Labour market, Poverty
5	2	M	PhD Student	Cliometrics and Growth theory
6	2	F	Associate Professor	Game theory
7	2	M	PhD Student	Economics of health
8	2	F	PhD Student	Economics of innovation

Finally, for the Draw task, we follow [Amabile, 1982]’s *consensual assessment technique*, in which we consider subjects creative (or not) when all external judges as long as some expert raters independently agree upon the evaluation. The juries were the same ones who evaluated the Alternative Uses task (see Table 3.4). This technique applies to sufficiently open-ended tasks that do not require special skills, and for which there is a wide variation in the target population (as is the case of the Draw task, which is open). Judges are external observers with experience in the targeted domain, using their own subjective judgments of creativity to rate the creative outputs independently. The drawings in our experiment were evaluated by four judges on a 1 (minimum) to 10 (maximum) scale, and were presented to them in different orders to control for potential ordering effects. In the end, our external judges were divided into two sets, one focusing on one sub-sample of subjects and a second focusing on the other sub-sample¹⁵. The value of inter-judge reliability is equal to 0.6567 for set 1 and 0.5596 for set 2. A reliable Cronbach’s alpha is between 0.6 and 0.7, but an alpha of 0.5 or greater is acceptable [Perry et al., 2004, Ursachi et al., 2015].

¹⁴Fluidity and flexibility are integrated into the computation of the Persistence indicator, and elaboration was not an acceptable and reliable measure with a Cronbach’s alpha of 0.4963.

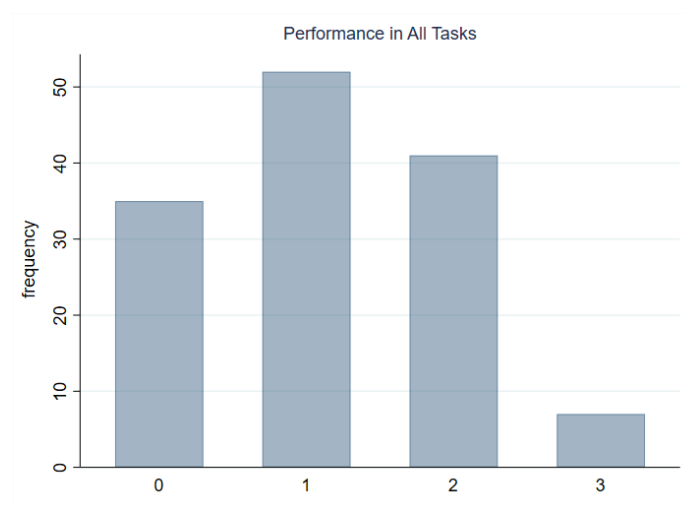
¹⁵The experimental session had two sets depending on the moment they took place. Set 1 refers to sessions in April 2021, juries 1 to 4, and Set 2 refers to sessions in December 2021 relates to juries 5 to 8.

As presented above, the creativity measures serve our analysis as a dependent variable. However, self-perceived creativity has also been measured. We distinguish four measures: (1) subjects' perceived general creativity, (2) subjects' perceived performance in the experiment as a whole, (3) subjects' perceived performance for each task, and (4) subjects' relative performance for each task (as they evaluate their performance compared to the other subjects).

2.4.2 First Insights on Creative Performance

Based on the creativity measures described above, the following graphs describe the subjects' performance in the three tasks. These initial insights into our analysis of subject performance aim, above all, at highlighting the disparity in performance across tasks and the fact that few subjects can perform well in each task, emphasising the importance of the difference in individual profiles. So, before moving on to a more detailed analysis of the determinants of their performance, we can observe the following elements.

Figure 2.1: Subjects' Performance as Number of Tasks They Succeed in

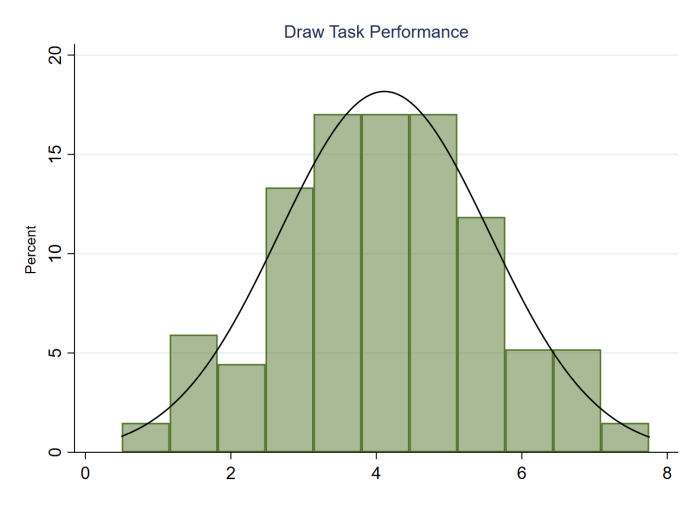


To begin with, Figure 2.1 illustrates the overall performance of the subjects during the experimental session. It reveals that a higher proportion of subjects were deemed creative in only one task (38.52%), meaning their performance exceeded that of the overall sample for that specific task. In contrast, the number of subjects identified as creative across two tasks was slightly lower (30.37%), and there was a significant decrease in creativity from two to three tasks, dropping to just 5.19%. More precisely,

if we look at the distribution within each category, among the subjects who performed in a single task, 23.08% performed in the AUT, 26.92% in the Tangram task, and 50.00% in the Draw task. Of the subjects who performed in two tasks, 17.07% performed in the Tangram and AUT, 36.59% in the Draw and AUT, and 46.34% in the Tangram and Draw Task. This graph also allows us to conclude that a noticeable number of subjects also failed in each task (25.93%). Although this graph does not tell us anything more about the ins and outs of performance on each individual task, it does confirm one of our intuitions that an individual's creative performance is not constant, whatever the task. However, one surprising result emerged as a significant number of subjects performed in both open and closed tasks. As a result, it becomes even more interesting to look at the characteristics of the subjects who performed these tasks to shed light on this graph.

If we now focus more closely on each task, we can still observe some interesting findings. For the Draw Task in Figure 2.2, the scores obtained follow a symmetrical distribution ($mean = 4.11$, $skewness = 0.001$). We encounter a statistically significant normal distribution ($p-value = 0.872$; Shapiro–Wilk and Shapiro–Francia tests for normality). Once again, this allows us to question the reasons for this distribution and to understand even more which subject profiles performed best.

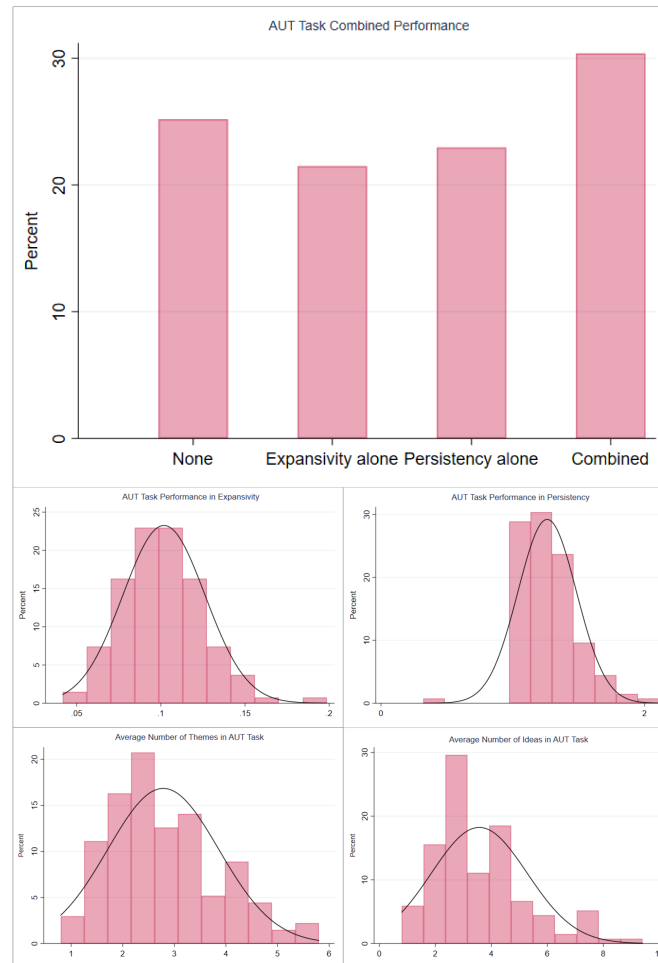
Figure 2.2: Subjects' Performance in the Draw Task



Moving on to the Alternative Uses task in Figure 2.3, we need to take a closer look at a number of graphics. Firstly, overall performance by comparing subjects who performed well on the Expansivity criterion, those who performed well on the

Persistency criterion, and even those who performed well on both criteria. We observed a similar percentage of subjects who performed well in Expansivity (21.48%) as those who performed in Persistency (22.96%), while 30.37% performed well in both criteria. Finally, a fairly large number of subjects did not perform in any of the criteria (25.19%).

Figure 2.3: Subjects' Performance in the Alternative Uses Task

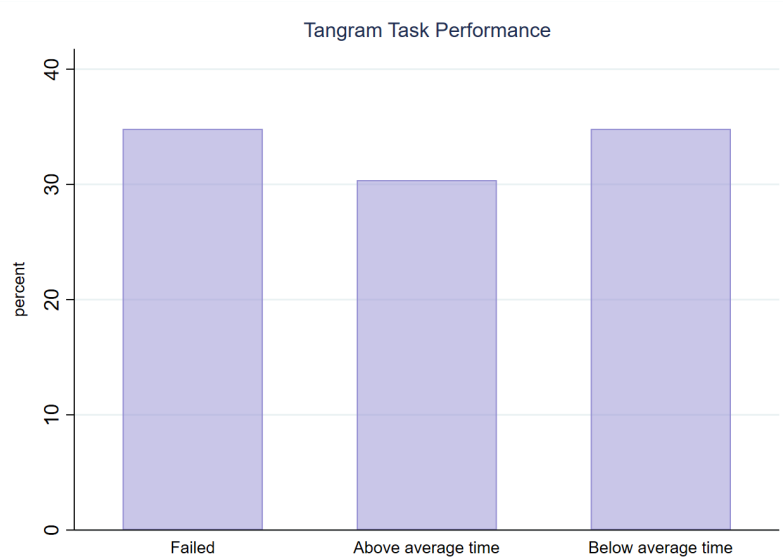


Due to the difference in performance between criteria, it becomes interesting to examine each of these two criteria. For Expansivity, we observe a slightly symmetric distribution ($mean = 0.1$, $skewness = 0.47$)¹⁶. But it is not supported statistically because we reject the null hypothesis of normal distribution ($p-value = 0.0092$;

¹⁶To be noted, Expansivity scores range from 0.042 to 0.199 where a higher performance in Expansivity translates in a score closer to 0.

Shapiro–Wilk and Shapiro–Francia tests for normality) for skewness ($Pr(Skewness) = 0.025$) and kurtosis ($Pr(Kurtosis) = 0.018$). Then, for Persistency, we do not observe a symmetric distribution because is switched to the greatest values ($mean = 1.26$, $skewness = 0.32$). Indeed, we provide statistical evidence that does not follow a normal distribution ($p-value = 0.0011$; Shapiro–Wilk and Shapiro–Francia tests for normality), because there is kurtosis ($Pr(Kurtosis) < 0.001$), but none for skewness ($Pr(Skewness) = 0.117$). To better understand this Persistency score, which is the ratio between the number of ideas and the number of themes, we also examined their distributions. For the average number of ideas, we observe a right-skewed distribution ($mean = 3.57$, $skewness = 1.05$)¹⁷, while for the average number of themes covered by subjects, the mean being higher than the median and the positive skewness value indicate that the distribution is moderately skewed to the right ($mean = 2.79$, $skewness = 0.63$)¹⁸.

Figure 2.4: Subjects’ Performance in the Tangram Task



Finally, Figure 2.4 shows how subjects performed in the Tangram Task. Although the number of subjects in each performance-category differs slightly, they

¹⁷The statistical test significantly reports, we reject the null hypothesis of normal distribution ($p-value < 0.001$; Shapiro–Wilk and Shapiro–Francia tests for normality) mainly by skewness ($Pr(Skewness) < 0.001$) and a slightly by kurtosis ($Pr(Kurtosis) = 0.059$).

¹⁸The statistical test significantly reports the we reject the null hypothesis of normal distribution ($p-value = 0.021$; Shapiro–Wilk and Shapiro–Francia tests for normality) only by skewness ($Pr(Skewness) = 0.004$) and no by kurtosis ($Pr(Kurtosis) = 0.853$).

are relatively similar. Each category (failure, low performance, and high performance) represents around a third of our population, respectively 34.81%, 30.37%, and 34.81%.

2.4.3 Regression analysis

In this section, we conduct a regression analysis to explore the drivers of individual creativity controlling for individual characteristics/profiles and the integration of different degrees of task openness¹⁹. Thanks to this investigation, we aim to test the hypothesis presented in Section 4.3.

Draw task

We first devote our attention to the determinants of the subjects' performance in an open task, namely the Draw task. The dependent variable in this first regression is a binary variable, which differentiates subjects who performed well (over the average performance) versus those who performed worse (under or equal to the average performance) in the Draw task. Using a Probit regression, we study subjects' creativity in an open task in relation to the following regressors: sociodemographic variables, social habits and attitudes, skills and training, and performance in creativity. Table 2.3 presents the summary results of our regression²⁰.

First, in line with the rest of the literature, sociodemographic variables do not significantly impact individuals' performance [Baer and Kaufman, 2008, Eder and Sawyer, 2007]. The only exception is education, as we find evidence that subjects with a Master's degree perform better than others ($p_{value} = 0.044$). Noting that this significance is not consistent across all models, nor across all higher levels of education. Therefore, we choose to be cautious and do not assert a strong conclusion about it. In conclusion, our results align with *Hypothesis 1*, remarking on the non-significant effect of sociodemographic information on creative performance in an open task.

Regarding social habits and attitudes, we approximated subjects' social life by gathering three related items in the literature: going out at night, meeting people,

¹⁹During the different analyses applied, we implement the terminology presented by Moffatt [2020], which states the following: there is *mild evidence* if $p < 0.10$; there is *evidence* if $p < 0.05$; there is *strong evidence* if $p < 0.01$.

²⁰The complete regression tables are available in Appendices 2.6.

Table 2.3: Summary Results for the Draw Task

	<i>Significativity</i>
Diploma (Master’s degree)	+ +
Social life	+
Risk	--
Expertise in dance	--
Tangram task performance (low)	-
Self-perceived creativity (Draw task)	+ +

We indicate a positive significant effect by + + + $p < 0.01$, + + $p < 0.05$,

+ $p < 0.1$. The negative effect by - - - $p < 0.01$, -- $p < 0.05$, - $p < 0.1$.

and going to bars/restaurants. We built this variable as a reliable and accepted measure of social life (focus on nocturnal social activities), which corroborates the Cronbach’s alpha equal to 0.7329. We observe mild evidence of a positive interplay of this social life on subjects’ performance ($p_{value} = 0.069$), showing that a more vivid social life correlates with subjects’ creative performance in an open task. However, we do not find a significant effect of the Diversity of Life Experiences (DOLE) on subjects’ performance. If the variable accounts for nocturnal social interactions, the second considers exposure to diverse individuals, cultures, or environments. Therefore, we can not support *Hypothesis 2a* on the role of exposure on performance but underline the role played by frequent social interactions and activities on it, in this case, only for nocturnal social activities. Then, for risk attitude, we observe that risk loving subjects performed less in the Draw task ($p_{value} = 0.014$). One reason behind this result can be found in Attanasi et al. [2021] suggesting that “*risk-averse subjects would be more accurate in satisfying the imposed constraints*” (p. 315). Although this is an open task with a very low level of constraint, the propensity of risk-averse individuals to thoroughly follow instructions may have enabled them to detach themselves from the “similar to Earth” dimension of their drawings. In the end, this result contradicts *Hypothesis 2b* stating that performance in an open task would benefit more from risk-loving attitudes.

About skills, we need first to distinguish skills relevant to the task itself and skills relevant to other areas of creativity. We do not find evidence for task-related expertise but observe evidence of a negative correlation between subjects’ performance and

expertise in dance ($p_{value} = 0.019$). This contradicts *Hypothesis 3a*, which states that expertise in creativity-related areas might promote individual creativity. Moreover, as we find no evidence of a significant role played by subjects' training in creativity on their performance, we can not support *Hypothesis 3b*.

Afterwards, subjects with self-perceived high creative performance in the Draw task did perform well in that task. This conclusion corroborates the *Hypothesis 4a*, which validates the positive interplay between self-perception and creative performance. This is also in line with the literature on the elicitation of self-perceived creative performance right after the task has been performed [Pretz and McCollum, 2014]. Finally, since different levels of task openness induce different thinking processes, individuals may not perform equally on all types of tasks depending on their creative capacities. For the Draw task, we observe evidence that performing above the average time in the Tangram task harms the performance. However, this is only the case for low performance, i.e., subjects completing the Tangram with a resolution time above the average; no significant effect is found for those performing well in the task with a resolution time inferior to the average one. Therefore, we do not find support for *Hypothesis 4b*.

Alternative Uses task

To explore the determinants of subjects' performance in the open-with-constraints task, we implemented two regressions, each designed to assess a different creativity criterion. Using the Alternative Uses task, we aim to test subjects' performance in Expansivity (originality measure controlling for external, compared to the whole sample, fixation effect) and Persistency (internal fixation effect). As with the previous regression, our dependent variable is a binary one stating whether the subject has performed below or above the average score of the entire sample. For more details on the definition of the dependent variables and the independent variables, see Section 2.4.1 and Appendices 2.6, respectively. We begin by paying attention to Expansivity, and Table 2.4 presents the summary results of our regression²¹

We focus first on sociodemographic variables, where we report evidence of a significant negative impact of age ($p_{value} = 0.023$) and gender ($p_{value} = 0.046$) on subjects' performance in the Expansivity criterion. More specifically, we observe

²¹The complete regression table is available in Appendices 2.6.

Table 2.4: Summary Results for the Alternative Uses Task in Expansivity

	<i>Statistical significance</i>
Gender	--
Age	--
Diploma (Doctoral degree)	++
Social life	---
Program	+
Domain	--
Expertise in inventions and technologies	+
Overconfidence (AUT)	--

We indicate a positive significant effect by +++ $p < 0.01$, ++ $p < 0.05$, + $p < 0.1$.

The negative effect by --- $p < 0.01$, -- $p < 0.05$, - $p < 0.1$.

evidence of a negative relationship between creative performance with women and older subjects. The pre-existing literature is ambiguous for the former as there are mixed results on the relationship between gender and creative performance [Attanasi et al., 2021, Baer and Oldham, 2006, Bradler, 2015, Charness and Grieco, 2019]. For the age of subjects, we need to stay cautious in our interpretation as the age is mostly concentrated around the average of 23-years-old with few extreme variables with a maximum age of 45. Thus, the negative effect of age might be driven by these extreme values. In addition, we observe evidence of the positive interplay of having a doctoral degree in Expansivity performance ($p_{value} = 0.045$). As was the case in the Draw task for Master's degrees, this specific result is not consistent across all education levels, which limits our ability to conclude on the effect of education on creativity. Even though, these significant relationships between performance and sociodemographic variables arise and go against the existing literature [Baer and Kaufman, 2008, Eder and Sawyer, 2007] and our *Hypothesis 1*.

Then, on social habits and attitudes, the only variable that arises significantly is again social life. Indeed, we find strong evidence of a negative correlation between social life and Expansivity ($p_{value} = 0.003$). Contrarily to the Draw task, having more nocturnal social interactions and activities is detrimental to the creative performance. While it helped subjects produce more original drawings of alien animals, it diminished their ability to provide more original ideas of Alternative Uses of objects compared to the entire sample. We interpret this as the fact that experiencing

social life makes creative ideas less original. If more social life might be interpreted as greater exposure to broader individuals' perspectives and experiences, we assume that it is also exposure to individuals from our social circles, who are often closer to us due to the predominance of homophilic behaviours favouring interactions with individuals similar to us [Louch, 2000, Spencer and Pahl, 2007]. Then, our variables also allow us to control for exposure to diversity using the DOLE scores. Unfortunately, we do not report any significant effect of DOLE on subjects' performance in Expansivity. Same for risk preference. Thus we do not find evidence supporting *Hypothesis 2a* and *2b* respectively.

Regarding subjects' expertise, we again need to distinguish skills relevant to the task itself and skills relevant to other areas of creativity. First, in line with *Hypothesis 3b*, we show mild evidence of the positive impact of training on creativity on Expansivity ($p_{value} = 0.058$). This shows the benefit provided by such training on the ability of subjects to provide more original ideas compared to others. We also find evidence of the negative interplay between the study domain of sciences and technology and Expansivity ($p_{value} = 0.010$). It indicates that students who specialized in these areas performed less in terms of the originality of their ideas than students in economics and management. However, individuals who declared expertise in inventions and technologies did perform better ($p_{value} = 0.098$) in Expansivity. These two last results are interesting as they stress a difference between two elements that may seem similar. For both, technologies may have a central role in individuals' interests. However, what matters is whether we consider it as a study domain or a domain of (self-declared) expertise. While the former is only the study domain of subjects, which does not provide any engagement or proficiency indicator, the latter integrates the fact that subjects are confident enough in their ability to attest to a certain degree of expertise in this area. This conclusion supports *Hypothesis 3a* and shows that the effect is highly dependent on the proximity between the task and the expertise area. Finally, we find mild evidence of a positive effect of creativity training on subjects' performance ($p_{value} = 0.058$), meaning that students benefiting from this creativity training are able to provide more original ideas. It supports *Hypothesis 3b*.

To conclude, we find evidence of a negative interplay between subjects' overconfidence in outperforming others in the Alternative Uses task and their actual performance in terms of Expansivity ($p_{value} = 0.021$). This suggests that, against *Hypothesis 4a*, a relative overconfidence of subjects' own performance leads to poorer

performance in terms of originality. This result is in line with the literature on students' self-perceived creativity and their tendency to overestimate it [Pesout and Nietfeld, 2021, Kruger and Dunning, 1999]. Regarding the relation with other tasks and criteria of individual creativity, no significant evidence is found. Thus, we do not support *Hypothesis 4b* for Expansivity in the open-with-constraints task because no significant effect of the performance in the other tasks is found.

Focusing now on Persistency, i.e., a subject's own fixation effect across the different ideas provided, our dependent variable is a binary variable differentiating over the mean and below/equal the mean performances. Table 2.5 presents the summary results of our Probit regression²²

Table 2.5: Summary Results for the Alternative Uses Task in Persistency

	<i>Statistical significance</i>
Going to theaters and films	--
Social life	---
DOLE	+++
Program	---
Domain	+++

We indicate a positive significant effect by +++ $p < 0.01$, ++ $p < 0.05$, + $p < 0.1$.

The negative effect by --- $p < 0.01$, -- $p < 0.05$, - $p < 0.1$.

In the case of sociodemographic variables, as suggested in *Hypothesis 1*, we do not report a significant effect on Persistency in the open-with-constraints task. This result stands out the fact that even being also evaluated for the same task, for Expansivity different sociodemographic variables, such as age, gender and doctoral education, appear to be relevant. Therefore, we can suggest that not only the degree of openness matters but also the criterion to measure creativity, at least in this open-with-constraints task for creativity.

Then, we also control for the role of social habits and attitudes on Persistency performance. We report strong evidence for a negative of Going to theatres and films ($p_{value} = 0.040$) and evidence for Social life ($p_{value} = 0.001$) on subjects' perfor-

²²The complete regression table is available in Appendices 2.6.

mance. On the one side, Going to theatres and films seems to negatively influence performance in terms of persistence, meaning that higher exposition to such forms of art does create a bigger tendency for fixation effect. On the other side, we apply the same interpretation of Social life as the one for its role in Expansivity, stating that individuals often encounter others from close social circles who tend to be more similar to them, minimizing their exposure to diversity and harming the possible originality of their ideas. In fact, when we devote attention to the DOLE, which considers exposure to diverse individuals, cultures, or environments, the sign of the effect is reversed and has a strong positive correlation with Persistency ($p_{value} = 0.006$). Thanks to this last result support the *Hypothesis 2a*. Finally, given that no significant effect arises for risk preference, we reject the *Hypothesis 2b*.

On the side of subjects' skills, we find a strong positive correlation between the domain students are specialized in and Persistency ($p_{value} = 0.000$). We find that students in economics and management perform less than students in sciences and technologies, while it was the opposite for Expansivity. However, apart from this university specialization, we find no evidence of an effect of creativity-related skills, preventing us from supporting *Hypothesis 3a*. Regarding training in creativity, we encounter strong evidence of a negative correlation with Persistency in the Alternative Uses task ($p_{value} = 0.000$). Trained individuals provide more original ideas but are not specifically able to provide ideas that cover different themes; i.e. they tend to provide ideas within the same category, increasing a possible fixation effect. Thus, we can not corroborate *Hypothesis 3b* for Persistency.

Furthermore, we do not find a significant effect of self-perceived creativity across measures and tasks of individual creativity, which makes us not validate *Hypothesis 4a* and *4b*, respectively.

Tangram task

To conclude our data analysis, we focus on our closed task, which is the Tangram task. Based on Ariely et al. [2009], we codify our dependent variable distinguishing low from high performance, but we also differentiate low performance from failing the task. We end up with three types of outcomes: whether subjects fail to complete the Tangram, completion time above the average (low performance) or below the average (high performance). We implement an Ordered Probit regression, the summary results which are presented in Table 2.6²³

²³The complete regression table is available in Appendices 2.6.

Table 2.6: Summary Results for the Tangram Task

	<i>Statistical significance</i>
Age	— — —
French speaker	+
Diploma (Bachelor's degree)	++
Foreign places	++
Risk	— —
Expertise in visual arts	—
Expertise in inventions and technologies	+
Expertise in scientific inquiries	—
Program	+++
Expansivity performance	—
Self-perceived creativity (Tangram task)	+++

We indicate a positive significant effect by +++ $p < 0.01$, ++ $p < 0.05$, + $p < 0.1$.

The negative effect by — — — $p < 0.01$, — — $p < 0.05$, — $p < 0.1$. The n.s. means no significant effect.

We begin by testing the impact of sociodemographic information on subjects' performance in the Tangram task. First, we find strong evidence of a negative effect of age ($p_{value} = 0.004$) on performance in the closed task. In line with the result encountered in the Alternative Uses task for Expansivity. While the effect of age is often associated with task complexity, as “more complex cognitive tasks place greater demands on a working-memory resource that declines with increased age” [Salthouse, 1992], it is hard to follow this interpretation since the age in our sample does not vary a lot with a minimum age at 19, maximum at 45, but a median age at 22. While, being a French speaker positively contributes to the performance ($p_{value} = 0.031$). While this variable was not significant for other tasks, it is with mild evidence for the Tangram task. One interpretation can be based on the complexity of the task since the instructions specify that subjects need to “make a square with all the pieces of wood at your disposal,” which means that no other forms or under-use of pieces are allowed. As a result, the understanding of the instructions by native speakers did benefit them due to the low cognitive load imposed on the completion of the Tangram. Finally, we find evidence of a positive effect of having a Bachelor's degree ($p_{value} = 0.014$). Again, this result is not consistent across all types of higher degrees, which limits the possible interpretation of this result. Similarly to Expansivity, we can not support *Hypothesis 1* due to the presence of significant sociodemographic variables.

Regarding social habits and attitudes, we find evidence of a positive relationship

between performance in the closed task and whether subjects who often go to places frequented by foreigners ($p_{value} = 0.035$). Here, subjects that responded positively to the question are more exposed to diverse perspectives, which benefit their creative performance. Although only one variable about exposure is significant, this positive effect on the performance in a closed task nonetheless calls into question our presupposition. We can not support *Hypothesis 2a*. Then, we report evidence of a negative interplay between risk-loving attitudes and individual creativity in the Tangram task ($p_{value} = 0.028$). Thus, as *Hypothesis 2b* suggests, subject risk averse are more focused on respecting constraints and perform more in a closed task (find a solution to a problem). This result is coherent with the literature [Attanasi et al., 2019b].

Additionally, we investigate the creativity-related skills. We find mild evidence of a negative correlation between expertise in visual arts and creative performance in the Tangram task ($p_{value} = 0.073$). Our interpretation is that visual arts is a creativity-related skill in line with our open task. As we found a negative relationship between performance in the Draw task from closed task, we find similar logic here with the same dichotomy. However, we do not observe a clear relationship because the Tangram task (closed task) does not raise significant performance on the Draw task (open task). On the contrary, we find mild evidence of a positive effect of expertise in inventions and technologies on creative performance in the closed task ($p_{value} = 0.086$) but a negative effect of expertise in scientific inquiries ($p_{value} = 0.051$). Even if, in both cases, we refer to pure sciences, we can state that they look for two different objectives. Indeed, science seeks to inquire to find the truth, whereas technology seeks utility. This is coherent with our results, as expertise in invention of technology looks to find the utility, in line with the closed task, find a solution to a problem, in this case, solving Tangram. While scientific inquiry may relate more to open thinking because the search for the truth can have several solutions, it is more blue-sky thinking opposite to convergent thinking, which explains the negative effect. Thus, we do not fully support to *Hypothesis 3a*, which remarks that expertise has a positive impact on performance. Additionally, we show a significant positive effect of training on creativity ($p_{value} = 0.004$), which confirms *Hypothesis 3b*.

Finally, we regard the dynamics between performance in the Tangram task and both self-perceived creativity and other creative performances. Regarding self-perceived creativity, as for the Draw task, we observe a positive interplay between the perceived performance and the effective performance ($p_{value} = 0.000$). This result corroborates

Hypothesis 4a, which conjectures the positive relationship between self-perception of creativity and actual performance. In contrast, we do not present a significant effect of the other measures of individual creativity on the closed task, except for mild evidence for Expansivity in the Alternative Uses task with a negative effect on the Tangram performance ($p_{value} = 0.092$). Thus, we do not fully support *Hypothesis 4b*.

2.5 Discussion and conclusion

This study investigates individual creativity by devoting our attention to the role of two main drivers: subjects' profile characteristics and the degree of openness of the creative task. On the one hand, we identify subject profiles by considering sociodemographics, social habits and attitudes, skills, and creative performance (whether self-perceived or task-based). As indicated in the introduction to this work, our approach to creativity aims to combine an individualistic as well as a socio-cultural vision, recognising the importance of studying individual creativity while integrating its determining elements in relation to the influence of the environment to which the individuals belong. On the other hand, we introduce the degree of openness of the task, considering three dimensions: open, open-with-constraints, and closed task [Attanasi et al., 2021]. The main conclusion of this study is that the drivers of individual creativity depend on the openness of the task and even, for one of our tasks, the chosen creativity criterion. Table 2.7 provides a summary of our results regarding each hypothesis across tasks; differentiating whether a hypothesis has been supported, invalidated²⁴, or if no significant evidence, either supporting or invalidating, has been found.

Firstly, our study confirms no clear and significant effect of sociodemographic variables on an individual's creative performance. Comparing our three tasks, we do not find a systematic pattern of one variable affecting a subject's performance. We support the hypothesis for the open task (Draw), but the opposite results are found for the closed task (Tangram). For the open-with-constraints (AUT) task, we have both directions that emerge depending on the chosen creativity criteria. More specifically, some variables arise as significant; however, any generalizations we could make about them would be too weak. Regarding age, even though a significant negative impact is found for two tasks, the age range in our sample is too concentrated

²⁴*Invalidated* means that the results obtained in our analyses have not only reject our hypothesis but contradict them.

Table 2.7: Summary of Results and Validation of Hypotheses

	Draw task	AUT (Expansivity)	AUT (Persistency)	Tangram task
Hypothesis 1: <i>Sociodemographic factors (age, gender, education, and culture) have no significant effect on individuals' creative performance regardless of task openness.</i>	Supported	Invalidated	Supported	Invalidated
Hypothesis 2a: <i>Individuals' exposure to diverse experiences positively impacts their performance in the context of more open tasks, while this has no significant effect on more closed tasks.</i>	Ø	Ø	Supported	Invalidated
Hypothesis 2b: <i>Individuals' propensity to take risks positively impacts more open-ended tasks, while risk-averse individuals perform better in closed tasks.</i>	Invalidated	Ø	Ø	Supported
Hypothesis 3a: <i>Individuals' creative performance is positively impacted by their expertise in creativity-related areas, where this positive impact is reinforced when considering task-relevant expertise.</i>	Ø	Supported	Invalidated	Invalidated
Hypothesis 3b: <i>Individuals with creativity training perform better regardless of task openness.</i>	Ø	Supported	Invalidated	Supported
Hypothesis 4a: <i>Individual's self-perceived creativity is a good predictor of their actual creative performance for more open tasks, less for closed tasks.</i>	Supported	Invalidated	Ø	Supported
Hypothesis 4b: <i>Individuals performing well in more open-ended tasks perform less in more closed tasks, and conversely.</i>	Ø	Ø	Ø	Invalidated

Notes: Ø indicates that no significant relationship has been found, preventing us from concluding on the specific hypothesis.

to truly conclude a possible effect of age on creativity, as it might be the case with the decrease of cognitive flexibility. In addition, no consistent pattern has been found for age and educational background. The only element that caught our attention is the role played by being a native speaker in the case of the Tangram task. Our first interpretation here was related to the complexity of the task being higher than that

of other tasks. Still, as the question of task complexity was not the main objective of our study, we think that this result requires additional investigation.

On social habits and attitudes, we assumed that exposure to diverse experiences would positively influence subjects' performance in an open task [Hardy and Gibson, 2017, Hey, 1998, Feist, 1998]. At the same time, it would have no specific impact on performance in a closed task. In the end, our study does not support this hypothesis. We find a clear relationship between exposure to diverse life experiences and performance in the Persistency creativity criterion in the Alternative Uses task. In other words, subjects manifested a lower fixation effect if exposed to more diverse experiences, but this result only arose in the open-with-constraints task. We do not find any evidence for the open task and only a weak significant effect on a closed task. One notable finding is the role played by frequent social interactions, which can not directly be linked to exposure to diversity but reveals the role played by the environment on a subject's creative performance. Besides these considerations on the role of exposure to diverse experiences, we also paid attention to the risk attitude of subjects. While we expected risk-loving attitudes to have a positive effect on performance in open tasks [Ma, 2009, Mueller et al., 2012], and risk-averse attitudes to benefit performance in closed tasks [Attanasi et al., 2019b], our final analysis reveals that risk-averse attitudes are more beneficial for both tasks. However, we do not find evidence for the intermediary task, the open-with-constraints one.

Afterwards, our analysis investigated the impact of subjects' expertise in certain domains, as well as the importance of creativity training. Both elements contribute to the set of skills from which subjects can benefit. In both cases, we do not find support for our hypotheses. However, we do not want to conclude that these elements do not affect creative performance either. We underline some limits of our study that might explain these non-significant results. First, regarding subjects' domains of expertise, one limitation is that these variables are self-assessed. Effective measures of expertise might provide completely different sets of expertise for each participant based on their true performance and not their subjective confidence. Moreover, the specific choice of domains might have influenced the results. Further work should continue to investigate the role of expertise in creativity by testing alternative measurement means and categories. Then, the mixed results on creativity training likely stem from the low number of subjects benefiting from such training, limiting the possible results in our regression analysis. As for domains of expertise, we call for further investigation into the role of creativity training, comparing its effects across tasks and

thinking processes. Especially since we found differences in the effect of creativity based on the chosen creativity criteria.

Finally, the last hypotheses formulated considered subjects' creative performance in two ways: self-perceived and cross-task performance. On the one hand, self-perceived creativity is a good predictor of subjects' effective performance for closed and open tasks. While it confirms our hypothesis across these two tasks, the same relationship does not appear for the open-with-constraints task. However, the results for the open-with-constraints task demonstrate the negative impact of subjects' overconfidence on their true performance [Kruger and Dunning, 1999, Pesout and Nietfeld, 2021]. Then, we compared the performance across tasks and could not conclude the effect of performing well in a more open-ended task versus a closed-ended task or conversely. It appears that only good performance in the open-with-constraints task, in terms of Expansivity, harmed the performance in the resolution of the Tangram. As no clear pattern emerges, we can not conclude on this matter.

For future research, we recommend addressing two of the main limitations of our study. First, individual creative performance is task-dependent. Thus, it will be interesting to test different tasks with the same degree of openness, but seeking different topics. For instance, in our case, subjects were asked to draw a creative output in the open task. Thus, the results might differ, such as the contribution of [Charness and Grieco, 2013] with different types of intelligence considering drawing, maths and verbal tasks, but controlling for the role of subjects' profile characteristics. Furthermore, we captured subjects' profiles using a certain breadth of questions/measures, but the literature is vast, and different methodologies should be tested. For example, we captured risk preference by a questionnaire, hence self-perceptions, but we could implement a lottery [such as, Holt and Laury, 2014] or even methodologies from other disciplines [Mathieu et al., 2017], such as psychology. The same applies to the domain of expertise, which should be measured through channels other than the self-assessment of subjects.

Ultimately, this study underlines the need to study individual creative performance in light of the task type and the chosen creativity criteria since determinants vary across them. While the classical dichotomy between closed and open tasks from Guilford [1950, 1967] is the most favoured in the literature, we reiterate that a more precise declination is required since our third intermediary category of open-with-constraints tasks, as presented by [Attanasi et al., 2021], presents clear differences to the other two and can not just be associated with one or another.

These observations regarding individual creativity are obviously not to be detached from a broader picture of creativity. It should be remembered that this understanding of individual creativity is a critical element of understanding creativity at higher levels of observation, as is the case for work on team creativity. A team is more than just a collection of individuals; it is a blend of their unique contributions, with their interactions playing a central role in shaping its dynamics. The study of individual creativity to identify creative profiles thus enriches a collective approach to creativity as much as it will be enriched by it. What is essential, then, is to consider these levels of observation as interdependent elements, the joint study of which enables us to understand creativity in its entirety.

2.6 Appendices

Instructions

Draw task

For this task, you are asked to imagine and draw an animal from another planet very different from planet Earth.

You will need to draw this animal from the front and from the side. To complete your drawing, you will also need to write a short description of the animal and give it a name.

There are no other constraints for this exercise apart from those mentioned above. You must complete it in no more than 15 minutes.

Alternative Uses task

For this task, you are asked to write down on the sheets of paper provided all the original uses of a given everyday object. There are certainly many common and unusual ways of using such an object, so write down only the unusual, creative, and

uncommon uses you can think of.

To help you understand what is expected of you, let's take the example of a can of soda. Common uses would be "to contain liquid", "to serve as a glass", "to preserve food", etc. Whereas the uses for which you could be credited would be its "use like a flowerpot," "a lantern," "a windmill" (after cutting), "a telephone" (when two are connected by a wire), and so on.

Here, there will be no constraints in terms of the shape, number or size of the object.

The task consists of five words and will take 15 minutes. You will have three minutes per word to find as many unusual uses as possible. The words are as follows:

1. *A brick*
2. *A cardboard box*
3. *An extension lead*
4. *A metal pipe*
5. *A t-shirt*

Tangram task

For this task, we ask you to solve a tangram puzzle in the 15 minutes allotted. You have to make a square with all the pieces of wood at your disposal. As soon as you have done this, we ask you to raise your hand so that we can validate the resolution of the puzzle.

Questionnaire

You just participated in three creative tasks, and we would like your feedback on this experience through this questionnaire.

1. On a scale of 1 to 5, where 1 means "not creative at all" and 5 means "very creative," how would you rate your level of creativity during this

- experience?
- 1 2 3 4 5
- For one of the tasks you were given, we asked you to come up with as many unusual uses for everyday objects as possible.
2. Did you encounter any difficulties with this task? If so, what were they?
3. On a scale of 1 to 5, where 1 means “not creative at all” and 5 means “very creative,” how would you rate your level of creativity in this task?
- 1 2 3 4 5
4. Do you think you were more creative than the average participant? (Yes/No)
- For one of the tasks you were given, we asked you to draw an animal from another planet.
5. Did you encounter any difficulties with this task? If so, what were they?
6. On a scale of 1 to 5, where 1 means “not creative at all” and 5 means “very creative,” how would you rate your level of creativity in this task?
- 1 2 3 4 5
7. Do you think you were more creative than the average participant? (Yes/No)
- For one of the tasks you were given, we asked you to solve a Tangram puzzle.
8. Did you encounter any difficulties with this task? If so, what were they?
9. On a scale of 1 to 5, where 1 means “not creative at all” and 5 means “very creative,” how would you rate your level of creativity in this task?
- 1 2 3 4 5
10. Do you think you were more creative than the average participant? (Yes/No)

Code book

Variable	Definition
Age	Variable measuring subjects age. In regressions, age is defined as a binary variable (1 = above average, 0 = below average).
Diploma	Categorical variable determining subjects formation which takes value 0 if subject maximum formation is “ <i>High school diploma</i> ,” 1 if “ <i>Bachelor’s degree</i> ,” 2 if “ <i>Master’s degree</i> ,” and 3 if “ <i>Doctoral degree</i> .”
Doing new things	Variable measuring how often subjects are doing new things. Categorical variable taking value 0 if “ <i>Never</i> ,” 1 if “ <i>Rarely</i> ,” 2 if “ <i>Sometimes</i> ,” 3 if “ <i>Often</i> ,” and 4 if “ <i>Very often</i> .”
DOLE	Variable measuring “Diversity of Life Experience” (DOLE) questionnaire [Douthitt et al., 1999] which is a validated measure, and adapt it to the French/European context. We used the average of the 5 categories inside the DOLE, which are the following: (1) Experiencing different cultures through travel [EDCTT], (2) Diversity of interests, likes, and attitudes [DILA], (3) Diversity of geographic residence(s) [DGR], (4) Relationships with parents/family environment [RPFE] and, (5) General relations with others/friends [GR-WOF]. The values can be negative or positive, and the over value of the DOLE is an aggregate of all dimensions.
Domain	Categorical variable taking value 1 if “ <i>Sciences et technologies</i> ,” and 0 if “ <i>Economics and management</i> .”

Variable	Definition
Draw performance	Binary variable stating if individuals performed above or below average in the Draw task.
Expansivity performance	Binary variable stating if individuals performed above or below average in the Alternative Uses task for the specific criteria of Expansivity.
Expertise	Binary variable stating if a subject has specific knowledge in a specific area or not. Areas of expertise are visual arts, music, dance, individual sports, architectural design, entrepreneurial projects, creative writing, inventions and technologies, scientific inquiries, theatres and films, and culinary arts.
Foreign friends	Variable measuring how many foreign friends the participant has. Variable taking 0 if “ <i>None</i> ”, 1 if “ <i>One</i> ”, 2 if “ <i>Two to three</i> ”, 3 if “ <i>Fourth to six</i> ”, and 4 if “ <i>More than six</i> .”
Foreign places	Binary variable stating whether subjects often go to places frequented by foreigners. Takes 1 if “Yes” and 0 if “No”.
French speaker	Binary variable taking the value 1 for native French speakers and 0 otherwise.
Gender	Binary variable taking value 1 for woman and 0 for man.
Go to theatres and films	Variable measuring how often subjects go to theatres, cinemas and/or concerts. Categorical variable taking value 0 if “ <i>Never</i> ,” 1 if “ <i>Rarely</i> ,” 2 if “ <i>Sometimes</i> ,” 3 if “ <i>Often</i> ,” and 4 if “ <i>Very often</i> .”

Variable	Definition
Introvert	Variable measuring how subjects consider themselves as more introverted or extroverted on a scale from 1 to 5, where 5 is fully introverted.
Doing new things	Variable measuring how often subjects do new things. Categorical variable taking value 0 if “ <i>Never</i> ,” 1 if “ <i>Rarely</i> ,” 2 if “ <i>Sometimes</i> ,” 3 if “ <i>Often</i> ,” and 4 if “ <i>Very often</i> .”
Overconfidence	Binary variable stating if subjects expected to perform better than others on one specific task.
Persistency performance	Binary variable stating if individuals performed above or below average in the Alternative Uses task for the specific criteria of Persistency.
Preference to work in group setting	Dummy variable considering subjects’ preference to work in a group setting or not. It takes a value of 1 if “Group” or 0 if “Individual”.
Program	Dummy variable considering subjects’ studies program that takes value 1 for students with a training in creativity, and 0 no training in creativity.
Risk	Variable measuring risk attitude. It assumes a value from 1 (extremely risk averse) to 10 (extremely risk loving), where 1 indicates “ <i>You always avoid risk</i> ” and 10 indicates “ <i>You like to take risks</i> .”
Self-perceived general creativity	Variable giving the self-perceived creativity (in general) of subjects on a scale from 0 to 5, where 5 is the highest degree.

Variable	Definition
Self-perceived creativity	Binary variable giving the self-perceived creativity of subjects compared to the average of the sample, 1 if above and 0 if below. It has been implemented for the overall performance during the experiment but also for each task separately.
Social life	Variable which we approximated subjects' social life by gathering three related items in the literature: going out on a night out, meeting people, and going to bars/restaurants. Therefore, we did the average and this categorical variable taking value 0 if " <i>Never</i> ," 1 if " <i>Rarely</i> ," 2 if " <i>Sometimes</i> ," 3 if " <i>Often</i> ," and 4 if " <i>Very often</i> ." The Cronbach's alpha of 0.7329 support this measure as reliable and acceptable.
Spoken languages	Binary variable stating if a subject speaks more languages (besides French) than others on average. 1 if more, 0 if less.
Tangram task performance	Categorical variable stating if individuals performed above or below average, or failed, in the Tangram task.

Regressions

Table 2.9: Probit Regression for the Draw Task

	Model 1	Model 2	Model 3	Model 4
Gender	-0.177 (0.227)	-0.244 (0.233)	-0.0300 (0.279)	0.150 (0.294)
Age	-0.141 (0.357)	-0.0106 (0.377)	-0.134 (0.403)	-0.345 (0.445)
Spoken languages	0.220 (0.267)	0.341 (0.286)	0.169 (0.321)	0.147 (0.337)
French speaker	0.120 (0.265)	0.106 (0.280)	0.101 (0.302)	0.0266 (0.311)
Diploma (High school degree)				
<i>Bachelor's degree</i>	0.439 (0.798)	0.245 (0.849)	0.562 (0.893)	0.845 (0.982)
<i>Master's degree</i>	0.981 (0.972)	0.932 (1.023)	1.249 (1.023)	2.222** (1.104)
<i>Doctoral degree</i>	0.776 (0.922)	0.563 (1.000)	0.310 (0.968)	0.656 (1.013)
Doing new things		-0.0649 (0.193)	-0.0838 (0.217)	-0.0188 (0.222)
Go to theaters and films		-0.0926 (0.140)	-0.180 (0.154)	-0.154 (0.165)
Social life		0.204 (0.179)	0.266 (0.181)	0.365* (0.200)
Introvert		0.0519 (0.153)	0.0264 (0.162)	0.00657 (0.162)
Preference to work in group setting		0.0635 (0.246)	0.0400 (0.268)	0.134 (0.282)
Risk		-0.135** (0.0644)	-0.146** (0.0700)	-0.188** (0.0767)
Foreign friends		0.0257 (0.105)	-0.0362 (0.117)	-0.0820 (0.124)

Continued on next page

Table 2.9: Probit Regression for the Draw Task (continued)

	Model 1	Model 2	Model 3	Model 4
Foreign places		0.123 (0.253)	0.124 (0.278)	0.125 (0.318)
DOLE		0.0149 (0.0196)	0.0225 (0.0214)	0.0131 (0.0222)
Program			0.771 (0.529)	0.960 (0.622)
Expertise				
<i>Visual arts</i>			-0.326 (0.296)	-0.284 (0.302)
<i>Music</i>			-0.0255 (0.325)	0.0692 (0.349)
<i>Dance</i>			-0.518* (0.273)	-0.648** (0.277)
<i>Individual sports</i>			0.0853 (0.390)	-0.0906 (0.394)
<i>Architectural design</i>			0.200 (0.318)	0.266 (0.322)
<i>Entrepreneurial projects</i>			-0.250 (0.283)	-0.0882 (0.315)
<i>Creative writing</i>			-0.131 (0.287)	-0.122 (0.297)
<i>Inventions and technologies</i>			0.119 (0.256)	0.182 (0.291)
<i>Scientific inquiries</i>			0.212 (0.482)	0.254 (0.463)
<i>Theaters and films</i>			0.373 (0.307)	0.483 (0.317)
<i>Culinary arts</i>			0.399 (0.260)	0.307 (0.290)
Expansivity performance				0.345 (0.263)

Continued on next page

Table 2.9: Probit Regression for the Draw Task (continued)

	Model 1	Model 2	Model 3	Model 4
Persistency performance				0.0667 (0.265)
Tangram task performance (fail to solve it)				
<i>Low performance</i>				-0.615* (0.358)
<i>High performance</i>				-0.105 (0.367)
Self-perceived general creativity				-0.358 (0.349)
Self-perceived creativity (experiment)				0.283 (0.375)
Self-perceived creativity (Draw task)				0.888** (0.382)
Overconfidence (Draw task)				-0.457 (0.340)
Constant	-0.566 (0.862)	-0.494 (1.132)	-0.846 (1.232)	-1.476 (1.392)
Observations	135	135	135	134
R^2	0.0169	0.0621	0.1289	0.1876

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.10: Probit Regression for the Alternative Uses Task in Expansivity

	Model 1	Model 2	Model 3	Model 4
Gender	-0.341 (0.226)	-0.342 (0.237)	-0.484* (0.270)	-0.579** (0.290)
Age	-0.405 (0.356)	-0.747** (0.362)	-0.815** (0.400)	-0.989** (0.436)
French speaker	0.145 (0.236)	-0.00583 (0.256)	-0.205 (0.299)	-0.164 (0.315)
Diploma (High school degree)				
<i>Bachelor's degree</i>	-0.331 (0.827)	-0.333 (0.755)	0.751 (0.946)	0.810 (1.078)
<i>Master's degree</i>	-0.687 (1.007)	-0.717 (0.952)	0.739 (1.169)	0.959 (1.305)
<i>Doctoral degree</i>	0.169 (0.942)	0.490 (0.911)	2.318** (1.182)	2.570** (1.282)
Doing new things		-0.0178 (0.194)	-0.00244 (0.231)	0.0369 (0.266)
Going to theaters and films		-0.158 (0.146)	-0.163 (0.152)	-0.182 (0.172)
Social life		-0.563*** (0.186)	-0.648*** (0.190)	-0.654*** (0.218)
Introvert		-0.0687 (0.149)	-0.111 (0.157)	-0.0526 (0.170)
Risk		0.0395 (0.0630)	0.0398 (0.0701)	0.0172 (0.0779)
Preference to work in group setting		-0.0381 (0.256)	-0.0502 (0.278)	-0.00642 (0.293)
Foreign friends		-0.0327 (0.115)	-0.0286 (0.130)	-0.0215 (0.141)
Foreign places		-0.0584 (0.260)	0.0609 (0.288)	0.156 (0.315)
DOLE		0.0277 (0.0199)	0.0273 (0.0221)	0.0242 (0.0255)

Continued on next page

Table 2.10: Probit Regression for the Alternative Uses Task in Expansivity
(continued)

	Model 1	Model 2	Model 3	Model 4
Program			1.019 (0.639)	1.193* (0.630)
Domain			-1.681* (0.874)	-2.206** (0.861)
Expertise				
<i>Visual arts</i>			-0.259 (0.294)	-0.142 (0.324)
<i>Music</i>			0.178 (0.344)	0.150 (0.375)
<i>Dance</i>			0.163 (0.270)	0.284 (0.287)
<i>Individual sports</i>			0.402 (0.371)	0.253 (0.385)
<i>Architectural design</i>			-0.384 (0.333)	-0.566 (0.347)
<i>Entrepreneurial projects</i>			-0.298 (0.287)	-0.212 (0.310)
<i>Creative writing</i>			0.430 (0.287)	0.477 (0.297)
<i>Inventions and technologies</i>			0.253 (0.267)	0.490* (0.296)
<i>Scientific inquiries</i>			-0.397 (0.510)	-0.520 (0.509)
<i>Theaters and films</i>			-0.216 (0.300)	-0.342 (0.319)
<i>Culinary arts</i>			-0.117 (0.256)	-0.170 (0.293)
Draw performance				0.281 (0.267)
Tangram task performance (fail to solve it)				

Continued on next page

Table 2.10: Probit Regression for the Alternative Uses Task in Expansivity
(continued)

	Model 1	Model 2	Model 3	Model 4
<i>Low performance</i>				-0.0907 (0.390)
<i>High performance</i>				-0.337 (0.362)
Self-perceived general creativity				-0.326 (0.365)
Persistency performance				0.112 (0.270)
Self-perceived creativity (experiment)				0.167 (0.353)
Self-perceived creativity (AUT)				0.448 (0.333)
Overconfidence (AUT)				-0.901** (0.390)
Constant	0.490 (0.875)	2.077** (1.027)	1.441 (1.245)	1.231 (1.476)
Observations	135	135	135	134
R^2	0.0309	0.1150	0.1753	0.2321

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.11: Probit Regression for the Alternative Uses Task in Persistency

	Model 1	Model 2	Model 3	Model 4
Gender	-0.205 (0.229)	-0.217 (0.246)	-0.177 (0.288)	-0.233 (0.297)
Age	0.533 (0.368)	0.316 (0.431)	0.558 (0.478)	0.542 (0.517)
Spoken languages	-0.0182 (0.236)	-0.118 (0.265)	-0.270 (0.308)	-0.183 (0.330)
Diploma				
<i>Bachelor's degree</i>	0.760 (0.489)	0.280 (0.564)	0.0593 (0.895)	0.212 (0.869)
<i>Master's degree</i>	-0.135 (0.678)	-0.685 (0.678)	-0.961 (0.768)	-0.822 (0.800)
<i>Doctoral degree</i>		-0.337 (0.206)	-0.217 (0.232)	-0.189 (0.247)
Go to theaters and films		-0.269* (0.146)	-0.386** (0.156)	-0.351** (0.171)
Social life		-0.611*** (0.186)	-0.616*** (0.191)	-0.678*** (0.211)
Introvert		0.212 (0.149)	0.172 (0.156)	0.209 (0.161)
Risk		-0.0765 (0.0639)	-0.0804 (0.0735)	-0.0762 (0.0806)
Preference to work in group setting		-0.223 (0.251)	-0.199 (0.286)	-0.292 (0.291)
Foreign friends		-0.0662 (0.108)	-0.0164 (0.119)	-0.0240 (0.121)
Foreign places		-0.110 (0.247)	-0.0547 (0.278)	-0.0508 (0.288)
DOLE		0.0549*** (0.0202)	0.0517** (0.0212)	0.0624*** (0.0229)
Program			-6.205*** (0.791)	-5.593*** (0.924)

Continued on next page

Table 2.11: Probit Regression for the Alternative Uses Task in Persistency
(continued)

	Model 1	Model 2	Model 3	Model 4
Domain			5.917***	5.372***
			(1.082)	(0.611)
Expertise				
<i>Visual arts</i>			-0.224	-0.0570
			(0.331)	(0.369)
<i>Music</i>			0.127	-0.0343
			(0.360)	(0.371)
<i>Dance</i>			-0.180	-0.0985
			(0.280)	(0.285)
<i>Individual sports</i>			0.164	0.00898
			(0.383)	(0.404)
<i>Architectural design</i>			0.0185	-0.142
			(0.322)	(0.346)
<i>Entrepreneurial projects</i>			-0.297	-0.311
			(0.292)	(0.314)
<i>Creative writing</i>			0.263	0.288
			(0.337)	(0.361)
<i>Inventions and technologies</i>			-0.347	-0.291
			(0.279)	(0.309)
<i>Scientific inquiries</i>			0.370	0.0410
			(0.468)	(0.533)
<i>Theaters and films</i>			0.0813	0.0198
			(0.339)	(0.349)
<i>Culinary arts</i>			-0.125	-0.0641
			(0.282)	(0.291)
Draw performance				0.0691
				(0.265)
Tangram task performance (fail to solve it)				
<i>Low performance</i>				0.0716
				(0.368)

Continued on next page

Table 2.11: Probit Regression for the Alternative Uses Task in Persistency
(continued)

	Model 1	Model 2	Model 3	Model 4
<i>High performance</i>				-0.265 (0.356)
Self-perceived general creativity				-0.0359 (0.336)
Expansivity performance				0.0494 (0.281)
Self-perceived creativity (experiment)				0.0662 (0.377)
Self-perceived creativity (AUT)				-0.210 (0.352)
Overconfidence (AUT)				-0.725 (0.450)
Constant	-0.598 (0.566)	1.987** (0.958)	2.177* (1.160)	2.394* (1.256)
Observations	132	132	132	131
R^2	0.0252	0.1394	0.2237	0.2494

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.12: Ordered Logit Regression for the Tangram Task

	Model 1	Model 2	Model 3	Model 4
Gender	0.111 (0.201)	0.0424 (0.209)	0.0232 (0.279)	-0.00980 (0.312)
Age	-1.176*** (0.338)	-1.121*** (0.344)	-1.571*** (0.448)	-1.670*** (0.578)
Spoken languages	0.163 (0.215)	0.290 (0.227)	0.502* (0.267)	0.607** (0.282)
Diploma				
<i>Bachelor's degree</i>	-0.236 (0.918)	0.271 (0.969)	1.446 (1.079)	2.839** (1.152)
<i>Master's degree</i>	-0.776 (1.233)	-0.479 (1.249)	-0.727 (1.398)	0.946 (1.752)
<i>Doctoral degree</i>	0.433 (1.009)	0.823 (1.098)	0.146 (1.183)	1.590 (1.526)
Doing new things		0.143 (0.173)	0.0584 (0.193)	0.0253 (0.213)
Going to theaters and films		-0.121 (0.125)	-0.105 (0.142)	-0.0233 (0.154)
Social life		0.0232 (0.154)	0.000331 (0.157)	0.0250 (0.212)
Introvert		0.182 (0.148)	0.230 (0.169)	0.197 (0.178)
Risk		-0.0529 (0.0575)	-0.0802 (0.0590)	-0.150** (0.0686)
Preference to work in group setting		-0.331 (0.217)	-0.178 (0.256)	0.0143 (0.273)
Foreign friends		0.0948 (0.0884)	0.0226 (0.104)	0.0156 (0.123)
Foreign places		0.206 (0.233)	0.374 (0.251)	0.543** (0.257)
DOLE		0.0110 (0.0186)	0.0240 (0.0194)	-0.0186 (0.0215)

Continued on next page

Table 2.12: Ordered Logit Regression for the Tangram Task (continued)

	Model 1	Model 2	Model 3	Model 4
Program			0.943*	1.582***
			(0.548)	(0.556)
Domain			1.354**	0.960
			(0.650)	(1.059)
Expertise				
<i>Visual arts</i>			-0.645**	-0.544*
			(0.305)	(0.303)
<i>Music</i>			0.192	0.203
			(0.337)	(0.333)
<i>Dance</i>			-0.0264	-0.0230
			(0.276)	(0.288)
<i>Individual sports</i>			-0.385	-0.551
			(0.389)	(0.356)
<i>Architectural design</i>			-0.0361	-0.0962
			(0.258)	(0.280)
<i>Entrepreneurial projects</i>			0.140	0.355
			(0.258)	(0.300)
<i>Creative writing</i>			0.438*	0.275
			(0.262)	(0.289)
<i>Inventions and technologies</i>			0.598**	0.470*
			(0.233)	(0.274)
<i>Scientific inquiries</i>			-0.917*	-0.944*
			(0.500)	(0.485)
<i>Theaters and films</i>			0.0388	0.176
			(0.310)	(0.301)
<i>Culinary arts</i>			0.0289	-0.0687
			(0.252)	(0.273)
Draw performance				0.00435
				(0.262)
Self-perceived general creativity				0.227
				(0.329)
Expansivity performance				-0.427*

Continued on next page

Table 2.12: Ordered Logit Regression for the Tangram Task (continued)

	Model 1	Model 2	Model 3	Model 4
				(0.253)
Persistency performance				-0.0866
				(0.276)
Self-perceived creativity (experiment)				0.472
				(0.309)
Self-perceived creativity (Tangram)				1.734***
				(0.273)
Observations	135	135	135	135
R^2	0.0670	0.1014	0.2008	0.3730

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Diversity, Interactions, and Team Creativity: An Experimental Perspective

This chapter was co-authored with

Sara GIL-GALLEN and Patrick LLERENA

Summary of the chapter

While team creativity has long been a subject of interest to researchers, few articles have been developed on the joint importance of diversity and interpersonal interactions on team creative performance. This study contributes to the literature by examining the role of diversity in team performance and unpacking networks' complex effects by distinguishing between in-situ and pre-existing networks. In addition, we explore the relationship between individual and team performance when both are measured experimentally. To do so, we implement a three-stage protocol allowing subjects' interactions without any form of intermediary. The final sample comprises 135 subjects for a total of 35 teams, and the findings reveal that the impact of diversity and networks on team creativity is contingent on the specific creativity criteria considered as either originality or feasibility.

3.1 Introduction

It was long believed that individuals pursued creativity on their own [John-Steiner, 2006]. Early research in this field echoed this view, concentrating predominantly on individual aspects of creativity [Barron and Harrington, 1981, Coldevin et al., 2018]. However, during the last decades, the literature suggests that creativity is inherently a collective endeavour, significantly shaped by interpersonal interactions. To grasp the creative phenomenon as a whole, this requires the study of both individual and team levels of creativity. Therefore, if creativity is defined as the generation of novel and useful ideas by individuals or teams [Amabile and Pratt, 2016], team creativity is one way for these ideas to emerge.

In addition, as innovations result from the successful implementation of novel and useful ideas, creativity becomes essential for innovation [Amabile et al., 1996, Sarooghi et al., 2015]. And so, given that innovation is central to organisations' performance - providing significant competitive advantages or simply survival capacities [Chatzoglou and Chatzoudes, 2018, Rousseau et al., 2015] - the crucial role of creativity for organisations should not be understated [Devanna and Tichy, 1990, Woodman et al., 1993]. Creativity, as a social and collective process, grows in importance as more and more activities within organisations are conducted in teams [Mathieu et al., 2017], which drives us to consider the profiles of team members and their interactions to understand what leads to higher team creative performance.

Notably, teams are more than just the sum of their members; they are “*inter-dependent collections of individuals who share responsibility for specified outcomes*” [Sundstrom et al., 1990, p. 120]. However, research on team diversity has been ongoing for years and findings are mixed, seeing diversity as both an opportunity and a threat to team performance. While “*contact with those who see the world differently is a logical prerequisite to seeing it differently ourselves*” [Kanter, 1996, p. 98], too great a distance between individuals can become an unmanageable challenge for working together. Moreover, the literature on the social dynamics and interactions between team members demonstrates their importance as they influence how these team members behave and how they perform. In the end, few studies, even more so in experimental economics, combine both team diversity and interactions to understand team creative performance.

As a result, this study enhances our understanding of team creativity in three ways: it examines the role of diversity in team performance, unpacks the complex

effects of networks on team performance by distinguishing between in-situ and pre-existing networks, and explores the relationship between individual and team performance when both are measured experimentally. Beyond this empirical contribution, we also want to emphasize the novelty of our work in terms of methodology. Especially since the protocol presented in this article is based on the specific case of collective experiments, “*experiments enabling the observation of collective processes by means of the social interactions of subjects without a systematic intermediary*”, which have never truly been conducted by experimentalists in economics, even less so in the case of a multi-stage protocol.

Since this study investigates the dynamics of team creativity, particularly the impact of diversity and social networks on it, using a multilevel approach, this leads us to formulate two main research questions. Firstly, to what extent does diversity in terms of team composition impact team creative performance? Secondly, how might individuals’ interactions, through pre-existing or in-situ networks, affect team creative performance?

To address these questions, a three-stage protocol has been implemented to gather comprehensive information on our subjects and measure creative performance at individual and team levels. Our final sample comprises 135 subjects divided into 35 teams, and the following study will mainly focus on the third stage of our protocol, the collective experiment. The findings highlight that the relationship between diversity, networks and team creativity strongly relies on the creativity criteria we focus on. Besides diversity and networks, our study also provides insights into the significant impact of creativity training, individual creative performance, cross-criteria performance, behavioural traits, and creative skills on team creative performance.

The remainder of this work is organized as follows: Section 3.2 reviews the relevant literature and key concepts. Section 3.3 details the experimental design and procedure. Section 3.4 presents the evaluation process, regression analysis, and findings of our study. Section 3.5 discusses the implications, and Section 3.6 offers concluding remarks.

3.2 Literature review

3.2.1 Individual and team creativity

For decades, creativity has gained more and more attention among diverse fields of research, including economics and management. Authors have contributed to this extensive body of literature by providing a wide array of definitions of this concept [Boden, 2004, Mednick, 1962, Ochse, 1990, Poincaré, 1908]. In this work, we rely on the definition proposed by Amabile and Pratt [2016], who characterize creativity as “*the production of novel and useful ideas by an individual or small group working together*” (p. 158). As mentioned earlier, while creativity has long been viewed as an individual process, it is also a collective pursuit. Indeed, creativity is a social process, and it is imperative to distinguish between individual and team creativity, as the latter is not solely a summation of the former. Instead, team creativity emerges from intricate and multifaceted social interactions among individuals and their environment [Amabile and Pratt, 2016, Csikszentmihalyi, 1997, Fischer et al., 2005, Hennessey and Amabile, 2010]. One thing to emphasize is how we define team creativity, which can be understood in many ways. Throughout this article, we will refer to the team creativity process as simultaneous co-creation [Fischer et al., 2005], which points to individuals “*jointly creating something at the same time*” (p. 485).

This raises the question of the method used to study creativity, both individual and collective. We employ the experimental approach and, more specifically, a collective experiment protocol to address this. It should be noted that a collective experiment is defined in our work as “*an experiment enabling the observation of collective processes by means of the social interactions of subjects without a systematic intermediary*”, where an intermediary is defined as “*the deliberate virtual or physical separation of subjects, which creates a sense of distance and artificially restricts their potential interactions*”, e.g., the computer interface that is widely used in experimental protocols. In essence, within this specific protocol, we aim to observe, in a controlled setting, face-to-face interactions among individuals who are part of a team and collaborate to generate a collective creative outcome. The term *collective experiment* will consistently refer to this definition throughout this work.

If we look at the already established experimental literature and its application to team creativity and collective experiments, it appears that very little work has been carried out on this subject. Meanwhile, several research articles have been published

on competition [Eckartz et al., 2012, Boudreau and Lakhani, 2011, Bradler, 2015, Bradler et al., 2016], cooperation [Baloche, 1994, Charness and Grieco, 2021, Chen et al., 2012], and coopetition [Elmoukhli, 2018, Zhao et al., 2016] processes, far too little attention has been paid to the co-creation aspect of creativity. It goes without saying that other disciplines, such as psychology, have explored creativity and developed experimental protocols involving multiple subjects from a co-creation perspective. And so, our observation of a scarcity of literature on the collective aspect of creativity in experimentation primarily holds for experimental economics. This gap is significant, given that methodological approaches and research questions vary widely across disciplines, leading to diverse and sometimes conflicting conclusions.

Back to the study of individual and team creativity, including outside the experimental field, we have decided to focus on the issue of creative performance. To address this, one should first distinguish the type of creativity involved. Here, we differentiate two facets of creativity, namely *ideation* and *implementation*. On the one side, creativity can be defined as having novel ideas. We consider it as creativity *per se*. On the other side, creativity can be considered as problem-solving oriented, requiring the implementation dimension in its evaluation, as ideas should be novel but also useful. This implementation-oriented creativity does not aim for a high quantity of ideas but instead finding the right idea to solve the problem. In the end, novelty and usefulness go hand in hand, and studying them jointly is necessary [Nijstad et al., 2010, Rietzschel et al., 2010].

In addition, attention has been paid to the link between individual and collective creative performance as a cross-level analysis. It might appear obvious that teams become more creative as their members demonstrate higher levels of individual creativity. However, upon closer examination, the relationship between individual and team performance is not trivial. There are two main perspectives on this relationship: the additive and the disjunctive model of team creativity [Triandis et al., 1963, van Knippenberg, 2017, Yuan et al., 2022]. On the one hand, the additive model considers team members to be a set of creative assets and that their combination will determine the team's performance as a whole. Each individual brings some knowledge, skills, or aptitudes to the team, but the discrepancy between their own individual performance does not really matter. What matters in the end is the sum of individuals' creativity. On the other hand, the disjunctive model considers that a team's creative performance is determined by the team member with the higher individual performance. In this sense, the most creative team member will lead the

team’s performance. While this literature has been widely studied, researchers found mixed evidence for both models [Yuan et al., 2022], requiring a stronger focus on the related moderators, mainly team characteristics, team atmosphere, and task characteristics.

However, creative performance also requires a specific focus on the elicited thinking processes. Two primary thinking processes stand out in the literature on creativity: convergent thinking and divergent thinking. Convergent thinking focuses on deriving a singular logical solution, whereas divergent thinking targets generating multiple logical solutions [Guilford, 1950, 1967]. Consequently, creative tasks may trigger different thinking processes or several at the same time. Drawing from this, creative tasks vary based on their degree of openness, characterized by individuals’ autonomy in solving them. This spectrum extends beyond the traditional *closed* versus *open* task dichotomy, encompassing a third category: tasks defined as *open with constraints* [Attanasi et al., 2021]. This is all the more important given that ideation is often linked with a divergent thinking process, while implementation is often linked with convergent thinking. However, these two types of thinking processes can not be totally separated, and most creative tasks require a combination of both, even if these tasks can be more divergent or convergent-thinking oriented [Guilford, 1950, 1967].

This distinction between different types of tasks also emphasizes that evaluating creative outputs can vary depending on the chosen creativity criteria. Creativity being multidimensional, it can not be assessed without comprehensive consideration. Guilford [1950] highlighted three primary criteria for assessing creativity: originality (*how infrequent a particular solution is*), fluency (*how many ideas were generated to solve a specific problem*), and flexibility (*how many themes cover the set of ideas generated*). What truly matters is the interconnectedness of these criteria, which should never be examined in isolation to prevent overlooking a significant aspect of the narrative underlying creative performance. Moreover, when we look at creativity at different levels, such as individual and team creativity, we need comparable measures and cross-level criteria [Hundschell et al., 2021].

So, what form should the evaluation of creative performance take? First, in a cross-level analysis context, a large body of literature decides to measure individual performances through questionnaires and scales instead of using experimental results of tasks effectively performed by subjects. Even though these questionnaires are recognized and validated [Baer and Oldham, 2006, Karwowski et al., 2018, Zhou and

George, 2001], they are mainly oriented towards divergent thinking measures [Stolaki et al., 2023]. Moreover, these questionnaires rely on the self-assessed evaluation of creativity, which by nature might be biased [Plucker et al., 2019]. To fully capture individual performance, researchers need to effectively measure it through different tasks. This implies that, in an individual-to-collective creative performance logic, subjects would perform individual and collective tasks within the same experimental protocol.

Finally, creativity assessment can be either objective [Simonton, 1999] or subjective [Amabile, 1982]. Thus, the remaining pivotal aspect after defining the task type and creativity criteria is the nature of the evaluator responsible for assessing the creativity of an output. Objective assessments do not necessitate specific guidelines or designated evaluators. Conversely, subjective evaluations, often preferred for open-ended tasks, demand the involvement of evaluators. Within the literature, two predominant options exist: peer evaluations (subjects from the same experiment) or external evaluators. In this work, we mainly use subjective measures of creativity, opting for the external judge option that guarantees lower biases in the creativity assessment.

3.2.2 Diversity

Heterogeneity, multiplicity, and variety are different terms to discuss diversity in the literature. Even though everybody has a clear idea of diversity, finding its unique and consensual definition in the literature is quite difficult. In this work, we rely on the definition provided by van Knippenberg et al. [2004], who defines diversity as “*differences between individuals on any attribute that may lead to the perception that another person is different from self*” (p. 1008). When individuals need to collaborate with others on a specific idea, task, or project, they need to adapt to everyone’s specificities. Diversity “*appears to be a double-edged sword, increasing the opportunity for creativity as well as the likelihood that group members will be dissatisfied and fail to identify with the group*” [Milliken and Martins, 1996, p. 403]. Several authors have defended the benefits of diverse teams [Amabile et al., 1994, Dutcher and Rodet, 2022, Kavadias and Sommer, 2009, Williams Phillips and O’Reilly, 1998]. However, empirical studies tend to show a more complex and mixed effect [Steiner, 1972]. In the end, the literature presents no clear and unquestionable consensus on whether diversity positively or negatively impacts teams’ performance and functioning because such results differ as long as we modify the types of diversity and the

context of the study [Spickermann et al., 2014]. Thus, on what variable of diversity should we focus? And if we consider several of them, how do we classify them?

If diversity is about individuals' specificities or characteristics, it is essential to differentiate their observational levels. We choose to distinguish the *surface* from the *deep-level* of observation. On the one hand, at the surface-level, variables respect three principles: immutability, immediate observation, and simple and valid measurement [Jackson and Wolinsky, 1996, Milliken and Martins, 1996]. We can consider surface variables such as age, gender, or even ethnicity. Harrison et al. [1998] used the term *heterogeneity at a surface level* interchangeably and defined it “*as differences among group members in overt, biological characteristics that are typically reflected in physical features.*” (p. 97). Beyond biological or physical features, authors also include a broader set of variables [Hundschell et al., 2021, Spickermann et al., 2014, Van et al., 2004], e.g. individuals' occupation. On the other hand, “*heterogeneity at a deep level includes differences among members' attitudes, beliefs, and values. Information about these factors is communicated through verbal and nonverbal behavior patterns and is only learned through extended individualized interaction and information gathering.*” [Harrison et al., 1998, p. 98]. These characteristics include, not exhaustively, personality traits, moods, or even life experiences. If so, deep-level variables can not verify the three concepts of immutability, immediate observation, or simple and valid measurement. As a result, measures become more challenging due to the complexity in collecting such information on individuals. However, even if we distinguish the surface from deep-level for diversity, it does not mean that these two categories should be analyzed separately. In fact, both are interrelated and are required to be studied simultaneously, even though the ways in which they are measured are different. For us, studying one without the other would always imply losing one side of the story.

The extensive literature on diversity remains inconclusive regarding its impact, whether positive or negative, on performance [Bell et al., 2011, Hundschell et al., 2021, van Dijk et al., 2012]. This literature outlines two conflicting facets of teamwork: a positive influence on decision-making, creativity, or performance versus a potential negative influence as reduced social integration [Stahl et al., 2010] or heightened conflicts [Jehn et al., 2010, Pelled et al., 1999]. This complex impact of diversity is also reflected in the nature of the relationships observed between the focus variables and the outcome. Researchers confirm a generally observed inverted U-shaped relationship between diversity and performance [Bahlmann, 2014, Cordero et al.,

1996, Godart et al., 2015, Huang and Chen, 2010, Li et al., 2016]. Ultimately, there is a lack of literature on diversity, especially when combined with creativity, within the context of a collective experiment, and we aim to address this gap.

Now that we have underscored the importance of diversity as a determinant of teams' performance and social dynamics, we still need to dig into the methodologies available for its measurement. Although many metrics have been proposed, scholars engage in an ongoing debate regarding the suitability of specific tools for particular analytical contexts [Harrison et al., 1998, Solanas et al., 2012]. In this discussion, we narrow our focus to two measures that offer distinct yet complementary analytical advantages. The first measure is the widely recognized Blau Index, which "*quantifies the probability that two members randomly selected from a population will be in different categories if the population size is infinite or if the sampling is carried out with replacement*" [Solanas et al., 2012, p. 7]. Thus, greater diversity is indicated by a higher Blau index. While the Blau Index provides insights into attributes' proportions within a team, it is important to note that it limits this observation to only one attribute at a time. Consequently, we need to complement the Blau Index with an alternative approach to gain a more comprehensive understanding of diversity within the team. This alternative category of measures relies on the notion of faultlines, which are "*hypothetical dividing lines that may split a group into subgroups based on one or more attributes*" [Lau and Murnighan, 1998, p. 328]. Adopting the concept of faultlines enables the consideration of multiple variables and their interrelations. Particularly, the concept of faultlines gives the capacity to quantify their strength, wherein "*faultlines become stronger as more attributes align themselves in the same way*" [Lau and Murnighan, 1998, p. 325]. Consequently, stronger faultlines imply greater homogeneity within the resulting divisions or subgroups. This approach enhances our ability to gauge the complexity of diversity dynamics within the group, transcending the limitations of one-dimensional attribute comparisons¹.

3.2.3 Social Interactions and Networks

As a social phenomenon, creativity demands that researchers understand the complex environment in which the creative process takes place, considering both the individuals involved and the interactions among them. As Brass [1995] stated, "*No*

¹See Appendices 3.7 for the computation of both diversity measures. Of note, our regression specifications only include the Blau Index, but a complementary analysis, provided in Appendices 3.7, also shows the results for the faultlines measures.

one is creative alone. You and I are only as creative as our social networks” (p. 94). As a result, it is imperative to delve deeper into the relationships and connections between individuals to unravel the complexity of their behaviours [Perry-Smith and Shalley, 2003]. Social network analysis has empowered researchers to dig into multifaceted behavioural phenomena, spanning from team cohesion to the emergence of conflicts, dynamics of trust, and beyond [Wechtler et al., 2023].

Additionally, besides these interactions between team members during the creative process, what also captivates our attention is the dynamics behind the emergence of new ideas. Creativity is about (re)combining multiple cognitive elements when these are pertinent to the problem at hand [Simonton, 1999]. And so, social interactions facilitate a greater number of possible combinations, elevating the probability of generating new ideas, by allowing information sharing between members.

This relationship between team members’ interactions and idea generation can not be fully detached from the team diversity mentioned earlier. Since diversity is often associated with a greater number of perspectives that contribute to the novelty of ideas, it can also have a detrimental impact on a team’s capacity to combine new elements and generate ideas since it might alter the quality of interactions, e.g., impeding team coordination Pollok et al. [2021].

While the existing literature highlights that an idea is a social construct, emphasizing the importance of individuals’ interactions in comprehending the flow and elaboration of ideas, it is crucial to put into perspective the interactions that emerge between members of the same team and the composition of that team. Our work goes in line with some authors’ reflections on the future of the study of team processes with the need to pursue research on the link between social networks and teamwork [Mathieu et al., 2017] or the combination of social network analysis and experimental economics presented as a comprehensive and rigorous approach to study such collective processes [Noh, 2022]. However, disentangling the role of team dynamics and direct social interactions within the creative process creates a challenge for experimentalists in economics, as the traditional recommendation on the proper conduct of an experiment often refers to the minimization or avoidance of face-to-face interactions — something that yet remains undeniable in real organisational practices.

3.3 Experimental design

3.3.1 A three-stage protocol

We collected our experimental data through a three-stage protocol involving a questionnaire session, an individual experiment, and a collective experiment. This protocol required subjects to visit the laboratory on three separate occasions². While the first phase of the protocol consisted of a comprehensive questionnaire on each subject's profile encompassing sociodemographic information, creativity, and social habits, as well as Diversity of Life Experiences [Douthitt et al., 1999] and domains of expertise [Carson et al., 2005, Dutcher and Rodet, 2022], the second phase comprised three individual creative tasks performed by each subject. Each task corresponds to a certain degree of task openness: the *Tangram Task* (as a closed task), the *Alternative Uses Task* (as an open with constraints task), and the *Draw Task* (as an open task). These three tasks were given to subjects to assess their individual creative performance as a reference point to provide us with a comparison of both individual and collective performance between the second and third stages of this protocol³.

In the third stage of the protocol, subjects participated in a collective experiment to evaluate the creative performance of teams composed of three to five subjects. In this experiment, subjects were tasked with developing and presenting an idea or project that could enhance their daily university life [Baruah and Paulus, 2011, Harvey, 2013, Kohn and Smith, 2011, Kohn et al., 2011]. The only instruction provided was that the most creative idea would be the one that deviated the most from existing university practices and also distinguished itself from ideas proposed by other teams. The collective experiment consisted of four distinct steps:

1. An *individual phase* during which subjects were asked to think about the problem by themselves. We provided subjects with scrap paper, and all subjects sat on their own with no possibility of talking to each other. This phase lasted 10 minutes, and the room's configuration was as displayed in *Layout A* of Figure 3.1.
2. A *strategic phase* where subjects were allowed to talk to other subjects under

²For control purposes, part of the subjects followed this sequence, while the other segment underwent the collective experiment before the individual one. This change was implemented to control for the possibility of any order effect.

³More details about the first two stages of the protocol and the instructions for the three individual tasks are available in Chapter 2 of this thesis.

certain conditions. Each subject received a set of four tokens. Imagine now that subject A and subject B are talking to each other. If A wants information about B's ideas, he has to give him a token (as payment for the information). Conversely, if B wants information about A's ideas, he has to give him a token. Depending on their strategy, subjects can refuse to receive a token (as a refusal to give information) and/or spend all of their tokens, part of them, or none. Of note, subjects could only use the tokens from their initial endowment but not those they got from other subjects in exchange for their ideas. This part lasted 15 minutes, and the room's configuration was as displayed in *Layout B* of Figure 3.1.

3. A *collective phase* where each team had its own area with a table, chairs, and a whiteboard. They could move and organize themselves as they wanted in this area. This step aimed to allow team members to exchange their ideas and converge toward one. It is also mentioned in the instructions that each team had to prepare a short presentation after converging on a specific idea. This part lasted 35 minutes, and the room's configuration was as displayed in *Layout C* of Figure 3.1.
4. A *presentation session* as a 5-minute speech exercise. Each team is completely free on the presentation form besides the time constraint. Each team then presented its idea for five minutes to the other teams, who were subsequently asked to evaluate their own performance in comparison to the others. To maintain control, teammates were not allowed to communicate with one another during this phase, preventing the order of presentations from influencing the other projects. The room's configuration is as displayed in *Layout C* of Figure 3.1.

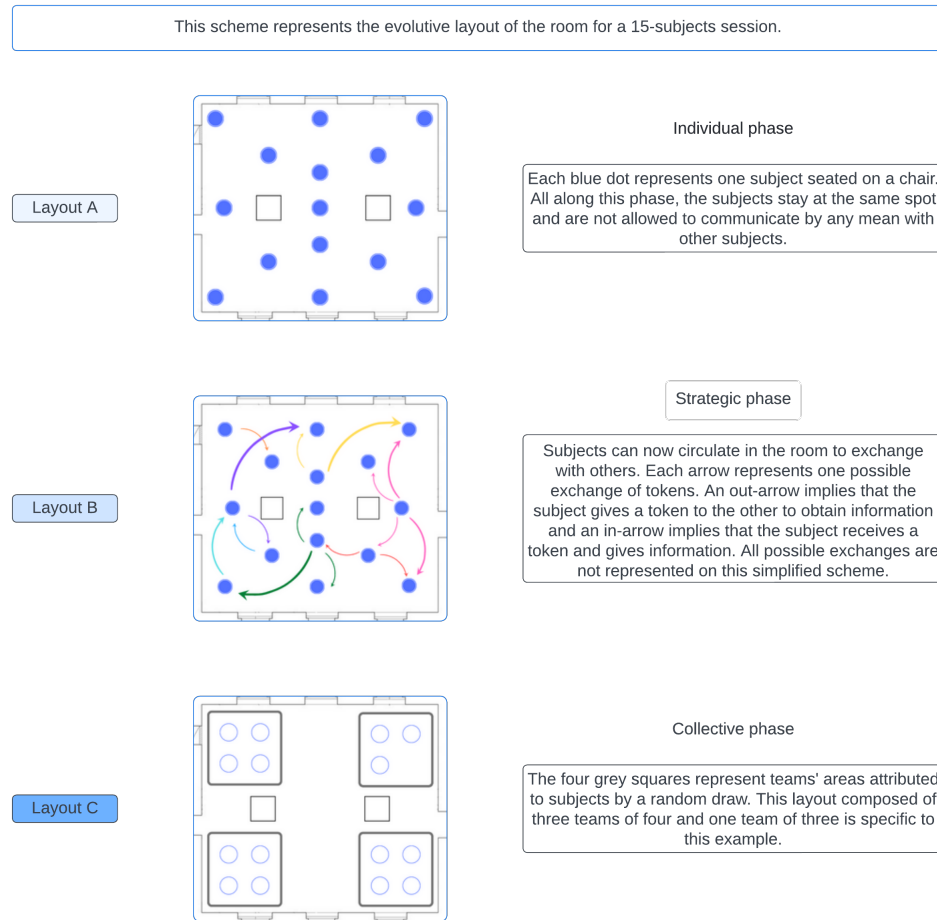
During the collective experiment, subjects were administered three distinct questionnaires to obtain information on their strategies, team experience, and pre-existing networks (see Appendices 3.7 for the detailed questionnaires).

Of note, an average of three months separated the initial from the third stage of the protocol. This deliberate spacing was implemented to minimise any potential effects of one stage on the others.

3.3.2 Experimental procedure

We recruited subjects from the University of Strasbourg (France). Our subject selection process involved two distinct categories of students depending on their back-

Figure 3.1: Experimental Layouts



ground specificities regarding creativity. We first selected a sample of Master's students majoring in economics. These students were chosen because their academic program did not emphasize or provide specific creativity skills or techniques training. For our second sample, we composed a group of students specializing in innovation management and student entrepreneurs⁴ as they all received specific training in creativity. The only difference between these two types of subjects benefiting from

⁴In the literature, students-entrepreneurs are defined as individuals carrying a venture creation activity besides their study activity [Bergmann et al., 2016, Longva, 2021, Nielsen and Gartner, 2017]. In our study, when we use the term of *students-entrepreneurs*, we refer to the specific status of French universities that was conceived as support providing credibility, security, and visibility to students' entrepreneurial projects.

creativity training is their actual involvement in entrepreneurial activities, as student entrepreneurs are working on a specific entrepreneurial project while others do not. In the end, we organised experimental sessions with teams that had not received creativity training and those that had, allowing us to clearly differentiate the effect of this training on team performance. Moreover, our reason for selecting students from the same academic programs rather than a random sample of subjects was to ensure a baseline level of network connectivity among subjects. While it was not imperative for all subjects to be acquainted, some degree of pre-existing connection was interesting to make networking strategies relevant and meaningful. Table 3.1 summarizes the characteristics of the academic programs from which our final sample of 135 subjects was drawn.

Table 3.1: Subjects Distribution by Specific Background

<i>Programs</i>	<i>N</i>	<i>Creativity</i>	<i>No creativity</i>
Economics	110		X
Innovation management and student-entrepreneurs	25	X	
Total	135		

The experimental sessions were conducted at the University of Strasbourg in the *Social Interactions Lab* (SIL). All experiments were conducted in French, and we organized a total of 13 sessions, with the following distribution of teams: 9 teams comprising 3 members, 22 teams comprising 4 members, and 4 teams comprising 5 members. Table 3.2 provides the descriptive statistics of our sample regarding socio-demographic information.

Finally, we adopted a flat payment system on the payment structure, meaning that subjects were paid a fixed amount rather than a performance-based or competition-based incentive. This choice was made after considering various factors, including the nature of the tasks and the mixed findings in the literature on the impact of incentives for creative tasks. While performance-based payments can be advantageous for closed tasks, they may have no or even a detrimental effect on open tasks. Amabile et al. [1996] has highlighted a crowding out effect of monetary incentives, which can diminish subjects' intrinsic motivation and, consequently, harm

Table 3.2: Summary of the Descriptive Statistics

Variables	N	mean	s.d	min	max
team	135	17.259	10.290	1	35
Gender	146	0.438	0.498	0	1
Age	146	23.123	3.961	19	45
Program	146	0.178	0.384	0	1
Languages	146	2.178	0.915	0	5
French speaker	146	0.651	0.478	0	1
French nationality	146	0.609	0.490	0	1

their creative performance. Eckartz et al. [2012] also concluded that performance-based payment had no significant effect on subjects' creative outcomes compared to flat payment. To maintain consistency and mitigate any potential monetary effect, we chose to apply the same payment scheme across all tasks. The payment structure for this experimental protocol is detailed in Table 3.3.

Table 3.3: Payment Scheme

Time	Payment	
Prequestionnaire	30 minutes	7€
Individual experiment	1 hour and a half	10 €
Collective experiment	2 hours	15 €

3.3.3 Hypotheses

The protocol outlined in the previous sections allows us to investigate four main hypotheses that will be depicted in the following paragraphs. These four hypotheses can be separated into four main categories: diversity, creativity training, cross-level performance, and social networks.

Our first hypothesis focuses on diversity and its impact on teams' creative performance. As acknowledged in the literature, diversity has a dual nature [Bahlmann, 2014, Cordero et al., 1996, Godart et al., 2015, Huang and Chen, 2010, Li et al., 2016], and its influence on creative performance yields mixed results [Attanasi et al., 2021, Hundschell et al., 2021]. As a consequence, diversity is often perceived as a double-edged sword. If diversity can lead to new perspectives, it can also lead to challenges related to conflicts or lack of cohesion, which are recognized to harm any social process and possible outcome. We aim to test different variables, both at the surface and deep-level of analysis and see how they might impact team performance. As the literature does not allow us to conclude on the direction of these relationships, we formulate two alternative hypotheses:

Hypothesis 1a: *A higher level of diversity within a team leads to greater creative performance.*

Hypothesis 1b: *A higher level of diversity within a team leads to lower creative performance.*

Our second hypothesis focuses on our subjects' creative training. Considering their profiles, we believe that one significant factor in subjects' performance is their prior training in creativity. As previously mentioned, our subjects can fall into two categories: these with no creativity training and these with creativity training. We intuitively anticipate that prior creativity training will enhance team creative performance. We enunciate our second hypothesis as:

Hypothesis 2: *Teams with training in creativity outperform teams with no training in creativity.*

Next, an underexplored area in the literature is the relationship between individual and team creativity, particularly from an experimental perspective. As mentioned earlier, it may seem intuitive that a team's performance depends on its members' performance. However, authors do not necessarily agree on the relationship between individual and collective performance. One stream assumes that what matters is the average performance of the members, while another stream observes the extremes (the most creative or the least creative of the team) [Mathieu et al., 2017]. Addi-

tionally, tasks can vary according to their degree of openness and, as a consequence, the type of thinking process they activate [Attanasi et al., 2021]. In line with our previous considerations, we expect that this relationship between individual and collective performance is stronger when considering similar types of tasks. That leads us to our third hypothesis:

Hypothesis 3a: *The higher the level of individual creative performance among each team member, the higher the team creative performance.*

Hypothesis 3c: *This relationship is stronger for the individual performance in the open task, which has the same degree of openness as the collective task.*

Finally, we assume that pre-existing links or relationships between our subjects impact their performance as a team. More connected teams would collaborate more effectively and, thus, perform better. In our analysis, we need to differentiate two types of networks: the pre-existing ones and the ones emerging in situ. We assume that both types of networks will have an impact but that the pre-existing network might have a greater impact than the short-term in situ network. Hence, we formulated our fourth hypothesis as such:

Hypothesis 4: *More connected teams perform better.*

Hypothesis 4a: *Team members with a denser in-situ network perform better.*

Hypothesis 4b: *Team members with a denser pre-existing network perform better.*

Section 3.4.1 will explain on how the creative tasks submitted to our subjects were assessed and then specify the specific criteria evaluated. With this, we will operationalise hypotheses 1-4, by associating to each hypothesis the criterion we used in order to assess creativity.

3.4 Results

In the following section, we present the results of our analysis, which are organized as follows. First, in section 3.4.1, we define our dependent variables, hence the assessment of creativity in selecting the reliable components. Then, in Section 3.4.2,

we provide an exhaustive regression analysis to test our hypotheses and confirm the results arising from our descriptive analysis.

3.4.1 Creativity evaluation

In this work, we gathered information on subjects' creativity through individual performance, team performance, and creativity self-assessments from diverse questionnaires conducted at each of the three stages of the protocol. In the end, measures of creativity are categorized into two components. On the one side, we have the explanatory variables that pertain to the individual performance of the subjects across three types of tasks, each corresponding to different levels of openness and thinking processes involved, and the self-assessed individual or collective performance variables. On the other side, the dependent variables relate to the collective performance of teams performing an open task.

To start with, the subjective assessment of task creativity relied on external evaluators chosen to ensure limited biases by remaining blind to the treatment. This choice aimed to secure more accurate evaluations, considering their detachment from experimental details. Table 3.4 presents the profiles of our evaluators.

Table 3.4: External Evaluators' Profiles

<i>Evaluator</i>	<i>Set</i>	<i>Gender</i>	<i>Position</i>	<i>Domain</i>
1	1	M	PhD Student	Cliometrics and Growth theory
2	1	F	Associate Professor	International innovation management, Values-based innovation management
3	1	M	CEO	Research and development, Health
4	1	F	PhD Student	Labour market, Poverty
5	2	M	PhD Student	Cliometrics and Growth theory
6	2	F	Associate Professor	Game theory
7	2	M	PhD Student	Economics of health
8	2	F	PhD Student	Economics of innovation

Then, as mentioned in Section 3.2.1, different criteria can be used to evaluate individual and collective creative performance. In addition to the widely used fluency,

flexibility, and originality measures, we complemented our analysis with four additional and complementary criteria. On the individual side, *expansivity*, as defined by Camarda et al. [2017], refers to “*the ability to provide solutions outside the fixation effect*” (p. 353) compared to all ideas generated by the entire sample. We did not compute the expansivity score as in Camarda et al. [2017]’s study. Instead, due to the diversity of themes and their individual dissimilarities, we opted to consider all ideas without categorization, basing scores solely on average frequency. Then, *persistence* is calculated as the ratio between fluency and flexibility, delineating the ability to offer solutions beyond a subject’s individual fixation effect. These two measures provide a more nuanced analysis of the originality present in subjects’ ideas. On the collective side, *feasibility* is defined as the practical possibility of implementing an idea and consists of a score attributed to teams and *elaboration* as the degree of detail provided for each idea. It is important to stress that the subjects did not have access to these criteria and were only given an evaluation of their creativity without any indication of the criterion involved.

In the end, in order to determine our collective dependent variables and their reliability, we based our final decisions on the *consensual assessment technique* Amabile [1982], which states that subjects are considered creative (or not) when all external evaluators (who are experts or have an acceptable experience in the target domain) rate them independently and agree upon their evaluation. This procedure is implemented for sufficiently open-ended tasks. The evaluators’ evaluation is subjective (from minimum 0 and maximum 10) and, in our case, divided into a taxonomy with the following components: feasibility, elaboration, and originality. Since we had two sets of experimental sessions, there were four evaluations for set 1, and the other four for set 2. But to define the selected measures of creativity, we need to obtain an acceptable inter-judge reliability as before, measured by Cronbach’s alpha [Ursachi et al., 2015]. For the set 1, the Cronbach’s alpha is feasibility ($\alpha = 0.68$), elaboration ($\alpha = 0.54$), and originality ($\alpha = 0.80$). While for set 2 of experiments and evaluators, we observe Cronbach’s alpha for feasibility ($\alpha = 0.59$), elaboration ($\alpha = 0.48$), and originality ($\alpha = 0.57$). The literature states as a reliable measure a Cronbach’s alpha over 0.6[Ursachi et al., 2015], but we can consider as acceptable an alpha of 0.5 or greater [Perry et al., 2004]. Therefore, we chose feasibility and originality as our dependent variables. Table 3.5 summarizes the relevant criteria used and their scales.

Table 3.5: Summary Criteria and Scale of Evaluation by Task

<i>Level</i>	<i>Type</i>	<i>Openness</i>	<i>Criteria</i>	<i>Possible values</i>	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>S.D.</i>
Individual	Tangram	Closed	Cognitive flexibility	1, 2 or 3	1	3	2	0.646
Individual	AUT	Open w/ constraints	Expansivity	0 or 1	0	1	0.309	0.463
	AUT	Open w/ constraints	Persistency	0 or 1	0	1	0.345	0.476
Individual	Draw	Open	Originality	0 to 10	0.5	7.75	4.111	1.447
Collective	Student life	Open	Feasibility	0 or 1	2	7.5	5.279	1.167
	Student life	Open	Originality	0 or 1	2.75	8.25	5.248	1.346

The possible values for the Tangram task are articulated as follows: 1 (fail), 2 (solved above the average time), or 3 (solved below the average time).

For the AUT task, for expansivity and persistency, we implement a dummy variable that takes the value 0 (above average) or 1 (below average).

3.4.2 Regression analysis

In this section, we present the results of our regression analysis in order to understand team creativity and its relationship to diversity and networks⁵. Specifically, we investigate the relationship between team creativity, divided into feasibility and originality, and other relevant explanatory variables. The analysis identified the following key categories of variables: diversity, creativity training, individual creative performance, networks, cross-criteria performance, behavioural traits, and creative skills. The codification and definition of the variables used in our analysis are depicted in the code book presented in Appendices 3.7. A first descriptive analysis of our data has been run to determine the bilateral relation between our variables of interest. This preliminary analysis serves to present the broad results that can be obtained through the multiplicity of information gathered across levels of observation. Specifically, this descriptive analysis provides some insights into the role of perceived team atmosphere and processes on performance. The following regression analysis stemming from this descriptive analysis focuses on the core variables aiming to test the hypotheses stated earlier. For more information on the descriptive analysis, see Appendices 3.7. The final results are summarized in Table 3.6⁶, where Tobit regressions have been chosen to account for the censoring of extreme values, particularly the minimal and maximal scores for each measure of creativity.

⁵The analysis was conducted using STATA, with some of the variables and indicators computed through R and Python depending on the required packages.

⁶See Appendices 3.7 for the complete regression tables.

Table 3.6: Summary Results of the Tobit Regressions

	<i>Feasibility</i>	<i>Originality</i>
Heterogeneous Women	n.s.	— — —
Heterogeneous Men	+ +	n.s.
Age diversity	+ +	n.s.
Diploma diversity	+	n.s.
Creativity Diversity	+ +	+ + +
Creativity training	—	n.s.
Additive model Drawing	+ +	n.s.
Disjunctive model Tangram	—	—
Disjunctive model Drawing	n.s.	+ +
Frequency network	n.s.	—
Friendship network	n.s.	+ + +
Originality	— — —	<i>not applicable</i>
Feasibility	<i>not applicable</i>	— —
Preference to work in group setting	n.s.	+ +
Trust	n.s.	+
Risk	n.s.	—
Visual arts	n.s.	+ +
Invention and technology	—	n.s.

We indicate a positive significant effect by + + + $p < 0.01$, + + $p < 0.05$, + $p < 0.1$. The negative effect by — — — $p < 0.01$, — — $p < 0.05$, — $p < 0.1$. The n.s. means no significant effect.

Diversity

To begin with, on the side of surface-level variables, the results of our regression show evidence⁷ of a significant positive effect of age diversity ($p_{value} = 0.020$) and mild evidence for diploma diversity ($p_{value} = 0.070$) on team performance in feasibility. First, more diverse teams in terms of age are offered broader sets of experience and perspectives that allow them to bring more resources to the teamwork, including a broader set of experiences in working as a team. Similar interpretations can be adopted for diploma diversity as, in general, a higher degree is highly correlated with the age of individuals. However, if it could be expected that age or diploma

⁷For this study, we implement the terminology presented by Moffatt [2020], which states the following: there is *mild evidence* if $p < 0.10$; there is *evidence* if $p < 0.05$; there is *strong evidence* if $p < 0.01$.

diversity would result in diverse perspectives nurturing originality, we do not find a significant result on this. These mixed results confirm Jansen and Searle [2021]’s interpretation that age diversity has a positive effect on team performance depending on specific contextual factors. If these factors were mainly related to team climate, we also conclude that one factor to explore is the specific creativity criteria chosen to measure creativity.

Then, we find evidence for a significant positive effect of gender diversity ($p_{value} = 0.033$) on feasibility scores. If this result shows that heterogeneous teams (with a majority of men) do perform better than a gender-balanced one, it is not the case for other types of unbalanced teams. However, for originality, we find only strong evidence of a negative effect of gender diversity, as a surface-level variable, on team performance ($p_{value} = 0.002$). Following the literature and the inconsistent relationship between gender and creativity [Hardy and Gibson, 2017], these results on gender diversity do not allow us to conclude on the effect of unbalanced gender compositions. As a result, we prefer to stay cautious in our conclusion.

On the side of deep-level variables, we find evidence of a positive effect of creativity diversity on team performance for feasibility and originality (*respectively* $p_{value} = 0.041$ and $p_{value} = 0.000$). The first thing to highlight is how this creativity has been assessed. The creativity diversity here is based on self-reported creativity and is not yet measured experimentally. However, self-perceived creativity is still an interesting measurement of creativity. In fact, since individuals’ own belief in their creative performance influences how they behave, “*people who feel more confident in their creative abilities would be more willing to engage in creative behavior*” [Sawyer and Henriksen, 2024, p. 182]. And this engagement will positively impact their performance. In the end, we do observe that more diverse teams perform better. “*After all, creative people aren’t creative in a general, universal way; they’re creative in a specific sphere of activity, a particular domain*” [Sawyer and Henriksen, 2024, p. 64]. One possible interpretation of the positive impact of creativity diversity, despite the expectation that less diverse teams with generally higher performance would excel, is that individual definitions of creativity influence their own perception. Individuals may focus on specific facets of creativity, such as divergent thinking, when defining it. As a result, diverse team profiles might be more advantageous, as they encompass a broader range of creative capacities, e.g., a more convergent thinking process as well. To investigate this element more closely, further elements in this regression will allow us to compare different types of individual performance that have been mea-

sured in an experimental context and disentangle the relationship between individual and team performance.

Training in creativity

We then draw our attention to the effect of having training in creativity or not on team performance. We find mild evidence of lower performance for teams composed of members with training in creativity ($p_{value} = 0.077$). We could have expected the reverse relationship, but, as before, this result has to be taken into account in comparison with the other facets of creativity. Our interpretation is that such training favours individuals' skills in finding novel ideas but does not invest enough in the implementability of these ideas. In our results, we do observe a positive effect of training on originality, but this relationship fades away as we add more variables, weakening the effect of it. If we stay cautious about the interpretation of these last elements due to the low number of observations of individuals who benefited from this training, we think that further attention should be paid to this relationship between training in creativity and the trade-off between originality and feasibility.

Contrary to the feasibility score, we do not find any significant relationship between the originality of an idea and the fact that team members already experienced training in creativity. This is all the more important since such training seems to harm feasibility while having no repercussions on originality. Again, these conclusions on the poor efficacy of such training have to be taken with caution since our sample only provides a low number of these profiles.

Individual creative performance

Going back to the question of individual performance, our three-stage experimental protocol allows us to observe the cross-level performance of our subjects, with creative performance measured both individually and collectively. Moreover, in line with the literature, we test both additive and disjunctive models of creativity to see whether the performance of the team is determined by its overall performance or driven by one leader. Quite interestingly, we find evidence for a positive effect of the additive model in the Draw task ($p_{value} = 0.037$) and mild evidence of a negative effect of the disjunctive model in the Tangram task ($p_{value} = 0.070$). As a reminder, in this third phase of the protocol, teams were asked to complete an open creative task. Thus, we observe that teams perform better in feasibility for an open task when the overall performance of team members in an individual open task is higher. However, teams

do perform less when there is one member with a high performance in a closed task. In other words, if the additive model is beneficial when the same thinking process is elicited, the disjunctive model harms the team performance once the thinking process is the opposite.

Comparing the additive and the disjunctive model of creativity for originality performance, there exists no significant relationship with the additive model. Nonetheless, we find mild evidence of a negative impact of the disjunctive model in the Tangram task of surprise on team performance ($p_{value} = 0.052$), while evidence for a positive impact of the disjunctive model in the Draw task ($p_{value} = 0.027$). As for feasibility, we see that higher performance in a similar elicited thinking process at the individual level is beneficial to the team, while the elicitation of an opposite thinking process is harming the team's performance. However, in the case of originality, team performance is always driven, positively or negatively, by the performance of the best member in one specific task, whether the Drawing or Tangram task. As a complement, we observe a dominance of the effect of the disjunctive model in the Tangram task over the Draw task due to its larger absolute marginal effect of 0.9631 against 0.6161.

Networks

While we do not find any significant relationships between network measures and the performance in feasibility, we find mild evidence of a negative effect for the subjects' frequency network ($p_{value} = 0.071$) and evidence of a positive effect for the subjects' friendship network ($p_{value} = 0.008$) on originality scores. As a reminder, team members belong to the same cohort of students even though the specific program they are involved in might differ. Thus, we observe that team members who declared denser links between them, in terms of how often they see each other outside the lab, performed less as a team. This might seem counter-intuitive because more frequent contact creates interpersonal bonds that are often considered beneficial to teamwork. Our interpretation is that the more frequent the contacts are, the less diverse the perspectives between team members as they will share more time together and more shared experiences, limiting how original the idea will be due to less diverse inputs. However, the same logic does not hold when it comes to friendship. In fact, this element is quite important as being in a team with friends allows individuals to feel in a safer environment where they will experience less judgement regarding their ideas, benefiting their overall originality. In the end, we do not find any significant

relationship between in-situ networks and our two creativity criteria. Our interpretation is that these interactions were too limited in time to truly affect the creative performance.

Cross-criteria performance

A “good” idea is at the same time of high originality and high feasibility [Rietzschel et al., 2010]. Since the literature shows a trade-off between the originality and the feasibility of an idea [Nijstad et al., 2010], it seems logical to test whether or not we find the same relationship in our results. Our results show that originality has a significant negative effect on feasibility ($p_{value} = 0.002$). Indeed, generating a more original idea, potentially riskier, harms the feasibility or implementability of this same idea. The same applies to originality, for which we report evidence of an inverse relationship with feasibility ($p_{value} = 0.017$). This means the implementability of an idea leads to a lower originality score of this same idea. Striving for an idea that could be effectively introduced may hinder teams from thinking outside the box. This result corroborates again the ones in the literature, stating the dichotomy between both components.

Behavioural traits and creative skills

To conclude, we introduced additional controls to our regressions to see the effect of behavioural traits and creative skills on team creative performance. On the one hand, we do not observe any significant effect of behavioural traits on team creative performance when it comes to the feasibility of ideas. However, we find evidence of a positive impact of a preference to work in a group setting ($p_{value} = 0.037$). This indicates a feeling of ease to work in a team context, which will naturally contribute positively to the team atmosphere and processes and thus performance. This result goes in line with the second one on trust, where we also find a positive effect of trust on team performance ($p_{value} = 0.065$) as individuals trust more easily others do contribute, as before, to the team atmosphere and processes in a positive way. Moreover, this willingness and trust might, similarly to the result of a friendship network, contribute positively to the psychological safety of team members, who are more open to sharing their most unusual ideas, benefiting the team’s originality [Edmondson, 1999, Pirola-Merlo, 2010]. Nonetheless, we find mild evidence of a negative effect of risk on team performance ($p_{value} = 0.083$). This result is rather counter-intuitive since we could expect that team members more willing to take risks

would perform better in originality as they may explore alternative ideas, focusing less on their realism.

On the other hand, we only find mild evidence of a negative effect of expertise in invention and technology of team members on feasibility ($p_{value} = 0.076$) and a positive impact of expertise in visual arts on originality ($p_{value} = 0.029$). First, even though related to creativity, it might be the case that expertise in invention and technology leads individuals to provide extremely novel ideas motivated by the disruption of the current state of inventions and technologies. As a result, really novel ideas will lack feasibility as their disruptive nature might harm their implementation. This goes back to the more traditional trade-off between novelty and usefulness of ideas mentioned earlier. Then, visual arts are mainly related to the arts and show a positive relationship between a specific individual's creative skill and team performance. This can be bridged with our previous result on the positive relationship between the individual performance in an open task, namely the Draw task, and the team performance in originality. This expertise in visual arts is directly linked to the quality of the creative outputs produced during the Draw task.

3.5 Discussion

The aim of this work is to better understand the dynamics of team creativity and its interaction with diversity and networks. To the best of our knowledge, this work is the first to study the effect of diversity and networks on creativity using a multilevel approach, considering a wider range of variables and providing creativity evaluations at both individual and team levels. In fact, until recently, “*the relationship between member creativity and team creativity was an issue that received only very modest research attention*” [van Knippenberg and J. Hoever, 2021, p. 60].

In line with the literature, our results show that feasibility and originality can be detrimental to each other. This questions the objective of any team or organisation that look for creative ideas and solutions. Is the objective to obtain more feasible (targeting implementability) or original (targeting novelty) ideas? Choosing a different target induces different strategies and team formation implications since the determinants of one or another differ.

Besides the relationship between the two dependent variables of our analysis, we focused our attention on diversity. If creativity diversity benefits both creativity criteria, we see that diversity related to socio-demographic variables only contributes

to feasibility performance. As such, when building a team, one needs to consider these relevant socio-demographic elements to ensure the good functioning of the team. If creativity is essential, providing the team with diverse profiles, in terms of age and diploma, will benefit the implementability of ideas.

However, as we said, creativity diversity benefits both originality and feasibility. The only limit of this result is that the creativity measure used is based on self-perceived scores and not a proper measurement of creative performance. In order to disentangle the relationship between individual and team creativity, our protocol provides precise individual and team measurements of creativity. Our results demonstrate that individual performance in an open task can predict the team performance in an open task as well, as there exists a positive relationship between them, but also that good individual performance in a closed task might harm team performance in an open task. This underlines the central role played by types of thinking processes that are differently elicited in a closed and open task. Especially if you take a closer look at the creativity model used. If feasibility is positively driven by the average performance of team members in the Draw task (additive model), it suffers from the good performance of the best team member in the Tangram task (disjunctive model). Originality follows this pattern in open and closed tasks but is sensitive in both cases to the best member in the Draw task and the Tangram task. This difference between additive and disjunctive models and the moderating role of task openness deserves further investigation.

At the same time, we observe interesting results for teams with creativity training, where we observe that such training negatively impacts their ability to provide feasible or implementable ideas. This is quite important as these programs might take this result into account to ensure that their training covers both aspects of creativity to ensure the ability to produce both novel and implementable ideas. Again, as our sample contains a limited number of subjects trained in creativity, we prefer to stay cautious about our conclusions. In addition, “*the effects of creativity training are [also] very domain specific*” [Baer, 2015, p. 169], which highlights the need to continue to explore the effect of creativity training on creative performance and maybe combining creativity criteria with domain-specific tasks.

Regarding networks, our results show that interactions have no significant effect on feasibility and, so, the way teams are able to produce implementable ideas. But they have an impact on how original their ideas will be. This is quite an interesting result since it shows two different directions. First, how often team members interact

with each other in the outside world is detrimental to teams' ability to provide original ideas. Conversely, how much they consider other team members as friends is crucial to providing these original ideas. The interpretation of the role of interactions in creative performance is two-sided. On the one hand, a high frequency of contact induces team members to often evolve in the same environment, which limits the benefits of greater exposure to diverse environments. On the other hand, being in a team with friends affects the way team members can share their ideas and, more specifically, share unusual ideas. Sharing (very) original ideas with friends is easier as they provide a safer environment with no judgement. In the end, a team is a clever blend of diverse environments and psychological safety, where the originality of ideas relies on the ability of team members to provide diverse inputs but also feel at ease to share them. Indeed, psychological safety is crucial, as well as trust, in order to provide a suitable environment for positive team processes and states to grow and benefit creativity and innovation [Clegg et al., 2002, Carmeli and Spreitzer, 2009].

Finally, our results focus on two types of controls: team members' behavioural traits and skills. Again, we observe a difference between feasibility and originality since none of our variables on individuals' skills affect feasibility except for the expertise in invention and technology. In the case of originality, we find a real impact of behavioural traits mainly. When constituting a team, the willingness of candidates to participate in a group setting should be ensured as this will improve team performance and their ability to trust others. Both elements are important as they will positively contribute to the team atmosphere and, again, team psychological safety. However, our results also show that team members' attitudes towards risk harm team originality. As the literature on the relationship between risk and creativity already presents mixed results [Crepaldi et al., 2024], this counter-intuitive one reasserts the need to investigate the relationship between risk and creativity.

Besides these results, our study still presents some limits. The first limit to address is the number of observations that have been gathered through this experimental protocol. Indeed, as a three-stage protocol can be highly demanding, our sample suffered from high attrition and could only reach 135 subjects for a total of 35 teams. This lack of observation limits some of our conclusions, e.g. the impact of creativity training on team performance, and further studies should be conducted with a larger sample of subjects.

Considering our contribution to social interactions and networks, two limitations must be considered. On the one side, direct social interactions in real life are gen-

erally repeated over time and not just one-shot experiences, which leads to different network developments and initiates a learning process. On the other side, once the networks are built, and the learning effect appears, the dynamics between creativity and network become bilateral. In fact, a significant segment of the existing literature investigates the link between networks and creativity from medium to long term. Our research takes a different route, focusing on the creative output of ad-hoc teams rather than scrutinizing social processes within established teams. Thus, we recognize that future work could be elaborated around a protocol similar to ours, including pre-existing teams and observing the possible stronger effect of diversity and networks on their collective creative performance. In addition, it is also important to investigate the process behind direct social interactions. According to Harrison et al. [1998], subjects in teams begin by assessing similarities and dissimilarities in terms of surface-level diversity, but greater interaction switches their attention to deep-level diversity. Indeed, deep-level diversity will have the most impact on team performance over time. Further research might also include this medium or long-term perspective on an experimental protocol and observe teams over time.

Finally, we strongly believe that the study of creativity, either individual or collective, needs to be based on a multi-criteria approach to its evaluation. If we focused here on two main criteria relevant to creativity and innovation, other criteria could be studied depending on the task submitted to individuals and teams. Additionally, while we focused on an open task for our collective experiment, research should be done on other types of thinking processes and their cross-level relationships.

3.6 Conclusion

Understanding how a team is composed and its members interact is essential to ensure performance in searching for novel and implementable ideas. The present work investigates the role of diversity and networks in team creativity.

First, this work reaffirms one important element of the literature on creativity. To obtain the greatest levels of creativity, a balance between feasibility and originality is needed. The most novel idea can be unfeasible for many reasons: the lack of economic resources, contacts to create the chain of production, or the necessary technology to implement them. Therefore, even if the idea is highly original, it is essential or even can be seen as a pre-requirement, the feasibility of that idea, as it will guarantee the evolution of the process transformation leading to innovation. Brilliant ideas

that are not achievable remain just as ideas. The ideas become innovative only after their proper implementation. Therefore, it is crucial to obtain a balance between components of creativity to have the best possible creative idea.

Our results present keen differences among the variables impacting team performance for both creativity criteria, feasibility (practical ideas), and originality (novelty) and address four research hypotheses. First, *Hypothesis 1a* is confirmed as we observe a positive effect of diversity on team performance (invalidating *Hypothesis 1b*). However, it has to be noticed that feasibility and originality are not covered by the same variables since sociodemographic variables only affect originality. On the contrary, *Hypothesis 2* on creativity training is not confirmed. If this training has no significant impact on originality, it does have a negative impact on feasibility. Regarding *Hypothesis 3a* on individual performance, we can only provide partial support, as we found evidence for both additive and disjunctive models of creativity, meaning that the determinant of the performance can be either the average score of the team or the performance of one team member as the main driver. The results also support *Hypothesis 3b*, which posits a positive relationship between individual and team performance in open tasks. Nonetheless, it also suggests an additional negative relationship with performance in closed tasks. Finally, *Hypothesis 4b* is only supported for pre-existing friendship networks in terms of originality performance. For frequency networks, the opposite relationship is observed as teams' originality is harmed. In the end, no significant relationship is observed for *Hypothesis 4b* and between feasibility and in-situ or pre-existing network.

To conclude, to the best of our knowledge, this is one of the first economic works studying the team creative process, as well as its relationship with diversity and social networks, implementing a three-stage protocol to grasp the complexity of cross-level performance. Notably, such a protocol demands a significant investment of time for its implementation and execution. Besides, we firmly believe it is important to understand the challenges associated with team creativity from an experimental perspective and explore the link between individual and team creativity. We hope our work inspires other researchers to develop projects with similar multi-stage protocols.

3.7 Appendices

Instructions

The following sub-sections present the different instructions given to the subjects during the third stage of the protocol: the collective experiment. We have decided to translate these instructions so all readers can understand them, but they are initially in French. The French instructions are available upon request. Moreover, readers interested in the instructions submitted during the first and second stages of the experimental protocol can also ask the authors to access them.

Individual phase

To begin with, we'll ask each of you to think individually about the problem posed. You will have 10 minutes to do this. You should write your ideas in a few sentences on the rough paper provided.

Reminder of the problem:

"Imagine an idea/project to improve your life at university".

Strategic phase

At this stage, we suggest that you share your ideas with each other. Remember that your final assessment will also depend on how different you are from the other teams.

To do this, you will have four coloured tokens. To ask another subject for information about the ideas they developed in the previous stage, you must give them one of your tokens. Conversely, if the other subject wants to know more about your ideas, they should give you one of their tokens. In this way, you can obtain information from four subjects of your choice. You can refuse to give out information, but you can also choose not to spend all your tokens.

Make sure you keep the tokens you have received and these you have not used.

You are invited to move around the room while respecting the rule of discussing only

in pairs.

You will have 15 minutes to do this. At the end, you will be asked to answer a questionnaire.

Reminder of the problem:

“Imagine an idea/project to improve your life at university”.

Collective phase

You will now work in teams to answer the question. The teams will be randomly assigned.

A work area will be allocated to each team, and you will be able to use it and the materials provided as you wish. Use of the equipment is not compulsory under any circumstances.

At the end of this stage, you should have prepared a five-minute oral presentation. Each team will then present its work to the other teams. Apart from the five-minute time limit, there are no other constraints.

You will have 35 minutes for the team reflection stage.

The presentations will then be filmed for assessment by an external jury.

Reminder of the problem:

“Imagine an idea/project to improve your life at university”.

Questionnaires

Intermediary questionnaire

- | | |
|---|---|
| 1. What colour of tokens were you initially assigned? | 2. How many tokens of this colour do you have left? |
|---|---|

- | | |
|---|--|
| <p>3. What other colours of tokens do you have?</p> <p>4. During the exchange phase, where you could discuss in pairs, what was your strategy in choosing your conversation partners? (You can select multiple answers)</p> | <p>best</p> <ul style="list-style-type: none"> • Talk to the most creative people • Talk to the most intelligent people • Randomly • Other(s): _____ |
|---|--|
- Talk to the people I know the
5. Why? _____

Pre-existing network questionnaire

Did you know this person before this experience?

Yes No Only by name

If yes, would you characterize your relationship as (indicate the importance of this relationship in parentheses):

Friendly (from 1 to 5: ...)

Professional (from 1 to 5: ...)

How often did you meet/talk?

Rarely A few times Several times Regularly

Final questionnaire

- | | |
|--|--|
| <p>1. On a scale from 1 to 5, how do you rate your individual level of creativity within your team?</p> <p>2. On a scale from 1 to 5, how do you rate the overall creativity level of your team?</p> | <p>3. Based on the presentations, on a scale from 1 to 5, how do you rate the creativity level of all subjects, regardless of teams?</p> <p>4. In your opinion, was your team more creative than the others in</p> |
|--|--|

- | general? | Yes | No | |
|---|-----|----|---|
| 5. On a scale from 1 to 7, where 1 means "Not at all" and 7 means "Absolutely," evaluate the following propositions regarding your team work TODAY: | | | 17. How often are there conflicts of ideas in your work unit? |
| 6. I am satisfied with working with this team. | | | 18. How often are there disagreements about who should do what in your work team? |
| 7. Our team acted in cohesion as a single team. | | | 19. How often do members of your work unit disagree on how to accomplish a collective task? |
| 8. To what extent was communication in your team open? | | | 20. To what extent did you feel comfortable delegating important functions to the members of your team? |
| 9. To what extent did you feel that your team had team spirit? | | | 21. On a scale from 1 to 5, where 1 means "Never" and 5 means "Very often," assess your reactions when facing negative or unpleasant events in everyday life. |
| 10. To what extent can you affirm that this team is a team in which it is good to work? | | | 22. I hope the problem will resolve itself. |
| 11. To what extent were the members of your team cooperative? | | | 23. When I feel negative emotions, I make sure not to express them. |
| 12. To what extent did you invest yourself in the task? | | | 24. I try to get the problem out of my head. |
| 13. What was the degree of friction among the members of your team? | | | 25. I control my emotions by not expressing them. |
| 14. To what extent are personality conflicts evident in your team? | | | 26. I tackle the problem head-on. |
| 15. How often do people get angry while working in your team? | | | 27. I let out my emotions to reduce stress. |
| 16. How often do people in your team disagree on opinions regarding the work done? | | | 28. I think about what I could do better to solve the problem. |

29. When I want to feel fewer negative situation.
emotions, I change how I view the

Code Book

Variable	Definition
Additive model Drawing	Variable measuring the additive model of creativity as it captures the average individual performance in the Draw task of the team members.
Additive model Expansivity	Variable measuring the additive model of creativity as it captures the average individual performance of the team members in the Alternative Uses task for the Expansivity criterion.
Additive model Persistency	Variable measuring the additive model of creativity as it captures the average individual performance of the team members in the Alternative Uses task for the Persistency criterion.
Additive model Tangram	Variable measuring the additive model of creativity as it captures the average individual performance of the team members in the Tangram task.
Age diversity	Variable defined by the Blau Index in terms of age of team members.
Creativity diversity	Variable defined by the Blau Index in terms of self-perceived creativity of team members.

Variable	Definition
Creative writing	Variable measuring the expertise in creative writing. If no expertise is equal to 0; expertise through learning is equal to 1; expertise through talent is 2; expertise through both learning and talent is equal to 3.
Diploma diversity	Variable defined by the Blau Index in terms of diploma of team members. Diploma is a categorical variable which takes value 0 if subject's degree is " <i>High school diploma</i> ;" 1 if " <i>Bachelor's</i> ;" 2 if " <i>Master's degree</i> ;" and 3 if " <i>PhD</i> ."
Disjunctive model Drawing	Variable measuring the disjunctive model of creativity as it captures the average individual performance in the Draw task of the team members.
Disjunctive model Expansivity	Variable measuring the disjunctive model of creativity as it captures the average individual performance of the team members in the Alternative Uses task for the Expansivity criterion.
Disjunctive model Persistency	Variable measuring the disjunctive model of creativity as it captures the average individual performance of the team members in the Alternative Uses task for the Persistency criterion.
Disjunctive model Tangram	Variable measuring the disjunctive model of creativity as it captures the average individual performance of the team members in the Tangram task.
Feasibility	Variable measuring the average feasibility from 0 to 10 for the four evaluators.
Frequency network	Variable measuring the edge density between team members in terms of frequency of their contacts.

Variable	Definition
Friendship network	Variable measuring the edge density between team members in terms of friendship.
Gender composition	Variable measuring the gender composition in each team, which is organized as follows: perfectly balanced teams is equal to 1; homogeneous team of women equal to 2; homogeneous team of men equal to 3; heterogeneous team with more women than men equal to 4; and heterogeneous team with more men than women equal to 5.
Invention and technology	Variable measuring the expertise in invention and technology. If no expertise is equal to 0; expertise through learning is equal to 1; expertise through talent is 2; expertise through both learning and talent is equal to 3.
Language diversity	Variable defined by the Blau Index in terms of first language of team members.
Originality	Variable measuring the average originality from 0 to 10 for the four evaluators.
Preference to work in team setting	Variable measuring the preference of team members to work in a team setting. It takes 1 if team members prefer to work in such setting, or 0 if team members prefer to work individually.
Risk	Variable measuring risk attitude. It assumes a value from 1 (extremely risk averse) to 10 (extremely risk loving).

Variable	Definition
Strategic phase network	Variable measuring the edge density between team members in terms of in-situ network during the strategic phase.
Trust	Variable measuring generalized trust of team members. It takes 0 if trust is below average and 1 above average.
Visual arts	Variable measuring the expertise in visual arts. If no expertise is equal to 0; expertise through learning is equal to 1; expertise through talent is 2; expertise through both learning and talent is equal to 3.

Diversity Measures

This Appendix offers an explanation of how diversity measures were computed in this work. The Blau Index is defined as $B = 1 - \sum_{i=1}^k p_i^2$ but requires corrections when comparing indexes based on different variables. The normalized Blau index becomes $B_N = \frac{B}{B_{max}}$, where $B_{max} = \frac{n^2(k-1)+a(a-k)}{kn^2}$ [Solanas et al., 2012]. In the end, if B_N is equal to zero, it means that all members of one team can be classified in the same category; there is no diversity. Conversely, the higher B_N is, the more dispersed team members are over the categories; a perfectly diverse team in a specific variable would have a B_N equal to 1.

Then, the strength of faultlines can be measured as $S(i) = \frac{b_i - a_i}{\max(a_i, b_i)}$ where $S(i)$ ranges from -1 to 1. With this hierarchical cluster analysis technique, researchers obtain the strength of faultlines of each team in an experiment but still need to decide on a cut-off value to define whether this value is truly significant. Following the procedure from Carton and Cummings [2013] and Meyer et al. [2015], they can calculate the cut-off value and discard the strength value for teams below the cut-off, ending up with a value of 0. In addition to the strength of faultlines, researchers can consider the average number of subgroups within each team and the average size of these subgroups. This work uses the *ASW cluster* package in R to compute these

measures based on Meyer and Glenz [2013].

Descriptive Analysis

We begin with the study of **surface-level variables**⁸, which refer to observable characteristics of the subjects [Jackson and Wolinsky, 1996, Milliken and Martins, 1996], mainly demographics, with regard to the two components of team creativity: feasibility (practical ideas which transformation process leads to innovation) and originality (novelty). For feasibility, we observe mild evidence⁹ of older subjects performing more poorly ($p=0.050$; Wilcoxon rank-sum test), and French nationality, inducing French speakers, triggering feasibility ($p=0.072$; Wilcoxon rank-sum test). While, for originality, the only relevant variable is inside the gender composition, where we find mild evidence that homogeneous women teams are less original ($p=0.050$; Wilcoxon rank-sum test). Hence, the practical ideas (feasibility) are fostered by younger French subjects, who, given the task, now better live in the French university to propose and improve. However, novelty (originality) decreases in homogeneous women teams, as encountered in several contributions in the literature, for women [Bradler, 2015, Charness and Grieco, 2019].

In addition, we analyze the **deep-level variables** separating “*Diversity of live experience*” (DOLE) [Douthitt et al., 1999], from personal traits¹⁰. On the one side, there are no significant effects of the DOLE (and the different components inside DOLE) on feasibility and neither originality. On the other side, for the personal traits, the results differ between the measure of creativity in terms of variables and signs. Regarding feasibility, we encountered evidence of a detrimental effect of having foreign friends ($\rho=-0.21$ and $p=0.014$; Spearman’s rank-order correlation). Alternatively, looking at originality, we observe evidence of a positive effect on having foreign friends ($\rho=0.19$ and $p=0.0236$; Spearman’s rank-order correlation), being a member of an association ($\rho=0.20$ and $p=0.023$; Spearman’s rank-order correlation), and mild evidence of having a learn/talent in humor ($\rho=0.15$ and $p=0.080$; Spearman’s rank-order correlation). Thus, the results suggest that novel ideas (originality) are a matter of open mind in terms of having foreign friends, participating

⁸The variables considered in surface diversity are gender, age, program, domain, diploma, French speaker, languages, French nationality, and revenue.

⁹For this study, we implement the terminology presented by Moffatt [2020], which states the following: there is *mild evidence* if $p<0.10$; there is *evidence* if $p<0.05$; there is *strong evidence* if $p<0.01$.

¹⁰We divide the personal traits into the following categories: self-assessment of creativity, idiosyncratic features, trust and risk, and hobby.

in associations, and learning or having a talent for humor. While, the practical ideas which the transformation process conduces to innovation (feasibility) are dismissed by the open mind in the sense of having foreign friends who offer different perspectives.

Moreover, we also control for the **team experience**. In the case of feasibility, we depict mild evidence of a positive effect stemming from the willingness to delegate within team dynamics ($\rho=0.18$ and $p=0.041$; Spearman's rank-order correlation). While for originality, we show evidence of a negative impact of subjects willing to get the problem out of their head ($\rho=-0.19$ and $p=0.025$; Spearman's rank-order correlation), being able to control their emotions ($\rho=-0.20$ and $p=0.018$; Spearman's rank-order correlation), and mild evidence for change perspective of problems in order to perceive less the negative emotion ($\rho=-0.15$ and $p=0.080$; Spearman's rank-order correlation). Also, we find a positive influence on originality, with mild evidence, in investing effort in the task ($\rho=0.16$ and $p=0.073$; Spearman's rank-order correlation), and evidence for subjects expecting problems to solve on their own ($\rho=0.18$ and $p=0.045$; Spearman's rank-order correlation). Furthermore, we control the asymmetry of the subject's perception of the team experience but do not discover any significant effect on feasibility and originality. Therefore, the practical ideas (feasibility) are fostered by a sort of cooperation driven by willingness to cooperate. In contrast, the novel ideas (originality) are reduced by attitudes of avoiding problems and over-control, but it triggers effort and calm in front of problems.

Besides, we also capture diversity by the **Blau index**, which allows us to look at the diversity in the team by variable. The higher the Blau index, the greater the diversity of the team members in the sense of the selected variable. For feasibility, we report strong evidence of a negative impact of the Blau index (diversity of team members) in terms of being a French speaker ($\rho=-0.24$ and $p=0.005$; Spearman's rank-order correlation) and evidence for expansivity in the individual AUT task ($\rho=-0.20$ and $p=0.019$; Spearman's rank-order correlation). On the contrary, we depicted a positive strong significant effect on the feasibility of greater diversity of the team members (Blau index) in age ($\rho=0.30$ and $p<0.001$; Spearman's rank-order correlation). For originality, we find strong evidence of a negative impact of the Blau index for age ($\rho=-0.26$ and $p=0.002$; Spearman's rank-order correlation) and evidence for the program as training on creativity ($\rho=-0.21$ and $p=0.016$; Spearman's rank-order correlation). Although, a positive strong effect is present for the Blau index for self-evaluation of general creativity ($\rho=0.39$ and $p<0.001$;

Spearman's rank-order correlation), diploma ($\rho=0.23$ and $p=0.008$; Spearman's rank-order correlation), expansivity in the AUT task ($\rho=0.23$ and $p=0.007$; Spearman's rank-order correlation), and evidence for the domain ($\rho=0.19$ and $p=0.024$; Spearman's rank-order correlation). These results suggest the relevant role of Blau index as a measure of diversity and remark on the distinct cognitive process between seeking for practical ideas that easily can be converted, after a process, to innovation (feasibility), and intrinsically novel ideas (originality). However, in Section 3.4.2, we control for the non-linear pattern given that in the literature is found an inverted-U shape of diversity.

In contrast, the **Faultlines** permits studying the diversity of team members by looking at several variables simultaneously, whereas the Blau index allows one variable each time. In this case, we consider faultlines surface-level diversity (includes the variables gender, age, program, French speaker, diploma, and domain), and deep-level (includes the variables creativity self-evaluation, introversion, preference for team or individual work, trust, risk, and DOLE scores). On one side, for feasibility, we do not find significant effects. On the other side, for originality, we observe strong evidence of a negative effect of the strength of deep-level ($\rho=-0.38$ and $p<0.001$; Spearman's rank-order correlation). Then, practical ideas converting to innovation (feasibility) seem to be influenced by faultlines, whereas it is detrimental to the generation of novel ideas (originality), as deep-level diversity among team members.

Furthermore, we study the change during the **intermediary phase** and strategies followed by the subjects to change information with others. We do not find a significant effect on feasibility, but we observe mild evidence of a positive impact of originality driven by subjects willing to change with people they know ($\rho=0.15$ and $p=0.083$; Spearman's rank-order correlation). In this manner, we suggest that subjects who were exchanging with people they know, which means looking at familiar subjects with which they feel comfortable to express themselves openly, were more likely, at the collective, to produce a novel idea (originality).

Moreover, we compare **individual creativity** (for open, open-with-constraints, and closed tasks) with team creativity (open task) by looking at both components. We do not discover any significant effect on feasibility. Nevertheless, when we consider the component of originality, we find evidence of a negative effect of the performance in the AUT in terms of persistency ($\rho=-0.20$ and $p=0.020$; Spearman's rank-order correlation), thus, internal fixation effect in individual creativity for the AUT task. Additionally, we encounter negative strong evidence of the relationship

between feasibility and originality ($\rho=-0.29$; $p<0.001$; Spearman’s rank-order correlation), demonstrating that each component activates a different cognitive process. This result confirms the distinction between seeking practical ideas that, under a transformation process, become innovation (feasibility), hence a necessary condition to reach innovation, whereas originality relates to the creativity per se of novel ideas that might not be implemented, because of the need of a new technology or large funding.

In addition, we study the complex dynamics between individual and collective creative performance following the additive and disjunctive model, based on the literature presented in section 3.2.1. In fact, we control for the three individual creative tasks and their different measures. We encounter significant evidence for the **additive model** but not the **disjunctive model**. Indeed, for the additive model, we observe that there exists mild evidence of a positive correlation between feasibility in the collective task with the average individual performance in the team for the tangram task ($\rho=0.16$ and $p=0.072$; Spearman’s rank-order correlation), whereas, arise evidence of a negative correlation between originality and the additive model for tangram task ($\rho=-0.22$ and $p=0.013$; Spearman’s rank-order correlation). Moreover, we depict strong evidence for an adverse correlation between originality in the collective task with the additive model in terms of persistency in the AUT task ($\rho=-0.27$ and $p=0.002$; Spearman’s rank-order correlation). These results suggest that the additive model matters mainly for the close task (tangram task), and in line with the previous results, the dichotomy between idea generation (originality) and practical ideas (feasibility) is still key and inverse. Also, this conclusion is in line with the literature, because the additive model is more predictive for low creativity requirements, hence, closed task [Yuan et al., 2022]. Furthermore, the adverse effect of the persistency additive model on originality might relate to the fact that a greater internal fixation effect among team members (on average) leads to lower team originality.

Finally, we devote our attention to looking at the pre-existing **networks** and the interactions created during the experiment and confront them with the two measures of team creativity. On the one hand, for feasibility, we encounter strong evidence of a negative effect of connections for subjects who frequently meet/talk ($\rho=-0.25$ and $p=0.003$; Spearman’s rank-order correlation), and for friends ($\rho=-0.35$ and $p<0.001$; Spearman’s rank-order correlation). In the same line, we observe a positive impact of the friend component (more disconnection with friends) with

strong evidence ($\rho=-0.36$ and $p<0.001$; Spearman's rank-order correlation), and evidence for subjects who frequently meet/talk ($\rho=-0.26$ and $p=0.003$; Spearman's rank-order correlation). On the other hand, we look at originality, where we have a negative impact where evidence is found for the friend component (disconnection with friends) ($\rho=-0.21$ and $p=0.014$; Spearman's rank-order correlation), strong evidence for the component referring to the frequency they meet/talk ($\rho=-0.01$ and $p=0.008$; Spearman's rank-order correlation). In the same fashion, we report a positive effect with originality with strong evidence for density (connections) of friends ($\rho=0.33$ and $p<0.001$; Spearman's rank-order correlation), and density for people they know previously ($\rho=0.31$ and $p<0.001$; Spearman's rank-order correlation). Consequently, we encounter that having more interactions with people that at least subjects know (or are friends with) is detrimental to generating practical ideas (feasibility), but increases the novel once (originality).

Regressions

Table 3.8: Tobit regression for Feasibility

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Homogeneous Women	0.447 (0.273)	0.632** (0.254)	0.589* (0.314)	0.610 (0.627)	0.604 (0.601)	0.663 (0.599)	0.605 (0.582)
Homogeneous Men	0.119 (0.278)	0.0408 (0.256)	-0.246 (0.280)	-0.247 (0.537)	-0.220 (0.394)	-0.231 (0.388)	-0.308 (0.400)
Heterogeneous Women	-0.239 (0.371)	0.592 (0.381)	1.011** (0.388)	1.051** (0.480)	0.223 (0.521)	0.143 (0.524)	0.121 (0.516)
Heterogeneous Men	0.521* (0.281)	0.903*** (0.270)	1.051*** (0.311)	1.082** (0.540)	1.080** (0.514)	1.112** (0.517)	1.099** (0.508)
Age diversity	1.503** (0.631)	2.253*** (0.600)	2.666*** (0.582)	2.173** (0.903)	1.833** (0.795)	1.809** (0.788)	1.785** (0.757)
Language diversity	-0.889*** (0.283)	-1.015*** (0.262)	-0.573** (0.282)	-0.277 (0.524)	-0.403 (0.431)	-0.418 (0.424)	-0.386 (0.421)
Diploma diversity	-0.570** (0.260)	0.844** (0.374)	0.966** (0.406)	0.813 (0.603)	0.809 (0.509)	0.869* (0.509)	0.947* (0.517)
Creativity diversity	0.871*** (0.281)	0.777*** (0.259)	0.318 (0.328)	0.399 (0.642)	1.207** (0.605)	1.279** (0.615)	1.229** (0.593)
Creativity training		-1.962*** (0.398)	-2.196*** (0.458)	-1.861** (0.896)	-1.505* (0.825)	-1.482* (0.817)	-1.455* (0.816)
Additive model Tangram			0.603** (0.295)	0.669 (0.489)	0.572 (0.449)	0.614 (0.443)	0.659 (0.444)
Additive model Persistency			-0.237 (0.436)	0.143 (1.104)	-0.0524 (1.083)	-0.0516 (1.081)	-0.0212 (1.065)
Additive model Expansivity			-0.0841 (0.513)	-0.0483 (1.073)	-0.159 (1.005)	-0.180 (0.994)	-0.365 (0.984)
Additive model Drawing			0.921*** (0.264)	1.073*** (0.394)	0.718** (0.360)	0.682* (0.368)	0.740** (0.351)
Disjunctive model Tangram			-0.533 (0.353)	-0.782 (0.633)	-0.982* (0.589)	-1.052* (0.600)	-1.085* (0.592)
Disjunctive model Persistency			-0.406	-0.0377	0.0390	0.0887	0.195

Continued on next page

CHAPTER 3. DIVERSITY, INTERACTIONS, AND TEAM CREATIVITY: AN EXPERIMENTAL PERSPECTIVE

Table 3.8: Tobit regression for Feasibility (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
			(0.341)	(0.798)	(0.748)	(0.732)	(0.726)
Disjunctive model Expansivity			3.755	4.450	1.796	1.493	0.757
			(5.864)	(16.45)	(16.09)	(16.04)	(15.84)
Disjunctive model Drawing			-0.469***	-0.507**	-0.199	-0.180	-0.226
			(0.165)	(0.256)	(0.237)	(0.245)	(0.233)
Frequency network				-0.583	-1.224	-1.383	-1.305
				(0.897)	(1.044)	(1.087)	(1.057)
Friendship network				-0.248	0.755	0.879	0.958
				(1.013)	(0.973)	(0.991)	(0.958)
Strategic phase network				0.769	0.604	0.570	0.670
				(1.076)	(0.941)	(0.923)	(0.925)
Originality score					-0.359***	-0.381***	-0.383***
					(0.130)	(0.127)	(0.123)
Preference to work in group setting						0.164	0.130
						(0.136)	(0.138)
Trust						0.134	0.146
						(0.107)	(0.0983)
Risk							-0.0407
							(0.0254)
Visual arts							-0.0413
							(0.0806)
Creative writing							0.0579
							(0.0627)
Invention and technology							-0.116*
							(0.0645)
Constant	3.724***	3.041***	2.193*	1.686	4.330*	4.474*	4.768**
	(0.644)	(0.609)	(1.249)	(2.227)	(2.343)	(2.319)	(2.306)
Observations	135	135	135	135	135	135	135
R^2	0.0570	0.1097	0.1520	0.1682	0.2150	0.2200	0.2294

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.9: Tobit regression for Originality

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Homogeneous women	-0.575*	-0.756**	-0.0515	-0.0174	0.215	0.344	0.208
	(0.311)	(0.296)	(0.345)	(0.548)	(0.573)	(0.557)	(0.555)
Homogeneous men	0.102	0.179	0.716**	0.0733	-0.0206	-0.0304	-0.167
	(0.316)	(0.299)	(0.308)	(0.706)	(0.565)	(0.568)	(0.582)
Heterogeneous women	-0.790*	-1.600***	-1.841***	-2.308***	-1.908***	-1.945***	-1.995***
	(0.422)	(0.443)	(0.427)	(0.689)	(0.658)	(0.647)	(0.633)
Heterogeneous men	-0.0847	-0.457	-0.0623	-0.00530	0.406	0.488	0.516
	(0.320)	(0.314)	(0.342)	(0.513)	(0.519)	(0.516)	(0.497)
Age diversity	-0.735	-1.466**	-1.531**	-0.948	-0.121	-0.101	-0.146
	(0.718)	(0.699)	(0.640)	(0.939)	(0.840)	(0.862)	(0.862)
Language diversity	-0.476	-0.353	-0.538*	-0.351	-0.457	-0.474	-0.451
	(0.322)	(0.305)	(0.310)	(0.540)	(0.449)	(0.446)	(0.435)
Diploma diversity	0.707**	-0.672	-0.428	-0.0101	0.299	0.415	0.507
	(0.296)	(0.435)	(0.446)	(0.678)	(0.593)	(0.578)	(0.589)
Creativity diversity	1.158***	1.250***	2.279***	2.254***	2.406***	2.461***	2.361***
	(0.320)	(0.302)	(0.360)	(0.358)	(0.363)	(0.346)	(0.349)
Creativity training		1.913***	1.380***	0.993	0.285	0.267	0.218
		(0.464)	(0.503)	(1.065)	(0.904)	(0.881)	(0.887)
Additive model Tangram			-0.561*	-0.271	-0.0164	0.0820	0.150
			(0.325)	(0.493)	(0.472)	(0.437)	(0.400)
Additive model Persistency			-0.741	-0.544	-0.489	-0.459	-0.440
			(0.479)	(0.781)	(0.795)	(0.808)	(0.797)
Additive model Expansivity			0.361	-0.308	-0.326	-0.360	-0.508
			(0.564)	(1.186)	(1.112)	(1.100)	(1.088)
Additive model Drawing			-0.858***	-0.989**	-0.581	-0.602	-0.553
			(0.290)	(0.466)	(0.505)	(0.492)	(0.500)
Disjunctive model Tangram			-0.438	-0.558	-0.856	-0.986*	-0.963*
			(0.388)	(0.513)	(0.525)	(0.511)	(0.490)
Disjunctive model Persistency			0.336	0.214	0.200	0.287	0.310
			(0.375)	(0.594)	(0.566)	(0.566)	(0.540)
Disjunctive model Expansivity			-4.563	-7.401	-5.707	-6.243	-7.251

Continued on next page

Table 3.9: Tobit regression for Originality (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
			(6.447)	(10.87)	(11.63)	(11.65)	(11.43)
Disjunctive model Drawing			0.709***	0.859***	0.666**	0.670**	0.616**
			(0.182)	(0.242)	(0.276)	(0.265)	(0.274)
Frequency network				-1.787*	-2.009*	-2.227*	-2.112*
				(1.059)	(1.143)	(1.173)	(1.159)
Friendship network				2.796**	2.702**	2.795**	2.944***
				(1.244)	(1.090)	(1.098)	(1.083)
Strategic phase network				-0.459	-0.166	-0.218	0.0180
				(1.087)	(0.947)	(0.937)	(0.932)
Feasibility score					-0.381**	-0.390**	-0.394**
					(0.169)	(0.162)	(0.163)
Preference to work in group setting						0.342**	0.305**
						(0.149)	(0.144)
Trust						0.189	0.222*
						(0.115)	(0.119)
Risk							-0.0457*
							(0.0261)
Visual arts							0.105**
							(0.0473)
Creative writing							-0.0346
							(0.0499)
Invention and technology							-0.0862
							(0.0783)
Constant	5.400***	6.066***	7.308***	7.373***	8.015***	8.030***	8.308***
	(0.733)	(0.709)	(1.373)	(1.959)	(1.958)	(1.966)	(1.946)
Observations	135	135	135	135	135	135	135
R^2	0.0601	0.0947	0.1673	0.2202	0.2631	0.2782	0.2853

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Chapter 4

Does Creativity Thrive on Plot Twists? Exploring the Role of Surprise on Team Creativity

This chapter was co-authored with

Sara GIL-GALLEN

Summary of the chapter

The creative process is inherently risky and uncertain. Unexpected events can arise, causing surprises, especially in organisational settings that demand adaptability and value flexibility. This relationship between team creativity and surprise is complex and warrants investigation. How does an unexpected event impact team creative performance according to the degree of surprise introduced in a task? How is the relationship between surprise and team creativity affected by team diversity, processes, and overall creativity? This study is the first to experimentally modulate the magnitude of surprise to explore its impact on creative performance, controlling for team members' characteristics. The experimental protocol involves 122 teams (366 subjects) working on two tasks. The first task is designed to create relevant schemata without surprise, while the second task varies in surprise degree between control and treatment groups. Our results are three-folded: surprise negatively affects team creative performance in terms of originality, team members' attitudes toward surprise moderate creative performance concerning feasibility, and positive experiences of surprise enhance feasibility and originality of ideas.

4.1 Introduction

The unexpected is an inescapable part of the creative process. By its very nature, creativity is not a fully routinized process, and requires individuals to take part in an uncertain and risky undertaking. Indeed, if creativity is a risky process due to the uncertainty of the success of its outcome, it becomes even more precarious due to the potential emergence of unexpected events [Fey and Kock, 2022] altering the conditions in which novel and useful ideas emerge. These unexpected events, varying in kind and importance, may cause surprise, and this inextricable link between surprise and creativity forces us to look at the impact of an unexpected event on people's creative performance.

Organisations operate in constantly changing environments that demand adaptability. Consequently, individuals in organisational settings are evermore encountering dynamic environments that highly value flexibility [Grant, 1996]. To put it differently, organisations are subject to uncertainty and unexpected events, and their sustainability relies on resilience and adaptation to those events. In fact, despite the ease and comfort provided by certainty or predictability, “*surprise is inevitable because it is part of the natural order of things and cannot be avoided, eliminated, or controlled*” [McDaniel et al., 2003, p. 266]. In addition, the fact that organisations are relying more and more on teamwork has been noticeable for many years, if not decades, and it does not look like this phenomenon is going to run out of steam [Mathieu et al., 2017]. As a result, understanding how teams within organisations will react and adapt to surprise is crucial.

As creativity is essential to an organisation's survival and performance [Amabile et al., 1996, Cook, 1998, Oldham and Cummings, 1996, Shalley, 1995, Woodman et al., 1993], teams' capacity to overcome surprise in a creative context becomes crucial to the study of team creativity. In a functional manner, when tackling a task, teams must consider potential changes as the task progresses because those changes bring challenges and so, team members must apprehend and integrate the surprising stimulus into a problem-reframing strategy to update the problem representation. The same logic applies to creative endeavours.

To explore the intricate relationship between team creativity and surprise, we need to investigate how unexpected events impact creative performance and how team members' perceptions and reactions to surprise influence this relationship. However, we also need to consider the type of unexpected event that teams would

have to face. In this study, there is no question of surprise about timing, procedure, environment setting, or outcome; it is about the complex question of problem-(re)framing. What happens once a team is faced with a change in constraints in the task they have to fulfil?

This leads us to formulate the following research questions: How will an unexpected event, caused by a change in constraints specification, impact teams creative performance according to the degree of surprise introduced in a task? And, how is this relationship between the resulting surprise and team creativity affected by other factors, such as team diversity, processes, and creative capacities?

To the best of our knowledge, this work is the first experimental protocol that modulates the magnitude of a surprising event to disentangle the relationship between creative performance and the emergence of a surprising event. More specifically, this study aims to test the effect of surprise depending on the elicited and perceived intensity and how this experience of surprise might have influenced the creative process and, by extension, the creative performance. The resulting experimental protocol involves face-to-face interactions within 122 teams (366 subjects) working on two specific tasks. The first task, common to each team, is an open task based on Baruah and Paulus [2011], Harvey [2013], Kohn et al. [2011], Maltese et al. [2023] that aims to create relevant schemata for subjects where surprise will never occur. While, the second task differs between control and treatment groups according to the elicited degree of surprise.

This investigation provides us with three main results. First, surprise negatively affects team creative performance regarding the originality or novelty of an idea. Second, team members' attitudes toward surprise only moderate the creative performance of a team in feasibility. Third, team members' positive experience of surprise moderates both team creative performance for feasibility and originality.

The remainder of this study is organized as follows: Section 4.2 presents the literature review on surprise, encompassing the definition of the concept and its link to creativity. Section 4.3 builds on the literature on creativity and surprise to present the different hypotheses that we aim to test in this study. Then, Section 4.4 focuses on the study methodology by describing the experimental design, procedure, and the resulting measures. Section 4.5 presents our results based on the creative assessment strategy, the descriptive analysis, and the regression analysis of our experimental data. Finally, Section 4.6 displays the discussion and concluding remarks about this work.

4.2 Literature review

4.2.1 Team creativity

Creativity is “*the production of novel and useful ideas by an individual or small group working together*” [Amabile and Pratt, 2016, p. 158]. This widely used definition emphasizes that creativity can emerge individually and collectively. For a long time, creativity was seen as a faculty of individuals endowed with a specific talent, alternatively attributed to divine inspiration [Gaut, 2012, Stokes and Paul, 2016] or to geniuses [Miller, 1996, Sawyer and Henriksen, 2012], but literature has evolved in the sense that creativity has become a faculty that every individual possesses and which should be seen as a muscle that can be trained. Moreover, the purely individual nature of the creative endeavour was set aside, and collective creativity was envisioned. Since creativity becomes universal, it allows us to look at the collective creativity that emerges from the complex and varied social interactions between individuals and the bidirectional relationship between them and their environment [Amabile and Pratt, 2016, Csikszentmihalyi, 1997, Fischer et al., 2005, Hennessey and Amabile, 2010]. This research refers to team creativity as the collective creative processes and products arising from teams where those teams are not simply the sum of the characteristics and abilities of their members but the “*interdependent collections of individuals who share responsibility for specified outcomes*” (p. 120) as described by Sundstrom et al. [1990].

This raises several questions about the heterogeneity of team members and their interactions. Furthermore, to grasp the complexities of team creativity, it is essential to examine different levels of observation, from the individual to the team or organisation, to understand how these variables interact to foster innovation [Woodman et al., 1993]. As a result, we focus on two key aspects: the impact of team composition and diversity and the emergence of processes and states within teams stemming from their collaborative endeavour.

Team diversity

Team heterogeneity has been studied for long and is often referred to as team diversity. Diversity is defined as “*differences between individuals on any attribute that may lead to the perception that another person is different from self*” [van Knippenberg et al., 2004, p. 1008]. Applied to the creativity context, the extensive literature on

diversity remains inconclusive regarding its impact, whether positive or negative [Bell et al., 2011, Hundschell et al., 2021, van Dijk et al., 2012]. While diversity is generally associated with the coexistence of different perspectives, which benefits the group by bringing in varied viewpoints and strengthening its ability to address specific situations or problems, it can also lead to perceiving others as different from ourselves. Since there is a preference for working with similar individuals, known as homophily, a tension arises between the benefits of diversity and the potential challenges of increased social distance. Ultimately, diversity must be considered cautiously as it can be a double-edged sword [Milliken and Martins, 1996, van Knippenberg, 2017].

To grasp the complexity of the effect of diversity on creativity we need first to distinguish the attributes observational levels with a distinction between the surface and deep-level of observation. The former is defined as “*differences among group members in overt, biological characteristics that are typically reflected in physical features*” also including more generally socio-demographic information, and the second as “*differences among members’ attitudes, beliefs, and values*” [Harrison et al., 1998, p. 97]. In addition to the type of variable concerned, these two categories of diversity are differentiated mainly by the channels through which they are measured. The same categories are often referred to as directly and non-directly observable variables since, unlike the surface-level, the deep-level information is only gathered through extended interactions. For the purposes of this study, this review focuses on demographic diversity at the surface level and personality diversity at the deep level.

The literature on demographic diversity is articulated around several variables, including gender and age. While we find mixed results on the effect of gender on creative performance [Attanasi et al., 2021], the same applies to gender diversity as opposite directions appeared mostly due to contextual elements, e.g., the complexity of the task [Bowers et al., 2000], the group size [Wegge et al., 2008], or the activated faultlines [Pearsall et al., 2008]. Regarding age, Jansen and Searle [2021] conclude in their literature review that age diversity is initially perceived to have a positive effect on team performance, with this relationship being highly dependent on contextual factors such as a positive team climate and lowered age discrimination. While our study focuses mainly on two demographic diversity variables, i.e. gender and age, the literature provides broader results that include other variables such as nationality or education.

For deep-level variables, this review mainly focuses on personality diversity, as ex-

tensive literature has been developed on the relationship between personality traits and creative performance. In fact, “*personality has more direct and powerful effects on group processes than other composition variables typically studied (e.g. age, race, gender, and information distribution)*” [Moynihan and Peterson, 2001, p. 328]. Following the Big-Five personality test [McCrae and Costa, 1987], individuals who demonstrate openness to experience can be qualified as curious, original, and non-conventional [Feist, 1998, Schilpzand et al., 2011]. Extraversion is another personality trait that positively impacts creative performance, which translates into active, ambitious, and assertive individuals [Feist, 1998]. Furnham and Bachtiar [2008] provide a possible explanation for the positive impact of openness to experience and extraversion on creative performance since individuals are inclined to accept more unusual ideas and “quirkiness”. Agreeableness seems to contribute negatively to the negative impact of personality traits on creative performance. If agreeableness is associated with trust, compliance, and gentleness, this willingness to maintain positive relationships might negatively affect creativity as individuals with low agreeableness contribute by being more proactive and detached from others’ judgment [Giancola et al., 2021]. To conclude, on the Big-Five personality test, conscientiousness and neuroticism do not present consistent conclusions on their relationship with creative performance [Hornberg and Reiter-Palmon, 2017]. If these relationships are more or less established at an individual level, we need to look at their collective level. As is often the case when focusing on diversity, Jansen and Searle [2021] underlines the role of contextual elements on the impact of personality diversity on team performance, e.g., creative confidence or task dependency. Thus, as extensive literature has been developed at the aggregated level of personality traits, almost all references do consider personality diversity as the mean of such traits within each team rather than a diversity measure in itself, with some exceptions applied to the literature on faultlines [Molleman, 2005].

To conclude on the effect of diversity on creative performance, beyond those surface and deep-level variables, we can also consider diversity in terms of task-related variables. In the specific context of our study, task-related variables are all the variables that directly might affect creativity. In fact, van Knippenberg and J. Hoever [2021] highlights that “*the relationship between member creativity and team creativity was an issue that received only very modest research attention*” (p. 60). In this study, we focus mainly on diversity in terms of individual creative performance. Knowing that the relationship between individual and team creativity is not straightforward,

with team creativity not being the sole summation of individuals' creativity, it becomes interesting to look at the heterogeneity of performance within a team, which the literature has not extensively addressed.

Team processes

Team creativity goes beyond the simple summation of individual outputs and is the result of the interactions between team members. These interactions are essential in the creative process as their multiple nature and forms result in a wide range of possible processes affecting both positively and negatively team performance. Among them are associative processes and states that have significant impacts on the creative process as they provide a fruitful environment for teamwork.

First, extended literature has been developed on cohesion. Above all, it should be noted that the cohesion of a team does not directly determine its creative performance but should be seen as a factor contributing to its creative capacity. As cohesion translates into the emergence of a social bond and sense of attraction, it benefits team performance [Forsyth, 2021, Schachter et al., 1951]. Moreover, cohesion is closely related to task commitment since task commitment is one of its determinants [Forsyth, 2021] and a catalyzer of team performance [Mullen and Copper, 1994]. The literature also brings out other processes and dynamics such as collaboration and cooperation [Drach-Zahavy and Somech, 2001, Mitchell et al., 2009], satisfaction [Lester et al., 2002], or communication [Caldwell and Everhart, 1998, Gilson and Shalley, 2004], which all contribute positively to the team performance, also in a creative context.

If we take a closer look at these different elements, we can see that they contribute to creating a team climate conducive to teamwork. An environment where teams can flourish reflects a certain level of psychological safety. This psychological safety fosters fluid and effective communication, which is beneficial to the team's creative process and performance. This notion of psychological safety is closely related to the trust that can emerge within a team and that fosters all the processes and states mentioned above in order to benefit creativity and innovation [Clegg et al., 2002, Carmeli and Spreitzer, 2009]. Creativity and innovation are then favoured by the complex interactions of those processes and states that nurture individual-level cognition and team-level sharing of information [De Dreu et al., 2011]. In a creativity context, by providing a safe environment to teams, the lack of fear of judgment resorts to be beneficial to the generation of novel ideas [Edmondson, 1999,

Pirola-Merlo, 2010]. Nevertheless, it should be noted that the temporal dimension is essential for the emergence of processes and states between members of the same team. While this may not be completely unachievable, those processes and states are more difficult to observe in an experimental environment where the time is highly constrained and might prevent their formation [Paulus et al., 2012].

In the end, team diversity and team processes are two elements that intermingle and respond to each other. Hülshager et al. [2009] in their comprehensive literature review provide evidence that team processes variables are strong support for creativity and innovation, as well as team composition, even though the relationship is weaker. Nonetheless, while the preceding paragraphs demonstrate the crucial role played by those variables, we must not overlook the importance of environmental factors in the creative process. This applies in particular to the subject of our study, which looks at the impact of an unexpected event on team creative performance.

4.2.2 Apprehending the unexpected

Certainty and predictability provide comfort and ease when it comes to apprehending the surrounding environment. In fact, “*much of what we experience is highly predicted [...] Occasionally, however, a surprising event occurs that violates our expectations*” [Bein et al., 2023, p. 1]. It becomes crucial to study unexpectedness once we accept that surprise is inevitable [McDaniel et al., 2003].

For centuries, the concept of surprise has intrigued scholars across various disciplines. Beginning with Aristotle, the term *explêxis* was described as a sudden and fear-like shock [Miller, 2015]. Similarly, philosophers of the modern era defined surprise as a pleasure and passion [Hume, 1739-40] or a violent and sudden change [Smith, 1795]. These early definitions, while insightful, did not delve into the underlying mechanisms or consequences of surprise. However, the scientific study of surprise advanced significantly in the late 19th century with the rise of behavioural approaches, notably through the work of Darwin [1872], linking surprise to physiological responses and attention. If the Darwinian conceptualization of surprise did not eclipse the work of emotion theorists, we had to wait for the advent of social and cognitive psychologists to further refine the concept of surprise and provide a reconciliation between the Darwinian and Emotion perspectives. The resulting new stream of literature associated surprise with the idea of unexpectedness of events and surprise was then conceived as the gap between anticipated and actual outcomes, where the unexpected triggers the reaction to the violation of expectations. As a

result, surprise can be defined as a *subjective, sudden, and unexpected phenomenon influenced by changes in expectations, encompassing physiological and cognitive dimensions*.

Thereafter, it is natural to question what causes surprise. Each of us experiences a continuous sequence of events that shape our daily lives. We all possess a set of beliefs and expectations to apprehend and make sense of the world. Thus, surprise arises when there is a discrepancy between our beliefs and expectations and what actually occurs as an event. Following the literature on schema theory [Pidduck et al., 2020], surprise results in the unexpected violation of expectations or beliefs [Grace and Maher, 2015]. What then becomes crucial is the ability to react and adapt to this discrepancy as surprise introduces a disruption of ongoing processes and information processing [Reisenzein et al., 2017]. The constant possibility to deal with the unexpected results in a need for resilience and adaptability because of the discomfort and urgency that unexpected events can create within individuals.

If surprise has been studied extensively and piqued the interest of so many disciplines, it is partly because it can manifest in various circumstances. Surprise is not confined to just a few areas of human life; a surprising event can emerge, whatever the activity in question. In the context of this work, we are particularly interested in creative endeavours and how surprise would impact the performance of teams once a surprising event occurs during the collective creative process. Again, this study focuses mainly on one type of surprise, which is a change in the specifications of constraints imposed on teams.

As already mentioned, organisations evolve through constant uncertainty and face unexpected events throughout their existence, facing sudden changes from internal and external factors [Fey and Kock, 2022]. As a result, an organisation’s sustainability is determined by its ability to overcome surprise and adapt its behaviour and decisions to the changes stemming from the occurrence of an unexpected event. This ability can be defined as resilience. The term resilience has many definitions, and the rest of this work mainly focuses on recovery resilience, which is defined as the “*bouncing back to a state of normalcy*” [Boin and Van Eeten, 2013, p. 431]. The crucial aspect of resilience is the need for the team to respond in a timely manner and adaptively to the stimulus [Maynard et al., 2015]. As such, resilience can be envisioned as the response to adversity, which is associated with difficult or unpleasant situations and might operate at both the individual and team level introducing the term team resilience as the “*capacity of a team to withstand and overcome stressors*”

in a manner that enables sustained performance; it helps teams handle and bounce back from challenges that can endanger their cohesiveness and performance” [Alliger et al., 2015, p. 177].

In order to achieve resilience, teams can rely on the elements already mentioned in the previous section: their diversity, but also their processes. “*Collective responses to adversity require effective communication, collaboration, and coordination among team members*” [Hartwig et al., 2020, p. 18]. Indeed, the answer to surprise by teams strongly relies on the relationship quality that exists between their members. In addition to the processes just mentioned, team states are also crucial with high importance for trust, cohesion, and the stemming psychological safety [Hartwig et al., 2020]. Having a good quality relationship between team members offers the possibility to enhance the aforementioned states that contribute to enduring the pressure of challenging circumstances [Meneghel et al., 2016].

Returning to the more specific question of creativity, what do we already know about this issue? Creativity corresponds to the generation of novel and useful ideas Amabile and Pratt [2016], Boden [2004]. In addition, creativity relies on problem-solving [Reiter-Palmon and R. Murugavel, 2020]. This emphasis on problem-solving is central as problem-solving is also a metacognitive process triggered by surprise [Grace and Maher, 2015, Schön, 1987]. It becomes interesting to look at the relationship between creativity and surprise from this perspective. In fact, surprise challenges the creative process from the moment it results in a necessary reframing of the problem at the core of this creative process. In fact, “*the act of trying to solve an unexpected problem forces one to break loose and give way to a completely different way of approaching the task.*” [McDaniel et al., 2003, p. 272]. Then, if we focus on the creativity of an individual or a team, the question arises not just of their ability to generate new and useful ideas for a given problem but also of their ability to reformulate the problem when a surprising event is introduced into the overall creative process. This ability can be again translated into adaptability or resilience, which are key in overcoming a surprising event in order to pursue the course of events, possibly modified to integrate this element of surprise. If surprise challenges the creative process, it is a “*spontaneous process that produces some very profitable disruptions*” [Guastello, 1995, p. 320]. But does this mean that surprise is always good for creativity? This calls into question the valence of surprise. Although this work does not go into further detail on the discussion of surprise as an emotion, it does not ignore the fact that the concept of surprise can be linked to both

positive and negative valence. The crucial point here is the question of the entity perceiving the surprise. In fact, “*a welcoming attitude toward surprise can help in loosening the frame of established and traditional ways of thinking and doing things.*” [McDaniel et al., 2003, p. 272]. However, the valence of surprise, as well as its degree of importance and its relevance to the problem, are essential parameters to include in any analysis to fully grasp the intricacies of the relationship between creativity and surprise and have, for now, been rarely addressed.

In line with those considerations in terms of team creativity and apprehension of the unexpected, the impact of surprise on the team’s creative process leads to the following research questions: How will an unexpected event, caused by a change in constraints specification, impact a team’s creative performance according to the degree of surprise introduced in a task? And, how is this relationship between the resulting surprise and the team’s creativity affected by other factors, such as the team’s diversity, processes, and creative capacities?

4.3 Hypotheses

This study aims to test certain hypotheses on the link between surprise and a team’s creative performance. More specifically, our four hypotheses focus on the effect of surprise on creative performance, considering the degree of surprise, the perception of surprise, and the moderating effects of both the attitudes towards surprise and the perceived valence of the surprise experience.

First, we assume that the unexpectedness of an event might have a different impact on a team’s performance depending on its degree or magnitude. A team faced with an unexpected problem-relevant event, such as a change in constraints specification, is forced to update its patterns to overcome this surprise and integrate the necessary changes. Surprise introduces a sudden change that modifies the problem to be solved, forcing members to reframe the initial problem. As “*introducing output constraints may alter the conditions under which solutions are generated and, in turn, spark unconventional thinking and exploration of novel ideas*” [Acar et al., 2019, p. 110], we expect first a positive effect of surprise if this surprise is at a low degree. As teams would only perceive a low level of surprise, we assume that they would more easily overcome it. However, this positive relationship holds true until a certain threshold. If surprise might serve as a motivational impetus by eliciting curiosity [Reisenzein et al., 2017], the attentional shift and interruption of ongoing

processing might negatively affect creativity, as it could create a longer delay or difficulties in enabling the evaluation and adaptation process to update existing schemas if the surprise is too large. In the end, we expect to observe higher performances in the treatment group with low surprise compared to the control group but a higher or equal performance in the control group compared to the treatment group with high surprise.

Hypothesis 1: *An inverted U-shaped relationship emerges between the degree of surprise and the team's creative performance.*

We also anticipate an effect regarding the team's subjective perception of the change affecting the creative process. Specifically, we differentiate how team members perceive the change in constraints specification (how important this change was) versus the magnitude of surprise introduced by the protocol as a treatment. We propose that teams whose members perceive the change as more significant will perform worse, as the challenge of reframing the problem and adapting to the new conditions will be greater.

Hypothesis 2: *Perceiving the change in constraints specification as more important hinders the team's creative performance.*

Then, we also assume that the team's attitudes toward surprise outside the experimental session might influence their performance as a team. In fact, for both treatment groups experiencing surprise, we assume team members' positive attitude toward it will contribute positively to the team's performance. The literature on surprise does not present a consensus on whether surprise has a clear valence [Noordewier and Breugelmans, 2013]. In some cases or for some individuals, surprise might have either a positive or negative valence, influencing their behaviours when facing it. One classic example is the one of a surprise birthday party. If, in general, most individuals would perceive it as a positive event, others might dislike it, as they have a tendency to feel uneasy about it. As a result, individuals perceiving surprise as a positive phenomenon might respond better to an unexpected event, helping them to overcome it more easily and perform better in the task once the surprising element has been integrated into the process. At the team level, team members' aggregated attitudes toward surprise become crucial to understanding its effect on

their performance.

***Hypothesis 3:** Team members' positive attitudes toward surprise positively influence the impact of surprise on the team's creative performance.*

Finally, we wonder how the surprise experienced by teams during the experimental session impacts their performance. As mentioned above, literature has been developed on the valence of surprise, whether individuals perceive surprise positively or negatively. More importantly, Tan and Qu [2015] considers that creativity is influenced not only by valence but also by appraisal occurring when individuals assess the situations they encounter and create plans based on their evaluations. In our specific context, this translates into how team members perceive their own management and adaptation to unexpected events as well as the ones of other team members. It is not only about team members' general attitudes toward surprise but also how they experience it once surprise occurs. Ultimately, a positive surprise that satisfies or reassures team members while avoiding conflicts is closely linked to the overall team experience, including key processes and states like communication, cohesion, and trust, which are known to enhance team creativity [Carmeli and Spreitzer, 2009, Forsyth, 2021, Gilson and Shalley, 2004].

***Hypothesis 4:** Team members' positive experience of surprise positively influence the impact of surprise on the team's creative performance.*

4.4 Method

4.4.1 Subjects and procedure

122 teams of 3 members participated in our experimental protocol. The sample consists of 366 subjects (58.47% female, 41.53% male) with an average age of 22 years ($s.d. = 3.84$; $range = [17; 63]$). Subjects were recruited via the university ORSEE platform and we organized a total of 37 experimental sessions that took place at the *Social Interactions Lab* (Strasbourg, France). The subjects were divided into three groups (one control group and two treatment groups), as detailed in the section below. Table 4.1 presents some descriptive statistics of those sessions, differentiating

control and treatments.

Table 4.1: Descriptive Statistics of Experimental Sessions

	mean	s.d.	min	max	total
<i>For control group</i>					
Subjects	8.769	2.862	6	12	114
Groups	2.923	0.954	2	4	38
<i>For treatment 1</i>					
Subjects	11.182	1.401	9	12	123
Groups	3.727	0.467	3	4	41
<i>For treatment 2</i>					
Subjects	9.923	2.253	6	12	129
Groups	3.308	0.751	2	4	43
Total					
Subjects	9.892	2.436	6	12	366
Groups	3.297	0.812	2	4	122

In terms of payment structure, subjects received 20 euros (including 7 euros of show-up fee) for participating in the experiment. If most experiments are commonly based on performance, it has been observed that creative tasks are a specific type of task. Indeed, paying subjects according to their performance might create a crowding-out effect on their creative performance due to strong intrinsic motivations when they have to be creative Amabile et al. [1996].

4.4.2 Tasks

This experimental protocol consists of two different tasks, with the teams remaining the same across both tasks. Each experimental session lasted one hour, with 20 minutes for the first task, 30 minutes for the second, and 10 minutes to complete additional questionnaires. To start with, we need to distinguish the two creative tasks presented to subjects as they do not serve the same purpose within the protocol.

The schema-setting task

Individuals experience surprise when an event exceeds a certain degree of discrepancy compared to their beliefs or expectations. We refer to these beliefs and expectations as schemas. In order to observe this schema discrepancy in an experimental setting, we have two main possibilities: either we assume that the relevant schema is already established in our subjects' minds, or we need to install or elicit this specific schema during the experiment [Reisenzein et al., 1996]. In this work, the choice has been made to use the second strategy and advocate a first task for eliciting a specific schema. This is mainly justified in cases where the mechanisms the subjects will have to deal with within the experimental setting are not common elements in their everyday lives, as here, in the case of ideation activities. The literature refers to a repetition-change paradigm to describe such a strategy where subjects are first exposed to a baseline to then, in a second phase, experience a schema-discrepant event where their expectations are disconfirmed. In the end, the aim of this schema-setting task is to present a task devoid of surprise or unexpectedness, making the assumption that, while facing an additional task, subjects might not expect a change in the dynamic of this task. In other words, the schema settled does not aim to fix the theme but the sequence of the task. Regarding the task itself, teams had 20 minutes to select and elaborate on an idea to improve their daily life in their university [Baruah and Paulus, 2011, Harvey, 2013, Kohn et al., 2011, Maltese et al., 2023]. During the task, teams had to fill out an *idea sheet* presenting the final project with a title, a description of the idea or project, the advantages and disadvantages of it, and the resources required to implement it¹. Control and treatment groups completed the exact same task with no difference in duration or instructions.

The surprise task

This second task lasted 30 minutes, and teams were asked to fill out an idea sheet similar to the one presented to them during the first task. As this format is not commonly used by subjects, the first task also aimed to familiarise them with it so that they could complete it by minimizing the effect of this new format on the substance of the idea presented for the second task. Unlike the first task, this second task does not aim to establish a specific schema in the subjects' minds but to test the effect of unexpectedness on treated teams' creative performance. Concerning the instruction

¹See Appendices 4.7 for more details on the idea sheet format.

of the task itself, the modalities changed according to control and treatment groups. First, control groups had to “*Imagine and describe the dwellings of a city in the future, built underwater*”. No surprise was introduced in this task, meaning subjects should not experience any schema discrepancy compared to the first task. Then, we need to distinguish the two treatment groups according to the level of surprise elicited. If both treatment groups experienced the same change, i.e., introducing an additional constraint to their final task, they differ in their possible anticipation of such a change. For treatment 1, teams received hints in the instructions of a possible change during the task (low surprise)². For treatment 2, no hint has been given to them (high surprise). Starting with “*Imagine and describe the dwellings of a city in the future*”, both treatment groups experienced a change where the “*built underwater*” part was added to the instructions halfway through the task. The ultimate difference resides in the degree of surprise or unexpectedness of this change, as some teams have been warned about a possible change while others have not.

Both tasks presented to subjects are categorized as open tasks in the creativity literature, as no specific restrictions are applied except the time constraint. Those tasks are often referred to as *outside the box* tasks [Charness and Grieco, 2019]. It is important to emphasise that a choice was made in the difference in duration of the second task between the different treatments. Although the treated teams had 15 minutes with the final instruction, compared with 30 minutes for the teams in the control group, we felt it was important, above all, to maintain the same duration between the different treatments.

In addition to the data collected from these two tasks, the subjects were given two questionnaires. Firstly, at the beginning of the experiment, a questionnaire on their socio-demographic information and idiosyncratic features. Afterwards, at the end of the experiment, a questionnaire covering various themes: creativity, reaction to surprise (in everyday life and during the experiment), group processes, and pre-existing networks between our subjects³.

²See Appendices 4.7 for more details on the instructions presented to subjects.

³See Appendices 4.7 for more details on the questionnaires.

4.4.3 Measures

Creativity

Creativity is a multifaceted concept that can be approached from many different angles. In particular, evaluating something as creative mainly relies on selected criteria. Across the literature, several criteria have emerged to study the number of ideas generated by individuals or teams or the different themes they gather [Guilford, 1967]. However, when it comes to one specific idea or project, it is possible to evaluate it according to what the team is aiming for. Going back to the very definition of the concept, creativity consists in “*the production of novel and useful ideas by an individual or small group working together*” [Amabile and Pratt, 2016, p. 158]. Then, we can distinguish two components of creativity: its novelty and its usefulness. While novelty is often approached as the originality of an idea, usefulness is subject to more debate in the literature and is generally linked to the appropriateness, utility, quality, and effectiveness of an idea [Magni et al., 2023]. As a result, we make a distinction in creative performance between the originality of an idea, as creativity *per se*, and the feasibility of an idea, as problem-solving-oriented creativity, depending on whether the idea is implementable or not.

Based on the extended literature on creativity, the creative assessment of our two creativity criteria was done by external judges following the *Consensual Assessment Technique* [Amabile, 1982]. This technique states that an idea is considered creative (or not) when all external judges evaluate it independently and agree upon this evaluation. This subjective assessment of creativity is mainly done for sufficiently open-ended tasks such as our schema-setting task and surprise task, and we need to obtain an acceptable inter-judge reliability measured by Cronbach’s alpha [Ursachi et al., 2015]⁴. In the end, *feasibility*, defined as the practical possibility of implementing an idea, and *originality*, defined as the degree of novelty of an idea, consist of scores ranging from 0 to 10 attributed to each team. In total ten external judges rated teams’ ideas according to the feasibility ($\alpha_{Schema} = 0.544$ and $\alpha_{Surprise} = 0.543$) and originality ($\alpha_{Schema} = 0.696$ and $\alpha_{Surprise} = 0.677$) criteria. Table 4.2 presents information about the external evaluators, indicating whether they assessed the schema-setting task (Task 1) or the surprise task (Task 2). We also controlled for the accurate evaluation of ideas by regressing evaluators’ scores on the format of the

⁴Perry et al. [2004] states that a reliable measure of Cronbach’s alpha is over 0.6 [Ursachi et al., 2015], but that we can consider as acceptable an alpha 0.5 or greater.

Table 4.2: External Evaluators' Profiles

<i>Evaluator</i>	<i>Task</i>	<i>Gender</i>	<i>Position</i>
1	1	M	PhD Student
2	1	F	Associate Professor
3	1	M	Associate professor
4	1	F	PhD Student
5	1	M	Alumni of the University
6	2	M	Consultant in Innovation
7	2	M	PhD Student
8	2	F	PhD Student
9	2	F	Consultant in innovation
10	2	M	Alumni of the University

idea sheet provided to the teams. The results allow us to conclude that evaluators did provide attention to the elicited criteria as evaluators rated more positively the originality of ideas when a title ($p_{value} = 0.034$) was provided as well as the idea's description ($p_{value} = 0.009$) and strengths ($p_{value} = 0.080$) where subjects could defend the originality of their ideas. For feasibility, the scoring was positively related to the title ($p_{value} = 0.009$) and description ($p_{value} = 0.000$) but also to the idea's weaknesses ($p_{value} = 0.034$) where subjects could present the idea but also acknowledge its limits. In addition, the scores were diminished when teams provided the needs ($p_{value} = 0.035$) to implement their ideas, as this would naturally hinder feasibility.

While our measure of creativity is based primarily on the creativity scores attributed to each idea, creativity can also be measured based on the creative capabilities of individuals or teams involved in the creative process. This is why our work will also be enriched by declarative measures of creativity based on the general self-perceived creativity of group members and the creativity of teams as perceived by their members ([1*], [2*], [6*] to [9*], [23*]⁵).

⁵To be noted, here and in the following paragraphs [.] signals the question number so that readers can refer to the Appendices 4.7 for more details. If * is affixed next to the question number, then the question belongs to the post-questionnaire. Otherwise, the questions belong to the preliminary questionnaire. Both the preliminary and post-questionnaires use a Likert scales ranging from 1 to 5.

Surprise

If the literature section focused on the definition and stakes of surprise, there remains the question of how to detect and measure it. Our analysis is done based on declarative measures of surprise, which are the most used in surprise detection [Reisenzein et al., 2017]. More specifically, subjects answered two categories of questions about surprise and unexpectedness. On the one hand, subjects have been asked to answer questions on their general attitude towards surprise [3*] [4*] [5*]. More precisely, these questions aimed to investigate their self-perceived valence of surprise and their adaptability in everyday life or work situations.

On the other hand, subjects were asked to answer questions on how they experienced surprise during the second task of the experiment (only for treatment groups). Those questions encompass the perceived degree of surprise [18*], the valence associated with it [19*], their adaptability to the change [20*], and the adaptability of the other members of the group [21*]. In addition, questions were asked on the importance of the change of instructions [22*] and, related to creativity, if their self-perceived creativity evolved once the change occurred [23*].

Team relevant variables

First, questions have been asked on individuals' characteristics and idiosyncratic features that allows us to compute measures related to team diversity. In fact, the preliminary questionnaire can then be divided into three categories: socio-demographic information ([1] to [6]), personality traits using the Big Five Inventory with 10 items adapted to French (BFI-10) [Courtois et al., 2020] ([7]), and social attitudes on trust, risk, and preference to work in a group setting ([8] to [10]). The Blau Index, which *“quantifies the probability that two members randomly selected from a population will be in different categories if the population size is infinite or if the sampling is carried out with replacement”* [Solanas et al., 2012, p. 7] has also been computed for the relevant variables of gender, age, domain, personality traits, and self-declared creativity.

In addition, we focused on specific team processes that are supposed to affect teams' creative performance. Subjects answered questions on commitment [10*], satisfaction [11*] and [16*], trust [12*], reliability [13*], effective communication [14*], and cooperation [15*].

4.5 Results

In the following section, we present the results of our analysis as follows. First, we present the descriptive statistics in Section 4.5.1 and our treatment testing in Section 4.5.2. Then, in Section 4.5.3, we provide an exhaustive regression analysis to test our hypotheses and confirm the results arising from our descriptive analysis. The codification and definition of the variables presented in our results are depicted in the code book presented in Appendices 4.7.

4.5.1 Descriptive statistics

Our preliminary investigation began with a review of the descriptive statistics and the results of a balance test in order to confirm that the results were driven by the treatment effect and not by an imbalanced distribution of the relevant variables between the treatments. In Table 4.3, we provide an overview of our sample by depicting the descriptive parameters of the sociodemographic variables while Table 4.4 reports the balance test. These tests conclude that there are no statistically significant distributional imbalances; therefore, the results are robust, and the randomization is considered successful. Nevertheless, exceptions for age, diploma, and domain were found, as they were not distributed equally between treatment conditions. The results suggest that these imbalances in the distribution seem to be driven by T0 (control). Despite the existence of these imbalances, a variable that is highly correlated with the outcome of interest may be far more important than a large and significant imbalance of a variable that is uncorrelated with the variable of interest. Thus, analyses were conducted to determine whether the main variables (i.e., originality and feasibility in tasks 1 and 2) were affected by imbalanced variables. No significant effect emerged for *age* (feasibility task 1, $p_{value} = 0.982$; originality task 1, $p_{value} = 0.515$; feasibility task 2, $p_{value} = 0.263$; originality task 2, $p_{value} = 0.986$; Kruskal-Wallis equality-of-populations rank test), *diploma* (feasibility task 1, $p_{value} = 0.731$; originality task 1, $p_{value} = 0.939$; feasibility task 2, $p_{value} = 0.437$; originality task 2, $p_{value} = 0.345$; Kruskal-Wallis equality-of-populations rank test), or *domain* (feasibility task 1, $p_{value} = 0.133$; originality task 1, $p_{value} = 0.464$; feasibility task 2, $p_{value} = 0.877$; originality task 2, $p_{value} = 0.153$; Kruskal-Wallis equality-of-populations rank test).

Table 4.3: Summary of the Descriptive Statistics

Variables	T	N	mean	s.d	min	max
Gender	T0	114	0.5614	0.4984	0	1
	T1	123	0.6098	0.4898	0	1
	T2	129	0.5814	0.4952	0	1
	Total	366	0.5847	0.4934	0	1
Age	T0	114	22.0301	3.8487	17	63
	T1	123	0.5285	0.5012	0	1
	T2	129	0.4651	0.5007	0	1
	Total	366	0.4672	0.4996	0	1
French speaker	T0	114	0.8158	0.3893	0	1
	T1	123	0.8455	0.3628	0	1
	T2	129	0.8372	0.3706	0	1
	Total	366	0.8333	0.3731	0	1
French nationality	T0	114	0.7895	0.4094	0	1
	T1	123	0.8211	0.3848	0	1
	T2	129	0.8682	0.3395	0	1
	Total	366	0.8279	0.3780	0	1
Diploma	T0	114	1.535	0.6935	1	4
	T1	123	1.813	0.8031	1	4
	T2	129	1.667	0.7637	1	3
	Total	366	1.675	0.7626	1	4
Domain	T0	114	2.947	1.174	1	5
	T1	123	3.341	1.3356	1	5
	T2	129	3.364	1.3046	1	5
	Total	366	3.227	1.286	1	5
Trust	T0	114	3.114	1.079	1	5
	T1	123	3.146	0.997	1	5
	T2	129	3	1.038	1	5
	Total	366	3.085	1.037	1	5
Risk	T0	114	3.061	1.083	1	5
	T1	123	2.870	1.007	1	5
	T2	129	2.953	1.0297	1	5
	Total	366	2.959	1.039	1	5

See Appendices 4.7 for more details on the variables.

4.5.2 Treatment testing

In this section, we test the *Hypothesis 1* presented in Section 4.3 that relates to the role of surprise as our treatment effect. We begin by depicting a violin plot,

Table 4.4: Balance Tests

<i>Variables</i>	All treatment		T0 vs T1		T1 vs T2		T0 vs T2	
	<i>Chi2</i>	<i>P-value</i>	<i>Chi2</i>	<i>P-value</i>	<i>Chi2</i>	<i>P-value</i>	<i>Chi2</i>	<i>P-value</i>
Gender (female=1)	0.58	0.749	0.57	0.451	0.21	0.647	0.10	0.754
Age	3.70	0.157	3.69	0.055*	1.01	0.316	0.93	0.335
French speaker	0.40	0.820	0.37	0.542	0.03	0.857	0.19	0.660
French nationality	2.69	0.261	0.38	0.539	1.06	0.303	2.66	0.103
Diploma	7.53	0.023**	7.61	0.006***	2.15	0.143	1.68	0.196
Domain	8.29	0.016**	5.85	0.016**	0.01	0.906	6.86	0.009***
Trust	1.18	0.554	0.02	0.891	1.03	0.309	0.69	0.408
Risk-loving	1.71	0.425	1.62	0.203	0.13	0.721	0.86	0.355

which shows the distribution and mean of team originality and feasibility for the surprise task in Figure 4.1, and we distinguish by treatments. First, for feasibility, we encounter mild evidence⁶ of a differences in performance between the control group and teams assigned to treatment 1 ($p_{value} = 0.0934$)⁷ and evidence of a difference in performance between the control group and teams assigned to treatment 2 ($p_{value} = 0.0119$). However, we do not find evidence of a difference in performance between the two treatments ($p_{value} = 0.3640$). Then, for originality, we found strong evidence of a difference in performance between the control group and teams assigned to treatment 1 ($p_{value} = 0.0001$), as well as compared to treatment 2 ($p_{value} = 0.0001$). However, we do not find either evidence of a difference in performance between the two treatments ($p_{value} = 0.4959$). Taking into account those results, one might wonder if this non-significant difference in mean between our two treatments, which is consistent over both criteria, is not due to the perception of subjects of those two levels of surprise as equivalent. By comparing the mean perceived degree of surprise, we find strong evidence ($p_{value} = 0.0001$) of a significant difference in the degree of perceived surprise, with subjects assigned to the high surprise treatment effectively perceiving a higher degree of surprise than subjects assigned to the low surprise

⁶All along our analysis, we implement the terminology presented by [Moffatt, 2020], which states the following: there is mild evidence if $p \leq 0.10$; there is evidence of $p \leq 0.05$; there is strong evidence if $p \leq 0.01$.

⁷The mean team feasibility in task 2 is: T0= 6.37, T1=5.93, and T2=5.81. While team means originality in task 2 is: T0=5.55, T1=4.32, and T2=4.42.

treatment ($mean_{low} = 2.748$ and $mean_{high} = 3.364$). In addition, when subjects are asked about the importance of the change they experienced, they perceive it as being equally important ($mean_{lowsurprise} = 3.967$ and $mean_{highsurprise} = 3.984$), reinforcing the initial assumption that the true effect is due to surprise rather than the change in constraint itself. In the end, we make the assumption that perceiving surprise did have a negative impact on teams' performance compared to teams that did not experience surprise but that the degree of surprise itself does not result in differences in performance. This first assumption will receive further attention in Section 4.5.3 as we will investigate the determinants of teams' performance in more detail.

Figure 4.1: Violin Plot of Feasibility and Originality Scores by Treatment



4.5.3 Regression Analysis

In this section, we present the results of our regression analysis⁸ that aims to investigate the relationship between surprise and creativity, divided into feasibility and originality, and several relevant explanatory variables based on the hypotheses formulated earlier and the relevant concepts found in the literature. To do so, we divided our analysis into two steps. First, two regressions aim to understand the treatment effect of the surprise on both feasibility and originality, moderated by our team-relevant variables. Then, focusing on the sub-sample of our population that did experience surprise, we analyze the effect of the experience of surprise on their

⁸The analysis was carried through the STATA software.

performance. In both cases, we decided to implement a Tobit model, which allows us to censor the extremes, in this case, the minimal and maximal score for each measure of creativity.

The effect of surprise

First, our regression shows that the surprise treatment, whether low or high, had no effect on the score awarded by the evaluators on teams' **feasibility** scores as such. If this effect appeared in our first investigation of the results, it seems that the mild evidence put forward has disappeared as a result of the more detailed analysis and the addition of the moderators in our regression. The following paragraph depicts our results according to the subsequent categories: attitude towards unexpected events, team processes, diversity, idiosyncratic features, and creative performance.

Considering attitude towards unexpected events, we observe evidence for a negative effect of the positive perception of unexpected events in a general context on the overall team performance ($p_{value} = 0.026$). While this result may seem counter-intuitive at first glance, we assume that an individual with a positive perception of an unexpected event tends to be open to the unexpected, which might impair his or her ability to properly assess the ins and outs of this event, diminishing the benefit of critical thinking. Paradoxically, this infatuation or surge of curiosity becomes counterproductive for teams when it comes to providing a feasible, implementable idea. We did not find additional evidence for the effect of attitudes towards surprise on ideas' feasibility.

Then, we look at the influence of team processes on teams' feasibility scores. To start with, we found evidence for cooperation to benefit feasibility ($p_{value} = 0.047$), which goes in line with the literature since cooperation allows the team to work conjointly on the task and ideate as one to target a good idea and improve it. However, other team processes were demonstrated to be detrimental to teams' feasibility scores. On the one hand, we observe mild evidence that team members who declared that other team members could rely on them contributed negatively to the performance of their team ($p_{value} = 0.090$). In other words, we observe a negative effect when team members consider themselves as more reliable. This result might seem counter-intuitive since greater reliability should benefit team processes and creativity as it provides more psychological safety. Our interpretation behind this result is that our measure is based on the subject's own perception of their reliability and not the evaluation from other team members, which might bias it. We need then

to be cautious while interpreting this negative relationship. On the other hand, we observe mild evidence that more openness of communication also hinders feasibility ($p_{value} = 0.095$) with the underlying assumption that too open of a conversation might impeach teams to converge towards one sole and unique idea to then enrich it and ensure its possible implementation. As explained by Marks et al. [2001], teams benefit from effective communication. In our case, however, we assume that team communication was not sufficiently effective due to being overly open. These results on team processes mitigate the results gathered over time in the literature, showing a generally positive effect on the performance of good-quality interactions. In our case, what really matters is the cooperation between team members, while reliability and openness of communication hinder feasibility. We did not find additional support for the effect of team processes on feasibility.

Continuing with diversity, we observe respectively mild evidence and evidence of two opposite effects of diversity as gender diversity harms feasibility ($p_{value} = 0.057$) while openness to experience does improve the teams' feasibility score ($p_{value} = 0.038$). These results are important as they do not measure the mean of the variable of interest for each team but the balance between the different profiles within this team. First, we find a negative relationship between gender diversity and creative performance. Quite interestingly, for gender diversity, we do find a significant effect on creative performance even though this difference was not accentuated by the protocol itself, going against the result of Pearsall et al. [2008], who underlined the need to activate faultlines for gender diversity to have a significant effect. This result allows us to conclude that more gender diversity in a team will harm its performance in feasibility. While gender diversity benefits team performance as it leverages their potential through information elaboration, Kearney et al. [2022] find a negative effect of time pressure on the impact of gender diversity as it causes team withdrawal, disengagement, and lower contribution to team discussions. As a result, it might be possible that the time allocated for the task resolution had an impact on this relationship between gender diversity and performance. Additionally, other avenues should be explored in the literature on conflict, where gender diversity might cause harm to teams as a result of disagreements between members. With no specific measure of conflict, we can not test this hypothesis. Then, for personality diversity, we only find significant results for openness to experience but no evidence for other factors of conscientiousness, extroversion, agreeableness, and neuroticism. This positive effect of openness to experience diversity demonstrates that more diverse teams in

this variable will perform better. This mitigates the literature on the effect of personality traits on creative performance, as it was first assumed that a higher mean of these attributes were associated with higher performance, with some nuances when looking at the diversity index. It seems that diversity, also in terms of personality, benefits teams as it brings different perspectives and positions of team members. Besides gender and openness to experience, we did not find other significant diversity effects across our three categories of socio-demographic, personality, and creativity variables.

Additional controls have been added to our analysis encompassing subjects' idiosyncratic features. We find strong evidence that individuals inclined to trust others perform better as a team ($p_{value} = 0.002$). We hypothesize that more trustful behaviour benefits the team by providing an initial inclination to cooperate with others, reinforcing the dynamic highlighted above. However, we also find evidence of a positive effect of risk-lover behaviour on feasibility ($p_{value} = 0.029$). While we might have expected to favour low-risk behaviour in order to encourage an idea to be implemented, we observe an inverse relationship, with evidence of a positive effect on the feasibility of individuals' propensity to take risks. Here, we assume that a preference for risk contributes to one's resilience as individuals are more at ease to evolve in an ambiguous environment, which contributes to their ability to overcome surprise.

Finally, if we look at the creative performance across tasks and criteria, we find a positive relationship between feasibility in the schema-setting task and the surprise task ($p_{value} = 0.050$), acknowledging that teams that provided more implementable ideas during the first task did perform well in the feasibility score in the second task. As feasibility partly resides in teams' characteristics and relies on ongoing processes, we observe a natural relationship between the ability of teams to provide implementable ideas in the first and the second tasks. Additionally, feasibility and originality evolve in the same direction ($p_{value} = 0.000$) as teams performing well in originality in the surprise did perform well in feasibility. This result goes against the traditional literature on creativity as it should appear in the opposite direction between feasibility and originality.

CHAPTER 4. DOES CREATIVITY THRIVE ON PLOT TWISTS? EXPLORING THE ROLE OF SURPRISE ON TEAM CREATIVITY

Table 4.5: Tobit Regression for the Effect of Treatment on Feasibility

	Model 1	Model 2	Model 3	Model 4	Model 5
Treatment (low surprise)	-0.439 (0.345)	-0.469 (0.337)	-0.312 (0.333)	-0.279 (0.310)	0.143 (0.296)
Treatment (high surprise)	-0.568 (0.345)	-0.600* (0.342)	-0.573* (0.328)	-0.598* (0.315)	-0.148 (0.309)
Valence on unexpected events	-0.178* (0.107)	-0.170 (0.108)	-0.144 (0.107)	-0.222* (0.115)	-0.240** (0.108)
Adaptation to unexpected events (daily life)	-0.0216 (0.104)	0.0218 (0.104)	0.00634 (0.101)	0.0113 (0.0986)	-0.00653 (0.0925)
Adaptation to unexpected events (work life)	0.0887 (0.114)	0.0719 (0.115)	0.0646 (0.116)	0.00955 (0.108)	-0.0211 (0.102)
Commitment		0.0106 (0.0997)	0.00272 (0.0950)	-0.0244 (0.103)	0.0118 (0.0957)
Recommendation		-0.0150 (0.109)	-0.0320 (0.103)	-0.0484 (0.0945)	-0.157 (0.0964)
Trust in teammates		0.213 (0.137)	0.220* (0.129)	0.176 (0.131)	0.148 (0.117)
Reliability		-0.213* (0.123)	-0.230** (0.109)	-0.242** (0.106)	-0.159* (0.0935)
Open communication		-0.335* (0.184)	-0.364** (0.173)	-0.370** (0.165)	-0.205* (0.123)
Cooperation		0.441** (0.172)	0.376** (0.168)	0.317* (0.163)	0.273** (0.137)
Satisfaction		0.0504 (0.168)	0.0541 (0.148)	0.0617 (0.141)	0.0985 (0.120)
BI_Gender			-0.790*** (0.266)	-0.702*** (0.262)	-0.500* (0.262)
BI_Age			-0.331 (0.314)	-0.273 (0.306)	-0.0610 (0.284)
BI_Domain			-0.268 (0.387)	-0.289 (0.382)	-0.258 (0.347)
BI_Openness to Experience			0.654* (0.387)	0.646* (0.382)	0.621** (0.347)

Continued on next page

CHAPTER 4. DOES CREATIVITY THRIVE ON PLOT TWISTS? EXPLORING THE ROLE OF SURPRISE ON TEAM CREATIVITY

Table 4.5: Tobit Regression for the Effect of Treatment on Feasibility (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5
			(0.367)	(0.366)	(0.299)
BI_Conscientiousness			0.0738	0.118	0.282
			(0.295)	(0.283)	(0.277)
BI_Extroversion			0.117	0.0812	0.215
			(0.271)	(0.260)	(0.229)
BI_Agreeableness			0.145	0.239	0.252
			(0.271)	(0.264)	(0.257)
BI_Neuroticism			-0.128	-0.127	-0.191
			(0.306)	(0.307)	(0.281)
BI_Creativity			0.350	0.291	0.437
			(0.324)	(0.314)	(0.302)
Individual creativity (Task 1)				0.177**	0.123
				(0.0863)	(0.0812)
Collective creativity (Task 1)				-0.120	-0.117
				(0.0877)	(0.0823)
Individual creativity (Task 2)				-0.00580	-0.00218
				(0.0865)	(0.0762)
Collective creativity (Task 2)				0.196*	0.121
				(0.111)	(0.0975)
Trust				0.189**	0.204***
				(0.0735)	(0.0640)
Risk				0.119	0.163**
				(0.0829)	(0.0743)
Preference to work in group				-0.240	-0.196
				(0.175)	(0.154)
Originality (Task 2)					0.407***
					(0.0853)
Feasibility (Task 1)					0.174**
					(0.0884)
Constant	6.647***	5.765***	6.395***	5.675***	1.656
	(0.488)	(0.698)	(1.049)	(1.030)	(1.208)

Continued on next page

Table 4.5: Tobit Regression for the Effect of Treatment on Feasibility (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5
Observations	364	360	360	358	358
R^2	0.0077	0.0170	0.0429	0.0521	0.1085
F	1.175	1.043	1.362	1.979	3.208

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We can then examine the effect of surprise on the teams' creative performance in terms of **originality**. Contrary to what was observed in the feasibility case, the surprise treatment affected the originality scores of the teams compared with the teams that did not experience surprise. More specifically, we observe strong evidence, for both treatments, of a negative impact of experiencing surprise on the creative performance of teams ($p_{value} = 0.000$). However, as has been highlighted earlier, there is no significant difference between the two treatments. In the end, it seems that the sole experience of surprise affects originality, but its degree does not significantly moderate the effect.

Compared to feasibility, we do not observe any impact of subjects' attitudes towards surprise on their performance in originality, with all variables of interest showing no significant results.

However, we do observe some evidence of a positive impact of team processes on creative performance in originality. More specifically, team members who declared that they could praise this group to their friends as an excellent team to work with contributed positively to the performance. This provides mild evidence of a positive effect of overall positive team processes on originality ($p_{value} = 0.061$).

Moreover, we do not observe any significant effect of variables stemming from diversity but we do observe mild evidence of a negative impact of trust on performance in terms of originality ($p_{value} = 0.055$). As for feasibility, the result on trust is counter-intuitive as we would have expected more trusting individuals to perceive a safe team environment more easily, contributing to expressing more unusual ideas. Here, we find the opposite relationship, with the potential explanation being that the tendency to trust others has a negative effect on the ability to act proactively in the team, which diminishes the investment of team members in the creative process and possibly hurts their intrinsic motivations to contribute which is essential to perform.

Finally, we do observe significant relationships of performances across tasks and criteria. We find mild evidence of a positive relationship between teams' performance in the schema-setting task and the surprise task ($p_{value} = 0.089$). Comparing feasibility and originality for the same task, we observe strong evidence of a positive relationship between the two criteria ($p_{value} = 0.000$).

Table 4.6: Tobit Regression for the Effect of Treatment on Originality

	Model 1	Model 2	Model 3	Model 4	Model 5
Treatment (low surprise)	-1.227*** (0.331)	-1.280*** (0.324)	-1.253*** (0.324)	-1.270*** (0.321)	-1.183*** (0.285)
Treatment (high surprise)	-1.138*** (0.322)	-1.208*** (0.319)	-1.206*** (0.314)	-1.223*** (0.307)	-1.112*** (0.268)
Valence on unexpected events	0.0265 (0.0987)	0.0270 (0.0985)	0.000296 (0.0938)	0.0240 (0.0996)	0.111 (0.0949)
Adaptation to unexpected events (daily life)	0.0720 (0.105)	0.0994 (0.101)	0.0755 (0.0930)	0.0804 (0.0920)	0.0688 (0.0829)
Adaptation to unexpected events (work life)	-0.00559 (0.111)	-0.0232 (0.111)	0.00180 (0.101)	-0.0122 (0.106)	-0.0426 (0.0999)
Commitment		-0.0160 (0.110)	-0.00221 (0.108)	-0.00850 (0.112)	0.00175 (0.0980)
Recommendation		0.258** (0.115)	0.230** (0.104)	0.198* (0.105)	0.191* (0.102)
Trust in teammates		0.0216 (0.117)	0.0393 (0.115)	0.0588 (0.116)	-0.0186 (0.113)
Reliability		-0.180 (0.116)	-0.185* (0.110)	-0.217* (0.111)	-0.116 (0.0979)
Open communication		-0.192 (0.182)	-0.212 (0.181)	-0.226 (0.182)	-0.0529 (0.139)
Cooperation		0.0528 (0.179)	0.0259 (0.186)	-0.000360 (0.189)	-0.194 (0.161)
Satisfaction		-0.0995 (0.166)	-0.122 (0.156)	-0.128 (0.154)	-0.115 (0.128)
BI_Gender			-0.527**	-0.494*	-0.287

Continued on next page

CHAPTER 4. DOES CREATIVITY THRIVE ON PLOT TWISTS? EXPLORING THE ROLE OF SURPRISE ON TEAM CREATIVITY

Table 4.6: Tobit Regression for the Effect of Treatment on Originality (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5
			(0.252)	(0.253)	(0.248)
BI_Age			-0.484*	-0.457	-0.334
			(0.292)	(0.287)	(0.267)
BI_Domain			0.00643	-0.0140	0.182
			(0.358)	(0.341)	(0.324)
BI_Openness to Experience			0.145	0.143	-0.0466
			(0.354)	(0.351)	(0.305)
BI_Extroversion			-0.192	-0.206	-0.264
			(0.281)	(0.283)	(0.253)
BI_Conscientiousness			-0.273	-0.260	-0.228
			(0.266)	(0.266)	(0.250)
BI_Agreeableness			-0.113	-0.0827	-0.143
			(0.336)	(0.336)	(0.311)
BI_Neuroticism			0.0976	0.131	0.133
			(0.298)	(0.295)	(0.271)
BI_Creativity			-0.0879	-0.140	-0.187
			(0.291)	(0.292)	(0.279)
Individual creativity (Task 1)				0.0907	0.0129
				(0.0796)	(0.0815)
Collective creativity (Task 2)				-0.00453	-0.00340
				(0.108)	(0.0895)
Individual creativity (Task 2)				0.0596	0.0733
				(0.0809)	(0.0735)
Collective creativity (Task 2)				0.105	0.0694
				(0.102)	(0.0892)
Trust				-0.0328	-0.116*
				(0.0700)	(0.0602)
Risk				-0.0617	-0.0868
				(0.0867)	(0.0774)
Preference to work in group				-0.0956	-0.000386
				(0.185)	(0.165)
Feasibility (Task 2)					0.371***

Continued on next page

Table 4.6: Tobit Regression for the Effect of Treatment on Originality (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5
Originality (Task 1)					(0.0839)
					0.222*
					(0.130)
Constant	5.223***	6.025***	7.441***	7.257***	4.047***
	(0.465)	(0.829)	(0.998)	(0.984)	(1.082)
Observations	364	360	360	358	358
R^2	0.0358	0.0443	0.0610	0.0640	0.1222
F	3.962	3.042	2.310	2.029	3.357

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The experience of surprise

Since we have examined the treatment effect of surprise on feasibility and originality, we need to devote our attention to the experience of surprise itself, focusing only on a sub-sample of our subjects. Comparing feasibility and originality again, we do observe some differences in the determinants of creative performance when it comes to the experience of surprise.

For feasibility, we find strong evidence of a positive effect of the reaction of team members to the unexpected event ($p_{value} = 0.006$). In fact, subjects who perceived other teammates' reactions as fast and adequate performed better as a group. This result is central in the analysis of the relationship between creativity and surprise since an unexpected event disrupts ongoing processes, and this interruption delays the continuation of the task and requires teams to adapt and update their schema and their understanding of the task to pursue it. In terms of interpersonal interactions, this result also hints that team members were reassured by the response and behaviour of other team members, which certainly facilitated the performance of the team. This perception of fast and adequate reactions is also related to trust in other team members, knowing that trust positively affects team performance. Going in the same direction, we find evidence of a positive effect of self-declared creativity ($p_{value} = 0.034$) as soon as subjects perceived their degree of creativity higher after the change of constraints stemming from the unexpected event. Here, team members

who were more confident in their creative performance after the surprise, i.e., those who positively perceived their performance post-surprise, performed better. This translates a positive experience of surprise since unexpectedness did not shadow the creative process by providing a negative feeling towards it.

Table 4.7: Tobit Regression for Feasibility and Originality (Experience of Surprise)

	Feasibility	Originality
Degree of surprise (treatment)	0.0401 (0.420)	-0.00520 (0.370)
Degree of surprise (perceived)	-0.0535 (0.124)	-0.0918 (0.101)
Valence	-0.0310 (0.0973)	-0.101 (0.104)
Reaction (individual)	-0.0838 (0.160)	0.189 (0.139)
Reaction (team)	0.382*** (0.138)	0.103 (0.138)
Importance of change	0.00564 (0.106)	0.0458 (0.114)
Creativity after surprise	0.608** (0.286)	0.445* (0.231)
Similarity of ideas	-0.258 (0.607)	-0.239 (0.472)
Constant	4.594*** (1.050)	3.476*** (0.935)
Observations	221	221
R^2	0.1643	0.1582
F	1.888	1.530

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Then, for originality, the only significant relationship occurs with the perception of subjects of their creativity after the change with mild evidence for a positive effect ($p_{value} = 0.056$) following the previous argumentation. However, no other variables

are significant in understanding the surprise experience.

4.6 Discussion and conclusion

As pointed out in the preamble to this article, surprise, although sometimes dreaded, is an integral part of our lives. This applies to all contexts, including when we must carry out tasks involving others. Among the collective activities of interest, we have decided to focus on collective creative endeavours and how a team will perform when put in front of an unexpected change in task constraints. As previously emphasized, creativity is essential to innovation, and teams are a fundamental element for creativity and innovation [Wuchty et al., 2007]. In addition, organisations increasingly prioritise employees' creative and innovative behaviours to develop a competitive advantage [Açıkgöz and Günsel, 2016, Bammens, 2016]. Therefore, it is crucial to understand the potential impact of surprise on creativity, as the reaction to surprise and unexpectedness becomes a cornerstone of an organisation's survival. This led us to the following research questions: How will an unexpected event, caused by a change in constraints specification, impact a team's creative performance according to the degree of surprise introduced in a task? And, how is this relationship between the resulting surprise and the team's creativity affected by other factors, such as the team's diversity, processes, and creative capacities?

To address these questions, we empirically contribute to the literature on surprise and creativity by implementing an experimental protocol, which compares the creative performance of teams experiencing no surprise, low surprise, or high surprise. This study allows us to conclude the following results. First of all, we observe that, depending on the creativity criteria selected, the impact of surprise will be different. In fact, if experiencing a surprising event significantly impacts team performance in terms of originality, it seems that this same event does not affect feasibility scores once we control for team characteristics and experience. Furthermore, surprise hinders originality with no distinction of the degree of surprise. As we see a significant effect of our treatments on the originality criteria, we conclude that the collective creative process of teams when it comes to finding a novel idea is affected negatively by surprise. Our interpretation behind these results is that an idea's feasibility or implementability relies more on teams' inherent capacity to collectively elaborate a creative solution and perform in a problem-solving activity. Whereas originality requires teams to take part in an advanced ideation process eliciting divergent

and convergent thinking and forcing them to avoid common solutions. We assume that surprise will have more impact on the originality part of the creative process as this will cause an interruption in the current ideation process but only partially affect the inherent capacity of teams to converge towards an implementable idea. As a result, *Hypothesis 1* is only partially supported as we do observe a negative impact of surprise on creativity but do not observe a first improvement justifying an inverted U-shaped relationship. Moreover, we can not differentiate the effect of surprise according to the degree of surprise since there is no significant effect of self-declared surprise and a possible negative relationship with performance. To conclude on the effect of surprise, *Hypothesis 2* examines whether differences in the perceived importance of changes introduced by surprise have a significant impact on team performance. As our regressions show no significant effect of the variable accounting for the importance of the change, we are not able to conclude on this matter. If this result suggests that the importance of change plays no specific role in the relationship between surprise and creative performance, we prefer to be cautious and highlight possible adjustments to our protocol to better capture this effect. In fact, we are convinced that further work should be led on the importance of change, introducing different types of changes and varying their possible importance to see if there is a possible significant impact on the creative performance of teams.

Result 1: *Surprise negatively affects a team's creative performance when it comes to the originality or novelty of an idea.*

Afterwards, we were interested in how subjects' attitudes towards surprise might affect their performance as a team. While *Hypothesis 3* stated that positive attitudes towards surprise would contribute to teams' performance, we found more nuanced results. On the one hand, we only found a significant effect of those attitudes on the feasibility criteria and not the originality criteria. On the other hand, within the feasibility criteria, we observe a negative impact on the feasibility of a positive attitude, i.e. seeing an unexpected event as something positive. Since we only found a significant effect of one variable, we will be cautious and refrain from asserting that attitudes have a general negative impact as a moderator of the effect of surprise and consider that there is no support for *Hypothesis 3*.

Result 2: *Team members' attitudes toward surprise only affect the creative performance of a team in feasibility.*

Finally, we checked how subjects experienced surprise (positively or negatively) and how it affected their creative performance as a team. Our results partly support *Hypothesis 4* as only a few variables arise as significant when we consider our sub-sample of subjects that experienced surprise. For feasibility, subjects that considered their teammates' reaction to surprise as fast and adequate did perform better as a group. In addition, subjects who positively perceived the effect of the change on their performance did perform better. However, only the last element is significant for originality. Again, we find mixed results between feasibility and originality, but more importantly, we do not find strong effects of the selected variables on team performance. As a result, we only find partial support for *Hypothesis 4*.

Result 3: *Team members' positive experience of surprise affects both team creative performance for feasibility and originality.*

Overall, these results lead us to believe that the effect of surprise on team creative performance is resolutely intertwined with the creativity criteria and, by extension, the objective chosen to assess teams' performance. A team's capacity to converge and find a feasible and implementable idea or solution mostly resides in internal factors that weaken the effect of a surprising event. On the contrary, a team's ability to provide original ideas is highly sensitive to surprise as it disrupts the ideation process. This perturbation is the central element, while other internal factors are therefore diminished.

Besides the aforementioned main results on the relationship between team creativity and surprise, our analysis also provides one additional result that would require closer attention in future work. Somewhat counterintuitively, we found a positive effect of a preference for risk on feasibility. If we could expect the reverse relationship with more cautious individuals to be able to provide more feasible ideas, it seems that more risky behaviour is valuable in teams targeting the implementation of their ideas. Our interpretation behind these results that go against the traditional literature promoting risk for divergent thinking and outside-the-box types of answer that would favor originality [Kleiman, 2008, Toh and Miller, 2016], is that the surprise element truly matters here. Within this already fuzzy literature on the relationship between risk and creativity [Crepaldi et al., 2024], we assume that more risk-taking behaviour translates into lower discomfort when facing uncertainty and ambiguity, like here with surprise. The risky behaviour becomes a strength in maintaining focus on the problem and alternative paths [Allwood and Selart, 2001, Tegano, 1990], and

the external factor of surprise overturns the expected negative relationship between feasibility and risk. We remain cautious in our conclusions since this relationship was not the core element that we were interested in, and specific experimental work should be done on this possible moderating effect of surprise, including the measurement of risk through specific tasks and not questionnaire-based measures.

All these results demonstrate high relevance for organisations since they evolve in “*highly volatile and uncertain times, [where they] need to develop a resilience capacity which enables them to cope effectively with unexpected events, bounce back from crises, and even foster future success*” [Duchek, 2020, p. 215]. In fact, one of the ways in which organisations evolve and develop is through their innovation strategies and creative activities. With this in mind, it is important to understand individuals’ and teams’ behaviours as a reaction to unexpectedness and surprise. As “*every organisation has to cope with the need to deviate from what is planned or expected in response to disruptive events*” [Cotta, 2021, p. 15], our results do provide insights on what organisations should be aware of when constituting creative teams that might, as any team, encounter unexpected events. First, depending on the objective of the creative sessions or projects, whether the organisation is more interested in favoring implementability or novelty, team managers should consider that an unexpected event will not have the same repercussions. Indeed, greater attention should be paid to the possible negative impacts of the surprising event on the originality of ideas. This does not prevent managers or leaders from ensuring the feasibility of the project by promoting the team processes that are crucial to the creative process. Moreover, teams should pay attention to each member’s experience with surprise prior to the teamwork and during teamwork. As the reaction to surprise is initially an individual and subjective process, understanding each member’s reaction and how those personal views interact is at the core of our apprehension of the relationship between creativity and surprise. Team managers or leaders might favor groups where negative experiences with surprise would be balanced by members with a more positive view of it.

Although this work contributes empirically to the literature on surprise, creativity, and their relationship, it also presents certain limitations. First, as mentioned earlier, our analysis is based on declarative measures of surprise, which are the most commonly used in surprise detection [Reisenzein et al., 2017]. However, this requires subjects to answer after the surprising event occurred and only when the task ended. This type of measure brings limitations since it is not objective and does not directly

collect information on surprise perception and reaction. Then, we did not observe a significant difference between our two treatments, i.e., low and high surprise. If one explanation is that teams are not affected by the degree of surprise but only by the surprise itself, then one of our interpretations is that the surprise should be greater. Our protocol uses the change in constraints specification as a surprising element. It would then be interesting to introduce a more significant change with heavier consequences on the resolution of the creative problem. For example, instead of adding the constraint that the city should be built underwater, testing the transition from a focus on dwellings to a focus on transport would be possible, questioning the malleability of the imposed constraint [Medeiros et al., 2014]. This would allow us to question the importance of change, delving into the role of change magnitude in the relationship between surprise and creativity. But, we could also imagine introducing other types of events, not only on constraints but also on resources or team membership. There are plenty of possible events that may create surprise. As an insight for future research, we consider three main types of possible unexpected events affecting team creativity: internal factors events, external factors events, and task-relevant. Moreover, this work has studied *ad-hoc* teams, not pre-existing ones. If ad-hoc teams were closer to the literature on temporary teams, there would still be important differences in how team members would interact. As a result, one of the limits of this study is that we do not capture the possible impact of medium or long-term relationships between team members that might affect, positively or negatively, their performance and the creative process itself. One assumption might be that teams that already worked on a similar type of activity would overcome a surprising event more easily as they might experience more associative processes. Finally, the effect of surprise on the team's creative performance is certainly strongly related to the chosen task. In this work, we provided teams with an open task that required them to complete an idea sheet as a detailed description of their idea. One possibility would be to adapt a similar protocol to other types of creative tasks that may vary according to their degree of complexity. For example, one could examine the impact of surprise in a prototyping task, adding more complexity and highlighting another phase to the creative process also involving more tangible creative outputs.

Given the lack of empirical results on the more general question of creativity and surprise, it seems important to us that future research should address the limitations set out at the end of this work but also explore new avenues. One salient element would be to consider at what moment the unexpected event would occur. As stated

by Alliger et al. [2015], team resilience can be articulated around three dimensions: minimizing behaviour, managing behaviour (i.e., coping and recovering from adversity), and mending behaviour. Also, the creative process can be articulated around the three main phases of preparation, response generation, and response validation or evaluation [Amabile, 1983]. Introducing surprises at different stages of the team and creative process and triggering different resilient behaviours would provide a deeper understanding of how organisations perceive unexpected events and how they succeed, or not, in overcoming them.

Although there are still many questions about the relationship between team creativity and surprise, this work has made an empirical contribution to the subject by distinguishing the effect of surprise on creative performance according to the creativity criterion selected and highlighting certain elements that moderate the possible negative effect of surprise on the originality of ideas. We hope this work will spark a broader discussion on this crucial subject in societies where continuous change and inevitable surprises are the norm.

4.7 Appendices

Code Book

Variable	Definition
Adaptation to unexpected events (daily life)	Variable measuring the reaction of adaptation to unexpected events in subjects' daily life. <i>How do you react to unexpected events in everyday life, 1 being "I don't know how to react" and 5 "I adapt very quickly"?</i>

Variable	Definition
Adaptation to unexpected events (work life)	Variable measuring the reaction of adaptation to unexpected events in subjects' work life. <i>What happens when you need to quickly change what you are doing at work, 1 being "I don't know how to react" and 5 "I adapt very quickly"?</i>
BI	Variable measuring the normalized Blau Index for gender, age, domain, openness to experience, conscientiousness, extroversion, agreeableness, neuroticism, and creativity. Values of 0 (non-diverse) and 1 (diverse).
Collective creativity	Variable measuring the self-reported individual creative performance during task 1 and task 2. <i>On a scale of 1 (min) to 5 (max), how would you rate your individual level of creativity within your group for the first (second) task?</i>
Commitment	Variable measuring the self-reported commitment to the team. <i>I felt very committed to my group, 1 being "I strongly disagree" and 5 being "I strongly agree".</i>
Cooperation	Variable measuring the self-reported cooperation in the team. <i>To what extent were team members cooperative? 1 being "I strongly disagree" and 5 being "I strongly agree".</i>
Creativity after surprise	Dummy variable defining if subjects felt more creative before or after the change. <i>Do you think you were more creative before or after the change in the problem statement?</i>
Degree of surprise (perceived)	Variable measuring the self-perceived degree of surprise. Values from 1 to 5.

Variable	Definition
Degree of surprise (treatment)	Dummy variable differentiating the two degrees of surprise elicited in the protocol. Values of 0 for low surprise and 1 for high surprise.
Feasibility	Variable measuring the creative performance of teams in the feasibility criteria (in task 1 and task 2). Values from 0 to 10.
Importance of change	Variable measuring the self-reported importance of the change in the surprise task. <i>Comparing the two instructions given to you during the second task, how would you rate the importance of the change in the problem, 1 being “a minor change” and 5 being “a major change”?</i>
Individual creativity	Variable measuring the self-reported collective creative performance during task 1 and task 2. <i>On a scale of 1 (min) to 5 (max), how would you rate the creativity level of your group as a whole for the first (second) task?</i>
Open communication	Variable measuring the self-reported openness of communication in the team. <i>Did team members communicate openly with each other? 1 being “I strongly disagree” and 5 being “I strongly agree”.</i>
Originality	Variable measuring the creative performance of teams in the originality criteria (in task 1 and task 2). Values from 0 to 10.
Preference to work in group	Variable measuring the self-reported preference to work in a group environment. Value of 0 if a preference for individual setting and 1 for group setting.

Variable	Definition
Reaction (individual)	Variable measuring the self-reported subjects' own reaction to the unexpected event. <i>How would you describe your reaction to this event, 1 being "I did not know how to react" and 5 being "I adapted very quickly"?</i>
Reaction (team)	Variable measuring the self-reported team members' reaction to the unexpected event. <i>How would you describe the reaction of other members of your group to this event, 1 being "They did not know how to react" and 5 being "They adapted very quickly"?</i>
Reliability	Variable measuring the self-reported reliability in the team. <i>To what extent did you feel your teammates could count on you for help? 1 being "I strongly disagree" and 5 being "I strongly agree".</i>
Recommendation	Variable measuring the self-reported likelihood of recommending the team as a desirable group to work with. <i>I could praise this group to my friends as an excellent working group, 1 being "I strongly disagree" and 5 being "I strongly agree".</i>
Risk	Variable measuring the self-reported risk attitude. <i>On a scale of 1 to 5, how do you evaluate your attitude towards risk (1 I hate risk and 5 I love risk)?</i>
Satisfaction	Variable measuring the self-reported satisfaction with the team. <i>I was very satisfied with working with this team, 1 being "I strongly disagree" and 5 being "I strongly agree".</i>

Variable	Definition
Similarity of ideas	Dummy variable defining if one team's idea before and after the change are similar. Values of 0 for no and 1 for yes.
Treatment	Factor variable differentiating the control and treatment groups. Value of 0 for the control group, 1 for the <i>low surprise</i> group , and 2 for the <i>high surprise</i> group.
Trust	Variable measuring the self-reported trust attitude. <i>On a scale of 1 to 5, to what extent do you see yourself as someone who trusts others (1 not at all and 5 completely)?</i>
Trust in team-mates	Variable measuring the self-reported trust in team-mates. <i>To what extent did you trust your teammates during this experience? 1 being "I strongly disagree" and 5 being "I strongly agree".</i>
Valence	Variable measuring the self-reported valence of subjects' on the experienced surprise. <i>What would you associate with this event, 1 being "something very negative" and 5 being "something very positive"?</i>
Valence on unexpected events	Variable measuring the self-reported general valence of subjects on unexpected events. <i>What would you associate with an unexpected event, 1 being "something very negative" and 5 "something very positive"?</i>

Instructions

The following sub-sections present the different instructions given to the subjects during each experimental session. We have decided to translate these instructions so that all readers can understand them, but they are originally in French. The French instructions are available upon request.

Schema-setting task

For control and treatment groups:

For this first task, as students at the University of Strasbourg, we ask you to think about and respond as a group to the following problem:
“Imagine an idea/project to improve your life at the university”.
You have 20 minutes per group to complete the idea sheet summarizing your idea/project.

Surprise task

For control group:

For this second task, following the reading of a ministerial report on the climate crisis in your country, your municipality invites you to reflect as a group on your vision of the city of the future. The issue of housing is a priority for the municipality. Therefore, you must respond to the following problem:
“Imagine and describe the homes of a city in the future built underwater”.
You have 30 minutes per group to complete the idea sheet summarizing your idea.

For treatment group (low surprise):

For this second task, your municipality invites you to reflect as a group on your vision of the city of the future. The issue of housing is a priority for the municipality. Therefore, you must respond to the following problem:

“Imagine and describe the homes of a city in the future”.

Please note that the current world is subject to numerous changes and mutations, and it is possible that your municipality’s interests may change. Thus, the instructions given to you may also be subject to change.

You have 30 minutes per group to complete the idea sheet summarizing your idea.

Following the reading of a ministerial report on the climate crisis in your country, your municipality decides to focus its efforts on a particular issue and now asks you to follow a new direction and reflect on the following problem:

“Imagine and describe the homes of a city in the future built underwater”.

You have 15 minutes per group to complete the idea sheet summarizing your idea in accordance with this new issue.

For treatment group (high surprise):

For this second task, your municipality invites you to reflect as a group on your vision of the city of the future. The issue of housing is a priority for the municipality. Therefore, you must respond to the following problem:

“Imagine and describe the homes of a city in the future”.

You have 30 minutes per group to complete the idea sheet summarizing your idea.

Following the reading of a ministerial report on the climate crisis in your country, your municipality decides to focus its efforts on a particular issue and now asks you to follow a new direction and reflect on the following problem:

“Imagine and describe the homes of a city in the future built underwater”.

You have 15 minutes per group to complete the idea sheet summarizing your idea in accordance with this new issue.

Idea sheet

Group	
Title	
Description	
Needs	
Strengths	
Weaknesses	

Questionnaires

The following sub-sections present the two questionnaires submitted to the subjects. We have decided to translate these questionnaires so all readers can understand them, but they are originally in French. The French questionnaires are available upon request.

Preliminary questionnaire

1. You are:
- Male
 - Female
2. How old are you?
3. What is your native language?
4. What is your nationality?
5. What is the highest degree you have obtained?
- *Brevet*
 - *Baccalauréat*
 - Bachelor's Degree
 - Master's Degree
 - Doctorate
 - Other: _____
6. What is your field of study?
- Arts, letters, and languages
 - Law, economics, management, and political and social sciences
 - Humanities and social sciences
 - Science and technology
 - Business
 - Other: _____
7. You will find below 10 statements that may or may not apply to you.
- Write down the number indicating how much you agree or disagree with each statement (1 completely disagree and 5 completely agree):
- (a) I see myself as someone who is reserved.
 - (b) I tend to criticize others.
 - (c) I work conscientiously.
 - (d) I am relaxed, handle stress well.
 - (e) I have a great imagination.
 - (f) I am sociable, outgoing.
 - (g) I generally trust others.
 - (h) I tend to be lazy.
 - (i) I am easily anxious.
 - (j) I am not very interested in artistic matters.
8. On a scale of 1 to 5, to what extent do you see yourself as someone who trusts others (1 not at all and 5 completely)?
9. On a scale of 1 to 5, how do you evaluate your attitude towards risk (1 I hate risk and 5 I love risk)?
10. Generally, do you prefer to work in a group or individually?
- in a group
 - individually

Post-questionnaire

General:

1. What is your definition of creativity?
2. On a scale of 1 to 5, how would you rate your creativity?

Everyday:

3. What would you associate with an unexpected event, 1 being “something very negative” and 5 “something very positive”?
4. How do you react to unexpected events in everyday life, 1 being “I don’t know how to react” and 5 “I adapt very quickly”?
5. What happens when you need to quickly change what you are doing at work, 1 being “I don’t know how to react” and 5 “I adapt very quickly”?

During today’s group session:

6. On a scale of 1 (min) to 5 (max), how would you rate your individual level of creativity within your group for the first task?
7. On a scale of 1 (min) to 5 (max), how would you rate the creativity level of your group as a whole for the first task?

8. On a scale of 1 (min) to 5 (max), how would you rate your individual level of creativity within your group for the second task?

9. On a scale of 1 (min) to 5 (max), how would you rate the creativity level of your group as a whole for the second task?

10. I felt very committed to my group, 1 being “I strongly disagree” and 5 being “I strongly agree”.

11. I could praise this group to my friends as an excellent working group, 1 being “I strongly disagree” and 5 being “I strongly agree”.

12. To what extent did you trust your teammates during this experience? 1 being “I strongly disagree” and 5 being “I strongly agree”

13. To what extent did you feel your teammates could count on you for help? 1 being “I strongly disagree” and 5 being “I strongly agree”.

14. Did team members communicate openly with each other? 1 being “I strongly disagree” and 5 being “I strongly agree”.

15. To what extent were team members cooperative? 1 being “I strongly

disagree” and 5 being “I strongly agree”.

16. I was very satisfied with working with this team, 1 being “I strongly disagree” and 5 being “I strongly agree”.

17. From a purely practical point of view, did completing the idea sheet seem difficult to you? (Yes/No)

Focusing on the Second Task:

18. How would you rate the degree of surprise you felt when a constraint was added in the second task of the experience, 1 being “not surprising at all” and 5 being “very surprising”?

19. What would you associate with this event, 1 being “something very negative” and 5 being “something very positive”?

20. How would you describe your reaction to this event, 1 being “I did not know how to react” and 5 being “I adapted very quickly”?

21. How would you describe the reaction of other team members to this event, 1 being “They did not know how to react” and 5 being “They adapted very quickly”?

22. Comparing the two instructions given to you during the second task, how would you rate the importance of the change in the problem, 1 being “a minor change” and 5 being “a major change”?

23. Do you think you were more creative before or after the change in the problem statement? (Before/After)

Pre-existing network questionnaire

1. Did you know this person before this experience?

Yes No Only by name

If yes, would you characterize your relationship as (indicate the importance of this relationship in parentheses):

Friendly (from 1 to 5: ...)

Professional (from 1 to 5: ...)

How often did you meet/talk?

Rarely A few times Several times Regularly

Chapter 5

Can AI Enhance its Creativity to Beat Humans?

This chapter was co-authored with

Pierre PELLETIER and Rémy GUICHARDAZ

Summary of the chapter

Creativity is a fundamental pillar of human expression and a driving force behind innovation that now stands at a crossroads. As artificial intelligence advances at an astonishing pace, the question arises: can machines match and potentially surpass human creativity? This study investigates the creative performance of artificial intelligence (AI) compared to humans by analyzing the effects of two distinct prompting strategies across three different tasks and several creativity criteria. Human external evaluators have evaluated creative outputs generated by humans and AI, and these subjective creative scores were complemented with objective measures based on quantitative measurements and NLP tools. The results indicate that AI generally outperforms humans in closed and open-with-constraints creative tasks, although this advantage is more nuanced in the case of open tasks. Additionally, performance varies based on the specific criteria chosen for evaluation. Ultimately, while AI demonstrates superior creative capabilities, our findings suggest that incorporating human feedback is essential for maximizing AI's creative potential.

5.1 Introduction

Artificial intelligence (AI) is now part of our daily lives in both private and professional spheres. The uses of such technology are manifold, and the fields of application are almost infinite. AI can be defined as “*a system’s ability to correctly interpret external data, to learn from such data, and to use these learnings to achieve specific goals and tasks through flexible adaptation*” [Kaplan and Haenlein, 2019, p. 15]. Different models and algorithms exist in the realm of AI. Among them are *Transformers*, which are innovative architectures that have radically transformed the field of AI, particularly in Natural Language Processing (NLP) and Image Generation. These models use what are known as “*attention mechanisms*” [Vaswani et al., 2017], an innovative technique that allows the model to focus on different parts of the input sequence when processing or generating each part of the output sequence. Large Language Models (LLMs), such as GPT (Generative Pre-trained Transformer), are a specific class of Transformer models and have become a powerful tool for creators in multiple domains who use algorithms to generate, among other things, works of art, music, writing suggestions, and other similar outputs [Bubeck et al., 2023]. GPTs are now accessible to everyone with widely known user-friendly interfaces such as the one used in this study, *GPT-4* and *DALL-E*.

One of the issues at the heart of our use of AI is its relationship with humans. There are three possible scenarios: humans undertaking a task without AI, humans collaborating with AI, and AI wholly supplanting humans. These scenarios can be relevant for many human activities, including the tasks performed in organisations, as AI has already demonstrated its ability to change them [Cockburn et al., 2018, Von Krogh, 2018]. Within this context, AI, functioning as a partially autonomous technological entity or agentic technology, has significantly enhanced logistics or decision-making processes. Our study aims to explore another key activity of organisations, namely creativity. Creativity relies on problem-solving and problem-framing, which are essential for organisations as these elements define how organisations operate [Brusoni, 2005, Miron-Spektor et al., 2018]. As a result, an organisation’s creative capacity is a crucial comparative advantage and source of innovation [Chatzoglou and Chatzoudes, 2018, Woodman et al., 1993]. Moreover, creative endeavours are highly dependent on the individuals in charge and are highly labour-intensive. Given that creativity is highly sensitive to problem framing and demands significant investments in human resources, might organizations consider

an alternative approach, such as AI? For an organisation to take up the issue of AI as a creative resource, there are two possible options: seeing artificial creativity as a substitute for human creativity or as a complement to it. In other words, AI can be either envisioned as a self-sufficient generator or a complementary creative tool.

The question arises: to what extent does AI compete with human minds when it comes to creative endeavours? While for a long time, the possibility of a machine being creative was dismissed out of hand, new works have emerged to demonstrate that it is a possibility and artificial creativity is now defined as “*the production of highly novel, yet appropriate, ideas, problem solutions, or other outputs by autonomous machines*” [Amabile, 2019, p. 3]. As soon as we uphold the premise that creativity is an inherently human endeavour, it opens the door to AI as an alternative to human creativity. With its absence of fatigue, lack of frustration, and faster execution times, AI might be considered a viable substitute. However, the possibility of AI substituting the work of humans might be a cause of concern since the speed of AI advancements might represent a massive and long-lasting displacement of workers. Thus, understanding the performance of AI in creative tasks is critical as organisations increasingly adopt AI technologies, which could change, positively or negatively, the value attributed to human creativity.

These concerns about humans-AI collaboration or substitution are not new as in the founding work of Minsky [1961] attesting that “*With these systems, it will at last become economical to match human beings in real-time with really large machines. This means that we can work toward programming what will be, in effect, thinking aids*” (p. 28). Von Krogh [2018] put into perspective the stakes deriving from using AI in organisations by considering it as task input, process, and output. It is worth noting that our work focuses mainly on AI as a task output in the sense of the generation of creative solutions. In the end, if the literature on substitution versus complementarity of humans and AI seems to argue for AI as a way to augment human behaviour in organisations, it is still crucial to compare them to understand the specific areas or skills that give rise to this possible complementarity, especially when we know that AI is a fast-growing technology that requires specific attention to track its development and consider the rapidly changing results in the literature.

Besides the nature of the agent involved in the creative process, comparing outputs requires an assessment of the creative performance of these. Since creativity is a multifaceted concept, it requires a deep understanding of several components to grasp the complexity of it all, and multiple new research avenues are open when

it comes to Generative AI. While the literature on the creative performance of AI is expanding, the variety of tasks tested is still limited and often restricted in the number of dimensions or criteria chosen for the creativity assessment. Moreover, only a few studies have examined the issue of prompting strategies in the specific case of creativity. Indeed, if AI has the capacity to generate ideas, it has no agency. Thus, this generation is always conditional on the request that has been addressed, reinforcing the crucial role played by problem-framing in the creative process. As a result, the research questions at the heart of this work revolve around two major points: the multidimensional nature of creativity and the role of prompting strategies. First, does AI outperform humans in terms of creative performance? On any task? Second, does the prompting strategy adopted to generate creative outputs impact AI creative performance? And what type of prompt engineering strategy proves most effective when aiming for exceptionally high creative performance?

To achieve this, an online experiment was conducted where human evaluators assessed the creative performance of outputs produced by both humans and AI, focusing on the first and third scenarios mentioned earlier. Ultimately, we examine the individual creativity of the two agents—humans and AI—without yet considering their collaboration. In total, 199 subjects rated creative outputs across three tasks (a Text task, a Draw task, and an Alternative Uses task) on multiple criteria. In addition to these evaluations, other measures based on NLP techniques have been used. Our study demonstrates a clearly better performance of AI over humans across all tasks and almost all criteria. However, if AI generally outperforms humans in creative tasks, its success heavily relies on the prompting strategy used. While AI demonstrates strong creative abilities, it faces challenges in more complex tasks without human intervention. Ultimately, the study underscores the importance of human-machine collaboration, as human input, particularly in refining prompts and framing problems, plays a vital role in enhancing the quality and effectiveness of AI-generated creative outputs.

The remainder of the chapter is organized as follows: Section 5.2 outlines the state of the art, introducing the relevant literature and concepts evoked in this work. Then, Section 5.3 describes the experimental design encompassing the tasks themselves, data generation and collection, experimental procedure, and hypotheses. Section 5.6 presents our study’s descriptive statistics, regression analysis, and results. While Section 5.7 and Section 5.8 respectively present the discussion and conclusion of this work.

5.2 Literature review

5.2.1 Human and Artificial creativity

Over centuries, many authors have discussed the origin of creativity. While some began by characterizing creativity as divine inspiration [Gaut, 2012, Stokes and Paul, 2016], others followed the romantic view, considering creative individuals as geniuses [Miller, 1996]. In both approaches, creativity was perceived as something unknown, mysterious, and fundamentally inexplicable. Moving away from this elitist approach to creativity, it has been envisioned as a more complex process and product that can emerge from any individual or group of individuals. Creativity is no longer confined to the scientific or artistic sphere and has been studied from the angle of more ordinary forms of creativity. Creativity can then be simply defined as “*the ability to come up with ideas or artefacts that are new, surprising and valuable*” [Boden, 2004, p. 1].

What remains striking in the literature, however, is the inextricable link between creativity and humans, as “*creativity is part of what makes us human*” [Sawyer and Henriksen, 2024, p. 3]. Amabile and Pratt [2016] defines creativity as “*the production of novel and useful ideas by an individual or small group working together*” (p. 158). Most creativity-related concepts are articulated around features of the human mind, translating it as a cognitive capacity [Boden, 1998] that everyone can train. This culminates when creativity is defined as a social process resulting in a bidirectional relationship between individuals and their environment [Amabile and Pratt, 2016, Csikszentmihalyi, 1997, Fischer et al., 2005]. However, if creativity is a human-made concept, that should not be a reason for it to remain a human-centric one [Guckelsberger et al., 2017]. Indeed, given that creativity is commonly defined as the ability to generate novel and useful ideas Hennessey and Amabile [2010], it seems that any entity capable of generating such novel and useful ideas would be creative. Consequently, new definitions have recently emerged. Artificial creativity is “*the production of highly novel, yet appropriate, ideas, problem solutions, or other outputs by autonomous machines*” [Amabile, 2019, p. 3]. Whether using the term artificial or computational creativity, a growing number of authors argue that machines are able to provide human-like creative output using computer systems and algorithms [Veale and Cardoso, 2019].

Since humans and machines are both capable of producing creative outputs, it

allows for comparison in performance. Comparison therefore requires a careful examination of how creative performance is assessed. Beyond the duality between humans and machines, one must be careful about the type of task evaluated. Indeed, creativity is a cognitive process, at least if we stick to its human-centric definition, that elicits different thinking processes, mainly categorized as divergent and convergent. While convergent thinking aims to produce a singular logical solution, divergent thinking targets the generation of multiple solutions [Guilford, 1950, 1967]. This distinction allows us to differentiate tasks according to their degree of openness, i.e. the degree of autonomy allocated to individuals when carrying out a task. We can then consider three types of tasks based on different degrees of openness: closed (mostly relying on a convergent thinking process and strong constraints), open (mostly relying on a divergent thinking process and few or no constraints), and open-with-constraints (which is a task closer to real-world circumstances with a mix of divergent and convergent thinking process with some constraints) [Charness and Grieco, 2019, Attanasi et al., 2021].

Moreover, creativity is a complex and multifaceted phenomenon that requires a multidimensional approach to its evaluation. Creative performance in a specific task might differ based on the criteria being considered. The literature relies on four main criteria: originality (*how infrequent a particular solution is*), fluency (*how many ideas were generated to solve a specific problem*), flexibility (*how many themes cover the set of ideas generated*), and elaboration (*how detailed an idea is*) [Casotti et al., 2016, Guilford, 1950, 1967]. Besides these traditional criteria, multiple measures of creativity have emerged, also based on the usefulness or appropriateness of ideas that deserve closer attention [Hubert et al., 2024] as an idea is a “good” idea when it embraces both high originality and high feasibility [Magni et al., 2023, Rietzschel et al., 2010]. Finally, when assessing the creativity of an idea, one has to question the nature of the evaluator as both objective and subjective measures might be envisioned. While objective assessments are mainly based on quantitative and precise measurements, subjective assessments rely on human evaluators’ own personal perception of the output according to the selected criteria. These two assessment methods, although different, are complementary.

In the end, the literature on artificial creativity and its comparison to humans is growing in importance, but the results remain blurred. On the side of task openness, Charness and Grieco [2024] find a higher performance of humans in an open task, while AI performs better in a closed task, highlighting the complementarity between

human and artificial creativity. However, some authors also investigated different criteria of creativity with which the distinction between humans and machines is not straightforward. Even though AI is able to produce human-like outputs, it fails to provide the same level of unexpectedness or novelty of ideas [Stevenson et al., 2022]. On the contrary, in [Koivisto and Grassini, 2023], AI outperforms humans in originality and elaboration except for the best of them. Nonetheless, these results need to be nuanced if we look at the technological progress in terms of the AI model at use. While humans were still outperforming GPT-3.5, Haase and Hanel [2023] observed an almost similar performance between humans and AI when introducing GPT-4. In addition, this difference between AI models is not constant over all selected criteria since GPT-4 seems to outperform GPT-3.5 on fluency but not elaboration [Vinchon et al., 2023]. As a consequence, there is no clear-cut on whether AI outperforms humans in creative tasks. For Haase and Hanel [2023] to conclude that “*yes [AI are creative], as much or as little as humans*” (p. 11).

5.2.2 The specific case of Transformers

LLM models are a form of transformational models, i.e. statistical models based on the probability of occurrence of the most frequent events. Following this predictive logic, this type of model is trained on a large ensemble of data and, based on the various inputs and the most frequently occurring events previously generated, provides a coherent and appropriate response. These types of model can be referred to as GPTs (Generative Pretrained Transformer), among which the well-known application from OpenAI is GPT-4. Released in March 2023, GPT-4 aims to generate human-like texts depending on the context, particularly in response to a prompt. Then, DALL-E 2 is an AI image generation tool developed by OpenAI in 2022 [Ramesh et al., 2022], which generates an image based on a text description. The system underlying DALL-E 2 is based on two technologies, namely CLIP and diffusion. CLIP, which stands for Contrastive Language-Image Pre-training, plays a crucial role in DALL-E 2, serving as the main bridge between text and images. It consists of two neural networks: a text encoder and an image encoder, both trained on a large dataset of image-text pairs. These encoders map their inputs into a shared vector space, effectively creating a “concept space” which allows for the translation of semantic information between text and images. In DALL-E 2, the diffusion model complements CLIP by generating images based on the embeddings produced by CLIP. CLIP first processes the input text to create a text embedding, which is then used to generate a

corresponding image. This embedding vector captures the semantic meaning of the text in a form that can be used for image generation. The diffusion model then takes the embedding and generates an image by iteratively refining noise into a coherent visual representation.

Based on attention mechanisms, GPTs aim to provide answers for a given context and serve as problem-solving tools. Problem-solving requires looking for solutions in a solution space using reasoning that begins with the formulation of the problem to apply it to the information acquired by the solver where there is no pre-determined operator or strategy [Öllinger and Goel, 2010, Simon, 1973]. Additionally, problem-solving is closely related to problem-framing. When evaluating AI’s problem-solving capabilities, it is crucial to consider how the problem is formulated and how this formulation impacts the resolution. In the case of Generative AI, this means looking at the exact request formulated by the user and addressed to the model. This request is referred to as a prompt, where “*prompting means prepending instructions to the input and pre-training the language model so that the downstream tasks can be promoted*” [Ge et al., 2023, p. 3]. Since AI lacks agency, making its creative potential reliant on human assistance to generate responses [Hubert et al., 2024], the use of prompts becomes the main driver for problem-solving.

This prompt-dependency justifies the need to reflect on what prompting strategy should be selected by users. Indeed, any user of LLMs has experienced that the obtained output is highly dependent on what they have asked for and that some changes in the phrasing or framing of a question or instruction might totally change the result. As a consequence, new skills have emerged in organisations and for users of AI that need now to excel in prompt engineering, i.e., the process of creating, refining, and optimizing prompts. In fact, you might prefer to ask a question directly to the AI interface without any context or give a complete and comprehensive context to the interface to ensure that AI enhances its creative capacities. In other words, one might choose different prompting strategies using LLMs when trying to maximize AI performance depending on the intended objective. As for any relevant phenomenon, literature has emerged on the concept of prompting strategies, highlighting the importance of adapting the prompt when it comes to specialized tasks [Baidoo-Anu and Ansah, 2023]. These strategies encompass prompts based on personality [Chen et al., 2023, Xu et al., 2022], on a context for a specific field [Ge et al., 2023], or Chain-of-Thought reasoning [Wei et al., 2022]. However, no clear consensus prevails on the effect of each strategy as results might differ according to the elicited

strategy as well as the chosen task.

While we recognize that AI is increasingly capable of solving ever more complex problems [Von Krogh, 2018], we must also acknowledge a number of shortcomings. In fact, the use of AI does not guarantee a zero error rate as soon as we admit that models can suffer from hallucinations. These hallucinations refer to unreliable and nonsensical text, image, audio, or video outputs generated by LLMs [Rawte et al., 2023]. In these cases, prompts again play a central role as they contribute to *dehallucinating* them as the inaccuracy of generated outputs decreases. However, hallucinations might be valuable when used for creative purposes. Indeed, as sources of unexpected answers, hallucinations might fuel creativity [Rawte et al., 2023]. While this change of perspective may seem curious, it nonetheless calls into question the link between AI and creativity when we consider the points made earlier about prompt dependency and the resulting prompting strategies. Thus, besides the earlier mentioned concerns on the criteria chosen to assess creative performance, it is also essential to question what determines an ideator’s creative performance when it comes to problem-framing. This study aims to tackle this question by comparing two different prompting strategies characterized as *Naive* and *Expert* prompting while comparing different types of tasks and the resulting creativity criteria.

5.3 The experiment

The description of our experimental protocol is divided into two main parts. The first part is dedicated to the collection and generation of creative outputs by humans and AI. The second part implements an online experiment dedicated to the evaluation of these outputs by human evaluators.

5.3.1 Data collection and generation

The creative outputs used in this study can be broken down into two components: experimental data produced by humans and AI-generated data. In the case of human data, these were collected through three different creative tasks conducted during two experimental protocols [Maltese et al., 2023, Guichardaz et al., 2024]. Table 5.3 summarises the tasks and their characteristics.

Table 5.1: Tasks' Characteristics

Tasks	Aim	Nature of the task	Main thinking process involved
Text task	Writing a text based on a list of predefined and compulsory words	Close-ended	Convergent
Alternative Uses task	Finding unusual uses for everyday objects	Open-ended with constraints	Divergent
Draw task	Draw an alien animal	Open-ended	Divergent

This sequence of tasks aims to offer a more comprehensive understanding of creative contexts than what is currently found in the existing literature on the classical open/closed dichotomy [Charness and Grieco, 2024]. It does so by capturing a broader spectrum of task openness, ranging from closed to open-with-constraints and fully open tasks. To start with, the Text task inspired by Charness and Grieco [2019] is a closed task that requires the subject to combine a set of elements acting as constraints, providing a clearer goal to reach. As a result, such a task elicits more of a convergent thinking process. This Text task allows us to observe the capacity of humans and AI to produce a creative text while thematic constraints have been introduced.

Next, the Alternative Uses Task requires subjects to find unusual applications for ordinary items [Torrance, 1966, Guilford, 1967]. This task falls under the category of an open-with-constraints task. Widely employed for evaluating creative capacities, the Alternative Uses task is credited as one of the most prevalent methods. The aim of such a task is to assess individuals' capacity to generate a multitude of novel ideas, relying on more divergent thinking processes while requiring a certain degree of usefulness or appropriateness of ideas, which requires an ability to converge towards feasible ideas.

Finally, the Draw task aims to evaluate subjects' creativity through an open task. Introduced by Ward [1994], this task was specifically chosen to represent and elicit a divergent thinking process. While selecting a common idea requires less cognitive effort, the additional effort of exploring unusual and original ideas enhances creative

performance. Ward [1994, 2004] refers to this as the “path of least resistance”. In this task, when individuals are asked to imagine and create an extraterrestrial animal, they often incorporate many terrestrial characteristics (such as bilateral symmetry or sensory organs) into their creations [Ward, 1994, ?]. This demonstrates a fixation on familiar animal concepts. However, more creative individuals distinguish themselves by moving beyond these familiar elements, though “*the ability to generate a creative idea begins with known concepts*” [Birdsell, 2019, p. 45].

The same tasks were then used to produce the AI creative outputs. These were generated using GPT-4 (Text and Alternative Uses), coupled with DALL-E 2 (Draw task). More specifically, AI was prompted to perform the same tasks as our human subjects, with the only difference being the use of two different prompting strategies. Firstly, an AI whose prompts are as neutral as possible, with no other indications than the creative task itself (*naive prompting*). Secondly, an AI that we prompted so that it might be more effective in the specific task (*expert prompting*). Regarding Expert AI, we have generated a system prompt designed to provide the best, most creative response to each task. This system prompt was, therefore, used upstream of our user prompt, which included the exact instructions for the task in question. For Naive AI, we did not use a system prompt but directly addressed the instruction of each task as a user prompt. The purpose of these two strategies is to capture the two extreme cases of human use of AI. The *Naive* AI corresponds to a minimalist use of AI, akin to what an uninformed individual in the field might use. Conversely, *Expert* AI corresponds to the situation where a user employs a more sophisticated approach to prompt, thereby pushing its capabilities further.

Throughout the rest of the paper, we will refer to the outputs as *Human*, *Naive AI*, and *Expert AI* to differentiate them. Appendices 5.9 present the instructions for each task and the related pre-prompt in GPT-4 for both strategies.

5.3.2 Creativity assessment

Once the creative outputs have been collected or generated, the next step is to assess the creativity of these outputs. Outputs are evaluated according to the specific task and the corresponding creativity criteria as presented in the literature.

To start with, evaluators were randomly assigned to pools to evaluate one specific task among the three described above. This choice of separating evaluators by task has been made in order to avoid contamination from one type of evaluation to the other. Additionally, due to the difference in cognitive load to evaluate each task,

evaluators were presented with either 15 outputs from the Text task, 15 from the Draw task, or 9 from the Alternative Uses task randomly drawn from the pool of generated creative outputs. The Alternative Uses task, in particular, required evaluators to assess individually each idea on the list of alternative uses provided by the human or AI ideator, making the overall evaluation process more labor-intensive compared to the other tasks. This explains the lower number of outputs presented to each evaluator. Table 5.2 provides information on the number of creative outputs generated by task and creative agent. In the end, each output has been evaluated by at least 2 randomly matched evaluators from the assigned pool (1.32% by 2, 27.34% by 3, 71.36% by 4 evaluators).

Table 5.2: Number of Creative Outputs Generated and Evaluated

	Text Task	Alternative Uses Task	Draw Task
Generated (humans)	97	135	135
Generated (Naive AI)	90	90	90
Generated (Expert AI)	90	90	90
Evaluated (humans)	84	54	90
Evaluated (Naive AI)	85	54	90
Evaluated (Expert AI)	83	54	90

Before proceeding with their scoring, the evaluators were provided with the instructions given to the creators of the outputs (whether human or AI) as presented in Appendices 5.9. Then, each evaluator assigned, based on Likert scales, a score ranging from 0 to 5 to each creativity criterion. Evaluators were not informed at any point that some of the outputs they were assessing might be generated by AI. They were simply asked to evaluate each output individually based on the creativity criteria outlined in Table 5.3.

Table 5.3: Creativity Criteria Assessed by Evaluators

Criteria	Definition	Text	Alternative Uses	Draw
Validity	<i>Adherence to instructions</i>	X	X	X
Form	<i>Style and writing or drawing quality of outputs</i>	X		X
Elaboration	<i>Degree of details</i>	X		X
Originality	<i>Unusualness of ideas</i>	X	X	X
Feasibility	<i>Viability of ideas</i>		X	

To conclude the experimental sessions, evaluators answered a final questionnaire encompassing sociodemographic information and questions related to AI detection, usage, and attitudes. The questionnaire is available in Appendices 5.9.

5.3.3 Experimental procedure

The experimental protocol was conducted online. In total, 199 evaluators¹ were recruited via the ORSEE platform from the LEES, the experimental economics lab at the University of Strasbourg, which primarily consists of student email addresses. As evaluators only evaluated specific tasks, 65 of them evaluated the Draw task, 70 the Text task, and 64 the Alternative Uses task. Evaluators received a flat payment of €15 for their participation, and the average total response time for completing the experiment was 27 minutes. Table 5.4 presents some descriptive statistics about the evaluators' sociodemographic information, and Table 5.5 shows the comparison of populations of evaluators across tasks.

5.4 Hypotheses

Our study aims to compare the creative performance of AI and humans and how this performance is influenced by the nature of the creative task at hand and the chosen prompting strategy for creative outputs generated by AI. We formulate three hypotheses based on the literature in Section 5.2 and stylized facts.

¹Only two evaluators were removed from the original pool of 201 evaluators because they dropped out before completing the experiment.

Table 5.4: Evaluators' Socio-Demographic Characteristics per Task

Variable	Text	Alternative Uses	Draw
Mean Age	22.814 (3.036)	22.094 (2.646)	22.031 (2.952)
Mean Gender	0.657 (0.475)	0.672 (0.47)	0.831 (0.375)
Mean Diploma Licence	0.414 (0.493)	0.484 (0.5)	0.415 (0.493)
Mean Diploma Master	0.443 (0.497)	0.469 (0.499)	0.508 (0.5)
Mean Diploma Doctorat	0.057 (0.232)	0 (0)	0.031 (0.173)
Mean Droit	0.057 (0.232)	0.047 (0.212)	0.015 (0.123)
Mean Economie Gestion	0.414 (0.493)	0.297 (0.457)	0.4 (0.49)
Mean Lettres Langues	0.057 (0.232)	0.047 (0.212)	0.046 (0.21)
Mean Sciences exactes	0.186 (0.389)	0.203 (0.403)	0.154 (0.361)
Mean Psycho Socio	0.071 (0.258)	0.062 (0.242)	0.092 (0.29)
Mean Sciences politiques	0.029 (0.167)	0.094 (0.292)	0.092 (0.29)
Mean Native speaker	0.8 (0.4)	0.906 (0.292)	0.923 (0.267)
Mean French skills	7.671 (1.481)	8.25 (1.415)	8.462 (1.039)
# Evaluators	70	64	65

Table 5.5: Comparison of Evaluators' Socio-Demographic Characteristics Across Tasks

Metric	Draw vs Alternative Uses	Text vs Alternative Uses	Text vs Draw
Age	-0.063	0.721***	0.784***
Gender	0.159***	-0.015	-0.174***
Native speaker	0.017	-0.106***	-0.123***
French skills	0.212***	-0.579***	-0.79***
Diploma			
Bachelor	-0.069**	-0.07**	-0.001
Master	0.039	-0.026	-0.065**
PhD	0.031***	0.057***	0.026***
Domain			
Law	-0.031***	0.01	0.042***
Economics and management	0.103***	0.117***	0.014
Humanities	-0.001	0.01	0.011
Exact sciences	-0.049**	-0.017	0.032
Psychology and sociology	0.03*	0.009	-0.021
Political sciences	-0.001	-0.065***	-0.064***

Notes: This table presents the results of the ANOVA and Tukey's HSD test for the specified variable and group. The coefficients represent the Mean differences between groups. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Unlike humans, AI operates without mental or physical fatigue and can generate and manipulate a vast array of elements to achieve its objectives. We might consider, for instance, the contrast in time investment between AI and humans to compose or revise a text or create a drawing. With elements such as grammar, syntax, and semantics, AI demonstrates a significant ability to produce relevant and coherent responses that align with the formal requirements dictated by linguistic conventions (or visual conventions in the case of DALL-E). Nonetheless, it is necessary to consider the quality of human and artificial production in more detail, depending on the task in question.

As noted earlier, transformer-based models excel in processing vast amounts of data and identifying patterns within this data, enabling them to generate precise, contextually appropriate responses to specific prompts. Indeed, transformer-based models allow to “*access and generate larger amounts of knowledge, which in turn results in more possible connections of problems and solutions*” [Bouschery et al., 2023, p. 142]. This proficiency suggests that AI systems may outperform humans in close-ended tasks, which benefit from the model’s ability to identify the most accurate or optimal solution based on its learned patterns.

However, creativity is not merely a matter of (re)combining ideas, but of making those (re)combinations novel. So, although the knowledge set might be larger for an AI, the ability to combine distant elements might be higher for humans. Indeed, in its configurations, a Generative AI model aims to generate relevant outputs, playing on the prediction of probable and closely connected consecutive elements. Some authors criticized these AI productions as they “*merely emulate cognitive processes and cannot substitute the great flexibility, adaptability, and generativity we associate with human intelligence*” [Von Krogh, 2018, p. 408]. Conversely, human cognitive processes are more inclined to explore novel ideas in drawing on personal experiences, emotions, and cross-disciplinary knowledge, aspects where AI currently shows limitations. Indeed, even though AI and, more specifically, transformer-based models provide a broader set of knowledge and possible combinations [Bouschery et al., 2023], the originality of ideas always seems to be in favour of human creativity [Koivisto and Grassini, 2023]. Even though GPTs’ hallucinations do exist, “*given that LLMs are designed to generate approximately the statistically most plausible sequence of text based on their training data, perhaps they generate less-novel ideas*” [Girotra et al., 2023, p. 7], which might harm their ability to perform in more open-ended tasks oriented towards a more divergent thinking process. Therefore, we propose the

following hypothesis:

Hypothesis 1: *AI outperforms humans in close-ended tasks, whereas humans outperform AI in open-ended tasks.*

Then, if we consider the comparative performance of humans and AI, we must also consider the possible effect of the prompting strategy on the performance of AI itself. The literature has recently emphasized the crucial role of prompting strategies in augmenting the performance of large language models. Prompt serves as a navigational tool, enabling users to steer the model towards desired outputs by providing structured cues and directives. Moreover, orienting prompts offer users a mechanism to refine model responses through the strategic incorporation of specific instructions, constraints, context, or examples. This strategic tailoring not only ensures the alignment of model outputs with intended objectives but also enhances model performance across diverse tasks and domains. In the context of this paper, prompting strategies refer to the *Expert* use of AI compared to the *Naive* use of AI to differentiate between a Generative AI model designed to perform well in creative tasks and a more neutral model. Therefore, we formulate the following hypothesis:

Hypothesis 2: *Expert AI outperforms Naive AI across all tasks and creativity criteria.*

5.5 Methodology

5.5.1 Output Generation

As described in section 5.3.1, our agents are of three types: Human, Naive AI and Expert AI. In this section, we describe the instructions given to GPT-4 to generate the two types of AI agents (Details of the instructions given have been translated in Appendices 5.9). GPT-4 can be queried via the API by submitting both a user prompt and a system prompt. The system prompt is used to give initial instructions or directives that help to shape the model’s responses. For the Naive AI, we simply submitted a user prompt to GPT-4 with the instructions given to each human (most similar to the use of GPT-4 on the OpenAI platform). To create the Expert AI, we proceeded in two steps: the first was to provide GPT-4 with the instruction, asking the AI to generate a system prompt that would amplify its creativeness.

The system prompt generated by GPT-4 will then complement the user prompt, which includes the task instruction, enabling GPT-4 to respond to the instruction while being already set on a creative path. Of note, the creative output generation was calibrated at a temperature of 0.9 to maximize the coherence of answers while minimizing their possible redundancy.

5.5.2 Metrics

To compare our different agents, we used three types of measurement. The first type is human-based and was described in section 5.3.2. The other two types are built using GPT-4 or the vector representation of different outputs and are detailed below.

Theme-based metrics: We constructed measures of variety, balance, diversity, minimum theme frequency and uncommonness of theme combination scores across all tasks when applicable. For the AUT, we gave GPT-4 the set of individuals' responses for each word and requested it to create a list of 25 themes for each of the words. We then gave GPT-4 each individual idea again, along with the 25 themes, and asked it to assign the most relevant category to each idea. For the Text task, we proceeded differently, as we could not give GPT-4 all the texts simultaneously. We, therefore, first gave each text to GPT-4 to assign several themes, then retrieved all these themes and asked GPT-4 to group them together to give us a total of 25 possible themes. Finally, we resubmitted each text in GPT-4 and asked it to assign the most pertinent themes to each text. In this way, we were able to construct three measures based on these themes, namely variety (the number of themes as a measure of *flexibility*), balance (the proportion with which themes are used in an individual's responses as a measure of fixation effect,² the diversity is expressed by the Shannon index, which considers both the richness (how many distinct themes) and evenness (how equally the ideas are distributed across these themes). We also computed the proportion with which themes are used in all responses and kept the minimum for each observation (Theme Frequency (neg.)³). Finally, we calculated an uncommonness score *à la* Lee et al. [2015] to provide a metric on the combination of themes⁴.

²Here, we took the standard deviation of the proportion; a value of 0 means that all themes are uniformly distributed.

³We present this proportion as negative for readability; a higher value means that the usage of the given theme across all observations is low.

⁴This indicator captures the ratio of the observed number of co-occurrences to the expected number and shows how the pairing of two concepts is unusual or common.

Embedding-based metrics: Other quantitative measures are based on the embedding of texts, ideas or images. We represented our textual data in a semantic space (separating both tasks) using embedding techniques based on transformers. Rather than employing embeddings through OpenAI’s API, we used state-of-the-art French text representation, i.e. CamemBERT,⁵ as, in general, it is preferable to use language-specific models rather than multilingual models to represent texts. For both Text and AUT, we have broken down texts into sentences or specific ideas, respectively giving each sentence from a text and each idea in a list its own representation in the task-related semantic space. This strategy allows us to understand how an individual agent (Human or AI) uses concepts that are distant from each other when responding to an instruction. We opted for this solution as a representation of all the ideas or texts of a specific agent would not reflect the extent to which the agent exploits the knowledge space. In a similar way, we represented images in a vector space using the vision transformer model called CLIP. In this way, we can also project the drawings of different agents into a space that captures similarities at the visual level,⁶. Although this doesn’t allow us to construct distance measures at the level of drawings, it does enable us to capture distances between drawings and thus understand the heterogeneity of responses given by different agents. We calculated a centroid for each agent type and compared the distances of the different outputs to this centroid (Cosine Distance to centroid). Due to their high dimensionality, these vectorial representations cannot be visualized directly. Therefore, in order to understand whether our texts exist in the semantic space in a somewhat different way, we have represented them graphically by reducing the dimensionality of this semantic space. A typical technique to reduce the dimensionality of a semantic space is to use t-Distributed Stochastic Neighbor Embedding (T-SNE)⁷ [Van der Maaten and

⁵CamemBERT is based on the Transformer architecture, which enables it to process texts bidirectionally and capture context efficiently. It is a BERT model trained on the French part of OSCAR (Open Super-large Crawled Aggregated Corpus). More specifically, we use specialized French sentence embedding models such as Sentence-CamemBERT-Large, which can represent the semantics and meaning of French sentences in the form of mathematical vectors [Reimers, 2019, Martin et al., 2020]. Using this pre-trained model, we can project our sentences into a 768-dimensional vector space and compare them by calculating semantic distances using cosine similarity (Cosine Distance).

⁶As explained in Section 5.2.2 this model mainly allows text and images to be projected into a similar space, but we focus here on the visual part.

⁷T-SNE is a technique used to visualize and understand high-dimensional data sets and is very effective at preserving the local structure of the data, meaning that points that are close in high-dimensional space remain close in low-dimensional space. Unlike linear methods (such as PCA),

Hinton, 2008].

Controls: Finally, we created control variables for each of the tasks, namely the number of words for the text task, the average number of words per agent idea for the AUT, and the proportion of black pixels present in each image.

5.5.3 Analysis

Lastly, our analysis is based primarily on comparing the means of the various metrics between the different agents. First, we performed an ANOVA with a parametric Tukey’s HSD test, which we complemented with a non-parametric Pairwise Wilcoxon test to compare all the metrics (human-scored, theme-based, embedding-based and our control variables). The measures given by humans are Likert scales. Therefore, we then performed Ordered Polynomial Logistic regressions to understand how agent type plays on performance in each dimension, taking into account our control variables. Finally, we also investigated how evaluators’ socio-demographic characteristics influenced their responses and used an Ordered Polynomial Logit as well.

5.6 Results

The results of our study are divided between the three creative tasks submitted for evaluation, incorporating both objective and subjective measures⁸.

5.6.1 Text task

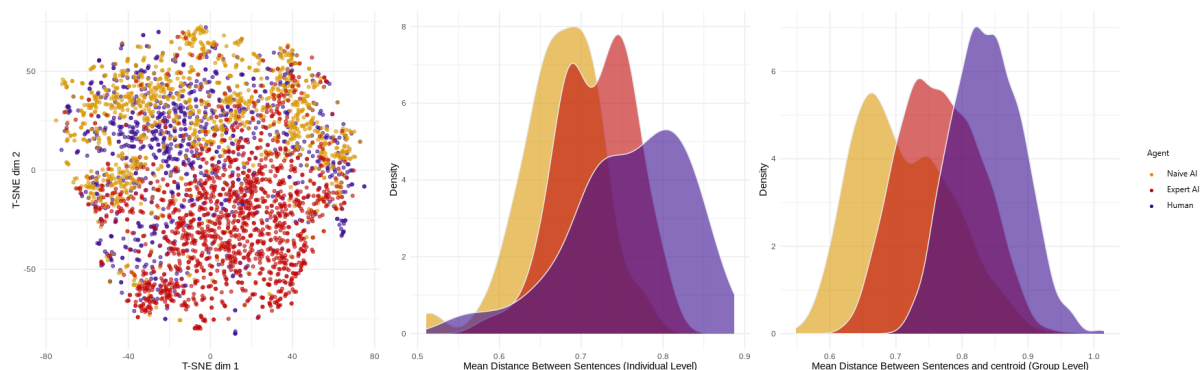
As a reminder, the Text task requires agents to write creative text based on a list of mandatory words. The objective measures of creative performance, presented in Table 5.6, show mixed results on whether AI outperforms humans or not in terms of performance. First, Naive AI outperforms Expert AI and humans in terms of *Variety* (the number of themes), while there is no significant difference between Expert AI and humans. We do not find any significant effect on *Theme frequency*. But, for the *Uncommonness* of theme combinations, Naive AI performs similarly to humans, and both outperform Expert AI. Finally, in terms of *Cosine distance*, humans outperform

t-SNE can capture non-linear relationships within the data.

⁸In addition, a principal component analysis, the results of which are available in Appendices 5.9, was carried out, demonstrating that our objective and subjective measures do indeed represent distinct elements of creative performance.

Expert and Naive AI, and Expert AI outperforms Naive AI. Taken together, these objective measures tell us something about the creative performance of our three agents, and in particular, that none of them is clearly outperforming. We observe that Naive AI is able to produce texts with a higher number of different themes but also to provide combinations of themes that are more unusual. While humans do not provide a more significant number of themes, they still offer unusual combinations. However, Expert AI, which was prompted to be more creative, provided fewer themes than Naive AI, and stayed behind in terms of the unusualness of theme combinations. On the cosine distance, humans provide texts with more semantic distance, meaning that their sequences of words are less predictable, and Expert AI finally outperforms Naive AI. Of note, while *Cosine distance* measures the semantic distance at the individual level (between the sentences of the same text), the measure of *Distance to centroid* represents the group-level semantic distance (between the sentences of one agent). Our results show that humans outperform Naive AI and Expert AI and Naive AI performs better than Expert AI.

Figure 5.1: Text Task Embeddings



Besides objective measures, we also analysed the creative performance of each agent according to the subjective evaluation provided by human external evaluators. Looking at the comparison between our three agents based on Tukey’s HSD test, we find a generalized outperformance of AI over humans on all four creativity criteria. The only differences reside in whether Expert AI also outperforms Naive AI. Table 5.12 presents the results based on our regression presented in Table 5.7, in which we use the Naive AI as the reference point and introduce a control on the text length to compare agents’ performance. The results show that on *Validity* and *Originality*, Naive AI and humans perform the same while outperforming Expert AI. However,

humans fail to provide texts that are as elaborated and high-quality in terms of form or style. Regardless of the criteria, as was the case for subjective measures, Expert AI seems to stay behind.

Table 5.6: Comparison of Agents' Performance for the Text task

Metric	Expert AI vs Naive AI	Human vs Naive AI	Human vs Expert AI
Variety	-2.401***	-2.918***	-0.517
Theme Frequency (neg.)	0	0.003	0.002
Uncommonness	-0.162***	-0.04	0.122***
Cosine Distance	0.043***	0.083***	0.04***
Distance to centroid	0.059***	0.127***	0.068***
Validity	-0.096	-0.339***	-0.242***
Form	-0.046	-0.856***	-0.81***
Elaboration	0.191**	-1.117***	-1.307***
Originality	0.221*	-0.534***	-0.755***
# Words	106.767***	-153.27***	-260.037***
# Sentences	6.408***	-6.139***	-12.547***
Mean length sentences	-1.449***	-0.277	1.172***

Notes: This table presents the results of the ANOVA and Tukey's HSD test for the specified variable and group. The coefficients represent the Mean differences between groups. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 5.7: Polynomial Logit Regression for the Text task

	<i>Dependent variable:</i>							
	Validity		Form		Elaboration		Originality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expert AI	-0.237 (0.182)	-0.606*** (0.207)	-0.561*** (0.173)	-0.776*** (0.187)	-0.533*** (0.175)	-0.653*** (0.191)	-0.270 (0.170)	-0.368** (0.177)
Human	-0.495** (0.207)	-0.370 (0.230)	-0.762*** (0.200)	-0.766*** (0.214)	-0.594*** (0.201)	-0.613*** (0.218)	0.177 (0.198)	0.281 (0.206)
# Words	0.001 (0.001)	0.003*** (0.001)	0.004*** (0.001)	0.006*** (0.001)	0.009*** (0.001)	0.012*** (0.001)	0.006*** (0.001)	0.008*** (0.001)
Evaluator FE	NO	YES	NO	YES	NO	YES	NO	YES
AIC	2257.55	1970.32	2704.33	2405.31	2613.2	2341.91	2985	2807
Observations	980	980	980	980	980	980	980	980

Notes: This table presents the coefficients reflecting the impact of agent type on the various creativity criteria scores. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using an Ordered Polynomial Logit.

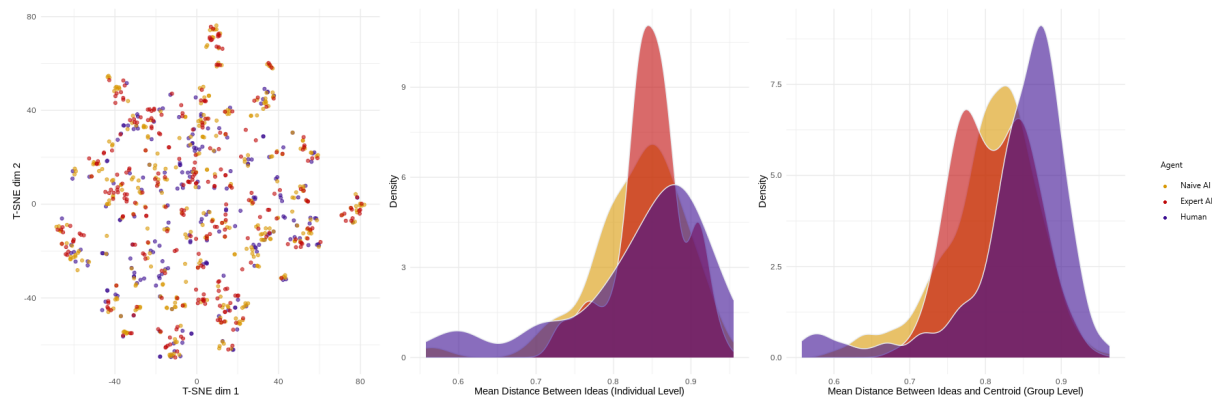
It is important to note that the relatively low performance of Expert AI is sensitive to the text length control introduced in our analysis. When this control is removed,

Expert AI performs better, surpassing both humans and Naive AI in *Elaboration* and *Originality*, while still lagging behind in *Validity* and *Form*. Then, the introduction of this control reveals that Expert AI’s initial outperformance was largely driven by the volume of text produced. Once this effect is accounted for, the content does not reflect superior performance. We interpret this as evidence that the sheer quantity of output from AIs may influence human evaluators, leading to higher scores. This control on the length of the texts is also important, as this explains the discrepancy between the results obtained from our regression in Table 5.7 and the parametric tests run in Table 5.6.

5.6.2 Alternative Uses task

The Alternative Uses task requires agents to find as many ideas as possible for unusual uses of everyday objects. Of note, all results obtained in Table 5.9 are in line with the parametric tests run in Table 5.8. Regarding objective measures, our results show a general outperformance of AI over humans. First, Naive AI surpasses Expert AI in terms of *Variety*, while both outperform humans. The same holds for *Diversity*. However, when considering *Theme frequency* and *Uncommonness*, there was no significant difference between Expert and Naive AI, while both still outperformed humans. While we found no significant effect of *Cosine distance* or *Balance* on our agents’ performance, we observe an outperformance of humans over AI agents on group-level semantic distance measured by *Distance to centroid*. These results show that AI provides texts with less frequent themes and more unusual combinations of themes, regardless of the prompting strategy. However, the Naive, which was not prompted to provide more creative texts, still performs the best in the number of themes and the diversity of texts. In any case, human texts never reach the performance of both AI agents except when comparing the semantic distance between agents’ pools of ideas.

Figure 5.2: Alternative Uses Task Embeddings



Focusing now on subjective measures, Expert and Naive AI clearly exceed humans in all three possible criteria: *Feasibility*, *Originality*, and *Validity*. In other words, AI provides ideas of unusual uses of everyday objects that are perceived as more fitting to the instructions and more implementable while also being more unique. Nonetheless, when comparing our two types of AI agents, Expert AI only performs better than Naive AI in the case of *Originality*⁹. At the same time, there is no difference between them for *Feasibility* and *Validity*. Here, prompting the AI with a specific requirement to be creative did influence its capacity to provide more original ideas. However, as there is no significant difference between our two prompting strategies for validity and feasibility, our interpretation is that an AI's own perception or definition of creativity is mainly related to originality and not the other criteria. Moreover, as LLM models are thought to provide the most appropriate ideas, it seems logical that both AI agents' performances are similar in feasibility and validity, which both confirm the goodness of fit of the idea to the instructions and the overall task purpose.

Lastly, introducing our control on the average number of words per idea shows a negative effect of longer answers on their originality, with evaluators preferring concision.

⁹This result is the only one differing from the ones obtained in the Tukey's HSD test, where there is no significant difference between Expert and Naive AI.

Table 5.8: Comparison of Agents' Performance for the Alternative Uses Task

Metric	Expert AI vs Naive AI	Human vs Naive AI	Human vs Expert AI
Variety	-0.815***	-2.519***	-1.704***
Balance (neg.)	0.003	-0.002	-0.005
Diversity	-0.137***	-0.636***	-0.499***
Theme Frequency (neg.)	0	-0.002**	-0.003***
Uncommonness	-0.139	-3.657**	-3.518**
Cosine Distance	0.017	0	-0.017
Distance to centroid	0.008	0.022***	0.015**
Validity	-0.063	-0.396***	-0.333***
Feasibility	-0.057	-0.438***	-0.38***
Originality	0.151	-0.516***	-0.667***
Mean # Words	1.7***	-1.295***	-2.995***

Notes: This table presents the results of the ANOVA and Tukey's HSD test for the specified variable and group. The coefficients represent the Mean differences between groups. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 5.9: Polynomial Logit Regression for the Alternative Uses Task

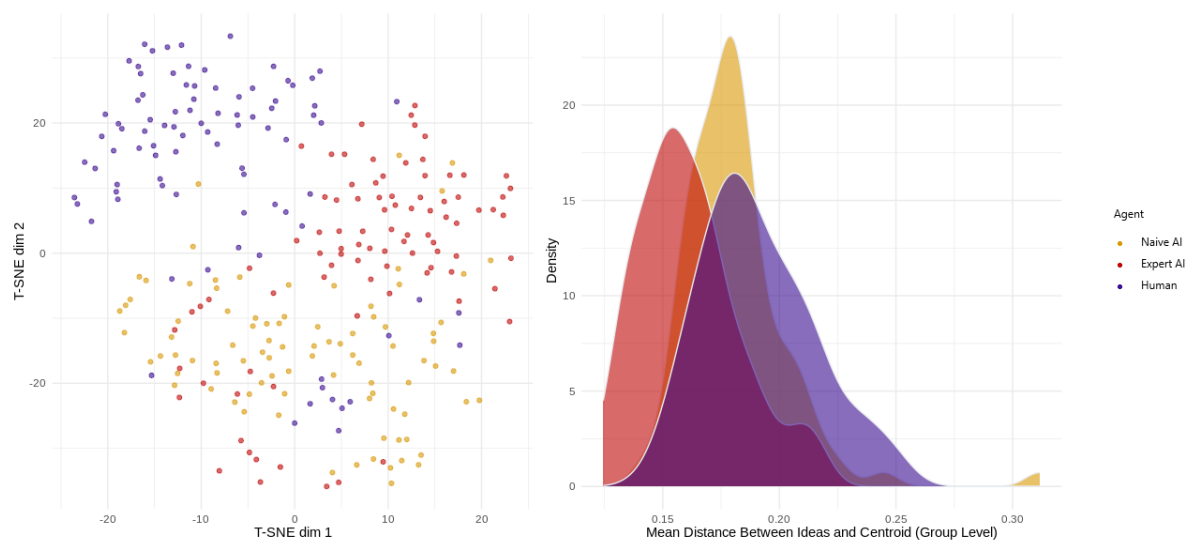
	<i>Dependent variable:</i>					
	Validity		Feasibility		Originality	
	(1)	(2)	(3)	(4)	(5)	(6)
Expert AI	0.079 (0.220)	-0.026 (0.265)	-0.234 (0.262)	-0.302 (0.282)	0.584*** (0.222)	0.817*** (0.241)
Human	-0.734*** (0.209)	-1.330*** (0.263)	-1.185*** (0.246)	-1.538*** (0.278)	-1.078*** (0.209)	-1.394*** (0.230)
Mean # Words	-0.088** (0.039)	-0.080* (0.048)	-0.081* (0.043)	-0.095** (0.048)	-0.129*** (0.042)	-0.169*** (0.046)
Evaluator FE	NO	YES	NO	YES	NO	YES
AIC	1253.72	942.4	959.38	905.48	1252.88	1140.79
Observations	576	576	576	576	576	576

Notes: This table presents the coefficients reflecting the impact of agent type on the various creativity criteria scores. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using an Ordered Polynomial Logit.

5.6.3 Draw task

Considering the Draw task, the instructions required to draw an alien animal coming from a planet different from Earth. First, we compared our three agents based on Tukey’s HSD tests across objective measures and found mixed results considering the possible outperformance of AI over humans. For *Variety* and *Theme frequency*, humans outperform Naive AI. However, there is no significant difference between Expert AI and the two others. The opposite holds for *Uncommonness*, as Naive AI outperforms humans while Expert AI still presents significant differences between the two others. In other words, humans can produce drawings with a larger number of themes and more unique themes, while Naive AI is able to create more uncommon combinations of themes. However, it seems that an AI prompted to be more creative lies in between these two performances with no outperformance of either humans or Naive AI, more as a compromise between them. While no individual level distance measure is computable, we measured group-level *Distance to centroid* and see that human drawings surpass Naive AI, both surpassing Expert AI.

Figure 5.3: Draw Task Embeddings



We also analysed the scores from human evaluators and observed an overall out-performance of Expert AI when generating an original, elaborated, qualitative, and valid drawing. Compared to the two previous tasks, we find a clear-cut between the performance of our two AI agents with a true impact of the prompting strategy pushing the model to be creative. However, focusing on the comparison between

Naive AI and Human, the difference is less clear. Considering the drawings *Form* and *Elaboration*, Naive AI outperforms humans. While for *Originality* and *Validity*, there is no significant difference between them. Again, all results obtained in Table 5.11 are in line with the parametric tests run in Table 5.10.

Table 5.10: Comparison of Agents' Performance for the Draw Task

Metric	Expert AI vs Naive AI	Human vs Naive AI	Human vs Expert AI
Variety	0.733	1.222***	0.489
Theme Frequency (neg.)	0.002	0.008**	0.006
Uncommonness	-0.012	-0.033*	-0.022
Distance to centroid	-0.02***	0.01***	0.03***
Validity	0.991***	0.111	-0.88***
Form	0.757***	-0.988***	-1.745***
Elaboration	0.923***	-0.769***	-1.692***
Originality	1.249***	0.006	-1.243***
% Pixel Used	0.018**	-0.143***	-0.161***

Notes: This table presents the results of the ANOVA and Tukey's HSD test for the specified variable and group. The coefficients represent the Mean differences between groups. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 5.11: Polynomial Logit Regression for the Draw Task

	<i>Dependent variable:</i>							
	Validity		Form		Elaboration		Originality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expert AI	1.283*** (0.147)	1.535*** (0.156)	1.117*** (0.145)	1.225*** (0.149)	1.206*** (0.143)	1.389*** (0.148)	1.605*** (0.148)	1.786*** (0.153)
Human	0.086 (0.167)	0.105 (0.174)	-1.307*** (0.174)	-1.530*** (0.181)	-1.101*** (0.171)	-1.194*** (0.176)	0.200 (0.167)	0.232 (0.170)
% Pixel Used	-0.342 (0.656)	-0.133 (0.701)	0.034 (0.649)	0.142 (0.693)	-0.425 (0.643)	-0.216 (0.671)	1.265* (0.651)	1.726** (0.677)
Evaluator FE	NO	YES	NO	YES	NO	YES	NO	YES
AIC	3105.8	2945.98	3103	3006.52	3190.1	3114.62	3232.35	3167.89
Observations	975	975	975	975	975	975	975	975

Notes: This table presents the coefficients reflecting the impact of agent type on the various creativity criteria scores. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using an Ordered Polynomial Logit.

It is noteworthy that our control of the complexity of the drawing measured by the share of pixels used in the image is only significant for originality, where a higher

number of pixels does contribute positively to the ratings in originality.

5.7 Discussion

This study examines the creative performance of Artificial Intelligence (AI hereafter) compared to humans and analyzing the impact of two different prompting strategies on creative outcomes across three different tasks. Based on the central role of creativity in our society and for organisations, the following results provide a deeper understanding of the possible substitution or complementarity of humans and machines when it comes to creative endeavours. In fact, our results give us interesting insights into the creative capacity of models of Generative AI compared to humans. In particular, our results show a tendency for AI to outperform humans in creative tasks. Nevertheless, the distinction between each task remains important as some nuances arise based on the chosen criteria and the objective or subjective nature of the stemming measures. Table 5.12 summarises the results obtained for our objective and subjective measures.

When focusing on objective measures, we observe mixed results as our comparison between the creative performance of our three agents depends on the task and the criteria we focus on. For the Text task, there is no outperformance of Expert AI, though it was specifically prompted for better performance. However, when comparing Naive AI and humans, the clearer difference resides in the semantic distance. This can be attributed to the way LMM models operate: they predict the next word based on the current context and probabilities learned from training data. This approach differs significantly from human language generation, which relies on complex neural processes, including synaptic connections and biochemical signals, stemming from emotions, social interactions, and cognitive development over time. Interestingly, the greater distance in human-generated texts does not correlate with better performance in subjective measures such as originality. In fact, as shown in the PCA analysis displayed in Appendices 5.9, this distance captures something distinct and possibly contradictory to subjective measures. Differing from Koivisto and Grassini [2023], we controlled for the correlation between our semantic distance measure and originality and found no significant relationship between these two. In other words, the distance observed in human texts does not correlate with more creative outputs according to the subjective criteria assessed by evaluators. This might suggest that

Table 5.12: Summary of Regression Results

	Text task	Alternative Uses task	Draw task
<i>Objective measures</i>			
Variety	$N > E = H$	$N > E > H$	$H > N^*$
Balance	\emptyset	<i>no significance</i>	\emptyset
Diversity	\emptyset	$N > E > H$	\emptyset
Theme Freq.	<i>no significance</i>	$E = N > H$	$H > N^*$
Uncommonness	$N = H > E$	$E = N > H$	$N > H^*$
Cosine distance	$H > E > N$	<i>no significance</i>	\emptyset
Distance to centroid	$H > E > N$	$H > E = N$	$H > N > E$
<i>Subjective measures</i>			
Validity	$N = H > E$	$E = N > H$	$E > N = H$
Form	$N > E = H$	\emptyset	$E > N > H$
Elaboration	$N > E = H$	\emptyset	$E > N > H$
Originality	$N = H > E$	$E = N > H$	$E > N = H$
Feasibility	\emptyset	$E > N > H$	\emptyset

Notes: * signals that there were no significant differences between Expert AI and the two other agents, Naive AI and humans. \emptyset indicates that the criterion is not computed or evaluated for this specific task.

the distance in human-generated texts is more likely a result of limited writing skills (in terms of syntax and/or vocabulary etc.) rather than an intentional creative strategy. For the two other tasks, we only see a clear outperformance of AI over humans when it comes to generate ideas of unusual uses of objects. For the Draw task, the most interesting result is the intermediary position of Expert AI that never performs more poorly than Naive AI or humans in all criteria.

Once we focus on the subjective measures, we can see that the pattern observed previously is not repeated. For the Text task, the evaluators also favoured Naive AI over Expert AI but were ambivalent about the evaluation of human outputs. In fact, for the creative measures specific to the degree of detail and the style of the texts (i.e. elaboration and form), the human performs more poorly than these

specific to originality and validity. This points to a duality between the structure and content of creative outputs. In other words, while AI challenges humans in terms of idea form or shape, humans still keep up the pace regarding the ideas' originality and appropriateness. However, we must insist on distinguishing tasks since this observation does not hold for all tasks, especially for the Alternative Use task in which we observe a lower performance of humans for each measure. In addition, for the first time, the Expert AI outperformed the other agents regarding ideas' originality. Finally, for the Draw task, whatever the criterion chosen, the Expert AI performed better than the other two agents, with a positive and significant effect of the prompting strategy pushing the AI model to perform in such a way as to maximize its creative performance. As far as human performance is concerned, the creative outputs are only maintained at the level of these of the Naive AI in terms of originality and validity, similar to the text task, but fails regarding form and elaboration. To summarize, humans appear to outperform AI on content-related criteria (such as originality and validity) but not on formal criteria (such as form and elaboration) only in text-based task, which is the most closed-ended in our study.

These findings challenge Hypothesis 1 and appear to contradict some previous parts of the literature suggesting that AI excels in closed-ended tasks (e.g., Charness and Grieco [2024]). Several factors could explain this discrepancy. First, in Charness and Grieco [2024], creativity is measured as a single dimension, with ratings of general creativity on a scale from 0 to 10. In contrast, our study employs a more nuanced measurement of creativity, capturing its multidimensional nature and demonstrating that AI's advantage over humans is only partial. Then, our findings underscore the importance of the criteria used to evaluate creativity.

While Koivisto and Grassini [2023], which focuses on the AUT, find that AI, on average, outperforms humans except for the best ones, our results show the opposite configuration where AI outperforms humans but is rarely as bad as them. When comparing the distributions of scores for each criterion across tasks and agents, we see that the lowest scores are mainly endowed by humans¹⁰. More specifically, for the Text task and the AUT, we observe that for all subjective measures, the proportion of humans with the lowest scores is more important than AI proportions. Including objective measures, we also observe a generally higher variability of scores for humans compared to AI. For the Draw task, the results differ as there are higher proportions

¹⁰Comparisons between our three agents per criterion and scores are available in Appendices 5.9.

of AI in the lower scores but the proportions are still smaller than humans. However, we observe a distinct pattern in the subjective measures: both humans and Naive AI exhibit an inverted U-shaped relationship in score proportions. At the same time, Expert AI shows a consistent increase in proportions, aligning with the direction of the scores. Ultimately, these findings are supported by our measures of semantic distance at the individual and group level, as humans provide texts, ideas, and drawings that are more spread in the semantic space compared to AI agents, which increases the uncertainty or unpredictability of their performance.

Our findings also challenge Hypothesis 2 regarding close-ended tasks (i.e. Text tasks) as Naive AI outperform Expert AI on subjective and objective measures but distance. One explanation for this surprising result might be related to how Expert AI was designed in our experiment. As explained in Section 5.5.1, Expert AI was generated by GPT-4 itself, with no human intervention. It seems that the prompts generated by this AI-based process are excessively focused on the task’s constraints, resulting in repetitive and stereotypical outputs featuring knights, princesses, and fairy-tale elements. Paradoxically, instructing GPT-4 to enhance its creative performance led it to overly concentrate on constraints, resulting in poorer scores on some creativity dimensions compared to humans and the naive use of the model. In other words, the Expert model’s rigid adherence to instructions led to a fixation effect, limiting its creativity. Consequently, rather than achieving the goal of an “Expert AI” capable of sophisticated and nuanced responses, the experiment revealed that the model was only able to design a “Meticulous AI” without effectively leveraging the constraints for creative purposes. This outcome underscores the importance of incorporating human feedback to fully utilize AI’s creative potential. Indeed, although we lack direct evidence, it is reasonable to conjecture that a more “human-in-the-loop” approach for Expert AI could significantly enhance its performance and potentially surpass human and Naive AI scores.

Once these results have been put into perspective, we can look at their consequences within organisations. Creativity is a central element in the activity of organisations. However, creativity is also risky because of the uncertainty inherent in the process. Therefore, an organisation’s rational decision is to focus on the most creative profiles to limit their risk. Given that our results show that our AI agents outperform human agents and that a comparison of the distribution of their performance by output shows that AI agents tend to perform better even in terms of minimums, it would be natural to see organisations turning away from human

ideators in favour of artificial ones. Moreover, we observe a higher uncertainty on the performance of humans based on a higher variability of their performance, which might favour the use of AI in order to minimize the risk for the creative process to fail. However, a second part of our results shows that while AI performs better than humans, the models have certain limitations that require interaction with a human agent to intervene in the prompting strategy. We would like to reiterate an important point concerning the text task. Although Generative AI models are black boxes with millions of parameters whose details we do not know, we do know that they are very sensitive to the way in which a query is conducted. In this study, we chose to let the model optimize this query, and we observed that in the Text task, the model continually created its own fixation effect due to the constraint of using a list of words from the same lexical field. Here, a human agent's intervention would contribute to controlling and avoiding such fixation effects to unlock AI's full potential.

To conclude our discussion, it is essential to underline that we controlled for the specific characteristics of our human evaluators to verify if any of their socio-demographic or AI-related information might have influenced their judgment¹¹. In the end, we could not identify any specific pattern and conclude that evaluators' profiles did not influence the evaluation scores used in our analysis.

5.8 Conclusion

This study aims to compare the creative performance of AI and humans to identify the strengths and weaknesses of each agent, human or AI. While there is a growing body of work on this subject, this work is the first to examine different tasks based on three levels of task openness and whose performance is assessed by different criteria, integrating two different prompting strategies for the AI agent. To carry out this study, creative outputs were generated by humans and by two distinct types of AI: a Naive AI (which received the same instructions as humans) and an Expert AI (that was prompted to be specifically creative). Once our three agents had generated the outputs, they were submitted for evaluation to external human evaluators who then assessed different creativity criteria for each task; objective measures of output creativity supplemented these subjective measures. In the end, this study on the

¹¹We also accounted for whether the evaluators detected that some outputs were AI-generated or not. Only a few evaluators noticed this, revealing no significant effect on the collected evaluations.

comparison between the creative capacities of humans and AI allows us to draw lessons both on the study of artificial creativity and the consequences for our society.

Firstly, further research should consider including more advanced AI evaluation of creativity. This last element, already discussed in the literature [Acar, 2023], should also consider which creativity criteria should be computed and whether new ones should be created to cover unexplored or under-explored areas of creativity assessment. In addition, further research should address one limit of our work in considering taking into account creative tasks closer to the real problems organisations face.

Secondly, our results indicate a clear outperformance of artificial creativity over human creativity. However, they also reveal flaws in Generative AI models in producing creative outputs depending on the prompting strategy adopted. While the first results could lead us to praise the replacement of human creativity by the machine, the second reinforces the idea of human-machine collaboration to maximise creative performance. As already stated for AI technologies in general, “*to fully exploit the potential of AI, human and machine intelligence must be tightly interwoven*” [Plasino and Purdy, 2018, p. 19]. The use of AI to generate creative outputs is then enriched by human intervention in fine-tuning this AI’s prompting to act on the problem-framing, then a central element of the creative process. This necessity of human intervention which becomes all the more important as the constraints intensify.

Finally, it is important to acknowledge that our sample consists exclusively of students, primarily undergraduates, who lack specialized skills in creativity or technical abilities in creative problem-solving. A promising direction for future research would be then to explore whether the observed outperformance of AI holds across specific sub-populations, such as entrepreneurs, managers, engineers, or artists. Investigating both standardized tasks (such as those examined in this study) and sector-specific creative activities could provide valuable insights into the extent of AI’s performance relative to human capabilities in various contexts.

5.9 Appendices

Instructions for data collection and generation

Text task

Instructions presented to Human and Naive AI¹²:

“In this first part, you are asked to write an interesting and original story using one or more personal memories or experiences with the list of words below. You must use all of the provided words, respecting their singular or plural forms, in addition to any other combination of words of your choice. The list of words is: ‘walls, bricks, towers, roof, keep, stones, rampart, door, window, flag.’ This task is time-limited, and you have a maximum of 15 minutes to complete it. Once submitted, your text will be evaluated by a jury composed of three other anonymous subjects as follows: your text will be compared with the text of another subject (randomly selected). Each member of the jury will then have to rank the two texts, placing the one they prefer first and the other second. The subject whose text is ranked first by at least two members of the jury will be awarded a ‘jury prize’ of €10. The subject whose text is ranked second will win nothing. The three members of the jury make their decisions in isolation and completely independently (they cannot communicate with each other). Each member must establish a ranking between the two texts. The final ranking is the aggregated ranking of the three jury members. Since the jury is composed of three members, there cannot be a tie between the two texts. The entire ranking procedure is completely anonymous. You will not know the identity of the jury members who evaluated you, nor the identity of the subject against whom you competed. Similarly, the jury members will not know your identity, nor that of the subject against whom you competed. Your ranking will be revealed to you at the end of the experiment.”

Prompt generated by GPT to address task instructions for Expert AI :

“Imagine that you are a storyteller from the Middle Ages, and you must narrate a captivating adventure where an unexpected hero uses elements from his everyday environment—walls, bricks, towers, roof, keep, stones, rampart, door, window, flag—to

¹²All the following instructions were translated from French to English. The original instructions are available upon request.

overcome a series of ingenious challenges. These objects must be central to the plot. You have 15 minutes to weave this story, ensuring it reflects a personal experience or a memorable event from your life, transposed into this medieval universe. Let your imagination run free, remembering that your story will be judged on its originality and ability to captivate an anonymous and independent jury.”

Alternative Uses task

Instructions presented to Human and Naive AI :

“For this task, we ask you to write down on the provided sheets all the original uses you can think of for a given everyday object. There are certainly common and unoriginal ways to use such an object; for this task, only write down the unusual, creative, and uncommon uses that come to mind. To help you better understand what is expected, let’s take the example of a soda can. Common uses would be “holding liquid,” “serving as a glass,” “preserving food,” etc. Whereas the uses for which you could receive credit might include “using it as a flower pot,” “a lantern,” “a windmill” (after cutting it), “a telephone” (when two are connected by a string), etc. Here, there will be no constraints in terms of the shape, number, or size of the object. The task includes five words and will last 15 minutes. You will have three minutes per word to find as many unusual uses as possible. The words will be as follows: A brick A cardboard box An extension cord A metal pipe A t-shirt”

Prompt generated by GPT-4 to address task instructions for Expert AI :

“Immerse yourself in a spirit of unbridled creativity and innovation. For each object on the following list—a brick, a cardboard box, an extension cord, a metal pipe, and a t-shirt—imagine surprising, unorthodox, and inventive uses that transcend their usual functions. Think of applications that would astonish, that would be full of ingenuity, or even poetic or humorous. Let your imagination run wild to transform the ordinary into the extraordinary, revealing unexpected facets of these everyday objects.”

Draw task

Instructions presented to Human and Naive AI :

“For this task, you are asked to imagine and draw an animal from a planet very different from Earth. You will need to draw this animal both from the front and in profile, and to complete your drawing, you will also need to write a short description of the animal and give it a name. For this exercise, you have no constraints other than those mentioned above. You must complete this in a maximum of 15 minutes.”

Prompt generated by GPT-4 to address task instructions for Expert AI :

“Draw a unique extraterrestrial animal that could exist on a planet very different from Earth. Consider strange and marvellous adaptations that would enable it to survive in unusual environments, such as a thick atmosphere, extreme temperatures, or variable gravity. Your creation should reflect boundless imagination and not be inspired by terrestrial life forms. Draw this animal using only a pencil. The goal is to emphasize the details and unique characteristics of the animal, employing a drawing technique that mimics the style and texture of pencil sketches.”

Examples of creative outputs

Text task

Human:

- *“Je me souviens de cette après-midi où nous avons construit une cabane dans le jardin avec l’aide de ma sœur et de mon père. Mon pauvre père ne savait pas encore dans quelle galère il venait de s’engager. En effet, avec ma sœur, nous aspirions à un château. Nous commençons par choisir un grand bosquet à l’intérieur duquel nous pourrions établir notre château. En guise de murs, nous n’avions pas de briques mais des arbres que mon père a soigneusement taillés pour y insérer des ouvertures qui allaient constituer respectivement une porte d’entrée et une fenêtre. Le feuillage des arbres établissait une belle toiture. Nous avons également laissé pousser un des*

arbres au milieu de la cabane pour y faire un donjon quelques années plus tard, lorsque l'arbre serait assez grand. Mon père n'en était pas au bout de ses peines, hélas ! Nous l'achevions en lui réclamant un rempart en pierres, avec deux tours, une pour chacune, tout autour de notre château. C'est à la fin de cette longue période de construction que nous organisons une cérémonie d'inauguration de notre magnifique château/cabane, durant laquelle nous plantons notre drapeau juste au-dessus de l'ouverture de la porte. C'était il y a 15 ans. Aujourd'hui, je repense à ces doux moments et me rends compte à quel point l'enfance est une période si courte et si belle... peut-être un peu plus éprouvante pour mon père..."

- *"Voilà ! On y était arrivé. C'était le début de l'après-midi, le mois d'août. Nous étions montés à la vieille ville de Lisbonne. Le soleil brûlait notre peau, fatiguée du trajet. Une partie du rempart de ce beau château était déjà visible, avec ses deux tours, une de chaque côté, un drapeau sur le sommet. On traverse l'entrée. Nous suivons le trajet défini sur le plan qui nous avait été donné. Nous nous trouvons face à un donjon. Nous montons. À l'intérieur, une petite fenêtre nous laisse voir une partie de la ville. Nous sommes finalement arrivés tout en haut du donjon. Et là, cela fut inoubliable. D'un côté la mer, de l'autre cette belle ville qu'on dirait perdue dans le temps. Après avoir profité de ces rayons de soleil et de cette belle vue, on redescend. Il y avait des pierres et des briques partout. Auraient-elles fait partie des murs et de la toiture qui ne sont plus là ? Qui les a construites ? Resteront-elles toujours là ? Nous devons pourtant partir. Quand nous étions sur le point de le faire, mon compagnon aperçoit une petite porte qui n'était pas dans le guide. Nous hésitions. Faut-il aller jeter un coup d'œil ? Enfin, après beaucoup d'hésitation, nous décidons d'y aller. Nous nous approchons de la porte. Nous l'ouvrons. Nous la traversons."*
- *"Il était un temps, à une époque lointaine où le drapeau était utilisé qu'en temps de guerre, les tours n'enfermaient pas de princesses, et le donjon était destiné à héberger le peuple, un monsieur, plus exactement un fabricant de briques faites à partir de pierres précieuses. Ces briques servaient à construire le rempart, la toiture, la fenêtre, les murs, la grande porte d'une grande maison qui un jour allait servir à tout être humain ayant eu un comportement irréprochable. Ce monsieur était un ange."*

Naive AI:

- *“Le soleil se couchait doucement, jetant une lueur dorée sur les murs du vieux château en ruine qui se nichait au sommet de la colline. Les briques, autrefois robustes et imposantes, étaient désormais érodées par le passage du temps, mais elles portaient encore la marque indélébile d’une époque révolue. Les tours dressées vers le ciel, bien qu’affaissées et effondrées par endroits, conservaient leur aura majestueuse. La toiture, jadis ornée d’élégants ardoises, avait cédé sous le poids des siècles, laissant apparaître un ciel parsemé d’étoiles à travers les interstices. L’ombre du donjon se dessinait de manière menaçante dans le lointain, ajoutant une touche mystique à l’ensemble du paysage. Je marchais lentement, émerveillé par l’histoire ancienne que racontait chaque pierre du rempart. La porte principale, bien que franks est délabrée, était encore suffisamment robuste pour agir en tant que gardien impénétrable du royaume oublié. La fenêtre à ma droite me donnait un aperçu de l’intérieur du château, où avait autrefois régné l’activité. Tout semblait en paix maintenant, à l’exception de la lueur du soleil couchant qui dansait sur les murs intérieurs, faisant revivre le château pendant un bref instant. Un drapeau battait doucement dans la brise au sommet du donjon, déchiré et coloré par le temps, mais il portait encore l’emblème qui avait autrefois été un signe de puissance et de gloire. Le château, avec toutes ses cicatrices, était un symbole frappant de la résilience et de l’endurance, un rappel silencieux que même en ruines, la beauté peut être trouvée. Cette promenade dans les ruines du château a été une expérience qui m’a profondément touché. Elle m’a rappelé que même si tout semble perdu, il y a toujours des traces d’espoir et de beauté qui demeurent. Ce château, malgré son état de délabrement, est resté debout, défiant le temps et l’histoire. C’est un symbole d’espoir et de résilience qui continuera à inspirer, même après mon départ.”*
- *“C’était un petit village niché au creux d’une vallée. L’héritage de ses glorieux ancêtres se reflétait dans chaque recoin : des murs de briques solides et épais en passant par les tours majestueuses qui surveillaient les montagnes environnantes. Il y avait une toiture en tuiles vernissées qui brillait sous le soleil d’été, et un imposant donjon qui se dressait en plein centre, comme un géant silencieux qui veillait sur ses habitants. Des pierres sculptées minutieusement marquaient le chemin vers le rempart, là où résidait le cœur et l’âme du village. C’était un lieu de rassemblement pour les jeunes et les vieux, où les nouvelles étaient échangées, où les enfants couraient librement, les échos de leurs rires résonnant contre les murs de Pierre. Au centre du rempart, il y avait une porte imposante, vieille de plusieurs siècles. Ses planches de bois étaient usées, mais elle se tenait toujours fièrement, comme une vieille dame qui avait vu*

passer les saisons. Elle avait accueilli des rois et des reines, avait été le témoin d'innombrables amours naissantes et de cœurs brisés, et avait survécu à des guerres et des famines. Les fenêtres du village étaient comme des yeux qui s'ouvraient sur le monde. Chaque soir, elles s'illuminaient de mille feux, dessinant un tableau vivant de couleurs chaleureuses. Derrière ces fenêtres, chaque habitant du village avait une histoire à raconter, un souvenir à partager. Et puis, il y avait le drapeau. Il flottait fièrement au sommet du donjon, agitant ses couleurs vives au gré du vent. Il était un symbole d'unité et de résilience, un rappel que ce village avait survécu à l'épreuve du temps et qu'il continuerait de le faire. C'est dans ce village que j'ai grandi, grandissant avec les histoires de mes ancêtres et les rêves d'un futur meilleur. C'est ici que j'ai appris l'importance de la communauté, l'importance de se souvenir de ses racines tout en regardant vers l'avenir. Chaque brique, chaque pierre, chaque tour, chaque morceau de cette toiture, chaque fenêtre, chaque porte, le drapeau, le donjon et le rempart... ils font tous partie de moi, et je fais partie d'eux."

- *"Laissez-moi vous raconter mon incroyable voyage à travers le temps, dans un château médiéval reculé, encerclé de hauts murs de briques solides. Le lieu avait été témoin de nombreux siècles, et chaque pierre racontait sa propre histoire. Les tours majestueuses du château se dressaient fièrement dans le ciel, dominant le paysage environnant avec une grâce sereine. La toiture du donjon, d'une architecture exquise, avait miraculeusement résisté au poids des ans. Il y avait quelque chose de véritablement fascinant dans la façon dont le temps semblait s'être arrêté en cette place. Le rempart du château, autrefois un symbole de puissance et de protection, avait subi l'érosion du temps, il était couvert de mousse mais n'en restait pas moins impressionnant. Je me suis dirigé vers la porte principale du château. Elle était grande, imposante, forgée en bois massif et renforcée de fer. Autrefois, cette porte devait être le premier rempart de défense contre les envahisseurs. À présent, elle était entrouverte, invitant le voyageur errant à découvrir les secrets qu'elle gardait. Je suis entré à l'intérieur, mes pas résonnant sur le sol de pierre. L'intérieur était un mélange d'ombre et de lumière, chaque fenêtre offrant une vue unique sur les terres environnantes. Le château était dépourvu de ses anciens habitants, il restait néanmoins doté d'une vie propre, une vie qu'on pouvait ressentir à chaque coin et recoin. Sur le sommet du donjon, flottant au vent, un drapeau déchiré portant le blason du château témoignait d'un passé oublié. Son gracieux ballet, bien qu'usé par le temps, portait en lui l'âme de ce lieu, un dernier hymne à la grandeur d'antan. Ce château, à la fois majestueux et mystérieux, était plus qu'un simple tas de pierres. C'était un livre d'histoire vivant, un testament silencieux de*

l'époque révolue. Et pendant un moment, j'ai eu la chance d'être une petite partie de cette histoire, un simple voyageur traversant les pages du temps."

Expert AI:

- *"Il était une fois un simple jardinier nommé Fernand, employé du château de Ker-Broc. C'était un homme ordinaire, mais son esprit vivant contrastait avec les murs massifs et les tours intimidantes du château. Chaque jour, il s'occupait des plantes, ramassait les pierres du chemin et admirait le drapeau flottant sur le donjon. Il était émerveillé par les merveilles que sa position humble lui offrait. Un jour, le seigneur du château annonça qu'un tournoi aurait lieu, et le gagnant deviendrait son successeur. Fernand fut surpris. Il n'était qu'un jardinier, sans aucune expérience en combat, comment pourrait-il rêver de remporter un tournoi? Pourtant, à chaque fois qu'il passait un rempart ou franchissait la porte grise du château, l'idée lui revenait. La veille du tournoi, Fernand ne put fermer l'œil. Il se leva et arpenta les murs en pierre, sophistiqués mais pourtant si familiers pour lui. Dans son esprit, il élaborait un plan audacieux. Il connaissait chaque brique et chaque tour du château, chaque fenêtre et chaque fente dans le rempart. Il connaissait les coins et recoins de la toiture du donjon si bien qu'il pourrait s'y déplacer les yeux bandés. Le jour du tournoi arriva. Fernand sauta dans l'arène, vêtu de sa tenue de jardinier et armé de son courage. Les rires moqueurs des chevaliers étaient étouffés lorsqu'il utilisa les briques du mur pour déséquilibrer son adversaire. Il grimpa à la tour pour éviter un autre, puis se cacha dans le rempart pour désarçonner le troisième. Il franchit la porte du donjon pour échapper aux autres, utilisa les toitures pour sauter par-dessus eux, puis se précipita vers le drapeau. Le public était dans l'incrédulité et émerveillé tandis que Fernand hissait le drapeau à la cime du donjon. Les chevaliers furent stupéfaits par le courage de ce jardinier et le seigneur, impressionné par son ingéniosité, lui donna la victoire. Fernand le jardinier était devenu Fernand le Seigneur. Cette histoire fait écho à ma propre expérience. Dans ma vie, j'ai dû subir des défis qui semblaient insurmontables. Mais comme Fernand, j'ai exploité mon environnement familier, mes compétences et mes connaissances pour les surmonter. Nous ne sommes peut-être pas des héros traditionnels, mais nous sommes des héros de notre propre histoire."*
- *"Une fois, dans un ancien royaume, vivait un jeune berger nommé Théo. Loin d'être un héros traditionnel, il passait la majorité de ses journées parmi les moutons et*

les pierres de l'immense rempart qui protégeait le village. À l'époque, je n'étais qu'un simple enfant, ami de Théo, émerveillé par l'immensité des murs de la cité, qui étaient faites de briques rouges robustes et se prolongeaient jusqu'à des tours menaçantes. Un jour, une prophétie s'est abattue sur notre humble village. La prédiction était claire : une énorme créature s'en prendrait à notre donjon et ne laisserait rien derrière elle. Théo, sans aucune compétence de chevalier, a décidé qu'il serait celui qui sauverait le village. Sa première épreuve était d'atteindre le donjon. Avec ingéniosité, il a utilisé les briques du rempart pour construire un escalier improvisé jusqu'à la toiture d'une maison. De là, il pourrait atteindre les tours et finalement arriver au donjon. En parcourant les murs, il a vu des fenêtres avec des familles terrifiées. Il a vu dans leurs yeux une lueur d'espoir car, pour la première fois, elles voyaient quelqu'un se battre pour elles. Cela l'a motivé encore plus. Déjouant la gravité, il a escaladé les tours, se rapprochant de plus en plus du donjon. Son voyage n'a pas été sans dangers. Les pierres du mur se détachaient parfois, rendant son ascension périlleuse. Pourtant, chaque fenêtre qu'il dépassait l'élevait plus haut, l'encourageant à continuer. Finalement, il a atteint le sommet du donjon, où le drapeau du royaume flottait fièrement. Il l'a remplacé par sa cape de berger, un symbole qu'il était temps pour un simple citoyen de défendre son royaume. Alors que la créature approchait, il a utilisé une corde pour descendre à la porte du donjon. Il a pris une pierre et l'a lancée sur la créature. Elle a rugi et a commencé à charger vers le donjon. Avec un sourire, il a fermé la porte du donjon. La créature, ne voyant pas l'obstacle, s'est heurtée à la porte et a été assommée. Les villageois ont acclamé Théo, le berger transformé en héros. Son exploit est devenu une histoire racontée encore et encore dans le royaume, en souvenir de l'ingéniosité et du courage d'un berger. Et pour moi, c'est l'histoire d'un ami qui, malgré son manque de force ou de formation, a utilisé les éléments de son environnement quotidien pour devenir le héros que notre village méritait. J'en ai tiré une leçon précieuse : nous sommes tous capables de grandes choses, à condition de faire preuve d'ingéniosité et de courage.”

- “Il était une fois, dans un village paisible du nom de Cluny, vivait un simple artisan du nom de Francis. Francis, un maçon de son état, vivait sa vie entre les murs de briques qu’il construisait et la douce chaleur de sa maison. Son destin, cependant, était bien plus grand que ce que ces murs pourraient jamais contenir. Une nuit, alors que Francis se reposait après une longue journée de labeur, le village fut attaqué par un seigneur féodal connu sous le nom de Baron Noir. Soudain, la quiétude du village s’évanouit, remplacée par le cri des tours de garde, le crépitement des flammes sur les

toitures et le grondement des roues du chariot du Baron Noir. Le Baron Noir avait un objectif : le donjon de Cluny, où se cachait un trésor mythique. Cependant, le donjon était bien protégé, par des remparts solides et une porte de fer presque impénétrable. Le baron, sûr de sa victoire, installa son drapeau noir au sommet de la colline, en signe de provocation. Francis, de son humble demeure, observait la vue par la fenêtre. Avec chaque pierre de son village qui tombait, son cœur se serrait de douleur. Il ne pouvait pas rester les bras croisés. Il devait agir. Armé de son expertise, il élaborait un plan. Avec les mêmes briques qu'il utilisait pour construire de simples murs, Francis commença à construire des pièges autour du donjon. Il renforça les remparts, ajouta des blocs de pierres pour créer un labyrinthe, installa une fausse porte pour tromper l'ennemi. La nuit tomba et le Baron Noir avança, sûr de sa victoire. Mais à chaque tour, à chaque recoin, il était confronté à un nouveau défi. Les murs qui semblaient fragiles étaient presque impénétrables, les remparts étaient insurmontables, et la porte qu'il pensait être l'entrée du donjon n'était rien d'autre qu'un piège. L'aube se leva et le Baron Noir n'était nulle part en vue. Le drapeau noir avait été remplacé par celui du village. Francis, de sa fenêtre, souriait en regardant son œuvre. Il avait réussi. Il avait utilisé les éléments les plus ordinaires de son quotidien pour protéger son village, pour être un héros. Cette histoire, mes amis, me rappelle ma propre expérience. Tout comme Francis, j'ai dû utiliser les outils de mon quotidien pour surmonter les défis qui se présentaient à moi. C'est une leçon que nous devrions tous retenir : peu importe combien une tâche peut paraître grande, avec de l'ingéniosité et de la détermination, rien n'est impossible."

Alternative Uses task

Human:

- Brique: *une arme de jet, une haltère, une marche pour faire du step, un chapeau, un instrument de musique percussion (adapté pour un public souffrant d'hyperacousie).*
- Boîte en carton: *jeu pour enfants ou chats, support peinture, décoration (après découpage), luge.*

Naive AI:

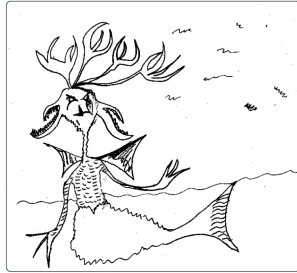
- Brique: *presser des vêtements en remplacement d'un fer à repasser, outil de cuisson pour le poulet rôti à la brique, support de livre, altère de fortune pour faire de l'exercice, bloc de refroidissement pour les ordinateurs portables, instrument de musique à percussion sabot de frein pour les vélos.*
- Boîte en carton: *maison pour chat ou cabane, décor de théâtres pour enfants, pot de plantation biodégradable, support d'écran pour créer un cinéma maison avec un smartphone, mur d'isolation phonique abat-jour pour lampe, chaise ou tabouret (si assez robuste), costume de robot pour Halloween.*

Expert AI:

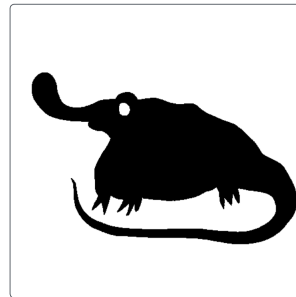
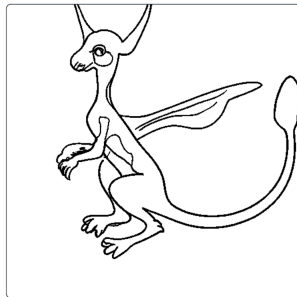
- Brique: *porte-livre sur une étagère, presse-papier de bureau, support de yoga pour des poses de flexibilité, four solaire miniature en la couvrant de feuille d'aluminium, moule à glace pour faire des glaçons géants, une percussion musicale en la frappant avec un marteau dans un concert de musique expérimentale.*
- Boîte en carton: *un théâtre d'ombre miniature pour des séances de conte nocturnes, une maison pour jouer pour les enfants, un espace de rangement secret sous le lit, un attrape-rêves décoratif une fois découpée et peinte. un bateau de marin d'eau douce dans une mare, un support pour des semis dans un potager urbain, une cabine à peindre pour libérer sa créativité artistique.*

Draw task

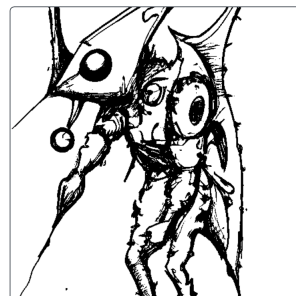
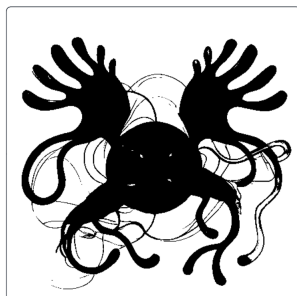
Human



Naïve AI



Expert AI



Instructions for creative assessment

General instructions

“Welcome to this social science experiment. We sincerely thank you for your participation. Before we begin, we would like to remind you of the importance of responding

seriously and thoroughly to the questions posed so that this experiment can yield valid and relevant results. All your responses will be collected anonymously. In this experiment, you will perform the role of an evaluator of various texts. You will be asked to assess a series of 15 texts according to different criteria. At the end of the experiment, you will also need to complete a questionnaire. For your participation in the entire experiment, you will be compensated 15€.

Text task

“For this task, a series of texts will appear on the screen one after the other, randomly. The texts have been created following the instructions below:

Instructions: You are asked to write an interesting and original story using one or more personal memories or experiences with the list of words below. You must use all of the proposed words, respecting their singular or plural forms, in addition to any other combinations of words of your choice. The list of words is “walls, bricks, towers, roof, keep, stones, rampart, door, window, flag.”

For each displayed text, you are asked to rate the text based on an evaluation grid consisting of four criteria:

- *Validity: adherence to instructions 0 to 5*
- *Form: writing style, syntax 0 to 5*
- *Elaboration: level of detail/description of the story 0 to 5*
- *Originality: 0 to 5*

A total of 15 texts will be displayed on the screen.

Alternative Uses task

“For this task, several lists of ideas for unusual uses of a common object will be displayed successively and randomly on the screen. The lists have been created following the instructions below:

Instructions: For this task, we ask you to write down on the sheets provided all the original uses of a given everyday object. There are certainly common and unoriginal

ways to use such an object; for this task, only write down the unusual, creative, and uncommon uses that you can think of. To help you better understand what is expected, let's take the example of a soda can. Common uses would be "containing liquid," "serving as a glass," "preserving food," etc. Whereas uses for which you could be credited might be: "using it as a flower pot," "a lantern," "a windmill" (after cutting), "a phone" (when two are connected by a wire), etc. There will be no constraints regarding the shape, number, or size of the object.

For each list displayed, you are asked to rate each idea using a scoring grid consisting of three criteria:

- *Validity: adherence to the instructions 0 to 5*
- *Feasibility: the suggested use of the object is doable/viable 0 to 5*
- *Originality: 0 to 5*

A total of 15 lists of ideas will be displayed on the screen."

Draw task

"For this task, a series of drawings will be displayed successively and randomly on the screen. The drawings were created following the instructions below:

Instructions: For this task, you are asked to imagine and draw an animal from a planet very different from Earth.

For each drawing displayed, you are asked to rate the drawing using an evaluation grid consisting of four criteria:

- *Validity: adherence to instructions 0 to 5*
- *Form: style of the drawing, graphic quality 0 to 5*
- *Elaboration: level of detail/precision of the drawing 0 to 5*
- *Originality 0 to 5*

A total of 15 drawings will be displayed on the screen."

Questionnaires

Socio-demographic questionnaire

- | | |
|---|---|
| <p>1. Your age: _____</p> <p>2. Your gender:</p> <p style="padding-left: 40px;">Male</p> <p style="padding-left: 40px;">Female</p> <p>3. The type of degree you are enrolled in:</p> <p style="padding-left: 40px;">Bachelor's</p> <p style="padding-left: 40px;">Master's</p> <p style="padding-left: 40px;">Doctorate</p> <p style="padding-left: 40px;">Other (specify) _____</p> <p>4. Your field of study:</p> <p style="padding-left: 40px;">Law Economics-Management</p> <p style="padding-left: 40px;">Literature-Languages</p> <p style="padding-left: 40px;">Exact Sciences</p> <p style="padding-left: 40px;">Psychology-Sociology</p> <p style="padding-left: 40px;">Political Science</p> <p style="padding-left: 40px;">Other (specify) _____</p> | <p>5. Is French your native language?</p> <p style="padding-left: 40px;">Yes</p> <p style="padding-left: 40px;">No (If no, which one?)</p> <p style="padding-left: 40px;">_____</p> <p>6. Among the following hobbies, check these you have or are currently practising:</p> <p style="padding-left: 40px;">Writing</p> <p style="padding-left: 40px;">Drawing</p> <p style="padding-left: 40px;">Painting</p> <p style="padding-left: 40px;">Reading</p> <p style="padding-left: 40px;">Visual Arts</p> <p style="padding-left: 40px;">Cinema/Theater</p> <p>7. On a scale of 1 to 10, how would you rate your proficiency in the French language (spelling, style, syntax, etc.)?</p> |
|---|---|

AI questionnaire

- | | |
|--|--|
| <p>1. During this experiment, we presented you with different productions to evaluate. These productions were created in two distinct ways. In your opinion, what are these two ways and how did you</p> | <p>identify them? (If you do not know how to answer this question, click "Next")</p> <p>2. During this experiment, we presented you with different productions to evaluate. Half of these pro-</p> |
|--|--|

- | | |
|--|---|
| <p>ductions were generated by humans and the other half by artificial intelligence. (Yes or No)</p> <p>3. If yes, did this influence your judgment?</p> <ul style="list-style-type: none"> • Yes, I think I judged the AI-generated productions more favorably • Yes, I think I judged the human-generated productions more favorably • It did not impact my judgment <p>4. How would you rate your knowledge of artificial intelligence? (1 being “I know absolutely nothing” and 5 being “I have advanced knowledge”)</p> <p>5. In your daily life (outside of work or studies), do you use artificial intelligence tools?</p> <p style="margin-left: 40px;">Never
Rarely
Often
Regularly</p> <p>6. Which applications?</p> <p style="margin-left: 40px;">GPT-4, Bing, Bard</p> | <p>Midjourney, DALL-E
DeepL
Others:</p> <p>7. In your work or studies, do you use artificial intelligence tools?</p> <p style="margin-left: 40px;">Never
Rarely
Often
Regularly</p> <p>8. Which applications?</p> <p style="margin-left: 40px;">GPT-4, Bing, Bard
Midjourney, DALL-E
DeepL
Others:</p> <p>9. AI applications are a good way to improve what I create. (From 1 “Strongly disagree” to 5 “Strongly agree”).)</p> <p>10. AI applications are a good way to generate relevant content. (From 1 “Strongly disagree” to 5 “Strongly agree”).)</p> <p>11. How would you rate your level of creativity? (From 1 to 5.)</p> <p>12. Do you think an AI can be creative? (Yes or No)</p> |
|--|---|

Descriptive statistics for each task

Table 5.13: Text Task Descriptive Statistics

Variable	Naive AI	Expert AI	Human
Mean Validity	4.418 (0.818)	4.322 (0.889)	4.080 (1.074)
Mean Form	4.012 (0.913)	3.966 (0.886)	3.156 (1.31)
Mean Elaboration	3.945 (0.856)	4.136 (0.849)	2.829 (1.311)
Mean Originality	3.506 (1.103)	3.728 (1.063)	2.972 (1.445)
# Evaluations	330	323	327
Mean # Words	380.424 (46.22)	486.735 (51.071)	226.845 (89.011)
Mean # Sentences	17.506 (3.383)	23.940 (3.782)	11.381 (5.346)
Mean Mean length sentences	22.106 (2.58)	20.612 (2.376)	21.737 (7.126)
Mean Cosine Distance	0.676 (0.049)	0.719 (0.046)	0.759 (0.072)
Mean Variety	7.847 (2.146)	5.446 (1.484)	4.929 (1.678)
Mean Theme Frequency (neg.)	-0.030 (0.018)	-0.030 (0.019)	-0.028 (0.019)
Mean Uncommonness	-0.310 (0.116)	-0.472 (0.194)	-0.350 (0.178)
# Observations	85	83	84

Table 5.14: Alternative Uses Task Descriptive Statistics

Variable	Naive AI	Expert AI	Human
Mean Validity	4.495 (0.85)	4.432 (0.947)	4.099 (1.272)
Mean Feasibility	4.729 (0.686)	4.672 (0.657)	4.292 (1.152)
Mean Originality	4.427 (0.841)	4.578 (0.697)	3.911 (1.24)
# Evaluations	192	192	192
Mean # word mean	6.345 (1.753)	8.073 (1.735)	5.022 (2.585)
Mean Cosine Distance	0.831 (0.064)	0.848 (0.045)	0.831 (0.096)
Mean Variety	5.796 (1.155)	4.981 (0.789)	3.278 (1.393)
Mean Balance (neg.)	0.952 (0.047)	0.955 (0.053)	0.950 (0.075)
Mean Diversity	1.698 (0.231)	1.561 (0.173)	1.062 (0.468)
Mean Theme Frequency (neg.)	-0.030 (0.018)	-0.030 (0.019)	-0.028 (0.019)
Mean Uncommonness	-0.310 (0.116)	-0.472 (0.194)	-0.350 (0.178)
# Observations	54	54	54

Table 5.15: Draw Task Descriptive Statistics

Variable	Naive AI	Expert AI	Human
Mean Validity	3.071 (1.559)	4.062 (1.203)	3.182 (1.54)
Mean Form	3.031 (1.303)	3.788 (1.218)	2.043 (1.319)
Mean Elaboration	2.692 (1.355)	3.615 (1.287)	1.923 (1.387)
Mean Originality	2.548 (1.449)	3.797 (1.268)	2.554 (1.516)
# Evaluations	325	325	325
Mean % Pixel Used	0.221 (0.115)	0.238 (0.095)	0.075 (0.042)
Mean Variety	6.811 (2.365)	7.544 (2.558)	8.033 (2.524)
Mean Theme frequency	-0.031 (0.022)	-0.029 (0.024)	-0.023 (0.017)
Mean Uncommonness	-0.704 (0.103)	-0.716 (0.094)	-0.737 (0.09)
# Observations	90	90	90

Pairwise Wilcoxon Test

Table 5.16: Comparison of Agents' Performance for the Text Task (non-parametric)

Metric	Expert AI vs Naive AI	Human vs Naive AI	Human vs Expert AI
Variety	-2.401***	-2.918***	-0.517**
Theme Frequency (neg.)	0	0.003	0.002
Uncommonness	-0.162***	-0.04	0.122***
Cosine distance	0.043***	0.083***	0.04***
Distance to centroid	0.059***	0.127***	0.068***
Validity	-0.096	-0.339***	-0.242**
Form	-0.046	-0.856***	-0.81***
Elaboration	0.191***	-1.117***	-1.307***
Originality	0.221**	-0.534***	-0.755***
# Words	106.767***	-153.27***	-260.037***
# Sentences	6.408***	-6.139***	-12.547***
Mean length sentences	-1.449***	-0.277***	1.172

Notes: This table presents the results of the Pairwise Wilcoxon test for the specified variable and group. The coefficients represent the Mean differences between groups. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 5.17: Comparison of Agents' Performance for the Alternative Uses Task (non-parametric)

Metric	Expert AI vs Naive AI	Human vs Naive AI	Human vs Expert AI
Variety	-0.815***	-2.519***	-1.704***
Balance (neg.)	0.003	-0.002	-0.005
Diversity	-0.137***	-0.636***	-0.499***
Theme Frequency (neg.)	0	-0.002	-0.003**
Uncommonness	-0.139	-3.657**	-3.518*
Distance to centroid	0.008	0.022***	0.015***
Validity	-0.062	-0.396***	-0.333**
Feasibility	-0.057	-0.438***	-0.38***
Originality	0.151	-0.516***	-0.667***
Mean # Words	1.7***	-1.295***	-2.995***

Notes: This table presents the results of the Pairwise Wilcoxon test for the specified variable and group. The coefficients represent the Mean differences between groups. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 5.18: Comparison of Agents' Performance for the Draw Task (non-parametric)

Metric	Expert AI vs Naive AI	Human vs Naive AI	Human vs Expert AI
Variety	0.733	1.222***	0.489
Theme Frequency (neg.)	0.002	0.008	0.006
Uncommonness	-0.012	-0.033**	-0.022
Distance to centroid	-0.02***	0.01***	0.03***
Validity	0.991***	0.111	-0.88***
Form	0.757***	-0.988***	-1.745***
Elaboration	0.923***	-0.769***	-1.692***
Originality	1.249***	0.006	-1.243***
% Pixel Used	0.018***	-0.143***	-0.161***

Notes: This table presents the results of the Pairwise Wilcoxon test for the specified variable and group. The coefficients represent the Mean differences between groups. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

PCA Analysis

Text task

Figure 5.4: Text Task PCA - All Criteria (Visual)

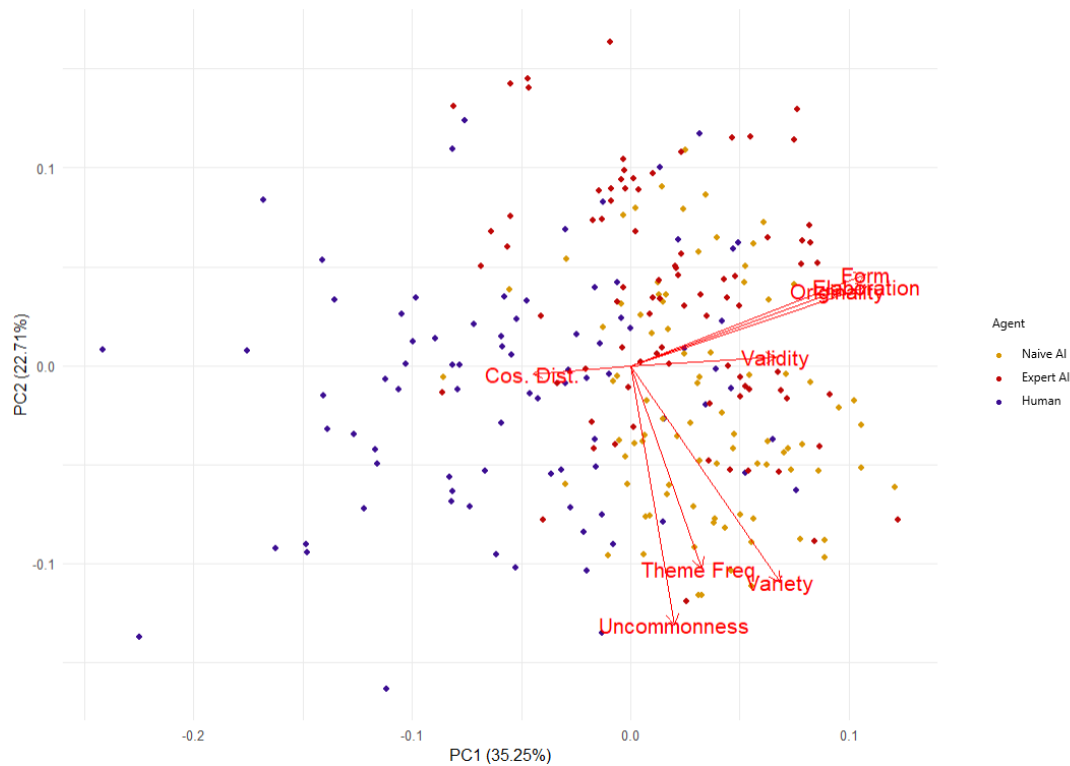


Table 5.19: Text Task PCA – All Creativity Criteria

name	Contribution dim 1	name	Contribution dim 2
Elaboration	26.09	Uncommonness	38.14
Form	25.68	Variety	26.74
Originality	20.16	Theme Freq.	23.43
Variety	10.42	Form	4.76
Validity	9.93	Elaboration	3.60
Cos. Dist.	4.49	Originality	3.25
Theme Freq.	2.35	Validity	0.04
Uncommonness	0.88	Cos. Dist.	0.04

Alternative Uses task

Figure 5.5: Alternative Uses Task PCA - All Criteria (Visual)

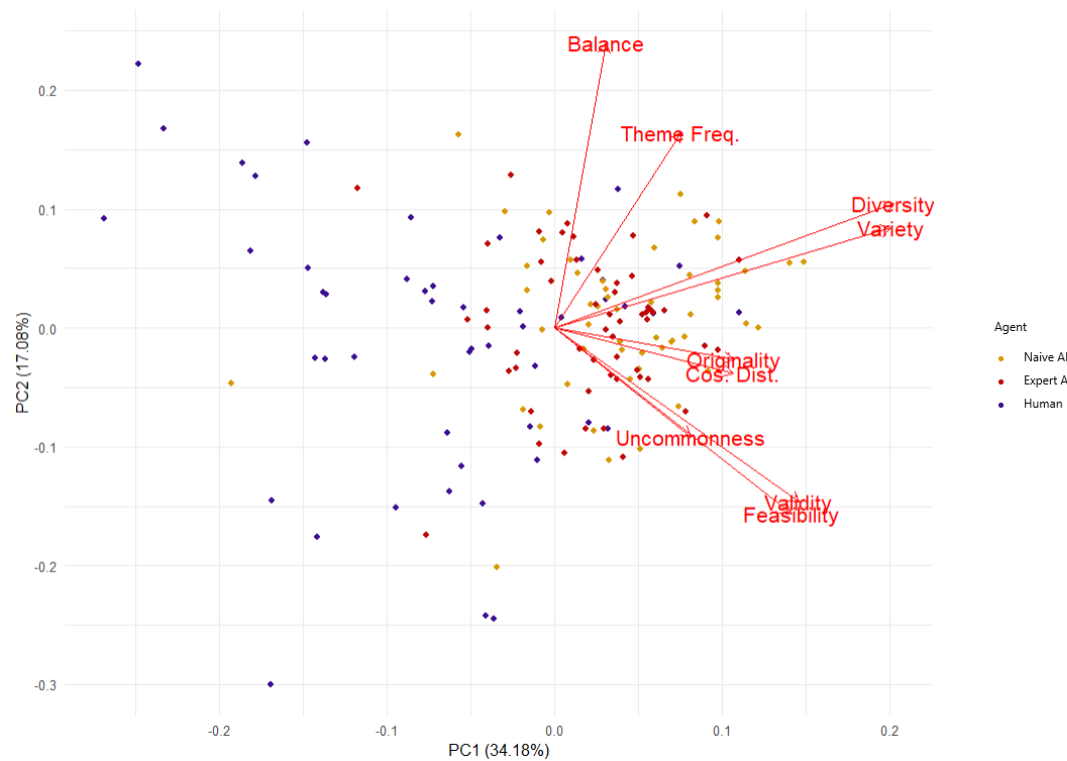


Table 5.20: Alternative Uses Task PCA – All Creativity Criteria

name	Contribution dim 1	name	Contribution dim 2
Diversity	25.83	Balance	36.14
Variety	25.36	Theme Freq.	17.03
Validity	13.44	Feasibility	15.41
Feasibility	12.60	Validity	13.44
Originality	7.24	Diversity	6.92
Cos. Dist.	7.18	Uncommonness	5.27
Uncommonness	4.19	Variety	4.46
Theme Freq.	3.56	Cos. Dist.	0.91
Balance	0.59	Originality	0.43

Draw task

Figure 5.6: Alternative Uses Task PCA - All Criteria (Visual)

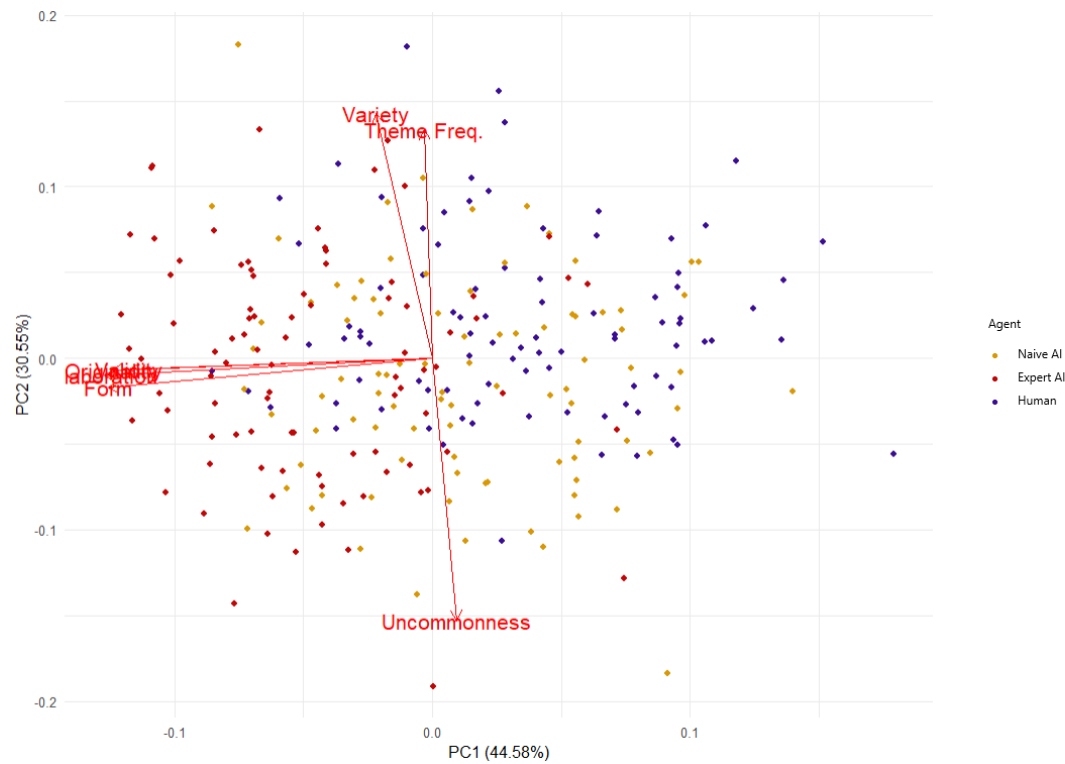


Table 5.21: Draw Task PCA – All Creativity Criteria

name	Contribution dim 1	name	Contribution dim 2
Elaboration	26.21	Uncommonness	37.79
Form	25.57	Variety	32.85
Originality	24.93	Theme Freq.	28.57
Validity	22.34	Form	0.47
Variety	0.79	Elaboration	0.16
Uncommonness	0.15	Originality	0.07
Theme Freq.	0.02	Validity	0.07

Additional tests and regressions

Differences in Evaluators' Characteristics across tasks

Table 5.22: Socio-Demographic Variables Descriptive Statistics

Variable	Text	Alternative Uses	Draw
Mean Age	22.814 (3.036)	22.094 (2.646)	22.031 (2.952)
Mean Gender	0.657 (0.475)	0.672 (0.47)	0.831 (0.375)
Mean Diploma Licence	0.414 (0.493)	0.484 (0.5)	0.415 (0.493)
Mean Diploma Master	0.443 (0.497)	0.469 (0.499)	0.508 (0.5)
Mean Diploma Doctorat	0.057 (0.232)	0 (0)	0.031 (0.173)
Mean Droit	0.057 (0.232)	0.047 (0.212)	0.015 (0.123)
Mean Economie Gestion	0.414 (0.493)	0.297 (0.457)	0.4 (0.49)
Mean Lettres Langues	0.057 (0.232)	0.047 (0.212)	0.046 (0.21)
Mean Sciences exactes	0.186 (0.389)	0.203 (0.403)	0.154 (0.361)
Mean Psycho Socio	0.071 (0.258)	0.062 (0.242)	0.092 (0.29)
Mean Sciences politiques	0.029 (0.167)	0.094 (0.292)	0.092 (0.29)
Mean Native speaker	0.8 (0.4)	0.906 (0.292)	0.923 (0.267)
Mean Writting	0.229 (0.42)	0.219 (0.414)	0.292 (0.455)
Mean Drawing	0.286 (0.452)	0.203 (0.403)	0.2 (0.4)
Mean Painting	0.157 (0.364)	0.219 (0.414)	0.138 (0.346)
Mean Reading	0.743 (0.437)	0.766 (0.424)	0.785 (0.411)
Mean Art	0.1 (0.3)	0.156 (0.363)	0.2 (0.4)
Mean Cinema	0.486 (0.5)	0.562 (0.497)	0.523 (0.5)
Mean French skills	7.671 (1.481)	8.25 (1.415)	8.462 (1.039)
Mean Self Awareness	0.414 (0.493)	0.281 (0.45)	0.308 (0.462)
Mean Dectect AI	0.114 (0.318)	0.031 (0.174)	0.092 (0.29)
Mean Jug favorable human	0.029 (0.167)	0 (0)	0.031 (0.173)
Mean Jug favorable ai	0.014 (0.119)	0 (0)	0 (0)
Mean Jug no impact	0.071 (0.258)	0.031 (0.174)	0.062 (0.24)
Mean AI knowledge	3.043 (0.801)	2.762 (0.771)	2.938 (0.927)
Mean AI usage daily	1.057 (0.827)	1.175 (0.847)	1.369 (0.97)
Mean AI app daily text	0.882 (0.322)	0.875 (0.331)	0.87 (0.336)
Mean AI app daily image	0.059 (0.235)	0.083 (0.277)	0.093 (0.29)
Mean AI app daily trad	0.608 (0.489)	0.562 (0.497)	0.593 (0.492)
Mean AI usage work	1.443 (0.822)	1.476 (0.871)	1.892 (0.914)
Mean AI app work text	0.933 (0.25)	0.926 (0.262)	0.934 (0.248)
Mean AI app work image	0.017 (0.128)	0.056 (0.229)	0.016 (0.127)
Mean AI app work trad	0.583 (0.493)	0.574 (0.495)	0.59 (0.492)
Mean Amelioration	3.314 (0.919)	3.254 (1.07)	3.277 (1.075)
Mean Pertinence	2.886 (0.767)	3.159 (0.947)	3.015 (0.985)
Mean Self Creativity	3.329 (0.751)	3.286 (0.951)	3.231 (0.941)
Mean AI Creativity	0.514 (0.5)	0.656 (0.475)	0.538 (0.499)
Mean Willingness	0.486 (0.5)	0.375 (0.485)	0.292 (0.455)
# Evaluators	70	64	65

Table 5.23: Comparison of Evaluators' Socio-Demographic Characteristics Across Tasks

Metric	Draw vs Alternative Uses	Text vs Alternative Uses	Text vs Draw
Age	-0.063	0.721***	0.784***
Gender	0.159***	-0.015	-0.174***
Diploma Licence	-0.069**	-0.07**	-0.001
Diploma Master	0.039	-0.026	-0.065**
Diploma Doctorat	0.031***	0.057***	0.026***
Droit	-0.031***	0.01	0.042***
Economie Gestion	0.103***	0.117***	0.014
Lettres Langues	-0.001	0.01	0.011
Sciences exactes	-0.049**	-0.017	0.032
Psycho Socio	0.03*	0.009	-0.021
Sciences politiques	-0.001	-0.065***	-0.064***
Native speaker	0.017	-0.106***	-0.123***
Writting	0.074***	0.01	-0.064***
Drawing	-0.003	0.083***	0.086***
Painting	-0.08***	-0.062***	0.019
Reading	0.019	-0.023	-0.042*
Art	0.044*	-0.056***	-0.1***
Cinema	-0.039	-0.077**	-0.037
French skills	0.212***	-0.579***	-0.79***
Self Awareness	0.026	0.133***	0.107***
Detect AI	0.061***	0.083***	0.022
Jug favorable human	0.031***	0.029***	-0.002
Jug favorable ai	0	0.014***	0.014***
Jug no impact	0.03**	0.04***	0.01
AI knowledge	0.177***	0.281***	0.104**
AI usage daily	0.195***	-0.117**	-0.312***
AI app daily text	-0.005	0.007	0.012
AI app daily image	0.009	-0.025	-0.034**
AI app daily trad	0.03	0.045	0.015
AI usage work	0.416***	-0.033	-0.449***
AI app work text	0.009	0.007	-0.001
AI app work image	-0.039***	-0.039***	0
AI app work trad	0.016	0.009	-0.007
Amelioration	0.023	0.06	0.037
Pertinence	-0.143***	-0.273***	-0.13***
Self Creativity	-0.055	0.043	0.098**
AI Creativity	-0.118***	-0.142***	-0.024
Willingness	-0.083***	0.111***	0.193***

Notes: This table presents the results of the ANOVA and Tukey's HSD test test for the specified variable and group. The coefficients represent the Mean differences between groups. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Effect of Evaluators' Characteristics on Evaluation**Text task**

Table 5.24: OLS – Average Score by Evaluators for the Text Task

	<i>Dependent variable:</i>			
	Validity (1)	Form (2)	Elaboration (3)	Originality (4)
Age	0.024 (0.016)	−0.023 (0.019)	−0.024 (0.021)	0.011 (0.020)
Gender	0.042 (0.100)	0.233* (0.122)	0.139 (0.132)	0.298** (0.127)
Native speaker	−0.183 (0.120)	−0.503*** (0.147)	−0.296* (0.159)	−0.329** (0.153)
Detect AI	−0.389** (0.175)	0.091 (0.214)	0.180 (0.232)	0.067 (0.223)
Jug favorable human	0.317 (0.333)	−0.166 (0.408)	−0.167 (0.442)	−0.730* (0.425)
Jug favorable ai	1.002** (0.436)	0.187 (0.534)	−0.271 (0.579)	−0.197 (0.556)
AI knowledge	−0.068 (0.062)	−0.063 (0.076)	−0.087 (0.082)	−0.031 (0.079)
AI usage daily	−0.083 (0.065)	−0.162** (0.079)	−0.219** (0.086)	−0.168** (0.083)
AI usage work	0.243*** (0.079)	0.146 (0.097)	0.081 (0.106)	0.052 (0.101)
Amelioration	−0.052 (0.069)	0.053 (0.084)	0.181** (0.091)	0.0001 (0.088)
Pertinence	0.139** (0.067)	0.113 (0.082)	0.198** (0.088)	0.325*** (0.085)
AI Creativity	−0.144 (0.095)	0.105 (0.117)	0.145 (0.127)	0.111 (0.122)
Willingness	−0.146 (0.101)	−0.161 (0.123)	−0.214 (0.134)	−0.126 (0.128)
Constant	3.722*** (0.434)	4.163*** (0.531)	3.565*** (0.577)	2.509*** (0.553)
Observations	210	210	210	210
R ²	0.125	0.145	0.146	0.171
Adjusted R ²	0.067	0.088	0.089	0.116
Residual Std. Error (df = 196)	0.629	0.769	0.834	0.801
F Statistic (df = 13; 196)	2.162**	2.559***	2.568***	3.119***

Notes: This table presents the coefficients reflecting the impact of evaluators' socio-demographic characteristics on the various creativity criteria scores. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using a OLS.

Table 5.25: Polynomial Logit – Evaluators Profiles for the Text Task

	<i>Dependent variable:</i>			
	Validity (1)	Form (2)	Elaboration (3)	Originality (4)
# Words	0.002*** (0.001)	0.005*** (0.001)	0.010*** (0.001)	0.005*** (0.001)
Age	0.039* (0.022)	−0.049** (0.022)	−0.063*** (0.021)	0.011 (0.020)
Gender	−0.197 (0.145)	0.344** (0.136)	0.191 (0.139)	0.469*** (0.135)
Native speaker	−0.220 (0.166)	−0.811*** (0.163)	−0.505*** (0.167)	−0.441*** (0.162)
Detect AI	−0.816*** (0.239)	0.320 (0.254)	0.535** (0.258)	0.281 (0.244)
Jug favorable human	0.526 (0.465)	−0.574 (0.464)	−0.612 (0.445)	−1.436*** (0.430)
Jug favorable ai	3.552*** (1.092)	0.283 (0.604)	−0.918 (0.559)	−0.602 (0.533)
AI knowledge	−0.181** (0.088)	−0.128 (0.083)	−0.200** (0.085)	−0.102 (0.083)
AI usage daily	−0.150 (0.092)	−0.326*** (0.089)	−0.471*** (0.090)	−0.232*** (0.086)
AI usage work	0.553*** (0.113)	0.356*** (0.109)	0.279** (0.109)	0.143 (0.105)
Amelioration	−0.157 (0.098)	0.146 (0.093)	0.400*** (0.094)	0.076 (0.091)
Pertinence	0.333*** (0.096)	0.232** (0.093)	0.454*** (0.093)	0.496*** (0.090)
AI Creativity	−0.371*** (0.138)	0.125 (0.129)	0.206 (0.132)	0.096 (0.128)
Willingness	−0.238 (0.145)	−0.481*** (0.142)	−0.650*** (0.144)	−0.346** (0.139)
AIC	2221	2659.55	2511.37	2916.83
Observations	980	980	980	980

Notes: This table presents the coefficients reflecting the impact of evaluators' socio-demographic characteristics on the various creativity criteria scores. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using an Ordered Polynomial Logit.

Alternative Uses task

Table 5.26: OLS – Average Score by Evaluators for the Alternative Uses Task

	<i>Dependent variable:</i>		
	Validity Max	Feasibility Max	Originality Max
Age	−0.061** (0.024)	−0.036 (0.022)	0.002 (0.016)
Gender	−0.169 (0.133)	−0.232* (0.120)	−0.012 (0.090)
Native speaker	−0.108 (0.218)	−0.494** (0.198)	−0.125 (0.148)
Detect AI	0.469 (0.374)	0.688** (0.339)	0.350 (0.253)
AI knowledge	0.109 (0.093)	0.063 (0.084)	−0.010 (0.063)
AI usage daily	−0.102 (0.106)	−0.304*** (0.096)	−0.090 (0.072)
AI usage work	−0.027 (0.104)	0.086 (0.094)	−0.062 (0.070)
Amelioration	−0.115 (0.085)	0.016 (0.077)	−0.013 (0.057)
Pertinence	0.271*** (0.083)	0.003 (0.076)	0.087 (0.056)
AI Creativity	−0.163 (0.140)	−0.004 (0.127)	−0.171* (0.095)
Willingness	0.132 (0.151)	0.265* (0.137)	−0.045 (0.102)
Constant	5.330*** (0.670)	5.564*** (0.608)	4.766*** (0.453)
Observations	189	189	189
R ²	0.127	0.124	0.101
Adjusted R ²	0.072	0.069	0.045
Residual Std. Error (df = 177)	0.773	0.701	0.522
F Statistic (df = 11; 177)	2.332**	2.267**	1.806*

Notes: This table presents the coefficients reflecting the impact of evaluators' socio-demographic characteristics on the various creativity criteria scores. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using a OLS.

Table 5.27: Polynomial Logit – Evaluators Profiles for the Alternative Uses Task

	<i>Dependent variable:</i>		
	Validity Max	Feasibility Max	Originality Max
	(1)	(2)	(3)
Mean # Words	−0.016 (0.036)	−0.006 (0.041)	0.022 (0.035)
Age	−0.123*** (0.037)	0.026 (0.047)	−0.065* (0.035)
Gender	−0.306 (0.199)	0.070 (0.223)	−0.397** (0.196)
Native speaker	−0.029 (0.330)	0.092 (0.381)	−0.918*** (0.343)
Detect AI	2.054* (1.081)	1.465 (1.104)	0.785 (0.524)
AI knowledge	0.144 (0.137)	−0.018 (0.168)	0.089 (0.136)
AI usage daily	−0.374** (0.170)	−0.228 (0.187)	−0.572*** (0.161)
AI usage work	0.080 (0.171)	−0.222 (0.185)	0.292* (0.156)
Amelioration	−0.334** (0.130)	−0.106 (0.152)	−0.054 (0.125)
Pertinence	0.738*** (0.131)	0.400*** (0.153)	0.095 (0.122)
AI Creativity	−0.434* (0.222)	−0.658** (0.273)	−0.168 (0.211)
Willingness	0.261 (0.225)	−0.092 (0.253)	0.529** (0.222)
AIC	1203.97	966.24	1298.24
Observations	567	567	567

Notes: This table presents the coefficients reflecting the impact of evaluators' socio-demographic characteristics on the various creativity criteria scores. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using an Ordered Polynomial Logit.

Draw task

Table 5.28: OLS – Average Score by Evaluators for the Draw Task

	<i>Dependent variable:</i>			
	Validity (1)	Form (2)	Elaboration (3)	Originality (4)
Age	−0.092*** (0.026)	−0.044 (0.028)	−0.062** (0.029)	−0.043 (0.027)
Gender	0.115 (0.211)	0.212 (0.228)	0.039 (0.230)	0.064 (0.220)
Native speaker	0.701*** (0.261)	0.863*** (0.281)	0.675** (0.284)	0.720*** (0.272)
Detect AI	−0.281 (0.292)	−0.183 (0.315)	−0.573* (0.318)	−0.163 (0.304)
Jug favorable human	−0.828* (0.490)	−1.183** (0.529)	−0.403 (0.533)	−0.909* (0.511)
AI knowledge	−0.063 (0.092)	−0.009 (0.100)	−0.070 (0.101)	−0.0004 (0.096)
AI usage daily	0.106 (0.084)	0.156* (0.091)	0.146 (0.091)	0.136 (0.088)
AI usage work	−0.128 (0.094)	−0.061 (0.101)	−0.086 (0.102)	−0.179* (0.098)
Amelioration	0.148 (0.101)	0.055 (0.110)	0.140 (0.110)	0.088 (0.106)
Pertinence	−0.059 (0.081)	−0.020 (0.088)	−0.001 (0.089)	−0.052 (0.085)
AI Creativity	−0.384** (0.174)	−0.179 (0.188)	−0.160 (0.189)	0.006 (0.181)
Willingness	−0.002 (0.195)	0.118 (0.210)	0.008 (0.212)	−0.187 (0.203)
Constant	4.947*** (0.738)	2.868*** (0.797)	3.310*** (0.803)	3.311*** (0.769)
Observations	195	195	195	195
R ²	0.144	0.141	0.117	0.118
Adjusted R ²	0.087	0.085	0.059	0.060
Residual Std. Error (df = 182)	0.917	0.991	0.999	0.956
F Statistic (df = 12; 182)	2.547***	2.494***	2.008**	2.023**

Notes: This table presents the coefficients reflecting the impact of evaluators' socio-demographic characteristics on the various creativity criteria scores. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using a OLS.

Table 5.29: Polynomial Logit – Evaluators Profiles for the Draw Task

	<i>Dependent variable:</i>			
	Validity (1)	Form (2)	Elaboration (3)	Originality (4)
% Pixel Cov.	1.550*** (0.512)	5.028*** (0.537)	4.088*** (0.522)	3.273*** (0.505)
Age	−0.129*** (0.024)	−0.072*** (0.023)	−0.083*** (0.023)	−0.064*** (0.023)
Gender	0.146 (0.178)	0.226 (0.179)	−0.016 (0.180)	0.021 (0.179)
Native speaker	0.814*** (0.231)	1.234*** (0.236)	0.897*** (0.231)	0.927*** (0.235)
Detect AI	−0.398 (0.255)	−0.356 (0.256)	−0.782*** (0.259)	−0.215 (0.257)
Jug favorable human	−0.917** (0.418)	−1.484*** (0.430)	−0.507 (0.433)	−1.091*** (0.420)
AI knowledge	−0.150* (0.084)	−0.070 (0.081)	−0.130 (0.082)	−0.045 (0.081)
AI usage daily	0.116 (0.075)	0.180** (0.073)	0.158** (0.072)	0.144** (0.073)
AI usage work	−0.105 (0.084)	−0.045 (0.081)	−0.077 (0.081)	−0.186** (0.081)
Amelioration	0.174* (0.091)	0.129 (0.090)	0.225** (0.088)	0.149* (0.089)
Pertinence	−0.064 (0.074)	−0.037 (0.073)	−0.007 (0.073)	−0.084 (0.072)
AI Creativity	−0.568*** (0.152)	−0.199 (0.150)	−0.155 (0.149)	0.027 (0.150)
Willingness	0.040 (0.169)	0.109 (0.170)	−0.033 (0.169)	−0.286* (0.169)
AIC	3152.81	3219.35	3317.55	3330.93
Observations	975	975	975	975

Notes: This table presents the coefficients reflecting the impact of evaluators' socio-demographic characteristics on the various creativity criteria scores. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Effects are estimated using an Ordered Polynomial Logit.

Comparison Between Agents per Criterion and Scores

Figure 5.7: Comparison between Agents of Objective Measures

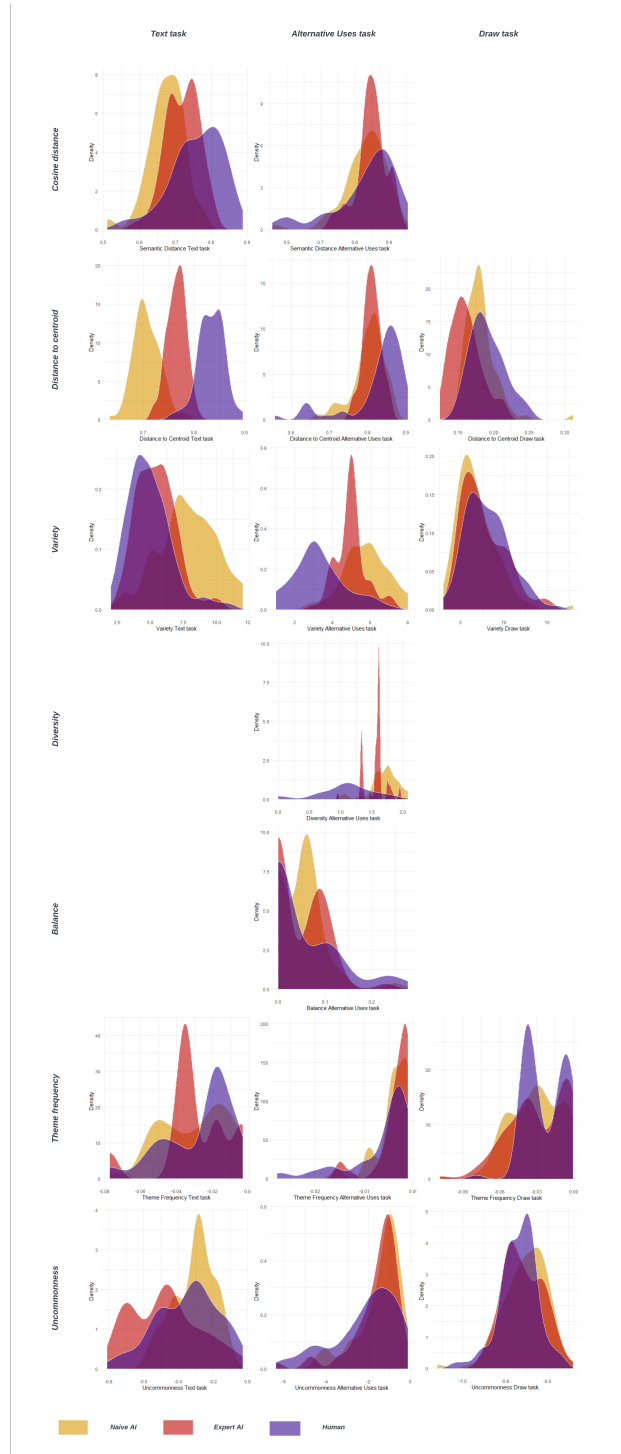
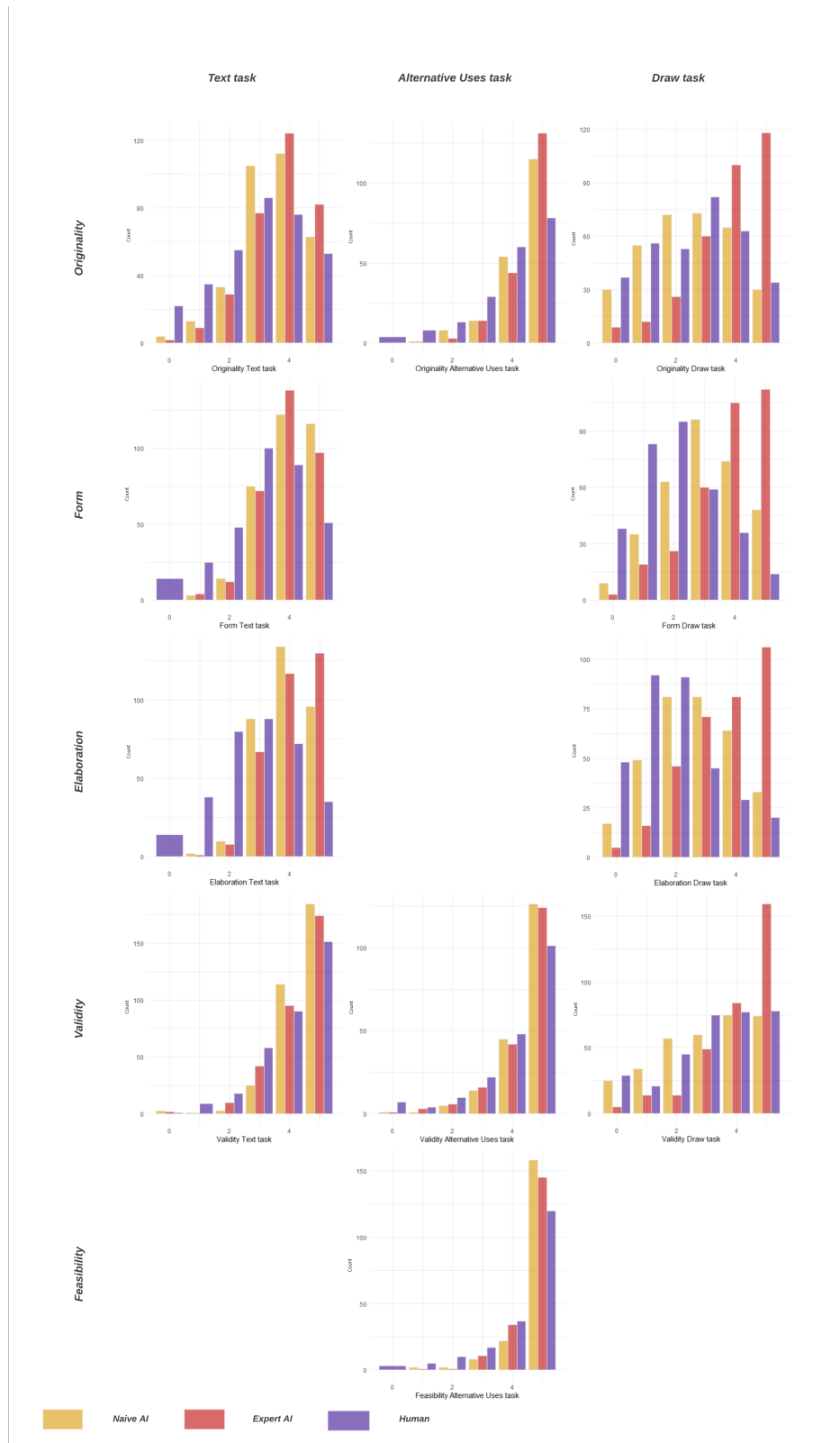


Figure 5.8: Comparison between Agents of Subjective Measures



General Conclusion

This thesis offers several insights into the determinants of individual, team, and artificial creativity. Throughout these five chapters, we have explored the multidimensional nature of creativity through the lens of creative performance, focusing on four main elements: the identity of the creative agent performing a creative task, the nature of the task evaluated, the multiple creativity criteria chosen to assess performance, and the intrinsic and environmental factors affecting it. This concluding chapter summarizes the key findings of each of the five chapters, presents the limitations of these works, and the stemming extensions to new research avenues they triggered.

First, Chapter 1, “*Rethinking laboratory experiments: The Case of Collective Experiments*”, emphasized the importance of rethinking laboratory experiments in studying collective processes. Despite its advancements over time, experimental economics did not invest fully in studying collective processes. The first contribution of this chapter is to settle the definition of a collective experiment, i.e., as “*any experiment enabling the observation of collective processes through the social interactions of subjects without a systematic intermediary*”. The chapter argued that the absence of *collective experiments* is rooted in the discipline’s practices’ historical, methodological, and technological context and advocated for advancing experimental methodologies to better capture collective processes through direct social interactions rather than relying on intermediaries. It reconsidered the role of intermediaries in experimental economics, as their presence can introduce bias in individuals’ behaviours and decisions during collective experiments due to an *intermediary interpretative layer*. Removing intermediaries challenges traditional experimental rules, such as anonymity and subject isolation, and calls for a re-evaluation of the tools used to study behaviour in controlled environments. By reviewing experimental economics’ historical and methodological evolution, this chapter also calls for adopting innovative tools and experimental designs, such as those exemplified by the Social

Interactions Lab (SIL), to enhance the study of collective processes and interactions.

Next, the second chapter entitled “*The Many Faces of Creative Profiles: Exploring Task Openness*”, explores the drivers of individual creative performance, focusing on the type of task, varying degrees of openness, and the multidimensional characteristics of ideators encompassing under-explored deep-level variables such as individuals’ skills or expertise. To do so, a within-subjects experimental protocol has been implemented to cover three levels of task openness. This study highlights the importance of examining individual creative performance in relation to the type of task and the criteria used for evaluating creativity, as these factors can significantly influence outcomes. Our findings underscore the impact of the surrounding environment, including social interactions or deep-level variables such as risk attitude and expertise, on creative performance. We also emphasize that while the traditional classification of tasks into closed and open categories is widely used, a more nuanced approach is needed. Specifically, the “open-with-constraints” tasks, as described by [Attanasi et al., 2021], demonstrate distinct characteristics that differ from the other task types and warrant further investigation. The primary limitation of this study lies in the nature of the tasks assigned to the subjects. Since our sample consisted of student subjects, we were unable to ask them to solve creative problems directly related to an organizational context. Future research could integrate tasks closer to real-life settings while continuing to investigate the profiles of creative individuals based on the thought processes involved, differentiating between open, open-with-constraints and closed tasks.

Moving from individual to team creativity, the third Chapter, “*Diversity, Interactions, and Team Creativity: An Experimental Perspective*”, built on the previous individual experimental protocol to gather information on subjects’ individual creative capacities to understand team creativity better. In addition to that, teams’ performance was analysed based on their diversity and both pre-existing and in-situ networks. This last combination provided insights into the intricacies of the effect of interactions within a creative team through a multilevel approach. The final results of our study conclude that the impact of diversity and networks on team creativity is contingent on the specific creativity criteria considered. It highlights the trade-off between feasibility and originality, emphasizing the need for teams to balance these aspects depending on their creative goals. Diversity, particularly in socio-demographic factors, enhances feasibility but not originality, while creativity training shows mixed results, negatively impacting feasibility. Additionally, social

networks influence originality, with pre-existing friendships boosting creative output. However, frequent interactions may limit creativity by reducing exposure to diverse environments. These conclusions contribute to the literature on team creativity by underlining the need to integrate more and more social network techniques into the experimental literature on creativity, introducing the first *collective experiment* as defined in Chapter 1. Moreover, it gives insights to organisations on the composition of creative teams based on the main creative objective, whether more original or feasible ideas are targeted. However, despite its contributions, this work also presents some limitations, and the final sample of 35 teams is one of them. Indeed, as a three-stage protocol, our experimental strategy required higher investment from subjects, which translated into higher attrition, limiting our final sample. If this limited number of teams does not prevent us from obtaining robust results, further works trying to have multi-stage protocols should aim for a larger number of teams. In addition, this work considers ad-hoc teams of students and no long-term or pre-existing teams, which limits us to consider the role played by team experience. This could also be completed with a medium or long-term perspective of an experimental protocol on team creativity by observing teams over several tasks and experimental sessions. All these limits have the benefit of opening new doors for future work on team creativity, including the multi-level approach by controlling, experimentally, for subjects' own creative performance. In the end, two main extensions of this work are envisioned — first, a similar exploration of the joint influence of diversity and networks on team creativity for pre-existing teams. Second is the investigation of the relationship between individual and team creativity by testing team creativity on tasks with varying degrees of constraints.

While Chapters 2 and 3 focused on intrinsic factors driving individual and team performance, the fourth chapter, entitled “*Does Creativity Thrive on Plot Twists? Exploring the Role of Surprise on Team Creativity*” concentrated on one specific environmental factor affecting team creativity: surprise. More specifically, it explores the impact of unexpectedness on a team performance. The experimental protocol is articulated around two tasks, a schema-setting task, and a surprise task that allowed us to observe teams' reaction to surprise and how they adapted to the sudden change in constraints specification. This study advances the literature at the crossroads of schema theory, surprise, and creativity. First, surprise only negatively affects the originality of teams' ideas, while it does not affect their feasibility once we control for team characteristics and experience. Second, team members' attitudes toward

surprise only affect a team's creative performance in terms of feasibility, not originality. Third, team members' positive experience of surprise affects both team creative performance for feasibility and originality. As a result, organisations need to pay particular attention to the dynamics of creative teams, especially when faced with unexpected events. The impact of a surprise can differ depending on the objective, whether it is to encourage novelty or strengthen implementability. Managers, therefore, need to assess these impacts carefully, particularly concerning the risks to the originality of ideas. Furthermore, as the reaction to a surprise is subjective and unique to each individual, it is crucial to understand how each member reacts and how these reactions interact. Team leaders should aim to create balanced teams, where members with a more positive attitude can mitigate the negative effects of surprises on others. Considering those elements will help organisations manage and overcome surprises when a sudden change occurs during the creative process. Nonetheless, this work presents some limitations. Firstly, it relies on declarative measures of surprise, where subjects report their surprise after the event and only at the end of the task. This method does not fully capture immediate surprise reactions. Of note, further improvements will follow for this work as we have access to the audio data from each team, which will provide rich and important insights into the creative process itself, including three main dimensions: the reaction to surprise, the creative process timeline and when the disturbance occurred, and team dynamics. Additionally, no significant difference was found between the low and high surprise conditions. It is possible that the surprise in our experiment has not been strong enough. Again, as was the case in Chapter 3, this study does not consider pre-existing routines might contribute to overcoming surprise differently. Finally, the study focused on a specific type of surprise, a change in constraints specification, which limits the generalization of our results to other types of surprise, e.g. a sudden turnover in team composition. Ultimately, all these limits provide new paths to investigate, from the team type to the degree and type of surprise. In addition, the limit on the measurement of surprise builds on the argument from Chapter 1 on the need to use more and more advanced technologies in an experimental setting where wearable devices to measure stress or other physiological data would provide insights into the reaction of each team member to surprise.

Lastly, the fifth chapter, *"Can AI Enhance its Creativity to Beat Humans ?"* extends our investigation into the determinants of creativity by not simply focusing on a human ideator but also incorporating artificial intelligence as a creative agent.

While Chapters 2, 3 and 4 focused on human creativity, this final chapter explored the potential of artificial creativity. Whereas creativity was a question of interaction between human agents, this chapter looked at the possible interaction between humans and machines. As a reminder, AI lacks agency, and the need for human intervention in its creative process can not be eliminated. In the relationship between humans and machines in creativity, this chapter not only compares their performance but also suggests ways of thinking about the importance of their working together to maximise joint performance. To do so, creative outputs collected during two experimental protocols, that of Chapter 2 and the one from Guichardaz et al. [2024], have been subjected to an evaluation alongside creative outputs generated by a Generative AI. External judges were asked to evaluate, according to different criteria, three types of task differing in to their degree of openness. The contributions in this chapter are based on two central points: the question of comparing the strengths and weaknesses of human and artificial creativity with regard to their multidimensionality, and the role played by the formulation of a problem through different prompting strategies in maximising the creative performance of an LLM model of the Chat-GPT4 (or DALL-E) type. The findings demonstrate that artificial creativity significantly outperforms human creativity. While humans compete with AI mainly on the Draw task, and only on specific criteria, our results show a clear dominance of AI over humans. It seems that AI models integrate what creativity is and succeed in performing over several criteria. The results are more vague on the dominance of the *Expert* strategy over *Naive*. However, we highlight a difference in the variability of performance, where the good performance of humans is more uncertain than the one of AI. Indeed, a higher proportion of humans, compared to AI, endorses lower creativity scores so the range of performance is more important than that of AI agents. In fact, these results also expose limitations in Generative AI models, which depend heavily on the chosen prompting strategy. While these results might initially suggest that AI could replace human creativity, they also highlight human-machine collaboration’s critical role in enhancing creative performance. As noted in the broader context of AI technologies, “*to fully exploit the potential of AI, human and machine intelligence must be tightly interwoven*” [Plastino and Purdy, 2018]. Therefore, the effectiveness of AI in generating creative outputs is greatly improved through human involvement, particularly in refining AI prompts and framing problems—a key aspect of the creative process. This human intervention becomes increasingly essential as task constraints become more demanding. The possible extensions emerging from this

work are two-fold. On the one hand, to continue the study of this human-machine collaboration and to define the conditions of this collaboration, more particularly in line with the results obtained on the fine-tuning role of prompts. On the other hand, investigating other prompting strategies to consider the most effective technique to enhance AI creativity.

Beyond the fact that this dissertation emphasises the necessity of studying creativity across various levels of observation — focusing on intrinsic and environmental factors — and advocating for a comprehensive multi-criteria evaluation to capture its scope fully, this dissertation also facilitates broader methodological conclusions and recommendations on economists' research agenda. First, this dissertation demonstrates the possibility of conducting new experiments in teamwork to improve our understanding of the economic process. Again, while the experimental method is increasingly popular today, its still young history has shown that inquiries into collective processes have often been marginalized. This work opens the discussion on the possibility of correcting this discrepancy and advancing the practice of experiments in economics from a methodological point of view. In addition, the field of economics must conduct a more comprehensive examination of creativity to enrich the existing literature, which is still too sparse from the perspective of levels of observation, creativity criteria, and the ideator's identity. As a reminder, a primary rationale for studying creativity lies in its strong connection to innovation. Innovation can only emerge when it is nourished by new, useful ideas that meet specific needs. Without creativity, the emergence of innovation is fundamentally unattainable. Therefore, understanding the determinants that facilitate the emergence of these ideas should be recognized as a critical focus for economists, enabling them to unravel the intricate relationship between creativity and innovation, particularly the crucial transition from one to the other. To do so, future research should continue to focus on the determinants of creative performance, possibly anchoring their laboratory investigations in an organisational context where the core relationship between creativity and innovation can be experimentally observed.

Conclusion Générale

Cette thèse propose plusieurs éclairages sur les déterminants de la créativité individuelle, d'équipe et artificielle. Tout au long de ces cinq chapitres, nous avons exploré la nature multidimensionnelle de la créativité à travers le prisme de la performance créative, en nous concentrant sur quatre éléments principaux : l'identité de l'agent créatif réalisant une tâche créative, la nature de la tâche évaluée, les multiples critères de créativité choisis pour évaluer la performance, ainsi que les facteurs intrinsèques et environnementaux qui l'affectent. Ce chapitre conclusif résume les principaux résultats de chacun de ces cinq chapitres, leurs limites de ces travaux et les possibles prolongements vers de nouvelles perspectives de recherche qu'ils ont suscités.

Tout d'abord, le Chapitre 1, intitulé *“Rethinking laboratory experiments : The Case of Collective Experiments”*, met l'accent sur l'importance de repenser les expériences en laboratoire dans l'étude des processus collectifs. Malgré ses avancées au fil du temps, l'économie expérimentale n'a pas pleinement investi l'étude des processus collectifs. La première contribution de ce chapitre est de définir l'expérience collective comme *“toute expérience permettant d'observer des processus collectifs à travers les interactions sociales des sujets sans intermédiaire systématique”*. Le chapitre soutient que l'absence d'*expériences collectives* s'enracine dans le contexte historique, méthodologique et technologique des pratiques de la discipline et plaide pour faire progresser les méthodologies expérimentales afin de mieux saisir les processus collectifs. Cela par des interactions sociales directes entre sujets plutôt que par le biais d'intermédiaires. Ce chapitre réexamine le rôle des intermédiaires dans l'économie expérimentale, car leur présence peut introduire un biais dans les comportements et les décisions des individus lors des expériences collectives en raison d'une *couche interprétative intermédiaire*. Éliminer ces intermédiaires remet en question les règles expérimentales traditionnelles, telles que l'anonymat et l'isolement des sujets, et appelle à une réévaluation des outils utilisés pour étudier les comportements dans des environnements contrôlés. En passant en revue l'évolution historique et méthodolo-

gique de l'économie expérimentale, ce chapitre appelle également à adopter des outils et des conceptions expérimentales innovants, comme ceux utilisés au sein du *Social Interactions Lab* (SIL), pour améliorer l'étude des processus et des interactions collectives.

Ensuite, le deuxième chapitre intitulé “*The Many Faces of Creative Profiles : Exploring Task Openness*”, explore les moteurs de la performance créative individuelle, en se concentrant sur le type de tâche, les différents degrés d'ouverture et les caractéristiques multidimensionnelles des idéateurs, en englobant des variables de "profondeur" encore sous-explorées telles que les compétences ou l'expertise des individus. Pour ce faire, un protocole expérimental intra-sujet a été mis en œuvre pour couvrir trois degrés d'ouverture de tâche. Cette étude souligne l'importance d'examiner la performance créative individuelle en fonction du type de tâche et des critères utilisés pour évaluer la créativité, car ces facteurs peuvent avoir un impact significatif sur les résultats. Nos analyses mettent en évidence l'impact des interactions sociales, mais aussi des variables de "profondeur" telles que l'attitude face au risque et l'expertise, sur la performance créative. Nous insistons également sur le fait que, bien que la classification traditionnelle des tâches en catégories fermées et ouvertes soit largement utilisée, une approche plus nuancée est nécessaire. En particulier, les tâches “ouvertes avec contraintes”, telles que décrites par Attanasi et al. [2021], présentent des caractéristiques distinctes qui diffèrent des autres types de tâches et méritent une investigation plus approfondie. La principale limite de cette étude réside dans la nature des tâches assignées aux participants. Étant donné que notre échantillon était composé de sujets étudiants, nous n'avons pas pu leur demander de résoudre des problèmes créatifs directement liés à un contexte organisationnel. Des recherches futures pourraient intégrer des tâches plus proches des situations réelles tout en continuant à explorer les profils des individus créatifs en fonction des processus de pensée impliqués, en différenciant les tâches ouvertes, ouvertes avec contraintes et fermées.

En passant de la créativité individuelle à la créativité d'équipe, le troisième chapitre, intitulé “*Diversity, Interactions, and Team Creativity : An Experimental Perspective*”, s'appuie sur le protocole expérimental individuel précédent pour recueillir des informations sur les capacités créatives individuelles des sujets afin de mieux comprendre la créativité en équipe. En plus de cela, la performance des équipes a été observée au travers d'une tâche collective, en contrôlant pour leur diversité ainsi que de leurs réseaux préexistants et en laboratoire. Cette dernière combinaison a permis de mieux comprendre les subtilités de l'effet des interactions au sein d'une

équipe créative à travers une approche multi-niveaux. Les résultats finaux de notre étude concluent que l'impact de la diversité et des réseaux sur la créativité en équipe dépend des critères de créativité spécifiques considérés. Il souligne le compromis à faire entre faisabilité et originalité, mettant en avant la nécessité pour les équipes de trouver un équilibre en fonction de leurs objectifs créatifs. La diversité, notamment en termes de facteurs socio-démographiques, améliore la faisabilité mais pas l'originalité, tandis que la formation à la créativité montre des résultats mitigés, ayant un impact négatif sur la faisabilité. De plus, les réseaux sociaux influencent l'originalité, les amitiés préexistantes augmentant la production créative. Cependant, des interactions fréquentes peuvent limiter la créativité en réduisant l'exposition à des environnements diversifiés. Ces conclusions contribuent à la littérature sur la créativité en équipe en soulignant la nécessité d'intégrer de plus en plus de techniques de réseaux sociaux dans la littérature expérimentale sur la créativité et en introduisant la première *expérience collective* telle que définie au Chapitre 1. De plus, cela donne des indications aux organisations sur la composition des équipes créatives en fonction de leur objectif créatif principal ; qu'il s'agisse de viser des idées plus originales ou plus réalisables. Cependant, malgré ses contributions, ce travail présente également certaines limites, et l'échantillon final de 35 équipes en est une. En effet, étant un protocole en trois étapes, notre stratégie expérimentale a nécessité un investissement plus important de la part des sujets, ce qui s'est traduit par un taux d'attrition plus élevé, limitant ainsi notre échantillon final. Bien que ce nombre limité d'équipes ne nous empêche pas d'obtenir des résultats robustes, les travaux futurs, visant à mettre en œuvre des protocoles en plusieurs étapes, devraient viser un nombre plus élevé d'équipes. De plus, ce travail prend en compte des équipes ad hoc composées d'étudiants mais aucune équipe de long terme ou préexistante, ce qui nous limite dans la prise en compte du rôle joué par l'expérience de l'équipe. Cela pourrait également être complété par une perspective à moyen ou long terme d'un protocole expérimental sur la créativité en équipe, en observant les équipes au cours de plusieurs tâches et sessions expérimentales. Toutes ces limites présentent l'avantage d'ouvrir de nouvelles voies pour les recherches futures sur la créativité en équipe, y compris une approche multi-niveaux en contrôlant, de manière expérimentale, la performance créative des sujets eux-mêmes. En fin de compte, deux extensions principales de ce travail sont envisagées : tout d'abord, une exploration similaire de l'influence conjointe de la diversité et des réseaux sur la créativité des équipes préexistantes. Ensuite, l'étude de la relation entre la créativité individuelle et la créativité d'équipe en testant cette

dernière sur des tâches avec des degrés de contraintes variables.

Alors que les Chapitres 2 et 3 se concentraient sur les facteurs intrinsèques qui déterminent la performance individuelle et celle des équipes, le quatrième chapitre, intitulé “*Does Creativity Thrive on Plot Twists ? Exploring the Role of Surprise on Team Creativity*” s’est focalisé sur un facteur environnemental spécifique affectant la créativité en équipe : la surprise. Plus précisément, il explore l’impact de l’inattendu sur la performance d’une équipe. Le protocole expérimental s’articule autour de deux tâches, une tâche visant à instaurer un schéma spécifique et une tâche de surprise, qui nous ont permis d’observer la réaction des équipes face à la surprise et leur adaptation à un changement soudain dans la spécification des contraintes. Cette étude fait avancer la littérature à l’intersection de la théorie des schémas, de la surprise et de la créativité. Premièrement, la surprise n’affecte négativement que l’originalité des idées, sans pour autant affecter leur faisabilité une fois les caractéristiques et l’expérience de l’équipe contrôlées. Deuxièmement, l’attitude des membres de l’équipe face à la surprise n’affecte la performance créative de l’équipe qu’en termes de faisabilité, pas d’originalité. Troisièmement, l’expérience positive des membres de l’équipe face à la surprise influence à la fois la faisabilité et l’originalité de la performance créative de l’équipe. Par conséquent, les organisations doivent prêter une attention particulière à la dynamique des équipes créatives, en particulier lorsqu’elles sont confrontées à des événements inattendus. L’impact d’une surprise peut différer en fonction de l’objectif, qu’il s’agisse de favoriser la nouveauté ou de renforcer la faisabilité. Les managers doivent donc évaluer ces impacts avec soin, notamment en ce qui concerne les risques pesant sur l’originalité des idées. De plus, étant donné que la réaction à une surprise est subjective et unique à chaque individu, il est crucial de comprendre comment chaque membre réagit et comment ces réactions interagissent. Les leaders d’équipe devraient chercher à créer des équipes équilibrées, où les membres ayant une attitude plus positive peuvent atténuer les effets négatifs des surprises sur les autres. Prendre en compte ces éléments aidera les organisations à gérer et à surmonter les surprises lorsqu’un changement soudain survient au cours du processus créatif. Néanmoins, ce travail présente certaines limites. Premièrement, il repose sur des mesures déclaratives de la surprise, où les sujets rapportent leur surprise après l’événement et uniquement à la fin de la tâche. Cette méthode ne capture pas pleinement les réactions immédiates à la surprise. Il est à noter que des améliorations ultérieures suivront pour ce travail, car nous avons accès aux données audio de chaque équipe, qui fourniront des informations riches et importantes sur le processus créatif lui-même, y compris trois

dimensions principales : la réaction à la surprise, la chronologie du processus créatif et le moment où la perturbation s’est produite, ainsi que la dynamique d’équipe. De plus, aucune différence significative n’a été trouvée entre les conditions de faible et de forte surprise. Il est possible que la surprise dans notre expérience n’ait pas été assez forte. Encore une fois, comme ce fut le cas dans le Chapitre 3, cette étude ne prend pas en compte les routines préexistantes qui pourraient contribuer à surmonter la surprise de manière différente. Enfin, l’étude s’est concentrée sur un type spécifique de surprise, à savoir un changement dans la spécification des contraintes, ce qui limite la généralisation de nos résultats à d’autres types de surprise, comme un changement soudain dans la composition de l’équipe. En fin de compte, toutes ces limites offrent de nouvelles voies à explorer, depuis le type d’équipe jusqu’au degré et au type de surprise. En outre, la limite liée à la mesure de la surprise s’appuie sur l’argument du Chapitre 1 concernant la nécessité d’utiliser des technologies de plus en plus avancées dans un cadre expérimental, où des dispositifs portables permettant de mesurer le stress ou d’autres données physiologiques fourniraient des informations sur la réaction de chaque membre de l’équipe à la surprise.

Enfin, le cinquième chapitre, intitulé *“Can AI Enhance its Creativity to Beat Humans ?”* élargit notre investigation sur les déterminants de la créativité en ne se concentrant pas uniquement sur un idéateur humain mais en intégrant également l’intelligence artificielle comme agent créatif. Alors que les chapitres 2, 3 et 4 se sont concentrés sur la créativité humaine, ce dernier chapitre explore le potentiel de la créativité artificielle. Il est important de rappeler que l’IA manque d’autonomie, et que l’intervention humaine dans son processus créatif ne peut être éliminée. Dans la relation entre humains et machines dans la créativité, ce chapitre ne se contente pas de comparer leurs performances, mais propose également des réflexions sur l’importance de leur collaboration, en termes de stratégies de promptage, pour maximiser la performance commune. Pour ce faire, les productions créatives recueillies lors de deux protocoles expérimentaux, celui du Chapitre 2 et celui de Guichardaz et al. [2024], ont été soumises à une évaluation aux côtés des productions créatives générées par une IA générative. Des juges externes ont été invités à évaluer, selon différents critères, trois types de tâches différant par leur degré d’ouverture. Les contributions de ce chapitre reposent sur deux points centraux : la question de la comparaison des forces et faiblesses de la créativité humaine et artificielle en regard de leur multidimensionnalité, et le rôle joué par la formulation d’un problème à travers différentes stratégies de promptage pour maximiser la performance créative d’un modèle de type

LLM, comme Chat-GPT4 (ou DALL-E). Les résultats montrent que la créativité artificielle dépasse de manière significative la créativité humaine. Alors que les humains rivalisent principalement avec l'IA sur la tâche de dessin, et uniquement sur certains critères, nos résultats révèlent une nette domination de l'IA sur les humains. Il semble que les modèles d'IA intègrent ce qu'est la créativité et réussissent à performer selon plusieurs critères. Les résultats sont plus flous quant à la supériorité de la stratégie *Expert* par rapport à *Naive*. Cependant, nous soulignons une différence dans la variabilité des performances, où la bonne performance des humains est plus incertaine que celle des agents IA. En effet, une proportion plus élevée d'humains, comparée à celle des IA, obtient des scores de créativité plus bas, ce qui rend l'éventail des performances humaines plus large que celui des agents IA. En réalité, ces résultats exposent également des limitations des modèles d'IA générative, qui dépendent fortement de la stratégie de promptage choisie. Bien que ces résultats puissent suggérer initialement que l'IA pourrait remplacer la créativité humaine, ils mettent également en lumière le rôle crucial de la collaboration homme-machine dans l'amélioration des performances créatives. Comme indiqué dans le contexte plus large des technologies de l'IA, “*to fully exploit the potential of AI, human and machine intelligence must be tightly interwoven*”[Plastino and Purdy, 2018]. Par conséquent, l'efficacité de l'IA dans la génération de productions créatives est considérablement améliorée grâce à l'implication humaine, en particulier dans l'affinement des instructions données à l'IA et dans la formulation des problèmes — un aspect clé du processus créatif. Cette intervention humaine devient de plus en plus essentielle à mesure que les contraintes de la tâche deviennent plus exigeantes. Les prolongements possibles de ce travail sont doubles. D'une part, poursuivre l'étude de cette collaboration homme-machine et définir les conditions de cette collaboration, plus particulièrement en lien avec les résultats obtenus sur le rôle de l'affinement des instructions. D'autre part, explorer d'autres stratégies de promptage pour identifier la technique la plus efficace pour améliorer la créativité de l'IA.

Au-delà du fait que cette thèse mette en avant la nécessité d'étudier la créativité à travers différents niveaux d'observation — en se concentrant sur les facteurs intrinsèques et environnementaux — et qu'elle prône une évaluation multicritère de la performance pour en saisir toute la portée, cette thèse permet également de tirer des conclusions méthodologiques plus larges et de formuler des recommandations pour l'agenda de recherche des économistes. Premièrement, cette thèse démontre la possibilité de mener de nouvelles expériences afin d'améliorer notre compréhension du

processus économique. Encore une fois, bien que la méthode expérimentale soit de plus en plus populaire aujourd'hui, son histoire encore jeune a montré que les investigations sur les processus collectifs ont souvent été marginalisées. Ce travail ouvre la discussion sur la possibilité de corriger cette lacune et de faire progresser la pratique des expériences en économie d'un point de vue méthodologique. De plus, le domaine de l'économie doit procéder à un examen plus approfondi de la créativité afin d'enrichir la littérature existante, encore trop limitée du point de vue des niveaux d'observation, des critères de créativité et de l'identité du créateur. Pour rappel, la principale justification de l'étude de la créativité réside dans son lien étroit avec l'innovation. L'innovation ne peut émerger que si elle est nourrie par des idées nouvelles et utiles qui répondent à des besoins spécifiques. Sans créativité, l'émergence de l'innovation est fondamentalement inatteignable. Par conséquent, comprendre les déterminants qui facilitent l'émergence de ces idées doit être reconnu comme un objectif crucial pour les économistes, leur permettant de démêler la relation complexe entre créativité et innovation, en particulier la transition cruciale de l'une à l'autre. Pour ce faire, les recherches futures devraient continuer à se concentrer sur les déterminants de la performance créative, en ancrant possiblement leurs investigations en laboratoire dans un contexte organisationnel où la relation fondamentale entre créativité et innovation pourrait être observée expérimentalement.

Bibliography

- O. A. Acar, M. Tarakci, and D. Van Knippenberg. Creativity and innovation under constraints: A cross-disciplinary integrative review. *Journal of management*, 45(1):96–121, 2019.
- S. Acar. Creativity assessment, research, and practice in the age of artificial intelligence. *Creativity Research Journal*, pages 1–7, 2023.
- A. Açıkgoz and A. Günsel. Individual creativity and team climate in software development projects: The mediating role of team decision processes. *Creativity and Innovation Management*, 25(4):445–463, 2016.
- K. Adams and G. Galanes. *Communicating in groups*. McGraw-Hill Publishing, 2011.
- P. Aghion and P. W. Howitt. *The economics of growth*. MIT press, 2008.
- J. W. Ahn, Y. Ku, and H. C. Kim. A novel wearable EEG and ECG recording system for stress assessment. *Sensors (Basel)*, 19(9), 2019.
- R. S. Albert and M. A. Runco. A history of research on creativity. 1999.
- M. Allais. Fondements d’une théorie positive des choix comportant un risque. *Colloques Internationaux du Centre National de la Recherche Scientifique (Econometrie)*, 40:127–140, 1953.
- G. M. Alliger, C. P. Cerasoli, S. I. Tannenbaum, and W. B. Vessey. Team resilience: How teams flourish under pressure. *Organizational Dynamics*, 44(3):176–184, 2015.
- C. M. Allwood and M. Selart. Social and creative decision making. In *Decision making: Social and creative dimensions*, pages 3–11. Springer, 2001.

- T. Amabile. The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, 45:357–376, 08 1983.
- T. Amabile. Guidepost: Creativity, artificial intelligence, and a world of surprises guidepost letter for academy of management discoveries. *Academy of Management Discoveries*, Feb. 2019. ISSN 2168-1007.
- T. M. Amabile. Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43:997–1013, 1982.
- T. M. Amabile. A model of creativity and innovation in organizations. *Research in Organizational Behavior*, 10:154, 1988.
- T. M. Amabile and N. D. Gyskiewicz. Assessing the environment for creativity: The work environment inventory. 1989.
- T. M. Amabile and M. G. Pratt. The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research in Organizational Behavior*, 36:157–183, 2016.
- T. M. Amabile, K. G. Hill, B. A. Hennessey, and E. Tighe. The work preference inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology*, 66:950–967, 1994.
- T. M. Amabile, R. Conti, H. M. Coon, J. Lazenby, and M. Herron. Assessing the work environment for creativity. *Academy of Management Journal*, 39:1154–1184, 1996.
- N. Anderson, C. K. De Dreu, and B. A. Nijstad. The routinization of innovation research: A constructively critical review of the state-of-the-science. *Journal of organizational Behavior*, 25(2):147–173, 2004.
- N. Anderson, K. Potočník, and J. Zhou. Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. *Journal of management*, 40(5):1297–1333, 2014.
- C. Andrade. Internal, external, and ecological validity in research design, conduct, and evaluation. *Indian Journal of Psychological Medicine*, 40(5):498–499, Sept. 2018.

- C. Andriopoulos. Determinants of organisational creativity: a literature review. *Management decision*, 39(10):834–841, 2001.
- D. Ariely and M. I. Norton. Psychology and experimental economics: A gap in abstraction. *Current Directions in Psychological Science*, 16(6):336–339, 2007.
- D. Ariely, U. Gneezy, G. Loewenstein, and N. Mazar. Large stakes and big mistakes. *The Review of Economic Studies*, 76(2):451–469, 2009.
- H. H. Asada, P. Shaltis, A. Reisner, S. Rhee, and R. C. Hutchinson. Mobile monitoring with wearable photoplethysmographic biosensors. *IEEE Eng. Med. Biol. Mag.*, 22(3):28–40, 2003.
- G. Attanasi, Y. Curci, P. Llerena, A. Pinate, M. Ramos-Sosa, and G. Urso. Looking at creativity from east to west: Risk taking and intrinsic motivation in socially and culturally diverse countries. Technical Report 2019-38, Bureau d’Economie Théorique et Appliquée, UDS, Strasbourg, 2019a.
- G. Attanasi, Y. Curci, P. Llerena, and G. Urso. Intrinsic vs. extrinsic motivators on creative collaboration: The effect of sharing rewards. Technical Report 2019-37, Bureau d’Economie Théorique et Appliquée, UDS, Strasbourg, 2019b.
- G. Attanasi, M. Chessa, S. Gil-Gallen, and P. Llerena. A survey on experimental elicitation of creativity in economics. *Revue d’économie industrielle*, 2021.
- J. Baer. The importance of domain-specific expertise in creativity. *Roeper Review*, 37(3):165–178, July 2015. ISSN 1940-865X. doi: 10.1080/02783193.2015.1047480. URL <http://dx.doi.org/10.1080/02783193.2015.1047480>.
- J. Baer and J. C. Kaufman. Gender differences in creativity. *The Journal of Creative Behavior*, 42(2):75–105, 2008.
- M. Baer and G. R. Oldham. The curvilinear relation between experienced creative time pressure and creativity: Moderating effects of openness to experience and support for creativity. *Journal of Applied Psychology*, 91(4):963–970, 2006. ISSN 0021-9010.
- M. D. Bahlmann. Geographic network diversity: How does it affect exploratory innovation? *Industry and Innovation*, 21(7-8):633–654, 2014.

- D. Baidoo-Anu and L. O. Ansah. Education in the era of generative artificial intelligence (ai): Understanding the potential benefits of chatgpt in promoting teaching and learning. *Journal of AI*, 7(1):52–62, 2023.
- L. Baloché. Creativity and cooperation in the elementary music classroom. *The Journal of Creative Behavior*, 28(4):255–265, 1994.
- Y. P. Bammens. Employees’ innovative behavior in social context: A closer examination of the role of organizational care. *Journal of Product Innovation Management*, 33(3):244–259, 2016.
- A. Bandura and S. Wessels. *Self-efficacy*. Cambridge University Press Cambridge, 1997.
- F. Barron and D. M. Harrington. Creativity, intelligence, and personality. *Annual Review of Psychology*, 32(1):439–476, 1981.
- J. Baruah and P. B. Paulus. Category assignment and relatedness in the group ideation process. *Journal of Experimental Social Psychology*, 47:1070–1077, 2011.
- O. Bein, C. Gasser, T. Amer, A. Maril, and L. Davachi. Predictions transform memories: How expected versus unexpected events are integrated or separated in memory. *Neuroscience & Biobehavioral Reviews*, page 105368, 2023.
- S. T. Bell, A. J. Villado, M. A. Lukasik, L. Belau, and A. L. Briggs. Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *Journal of Management*, 37(3):709–743, 2011.
- H. Bergmann, C. Hundt, and R. Sternberg. What makes student entrepreneurs? on the relevance (and irrelevance) of the university and the regional context for student start-ups. *Small Business Economics*, 47:53–76, 2016.
- B. J. Birdsell. Creative cognition: conceptual blending and expansion in a generative exemplar task. *IAFOR Journal of Psychology & the Behavioral Sciences*, 5(SI): 43–62, 2019.
- M. A. Boden. Creativity and artificial intelligence. *Artificial Intelligence*, 103(1): 347–356, 1998. ISSN 0004-3702. Artificial Intelligence 40 years later.
- M. A. Boden. *The creative mind: Myths and mechanisms*. Routledge, 2004.

- A. Boin and M. J. Van Eeten. The resilient organization. *Public Management Review*, 15(3):429–445, 2013.
- K. J. Boudreau and K. R. Lakhani. Field experimental evidence on sorting, incentives and creative worker performance. *Harvard Business School*, pages 11–107, 2011.
- S. G. Bouschery, V. Blazevic, and F. T. Piller. Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language models. *Journal of Product Innovation Management*, 40(2):139–153, 2023.
- C. A. Bowers, J. A. Pharmer, and E. Salas. When member homogeneity is needed in work teams: A meta-analysis. *Small group research*, 31(3):305–327, 2000.
- C. Bradler. How creative are you? – an experimental study on self-selection in a competitive incentive scheme for creative performance. *ERN: Experimental Economics (Topic)*, 2015.
- C. Bradler, S. Neckermann, and A. J. Warnke. Incentivizing creativity: A large-scale experiment with tournaments and gifts. *Behavioral & Experimental Economics eJournal*, 2016.
- D. J. Brass. Creative action in organizations: Ivory tower visions & real world voices, Aug 1995. Creative Action in Organizations: Ivory Tower Visions & Real World Voices.
- S. Brusoni. The limits to specialization: problem solving and coordination in ‘modular networks’. *Organization studies*, 26(12):1885–1907, 2005.
- S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, and Y. Zhang. Sparks of artificial general intelligence: Early experiments with gpt-4, 2023.
- R. Buda. Un bref historique de l’économie expérimentale. 02 2000.
- J. Burkell. Anonymity in behavioural research: Not being unnamed, but being unknown. *U. Ottawa L. & Tech. J.*, 3:189, 2006.
- J. F. Cabra and C. D. Guerrero. Regional creativity: Cultural and socio-economic differences. *Journal of Creativity*, 32(2):100022, 2022.

- B. S. Caldwell and N. C. Everhart. Information flow and development of coordination in distributed supervisory control teams. *International Journal of Human-Computer Interaction*, 10(1):51–70, 1998.
- A. Camarda, G. Borst, M. Agogu  , M. Habib, B. Weil, O. Houd  , and M. Cassotti. Do we need inhibitory control to be creative? evidence from a dual-task paradigm. *Psychology of Aesthetics, Creativity and the Arts*, 10 2017.
- C. Camerer and R. Hogarth. The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19:7–42, 01 1999.
- A. Carmeli and G. M. Spreitzer. Trust, connectivity, and thriving: Implications for innovative behaviors at work. *The Journal of Creative Behavior*, 43(3):169–191, 2009.
- S. H. Carson, J. B. Peterson, and D. M. Higgins. Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal*, 17:37 – 50, 2005.
- A. M. Carton and J. N. Cummings. The impact of subgroup type and subgroup configurational properties on work team performance. *Journal of Applied Psychology*, 98(5):732–758, Sept. 2013. ISSN 0021-9010.
- M. Cassotti, A. Camarda, N. Poirel, O. Houd  , and M. Agogu  . Fixation effect in creative ideas generation: Opposite impacts of example in children and adults. *Thinking Skills and Creativity*, 19:146–152, Mar. 2016. ISSN 1871-1871.
- E. H. Chamberlin. *"Note on Deviations from Equilibrium", The Theory of Monopolistic Competition*. Harvard University Press, 5th edition, 1946.
- E. H. Chamberlin. An experimental imperfect market. *Journal of political economy*, 56(2):95–108, 1948.
- W.-C. Chang and Z.-H. Chiang. A study on how to elevate organisational creativity in taiwanese design organisation. *International Journal of Innovation Management*, 12(04):699–723, 2008.
- G. Charness and D. Grieco. Individual Creativity, Ex-ante Goals and Financial Incentives. University of California at Santa Barbara, Economics Working Paper Series qt4mr6p1d5, Department of Economics, UC Santa Barbara, Feb. 2013.

- G. Charness and D. Grieco. Creativity and Incentives. *Journal of the European Economic Association*, 17(2):454–496, 2019.
- G. Charness and D. Grieco. Creativity and corporate culture. *Available at SSRN 3723865*, 2021.
- G. Charness and D. Grieco. Creativity and AI. *Available at SSRN 4686415*, 2024.
- G. Charness, A. Samek, and J. van de Ven. What is considered deception in experimental economics? *Experimental Economics*, 25(2):385–412, 2022.
- P. Chatzoglou and D. Chatzoudes. The role of innovation in building competitive advantages: an empirical investigation. *European Journal of Innovation Management*, 21(1):44–69, Jan. 2018.
- C. X. Chen, M. G. Williamson, and F. H. Zhou. Reward system design and group creativity: An experimental investigation. *The Accounting Review*, 87(6):1885–1911, 2012.
- Y. Chen, C. Wong, H. Yang, J. Aguenza, S. Bhujangari, B. Vu, X. Lei, A. Prasad, M. Fluss, E. Phuong, et al. Assessing the impact of prompting, persona, and chain of thought methods on chatgpt’s arithmetic capabilities. *arXiv preprint arXiv:2312.15006*, 2023.
- J. N. Choi. Individual and contextual predictors of creative performance: The mediating role of psychological processes. *Creativity Research Journal*, 16(2–3):187–199, Aug. 2004. ISSN 1532-6934.
- C. Clegg, K. Unsworth, O. Epitropaki, and G. Parker. Implicating trust in the innovation process. *Journal of occupational and organizational psychology*, 75(4):409–422, 2002.
- I. M. Cockburn, R. Henderson, and S. Stern. The impact of artificial intelligence on innovation: An exploratory analysis. In *The economics of artificial intelligence: An agenda*, pages 115–146. University of Chicago Press, 2018.
- G. H. Coldevin, A. Carlsen, S. Clegg, T. S. Pitsis, and E. P. Antonacopoulou. Organizational creativity as idea work: Intertextual placing and legitimating imaginings in media development and oil exploration. *Human Relations*, 72(8):1369–1397, Dec. 2018.

- P. Cook. The creativity advantage-is your organization the leader of the pack? *Industrial and commercial training*, 30(5):179–184, 1998.
- R. Cordero, N. Ditomaso, and G. F. Farris. Gender and race/ethnic composition of technical work groups: Relationship to creative productivity and morale. *Journal of Engineering and Technology Management*, 13(3):205–221, 1996. ISSN 0923-4748.
- C. Coricelli, U. Toepel, M.-L. Notter, M. M. Murray, and R. I. Rumiati. Distinct brain representations of processed and unprocessed foods. *European Journal of Neuroscience*, 50(8):3389–3401, 2019.
- L. Corry. Linearity and reflexivity in the growth of mathematical knowledge. *Science in Context*, 3(2):409–440, 1989.
- A. L. Cot and S. Ferey. La construction de « faits » économiques d’un nouveau type : éléments pour une histoire de l’économie expérimentale. *L’Actualité économique*, 92(1–2):11, 2016. ISSN 1710-3991.
- D. Cotta. Infrastructuring resilience through collective mind: Opportunity for smes. In *2021) Infrastructuring for organizational resilience: A Workshop Report*, volume 18, pages 15–19, 2021.
- R. Courtois, J.-M. Petot, O. Plaisant, B. Allibe, B. Lignier, C. Réveillère, G. Lecocq, and O. John. Validation française du big five inventory à 10 items (bfi-10). *L’Encéphale*, 46(6):455–462, Dec. 2020. ISSN 0013-7006.
- M. Crepaldi, G. Fusi, P. Iannello, M. L. Rusconi, et al. The bidirectional relationship between risk and creativity: A systematic review. *TPM: Testing, Psychometrics, Methodology in Applied Psychology*, 31(1), 2024.
- D. Cropley and A. Cropley. Functional creativity. *Camb. Handb. Creat*, pages 301–318, 2010.
- M. Csikszentmihalyi. Flow and the psychology of discovery and invention. *Harper-Perennial, New York*, 39:1–16, 1997.
- S. Da Costa, D. Páez, F. Sánchez, M. Garaigordobil, and S. Gondim. Personal factors of creativity: A second order meta-analysis. *Revista de Psicología del Trabajo y de las Organizaciones*, 31(3):165–173, 2015.

- C. Darwin. *The expression of the emotions in man and animals*. John Murray, 1872.
- D. D. Davis and C. A. Holt. *Experimental economics*. Princeton university press, 1993.
- C. K. De Dreu, B. A. Nijstad, M. N. Bechtoldt, and M. Baas. Group creativity and innovation: A motivated information processing perspective. *Psychology of aesthetics, creativity, and the arts*, 5(1):81, 2011.
- M. A. Devanna and N. Tichy. Creating the competitive organization of the 21st century: The boundaryless corporation. *Human Resource Management*, 29(4): 455–471, 1990.
- J. Diamond. Laboratory experiments, field experiments, and natural experiments. *Community ecology*, pages 3–22, 1986.
- S. S. Douthitt, L. T. Eby, and S. A. Simon. Diversity of life experiences: The development and validation of a biographical measure of receptiveness to dissimilar others. *International Journal of Selection and Assessment*, 7:112–125, 1999.
- A. Drach-Zahavy and A. Somech. Understanding team innovation: The role of team processes and structures. *Group Dynamics: Theory, Research, and Practice*, 5(2): 111, 2001.
- S. Duchek. Organizational resilience: a capability-based conceptualization. *Business research*, 13(1):215–246, 2020.
- W. Duff. *An essay on original genius: and its various modes of exertion in philosophy and the fine arts, particularly in poetry*. E. and C. Dilly, 1767.
- G. Dutcher and C. S. Rodet. Which two heads are better than one? uncovering the positive effects of diversity in creative teams. *Behavioral & Experimental Economics eJournal*, 2022.
- A. Dzedzickis, A. Kaklauskas, and V. Bucinskas. Human emotion recognition: Review of sensors and methods. *Sensors*, 20(3):592, 2020. ISSN 1424-8220.
- N. Eber and M. Willinger. *L'économie expérimentale*. Number 423 in Repères (Maspero). Editions la Découverte, 2005.

- K. Eckartz, O. Kirchkamp, and D. W. Schunk. How do incentives affect creativity? *Labor: Personnel Economics eJournal*, 2012.
- P. Eder and J. E. Sawyer. A meta-analytic examination of employee creativity. In *22nd annual conference, society of industrial and organizational psychology (SIOP), New York, NY, April*, 2007.
- A. Edmondson. Psychological safety and learning behavior in work teams. *Administrative science quarterly*, 44(2):350–383, 1999.
- M. Egger, M. Ley, and S. Hanke. Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*, 343:35–55, 2019. ISSN 1571-0661.
- R. Eisenberger, F. Haskins, and P. Gambleton. Promised reward and creativity: Effects of prior experience. *Journal of Experimental Social Psychology*, 35(3):308–325, 1999.
- M. Elmoukhli. *Co-opétition et performance créative dans le crowdsourcing d'idées*. PhD thesis, 2018. Thèse de doctorat dirigée par Balagué, Christine Sciences de gestion Université Paris-Saclay (ComUE) 2018.
- G. J. Feist. A meta-analysis of personality in scientific and artistic creativity. *Personality and social psychology review*, 2(4):290–309, 1998.
- L. Festinger. *The human legacy*. Columbia University Press, 1983.
- S. Fey and A. Kock. Meeting challenges with resilience—how innovation projects deal with adversity. *International Journal of Project Management*, 40(8):941–950, 2022.
- S. M. Fiore, M. A. Rosen, K. A. Smith-Jentsch, E. Salas, M. Letsky, and N. Warner. Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. *Human factors*, 52(2):203–224, 2010.
- G. Fischer, E. Giaccardi, H. Eden, M. Sugimoto, and Y. Ye. Beyond binary choices: Integrating individual and social creativity. *Int. J. Hum. Comput. Stud.*, 63:482–512, 2005.
- M. M. Flood. Some experimental games. *Management Science*, 5(1):5–26, 1958.

- P. W. Foos and D. Boone. Adult age differences in divergent thinking: It's just a matter of time. *Educational Gerontology*, 34(7):587–594, June 2008. ISSN 1521-0472.
- D. R. Forsyth. Recent advances in the study of group cohesion. *Group Dynamics: Theory, Research, and Practice*, 25(3):213, 2021.
- D. Friedman and S. Sunder. *Experimental Methods*. Cambridge Books, Cambridge University Press, 9780521456821 edition, 1994.
- A. Furnham and V. Bachtiar. Personality and intelligence as predictors of creativity. *Personality and individual differences*, 45(7):613–617, 2008.
- A. Gajda. The relationship between school achievement and creativity at different educational stages. *Thinking Skills and Creativity*, 19:246–259, Mar. 2016. ISSN 1871-1871.
- N. Gao, W. Shao, M. S. Rahaman, and F. D. Salim. n-gage: Predicting in-class emotional, behavioural and cognitive engagement in the wild, 2020. URL <https://arxiv.org/abs/2007.04831>.
- B. Gaut. Creativity and rationality. *The Journal of Aesthetics and Art Criticism*, 70(3):259–270, 2012.
- C. Ge, R. Huang, M. Xie, Z. Lai, S. Song, S. Li, and G. Huang. Domain adaptation via prompt learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
- A. S. Gerber and D. P. Green. *Field Experiments*. WW Norton, New York, NY, May 2012.
- M. Giancola, M. Palmiero, L. Piccardi, and S. D'Amico. The contribution of planning to real-world creativity: The moderating role of agreeableness. *Thinking Skills and Creativity*, 41:100890, 2021.
- L. L. Gilson and C. E. Shalley. A little creativity goes a long way: An examination of teams' engagement in creative processes. *Journal of management*, 30(4):453–470, 2004.

- K. Girotra, L. Meincke, C. Terwiesch, and K. T. Ulrich. Ideas are dimes a dozen: Large language models for idea generation in innovation. *SSRN Electronic Journal*, 2023. ISSN 1556-5068.
- V. P. Glăveanu. Developing society: Reflections on the notion of societal creativity. *Creativity, culture, and development*, pages 183–200, 2015.
- D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer. Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing*, 2(2):106–118, 2011. ISSN 1949-3045.
- V. P. Glăveanu and J. C. Kaufman. *Creativity: A Historical Perspective*, page 9–26. Cambridge Handbooks in Psychology. Cambridge University Press, 2019.
- F. C. Godart, W. W. Maddux, A. V. Shipilov, and A. D. Galinsky. Fashion with a foreign flair: Professional experiences abroad facilitate the creative innovations of organizations. *Academy of Management Journal*, 58(1), 2015.
- K. Grace and M. L. Maher. Surprise and reformulation as meta-cognitive processes in creative design. In *Proceedings of the third annual conference on advances in cognitive systems ACS*, page 8, 2015.
- R. M. Grant. Prospering in dynamically-competitive environments: Organizational capability as knowledge integration. *Organization science*, 7(4):375–387, 1996.
- S. Grüner. Sample size calculation in economic experiments. *Jahrbücher für Nationalökonomie und Statistik*, 240(6):791–823, Feb. 2020.
- F. Guala. *experimental economics, history of*, pages 99–106. Palgrave Macmillan UK, London, 2010. ISBN 978-0-230-28078-6.
- F. Guala. Experimentation in economics. In U. Mäki, editor, *Philosophy of Economics*, Handbook of the Philosophy of Science, pages 597–640. North-Holland, Amsterdam, 2012. doi: <https://doi.org/10.1016/B978-0-444-51676-3.50021-X>.
- S. J. Guastello. *Chaos, catastrophe, and human affairs: Applications of nonlinear dynamics to work, organizations, and social evolution*. 1995.
- C. Guckelsberger, C. Salge, and S. Colton. Addressing the “why?” in computational creativity: A non-anthropocentric, minimal model of intentional creative agency.

- In *International Conference on Computational Creativity 2017*. Association for Computational Creativity (ACC), 2017.
- R. Guichardaz, M. Lefebvre, H. Igersheim, and J. Pénin. Personality, creativity and adherence to intellectual property: A lab experiment on copyright. 2024.
- J. P. Guilford. Creativity. *American Psychologist*, 5:444–454, 1950.
- J. P. Guilford. The structure of intellect. *Psychological bulletin*, 53(4):267, 1956.
- J. P. Guilford. *The nature of human intelligence*. New York: McGraw-Hill, 1967.
- C. Guo, Y. V. Chen, Z. C. Qian, Y. Ma, H. Dinh, and S. Anasingaraju. *Designing a Smart Scarf to Influence Group Members’ Emotions in Ambience: Design Process and User Experience*, page 392–402. Springer International Publishing, 2016. ISBN 9783319402444. doi: 10.1007/978-3-319-40244-4_38. URL http://dx.doi.org/10.1007/978-3-319-40244-4_38.
- J. Haase and P. H. Hanel. Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. *Journal of Creativity*, 33(3):100066, 2023.
- J. H. Hardy and C. Gibson. Gender differences in the measurement of creative problem-solving. *The Journal of Creative Behavior*, 51(2):153–162, 2017.
- J. H. Hardy, A. M. Ness, and J. Mecca. Outside the box: Epistemic curiosity as a predictor of creative problem solving and creative performance. *Personality and Individual Differences*, 104:230–237, Jan. 2017. ISSN 0191-8869.
- A. B. Hargadon and B. A. Bechky. When collections of creatives become creative collectives: A field study of problem solving at work. *Organization science*, 17(4):484–500, 2006.
- D. A. Harrison, K. H. Price, and M. P. Bell. Beyond relational demography: Time and the effects of surface- and deep-level diversity on work group cohesion. *Academy of Management Journal*, 41:96–107, 1998.
- A. Hartwig, S. Clarke, S. Johnson, and S. Willis. Workplace team resilience: A systematic review and conceptual development. *Organizational Psychology Review*, 10(3-4):169–200, 2020.

- S. Harvey. A different perspective: The multiple effects of deep level diversity on group creativity. *Journal of Experimental Social Psychology*, 49:822–832, 2013.
- B. A. Hennessey and T. M. Amabile. The conditions of creativity, 1988.
- B. A. Hennessey and T. M. Amabile. Creativity. *Annual Review of Psychology*, 61(1):569–598, Jan. 2010.
- R. Hertwig and A. Ortmann. Economists’ and psychologists’ experimental practices: How they differ, why they differ, and how they could converge. In *The psychology of economic decisions*, pages 253–272. Oxford University Press, 2003.
- J. D. Hey. Experimental economics and deception: A comment. *Journal of Economic Psychology*, 19(3):397–401, 1998.
- A. C. Hoggatt. An experimental business game. *Behavioral Science*, 4(3):192–203, 1959.
- C. A. Holt and S. K. Laury. Chapter 4 - assessment and estimation of risk preferences. In M. Machina and K. Viscusi, editors, *Handbook of the Economics of Risk and Uncertainty*, volume 1 of *Handbook of the Economics of Risk and Uncertainty*, pages 135–201. North-Holland, 2014. doi: <https://doi.org/10.1016/B978-0-444-53685-3.00004-0>. URL <https://www.sciencedirect.com/science/article/pii/B9780444536853000040>.
- J. Hornberg and R. Reiter-Palmon. Creativity and the big five personality traits: is the relationship dependent on the creativity measure? 2017.
- Y.-F. Huang and C.-J. Chen. The impact of technological diversity and organizational slack on innovation. *Technovation*, 30(7):420–428, 2010. ISSN 0166-4972.
- K. F. Hubert, K. N. Awa, and D. L. Zabelina. The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Scientific Reports*, 14(1):3440, 2024.
- U. Hülshager, N. Anderson, and J. Salgado. Selecting for innovation: What’s good for job performance doesn’t need to be good for innovative performance. In *14th Congress of the European Association of Work and Organizational Psychology Santiago*, pages 13–16, 2009.
- D. Hume. *A treatise of human nature*. Oxford University Press, 1739-40.

- A. S. Hundschell, S. Razinskas, J. Backmann, and M. Hoegl. The effects of diversity on creativity: A literature review and synthesis. *Applied Psychology*, 2021.
- G. Iadarola, A. Poli, and S. Spinsante. Analysis of galvanic skin response to acoustic stimuli by wearable devices. In *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, 2021. doi: 10.1109/memea52024.2021.9478673. URL <http://dx.doi.org/10.1109/MeMeA52024.2021.9478673>.
- A. L. Ilha Villanova and M. Pina e Cunha. Everyday creativity: A systematic literature review. *The Journal of Creative Behavior*, 55(3):673–695, 2021.
- A. Innocenti. The early development of experimental economics: The influence of game theory. 2000.
- M. O. Jackson and A. Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71:44–74, 1996.
- N. Jacquemet and O. L’Haridon. *Experimental Economics: Method and Applications*. Cambridge University Press, 2018.
- D. M. Jankowska and M. Karwowski. Family factors and development of creative thinking. *Personality and Individual Differences*, 142:202–206, 2019. ISSN 0191-8869.
- A. E. Jansen and B. J. Searle. Diverse effects of team diversity: a review and framework of surface and deep-level diversity. *Personnel Review*, 50(9):1838–1853, 2021.
- K. A. Jehn, S. Rispens, and S. M. B. Thatcher. The effects of conflict asymmetry on work group and individual outcomes. *Academy of Management Journal*, 53: 596–616, 2010.
- V. John-Steiner. *Creative Collaboration*. Oxford University Press, New York, NY, Sept. 2006.
- G. K. Kalisch, J. W. Milnor, J. F. Nash, and E. D. Nering. Some experimental n-person games. 1954.

- R. M. Kanter. When a thousand flowers bloom: Structural, collective, and social conditions for innovation in organizations. In *Knowledge Management and Organisational Design*, pages 93–131. Elsevier, 1996.
- A. Kaplan and M. Haenlein. Siri, siri, in my hand: Who’s the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1):15–25, 2019. ISSN 0007-6813.
- M. Karwowski, D. M. Jankowska, J. Gralewski, A. Gajda, E. Wiśniewska, and I. Lebuda. Greater male variability in creativity: A latent variables approach. *Thinking Skills and Creativity*, 22:159–166, Dec. 2016. ISSN 1871-1871.
- M. Karwowski, I. Lebuda, and E. Wisniewska. Measuring creative self-efficacy and creative personal identity. *The Journal of Creativity and Problem Solving*, 28:45–57, 05 2018.
- A. Kaufmann, R. Faure, and A. L. Garff. *Les jeux d’entreprises*. PUF, Col. Que sais-je ?, 1976.
- S. Kavadias and S. C. Sommer. The effects of problem structure and team diversity on brainstorming effectiveness. *Manag. Sci.*, 55:1899–1913, 2009.
- E. Kearney, S. Razinskas, M. Weiss, and M. Hoegl. Gender diversity and team performance under time pressure: The role of team withdrawal and information elaboration. *Journal of Organizational Behavior*, 43(7):1224–1239, 2022.
- S. H. Kim, D. W. Ryoo, and C. Bae. U-healthcare system using smart headband. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2008. doi: 10.1109/iembs.2008.4649467. URL <http://dx.doi.org/10.1109/iembs.2008.4649467>.
- A. J. Kimmel. In defense of deception. 1998.
- P. Kleiman. Towards transformation: conceptions of creativity in higher education. *Innovations in education and teaching international*, 45(3):209–217, 2008.
- A. Kleinsmith and N. Bianchi-Berthouze. Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing*, 4(1):15–33, 2013.

- P. Koellinger. Why are some entrepreneurs more innovative than others? *Small Business Economics*, 31:21–37, 2008.
- N. W. Kohn and S. M. Smith. Collaborative fixation: effects of others’ ideas on brainstorming. *Applied Cognitive Psychology*, 25:359–371, 2011.
- N. W. Kohn, P. B. Paulus, and Y. Choi. Building on the ideas of others: An examination of the idea combination process. *Journal of Experimental Social Psychology*, 47:554–561, 2011.
- M. Koivisto and S. Grassini. Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific reports*, 13(1):13601, 2023.
- N. Kosmyna, C. Morris, U. Sarawgi, T. Nguyen, and P. Maes. Attentivu: A wearable pair of eeg and eog glasses for real-time physiological processing. In *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pages 1–4, 2019. doi: 10.1109/BSN.2019.8771080.
- S. W. Kozlowski and D. R. Ilgen. Enhancing the effectiveness of work groups and teams. *Psychological science in the public interest*, 7(3):77–124, 2006.
- J. Kruger and D. Dunning. Unskilled and unaware of it: how difficulties in recognizing one’s own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6):1121, 1999.
- D. C. Lau and J. K. Murnighan. Demographic diversity and faultlines: The compositional dynamics of organizational groups. *Academy of Management Review*, 23: 325–340, 1998.
- O. Lederman, A. Mohan, D. Calacci, and A. S. Pentland. Rhythm: A unified measurement platform for human organizations. *IEEE MultiMedia*, 25(1):26–38, 2018. doi: 10.1109/MMUL.2018.112135958.
- Y.-N. Lee, J. P. Walsh, and J. Wang. Creativity in scientific teams: Unpacking novelty and impact. *Research policy*, 44(3):684–697, 2015.
- S. W. Lester, B. M. Meglino, and M. A. Korsgaard. The antecedents and consequences of group potency: A longitudinal investigation of newly formed work groups. *Academy of Management Journal*, 45(2):352–368, 2002.

- A. K.-y. Leung, W. W. Maddux, A. D. Galinsky, and C.-y. Chiu. Multicultural experience enhances creativity: the when and how. *American psychologist*, 63(3): 169, 2008.
- Y. Li, F. Fu, J.-M. Sun, and B. Yang. Leader–member exchange differentiation and team creativity: An investigation of nonlinearity. *Human Relations*, 69(5):1121–1138, 2016.
- J. List. Field experiments: A bridge between lab and naturally occurring data. *Topics in Economic Analysis Policy*, 6, 2007.
- J. A. List, S. Sadoff, and M. Wagner. So you want to run an experiment, now what? some simple rules of thumb for optimal experimental design. *Experimental Economics*, 14(4):439–457, Mar. 2011.
- K. K. Longva. Student venture creation: developing social networks within entrepreneurial ecosystems in the transition from student to entrepreneur. *International Journal of Entrepreneurial Behavior & Research*, 2021.
- H. Louch. Personal network integration: transitivity and homophily in strong-tie relations. *Social networks*, 22(1):45–64, 2000.
- T. Lubart, C. Mouchiroud, S. Tordjman, and F. Zenasni. *Psychologie de la créativité-2e édition*. Armand Colin, 2015.
- H.-H. Ma. The effect size of variables associated with creativity: A meta-analysis. *Creativity Research Journal*, 21(1):30–42, 2009.
- D. W. MacKinnon. What makes a person creative? *Theory Into Practice*, 5(4): 152–156, 1966. ISSN 00405841, 15430421.
- F. Magni, J. Park, and M. M. Chao. Humans as creativity gatekeepers: Are we biased against ai creativity? *Journal of Business and Psychology*, pages 1–14, 2023.
- K. Malhi, S. C. Mukhopadhyay, J. Schnepfer, M. Haefke, and H. Ewald. A zigbee-based wearable physiological parameters monitoring system. *IEEE Sensors Journal*, 12(3):423–430, 2012. doi: 10.1109/JSEN.2010.2091719.

- A.-G. Maltese, S. Gil-Gallen, and P. Llerena. Disentangling the role of surface and deep-level variables on individuals' and groups' creative performance: A cross-level experimental evidence. 2023.
- M. A. Marks, J. E. Mathieu, and S. J. Zaccaro. A temporally based framework and taxonomy of team processes. *The Academy of Management Review*, 26(3):356, July 2001. ISSN 0363-7425.
- L. Martin, B. Muller, P. J. O. Suárez, Y. Dupont, L. Romary, É. V. de la Clergerie, D. Seddah, and B. Sagot. Camembert: a tasty french language mode. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020.
- P. Massimiliano. The effects of age on divergent thinking and creative objects production: a cross-sectional study. *High Ability Studies*, 26(1):93–104, Jan. 2015. ISSN 1469-834X.
- J. E. Mathieu, J. R. Hollenbeck, D. van Knippenberg, and D. R. Ilgen. A century of work teams in the journal of applied psychology. *Journal of Applied Psychology*, 102(3):452–467, 2017. ISSN 0021-9010.
- U. Maurer, A. G. Rowe, A. Smailagic, and D. P. Siewiorek. ewatch: a wearable sensor and notification platform. *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*, pages 4 pp.–145, 2006. URL <https://api.semanticscholar.org/CorpusID:16567032>.
- M. T. Maynard, D. M. Kennedy, and S. A. Sommer. Team adaptation: A fifteen-year synthesis (1998–2013) and framework for how this literature needs to “adapt” going forward. *European Journal of Work and Organizational Psychology*, 24(5): 652–677, 2015.
- R. R. McCrae and P. T. Costa. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, 52(1):81, 1987.
- J. C. McCullers. Issues in learning and motivation. In *The hidden costs of reward*, pages 5–17. Psychology Press, 1978.
- R. R. McDaniel, M. E. Jordan, and B. F. Fleeman. Surprise, surprise, surprise! a complexity science view of the unexpected. *Health Care Management Review*, 28(3):266–278, 2003. ISSN 03616274, 15505030.

- K. O. McGraw. The detrimental effects of reward on performance: A literature review and a prediction model. *The hidden costs of reward*, pages 33–60, 1978.
- K. E. Medeiros, P. J. Partlow, and M. D. Mumford. Not too much, not too little: The influence of constraints on creative problem solving. *Psychology of Aesthetics, Creativity, and the Arts*, 8(2):198, 2014.
- S. A. Mednick. The associative basis of the creative process. *Psychological review*, 69:220–32, 1962.
- I. Meneghel, I. M. Martínez, and M. Salanova. Job-related antecedents of team resilience and improved team performance. *Personnel review*, 45(3):505–522, 2016.
- B. Meyer and A. Glenz. Team faultline measures: A computational comparison and a new approach to multiple subgroups. *Organizational Research Methods*, 16(3):393–424, Apr. 2013. ISSN 1552-7425.
- B. Meyer, M. Shemla, J. Li, and J. Wegge. On the same side of the faultline: Inclusion in the leader’s subgroup and employee performance. *Journal of Management Studies*, 52(3):354–380, Feb. 2015. ISSN 1467-6486.
- J. S. Mill. On the definition of political economy; and on the method of investigation proper to it. *London and Westminster Review*, 4(October):120–164, 1836.
- A. I. Miller. *Insights of genius*. Springer, New York, NY, Sept. 1996.
- C. R. Miller. *Surprise: The poetics of the unexpected from Milton to Austen*. Cornell University Press, 2015.
- F. J. Milliken and L. L. Martins. Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of Management Review*, 21:402–433, 1996.
- M. Minsky. Steps toward artificial intelligence. *Proceedings of the IRE*, 49(1):8–30, 1961.
- E. Miron-Spektor, A. Ingram, J. Keller, W. K. Smith, and M. W. Lewis. Micro-foundations of organizational paradox: The problem is how we think about the problem. *Academy of management journal*, 61(1):26–45, 2018.

- R. Mitchell, B. Boyle, and S. Nicholas. The impact of goal structure in team knowledge creation. *Group processes & intergroup relations*, 12(5):639–651, 2009.
- W. C. Mitchell. Quantitative analysis in economic theory. *The American Economic Review*, 15(1):1–12, 1925.
- L. Mittone, A. Morreale, et al. What drives innovative behavior?-an experimental analysis on risk attitudes, creativity and performance. *Journal of Behavioral and Experimental Economics*, 98:101868, 2022.
- P. Moffatt, C. Starmer, R. Sugden, N. Bardsley, R. Cubitt, and G. Loomes. *Experimental economics: Rethinking the rules*. Princeton University Press, 2009.
- P. G. Moffatt. *Experiments: Econometrics for Experimental Economics*. New York: McGraw-Hill, 2020.
- E. Molleman. Diversity in demographic characteristics, abilities and personality traits: Do faultlines affect team functioning? *Group decision and Negotiation*, 14(3):173–193, 2005.
- A. Mollière. *Métaphysique de l’art*. Gaume Frères, 1849.
- M. Morgan. Morgan (2003) economics (porter ross cambridge history of science vol 7, book chap). 06 2023.
- O. Morgenstern. *Economic activity analysis*. 1954.
- F. Mosteller and P. Noguee. An experimental measurement of utility. *Journal of political economy*, 59(5):371–404, 1951.
- L. M. Moynihan and R. S. Peterson. A contingent configuration approach to understanding the role of personality in organizational groups. *Research in organizational behavior*, 23:327–378, 2001.
- J. S. Mueller, S. Melwani, and J. A. Goncalo. The bias against creativity: Why people desire but reject creative ideas. *Psychological science*, 23(1):13–17, 2012.
- B. Mullen and C. Copper. The relation between group cohesiveness and performance: An integration. *Psychological bulletin*, 115(2):210, 1994.

- T. H. Naylor, D. S. Burdick, and W. E. Sasser. Computer simulation experiments with economic systems: the problem of experimental design. *Journal of the American Statistical Association*, 62(320):1315–1337, 1967.
- S. L. Nielsen and W. B. Gartner. Am i a student and/or entrepreneur? multiple identities in student entrepreneurship. *Journal of Education and Training*, 59: 135–154, 2017.
- B. A. Nijstad, C. K. De Dreu, E. F. Rietzschel, and M. Baas. The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European review of social psychology*, 21(1):34–77, 2010.
- G. Noh. Effects of network size, diversity, and interaction frequency on individual creativity: A study from south korea. *Social Sciences*, 11(4):170, Apr. 2022.
- M. K. Noordewier and S. M. Breugelmans. On the valence of surprise. *Cognition & emotion*, 27(7):1326–1334, 2013.
- R. Ochse. *Before the gates of excellence: The determinants of creative genius*. CUP Archive, 1990.
- G. R. Oldham and A. Cummings. Employee creativity: Personal and contextual factors at work. *Academy of management journal*, 39(3):607–634, 1996.
- M. Öllinger and V. Goel. Problem solving. *Towards a theory of thinking: Building blocks for a conceptual framework*, pages 3–21, 2010.
- S. Pal, S. Mukhopadhyay, and N. Suryadevara. Development and progress in sensors and technologies for human emotion recognition. *Sensors*, 21(16):5554, 2021. ISSN 1424-8220.
- W. Pang, Y. Lu, H. Long, Q. Wang, and L. Lin. Three-generational families: Are they beneficial to chinese children’s creativity? *Thinking Skills and Creativity*, 35: 100623, Mar. 2020. ISSN 1871-1871.
- P. B. Paulus, M. Dzindolet, and N. W. Kohn. Collaborative creativity—group creativity and team innovation. In *Handbook of organizational creativity*, pages 327–357. Elsevier, 2012.

- M. J. Pearsall, A. P. Ellis, and J. M. Evans. Unlocking the effects of gender faultlines on team creativity: Is activation the key? *Journal of Applied Psychology*, 93(1): 225, 2008.
- L. H. Pelled, K. M. Eisenhardt, and K. R. Xin. Exploring the black box: An analysis of work group diversity, conflict and performance. *Administrative Science Quarterly*, 44(1):1–28, 1999.
- R. H. Perry, B. Charlotte, M. Isabella, and C. Bob. *SPSS Explained*. London Routledge, 2004.
- J. E. Perry-Smith and C. E. Shalley. The social side of creativity: A static and dynamic social network perspective. *The Academy of Management Review*, 28(1): 89–106, 2003. ISSN 03637425.
- O. Pesout and J. Nietfeld. How creative am i?: Examining judgments and predictors of creative performance. *Thinking Skills and Creativity*, 40:100836, 2021.
- R. J. Pidduck, L. W. Busenitz, Y. Zhang, and A. G. Moulick. Oh, the places you’ll go: A schema theory perspective on cross-cultural experience and entrepreneurship. *Journal of Business Venturing Insights*, 14:e00189, 2020.
- A. Pirola-Merlo. Agile innovation: The role of team climate in rapid research and development. *Journal of occupational and organizational psychology*, 83(4):1075–1084, 2010.
- E. Plastino and M. Purdy. Game changing value from artificial intelligence: eight strategies. *Strategy & Leadership*, 46(1):16–22, 2018.
- J. A. Plucker, R. A. Beghetto, and G. T. Dow. Why isn’t creativity more important to educational psychologists? potentials, pitfalls, and future directions in creativity research. *Educational psychologist*, 39(2):83–96, 2004.
- J. A. Plucker, M. C. Makel, and M. Qian. *Assessment of Creativity*, page 44–68. Cambridge Handbooks in Psychology. Cambridge University Press, 2019.
- H. Poincaré. *A description of his own creativity*. Trans. G.B. Halstead, 1908.
- P. Pollok, A. Amft, K. Diener, D. Lüttgens, and F. T. Piller. Knowledge diversity and team creativity: How hobbyists beat professional designers in creating novel board games. *Research Policy*, 50(8):104174, 2021. ISSN 0048-7333.

- J. E. Pretz and J. C. Kaufman. Do traditional admissions criteria reflect applicant creativity? *The Journal of Creative Behavior*, 51(3):240–251, 2017.
- J. E. Pretz and V. A. McCollum. Self-perceptions of creativity do not always reflect actual creative performance. *Psychology of Aesthetics, Creativity, and the Arts*, 8(2):227, 2014.
- A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022.
- R. I. Ramos-Garcia, M. H. Imtiaz, E. Sazonov, and S. T. Tiffany. Evaluation of rip sensor calibration stability for daily estimation of lung volume. In *2017 Eleventh International Conference on Sensing Technology (ICST)*, pages 1–5, 2017. doi: 10.1109/ICSensT.2017.8304419.
- V. Rawte, A. Sheth, and A. Das. A survey of hallucination in large foundation models. *arXiv preprint arXiv:2309.05922*, 2023.
- N. Reimers. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.
- R. Reisenzein, W.-U. Meyer, and A. Schuetzwohl. Reactions to surprising events: A paradigm for emotion research. *Proceedings of the 9th Conference of the International Society for Research on Emotions*, pages 292–296, 01 1996.
- R. Reisenzein, G. Horstmann, and A. Schützwohl. The cognitive-evolutionary model of surprise: A review of the evidence. *Topics in Cognitive Science*, 11(1):50–74, Sept. 2017. ISSN 1756-8765.
- A. Reiss and O. Amft. *Design Challenges of Real Wearable Computers*, pages 583–618. 2015. ISBN 978-1-4822-4350-5.
- R. Reiter-Palmon and V. R. Murugavel. *Problem Finding*, page 389–393. Elsevier, 2020.
- R. Reiter-Palmon, B. Wigert, and T. de Vreede. Team creativity and innovation: The effect of group composition, social processes, and cognition. In *Handbook of organizational creativity*, pages 295–326. Elsevier, 2012.
- R. Richards. *Everyday creativity: our hidden potential*. 2007.

- E. F. Rietzschel, B. A. Nijstad, and W. Stroebe. The selection of creative ideas after individual idea generation: Choosing between creativity and impact. *British journal of psychology*, 101(1):47–68, 2010.
- R. Rosso, G. Munaro, O. Salvetti, S. Colantonio, and F. Ciancitto. Chronious: An open, ubiquitous and adaptive chronic disease management platform for chronic obstructive pulmonary disease (copd), chronic kidney disease (ckd) and renal insufficiency. 2010. doi: 10.1109/iembs.2010.5626451. URL <http://dx.doi.org/10.1109/iembs.2010.5626451>.
- A. E. Roth. Laboratory experimentation in economics. *Economics amp; Philosophy*, 2(2):245–273, 1986.
- A. E. Roth, editor. *Laboratory Experimentation in Economics: Six Points of View*. Cambridge University Press, 1987.
- A. E. Roth. The early history of experimental economics. *Journal of the History of Economic Thought*, 15(2):184–209, 1993.
- A. E. Roth. Introduction to experimental economics. *The handbook of experimental economics*, 1:3–109, 1995.
- A. E. Roth and M. A. O. Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press, 1990. ISBN 9781139052214.
- M. Rousseau, B. Mathias, Ph.D., L. Madden, and T. Crook. Innovation, firm performance, and appropriation:. *International Journal of Innovation Management*, 20:1650033, 11 2015.
- J.-L. Rullière. D’adam smith à vernon l. smith la main invisible observée à travers les comportements expérimentaux. *Revue d’économie politique*, 113(3):309–321, 2003.
- M. A. Runco and R. S. Albert. *Creativity research: A historical view*. Cambridge University Press, 2010.
- S. Saganowski, B. Perz, A. G. Polak, and P. Kazienko. Emotion recognition for everyday life using physiological signals from wearables: A systematic literature review. *IEEE Transactions on Affective Computing*, 14(3):1876–1897, 2023.

- T. A. Salthouse. Why do adult age differences increase with task complexity? *Developmental Psychology*, 28(5):905, 1992.
- T. A. Salthouse. 31 Relations between Age and Cognitive Functioning. In *Major Issues in Cognitive Aging*. Oxford University Press, 12 2009. ISBN 9780195372151.
- P. A. Samuelson and W. D. Nordhaus. *Economics*. McGraw-Hill, 1985.
- P. A. Samuelson and W. D. Nordhaus. *Economics*. McGraw-Hill, 1992.
- H. Sarooghi, D. Libaers, and A. Burkemper. Examining the relationship between creativity and innovation: A meta-analysis of organizational, cultural, and environmental factors. *Journal of Business Venturing*, 30(5):714–731, 2015. ISSN 0883-9026.
- H. Sauermann and R. Selten. Ein oligopolexperiment. *Zeitschrift für die gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics*, (H. 3): 427–471, 1959.
- R. K. Sawyer and D. Henriksen. *Explaining creativity: The science of human innovation*. Oxford university press, 2012.
- R. K. Sawyer and D. Henriksen. *Explaining creativity: The science of human innovation*. Oxford university press, 2024.
- S. N. M. Sayed Ismail, N. A. Ab. Aziz, and S. Z. Ibrahim. A comparison of emotion recognition system using electrocardiogram (ecg) and photoplethysmogram (ppg). *Journal of King Saud University - Computer and Information Sciences*, 34(6, Part B):3539–3558, 2022. ISSN 1319-1578.
- S. Schachter, N. Ellertson, D. McBride, and D. Gregory. An experimental study of cohesiveness and productivity. *Human Relations*, 4(3):229–238, 1951.
- M. C. Schilpzand, D. M. Herold, and C. E. Shalley. Members’ openness to experience and teams’ creative performance. *Small Group Research*, 42(1):55–76, 2011.
- P. Schmidt, A. Reiss, R. Dürichen, and K. V. Laerhoven. Wearable-based affect recognition—a review. *Sensors*, 19(19):4079, 2019. ISSN 1424-8220.
- M. A. Schmuckler. What is ecological validity? a dimensional analysis. *Infancy*, 2(4):419–436, Oct. 2001.

- D. A. Schön. Educating the reflective practitioner. 1987.
- R. Selten. Emergence and future of experimental economics. In *Observation and experiment in the natural and social sciences*, pages 63–70. Springer, 2003.
- R. Selten. Biographical. NobelPrize.org. Nobel Prize Outreach AB, 2024. URL <https://www.nobelprize.org/prizes/economic-sciences/1994/selten/biographical/>. Accessed March 28, 2024.
- D. Serra. Un aperçu historique de l'économie expérimentale : des origines aux évolutions récentes. *Revue d'économie politique*, Vol. 122(5):749–786, Nov. 2012a.
- D. Serra. Principes méthodologiques et pratiques de l'économie expérimentale: une vue d'ensemble. *Revue de philosophie économique*, (1):21–78, 2012b.
- D. Serra. La "révolution" expérimentale en économie. une histoire des courants de recherche qui l'incarnent. Technical report, 2022.
- C. E. Shalley. Effects of coaction, expected evaluation, and goal setting on creativity and productivity. *Academy of Management journal*, 38(2):483–503, 1995.
- C. E. Shalley, L. L. Gilson, and T. C. Blum. Interactive effects of growth need strength, work context, and job complexity on self-reported creative performance. *Academy of Management journal*, 52(3):489–505, 2009.
- S. Siegel and L. E. Fouraker. Bargaining and group decision making: Experiments in bilateral monopoly. 1960.
- H. A. Simon. The structure of ill structured problems. *Artificial intelligence*, 4(3-4): 181–201, 1973.
- D. K. Simonton. *Origins of genius: Darwinian perspectives on creativity*. Oxford University Press, 1999.
- A. Smith. *Essays on philosophical subjects*. London: T. Cadell Jun. and W. Davies, 1795.
- V. L. Smith. An experimental study of competitive market behavior. *Journal of political economy*, 70(2):111–137, 1962.
- V. L. Smith. Microeconomic systems as an experimental science. *The American economic review*, 72(5):923–955, 1982.

- V. L. Smith. Experimental economics at purdue. *Papers in Experimental Economics*. Cambridge University Press, Cambridge, England, 1991a.
- V. L. Smith. Rational choice: The contrast between economics and psychology. *Journal of Political Economy*, 99(4):877–897, 1991b.
- A. Solanas, R. M. Selvam, J. Navarro, and D. Leiva. Some common indices of group diversity: Upper boundaries. *Psychological Reports*, 111:777 – 796, 2012.
- R. Spears. Social influence and group identity. *Annual review of psychology*, 72: 367–390, 2021.
- L. Spencer and R. Pahl. *Rethinking friendship: Hidden solidarities today*. Princeton University Press, 2007.
- A. Spickermann, M. Zimmermann, and H. A. Gracht. Surface- and deep-level diversity in panel selection — exploring diversity effects on response behaviour in foresight. *Technological Forecasting and Social Change*, 85:105–120, 2014.
- R. J. Spiro et al. Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. technical report no. 441. 1988.
- G. Stahl, M. Maznevski, A. Voigt, and K. Jonsen. Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies*, 41:690–709, 05 2010.
- L. D. Steiner. *Group Process and Productivity*. Social Psychological Monograph. Academic Press, San Diego, CA, Sept. 1972.
- R. J. Sternberg and T. I. Lubart. *Defying the crowd: Cultivating creativity in a culture of conformity*. Free press, 1995.
- R. J. Sternberg and T. I. Lubart. Investing in creativity. *American psychologist*, 51 (7):677, 1996.
- A. D. Stevens. *Social problem-solving and cognitive flexibility: Relations to social skills and problem behavior of at-risk young children*. Proquest, Umi Dissertation Publishing, Sept. 2011.
- C. Stevenson, I. Smal, M. Baas, R. Grasman, and H. van der Maas. Putting gpt-3’s creativity to the (alternative uses) test. *arXiv preprint arXiv:2206.08932*, 2022.

- D. Stokes and E. S. Paul. Naturalistic approaches to creativity. *A Companion to Experimental Philosophy*, pages 318–333, 2016.
- A. Stolaki, M. Satratzemi, and S. Xinogalos. Examining the effects of creativity, collaboration, creative diversity and autonomy on team creative performance. *Thinking Skills and Creativity*, 50:101415, Dec. 2023. ISSN 1871-1871.
- R. Sugden. Hume’s non-instrumental and non-propositional decision theory. *Economics & Philosophy*, 22(3):365–391, 2006.
- E. Sundstrom and etc. *Supporting work team effectiveness*. Jossey-Bass, London, England, Oct. 1998.
- E. Sundstrom, K. P. De Meuse, and D. Futrell. Work teams: Applications and effectiveness. *American psychologist*, 45(2):120, 1990.
- A. Svorenčík. Creating economics in the lab: from physical place to laboratory space. In *Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise*, pages 163–177. Emerald Publishing Limited, 2018.
- A. Svorenčík. The role of computers in the emergence of experimental economics laboratories: Material culture and moral economy. *Æconomia. History, Methodology, Philosophy*, (13-3):769–799, 2023.
- C.-S. Tan and L. Qu. Affect and creativity: An old topic and new direction. *Creativity, culture, and development*, pages 63–81, 2015.
- C. W. Taylor. Various approaches to and definitions of creativity. In R. J. Sternberg, editor, *The Nature of Creativity: Contemporary Psychological Perspectives*, pages 99–121. Cambridge University Press, Cambridge, 1988.
- D. W. Tegano. Relationship of tolerance of ambiguity and playfulness to creativity. *Psychological reports*, 66(3):1047–1056, 1990.
- D. Teichmann, D. De Matteis, T. Bartelt, M. Walter, and S. Leonhardt. A bendable and wearable cardiorespiratory monitoring device fusing two noncontact sensor principles. *IEEE Journal of Biomedical and Health Informatics*, 19(3):784–793, 2015. doi: 10.1109/JBHI.2015.2417760.
- J. R. Thorpe, B. H. Forchhammer, and A. M. Maier. Development of a sensor-based behavioral monitoring solution to support dementia care. *JMIR mHealth*

- and uHealth*, 7(6):e12013, 2019. ISSN 2291-5222. doi: 10.2196/12013. URL <http://dx.doi.org/10.2196/12013>.
- L. L. Thurstone. The indifference function. *The Journal of Social Psychology*, 2(2): 139–167, 1931.
- C. A. Toh and S. R. Miller. Creativity in design teams: the influence of personality traits and risk attitudes on creative concept selection. *Research in Engineering Design*, 27:73–89, 2016.
- E. P. Torrance. Torrance tests of creative thinking. *Educational and Psychological Measurement*, 1966.
- H. Triandis, A. Bass, R. Ewen, and E. Mikesell. Team creativity as a function of the creativity of the members. *Journal of Applied Psychology*, 47:104–110, 04 1963.
- G. Ursachi, I. A. Horodnic, and A. Zait. How reliable are measurement scales? external factors with indirect influence on reliability estimators. *Procedia. Economics and finance*, 20:679–686, 2015.
- N. Vallois and D. Jullien. A history of statistical methods in experimental economics. *The European Journal of the History of Economic Thought*, 25(6):1455–1492, 2018.
- A. E. M. Van, I. E. De Pater, A. L. Kristof-Brown, and E. C. Johnson. Fitting in: Surface- and deep-level cultural differences and expatriates’ adjustment. *Academy of Management Journal*, 47(5):697–709, Oct. 2004. ISSN 1948-0989.
- L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.
- H. van Dijk, M. L. van Engen, and D. van Knippenberg. Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes*, 119(1):38–53, 2012. ISSN 0749-5978.
- D. van Knippenberg. Team innovation. *Annual Review of Organizational Psychology and Organizational Behavior*, 4, 04 2017.
- D. van Knippenberg and I. J. Hoefer. *Team creativity and innovation*. Edward Elgar Publishing, Oct. 2021. ISBN 9781788977265.

- D. van Knippenberg, C. D. D. Dreu, and A. C. Homan. Work group diversity and group performance: an integrative model and research agenda. *The Journal of applied psychology*, 89 6:1008–22, 2004.
- J. F. Vargas, B. Zhou, H. Bello, and P. Lukowicz. Brainwear: Towards multi-modal garment integrated eeg. UbiComp '21, 2021. doi: 10.1145/3460421.3478814. URL <http://dx.doi.org/10.1145/3460421.3478814>.
- K. N. Vasilaky and J. M. Brock. Power (ful) guidelines for experimental economists. *Journal of the Economic Science Association*, 6(2):189–212, 2020.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need, 2017.
- T. Veale and F. A. Cardoso, editors. *Computational creativity*. Computational Synthesis and Creative Systems. Springer International Publishing, Cham, Switzerland, 1 edition, Aug. 2019.
- J. Villion. L'économie expérimentale : une variété de rôles. *Idées économiques et sociales*, N° 161(3):4–5, 2010. ISSN 2257-5111.
- F. Vinchon, V. Gironnay, and T. Lubart. The creative ai-land: Exploring new forms of creativity. 2023.
- G. Von Krogh. Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*, 4(4):404–409, 2018.
- J. von Neumann and O. Morgenstern. Theory of games and economic behaviour. 1964.
- J. von Neumann, O. Morgenstern, and A. Rubinstein. *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton University Press, 1944. ISBN 9780691130613.
- T. B. Ward. Structured imagination: the role of category structure in exemplar generation. *Cognitive Psychology*, 27:1–40, 1994.
- T. B. Ward. Cognition, creativity, and entrepreneurship. *Journal of business venturing*, 19(2):173–188, 2004.

- T. B. Ward and C. M. Sifonis. Tosh demands and generative thinking: What changes and what remains the same? *The Journal of Creative Behavior*, 31(4):245–259, 1997.
- T. B. Ward, M. J. Patterson, and C. M. Sifonis. The role of specificity and abstraction in creative idea generation. *Creativity Research Journal*, 16(1):1–9, 2004.
- H. Wechtler, J. K. Marodin, and M. J. Lehtonen. Systematic literature review on networks of innovative teams: Current trends and future research avenues. *European Management Journal*, 2023. ISSN 0263-2373.
- J. Wegge, C. Roth, B. Neubach, K.-H. Schmidt, and R. Kanfer. Age and gender diversity as determinants of performance and health in a public organization: the role of task complexity and group size. *Journal of Applied Psychology*, 93(6):1301, 2008.
- J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.
- R. Weiner. *Creativity and beyond: Cultures, values, and change*. SUNY Press, 2000.
- M. Weisberg and R. Muldoon. Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*, 76(2):225–252, Apr. 2009.
- K. Williams Phillips and C. O'Reilly. *Demography and Diversity in Organizations: A Review of 40 Years of Research*, volume 20, pages 77–140. 01 1998.
- R. C. Wilson, J. P. Guilford, P. R. Christensen, and D. J. Lewis. A factor-analytic study of creative-thinking abilities. *Psychometrika*, 19:297–311, 1954.
- R. Woodman, J. Sawyer, and R. Griffin. Toward a theory of organizational creativity. *Academy of Management Review*, 18:293–321, 04 1993.
- S. Wuchty, B. F. Jones, and B. Uzzi. The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039, 2007.
- W. Wundt. Principles of physiological psychology. 1948.
- X. Xu, Z. Gou, W. Wu, Z.-Y. Niu, H. Wu, H. Wang, and S. Wang. Long time no see! open-domain conversation with long-term persona memory. *arXiv preprint arXiv:2203.05797*, 2022.

- C. Yang, C. Garcia, J. Rodriguez-Andina, J. Farina, A. Iñiguez, and S. Yin. Using ppg signals and wearable devices for atrial fibrillation screening. *IEEE Transactions on Industrial Electronics*, 66:8832–8842, 2019. doi: 10.1109/TIE.2018.2889614.
- Y. Yuan, S. E. Humphrey, and D. van Knippenberg. From individual creativity to team creativity: A meta-analytic test of task moderators. *Journal of Occupational and Organizational Psychology*, 95(2):358–404, Jan. 2022. ISSN 2044-8325.
- Z. Zhao, D. Renard, M. Elmoukhli, and C. Balague. What affects creative performance in idea co-creation: competitive, cooperative or co-competitive climate ? *International Journal of Innovation Management*, 20(04):1640002, May 2016.
- J. Zhou and J. M. George. When job dissatisfaction leads to creativity: Encouraging the expression of voice. *The Academy of Management Journal*, 44(4):682–696, 2001. ISSN 00014273.

List of Figures

1.1	Categorization of Collective Experiments	65
1.2	Collective Experiment Variables Matrix	72
1.3	Measuring Tools Setting	87
2.1	Subjects' Performance as Number of Tasks They Succeed in	115
2.2	Subjects' Performance in the Draw Task	116
2.3	Subjects' Performance in the Alternative Uses Task	117
2.4	Subjects' Performance in the Tangram Task	118
3.1	Experimental Layouts	163
4.1	Violin Plot of Feasibility and Originality Scores by Treatment	220
5.1	Text Task Embeddings	267
5.2	Alternative Uses Task Embeddings	270
5.3	Draw Task Embeddings	272
5.4	Text Task PCA - All Criteria (Visual)	298
5.5	Alternative Uses Task PCA - All Criteria (Visual)	299
5.6	Alternative Uses Task PCA - All Criteria (Visual)	300
5.7	Comparison between Agents of Objective Measures	309
5.8	Comparison between Agents of Subjective Measures	310

List of Tables

2.1	Summary of the Descriptive Statistics	109
2.2	External Judges' Profiles	114
2.3	Summary Results for the Draw Task	120
2.4	Summary Results for the Alternative Uses Task in Expansivity	122
2.5	Summary Results for the Alternative Uses Task in Persistency	124
2.6	Summary Results for the Tangram Task	126
2.7	Summary of Results and Validation of Hypotheses	129
2.9	Probit Regression for the Draw Task	139
2.9	Probit Regression for the Draw Task (continued)	140
2.9	Probit Regression for the Draw Task (continued)	141
2.10	Probit Regression for the Alternative Uses Task in Expansivity	142
2.10	Probit Regression for the Alternative Uses Task in Expansivity (continued)	143
2.10	Probit Regression for the Alternative Uses Task in Expansivity (continued)	144
2.11	Probit Regression for the Alternative Uses Task in Persistency	145
2.11	Probit Regression for the Alternative Uses Task in Persistency (continued)	146
2.11	Probit Regression for the Alternative Uses Task in Persistency (continued)	147
2.12	Ordered Logit Regression for the Tangram Task	148
2.12	Ordered Logit Regression for the Tangram Task (continued)	149
2.12	Ordered Logit Regression for the Tangram Task (continued)	150
3.1	Subjects Distribution by Specific Background	164
3.2	Summary of the Descriptive Statistics	165
3.3	Payment Scheme	165

3.4	External Evaluators' Profiles	168
3.5	Summary Criteria and Scale of Evaluation by Task	170
3.6	Summary Results of the Tobit Regressions	171
3.8	Tobit regression for Feasibility	194
3.8	Tobit regression for Feasibility (continued)	195
3.9	Tobit regression for Originality	196
3.9	Tobit regression for Originality (continued)	197
4.1	Descriptive Statistics of Experimental Sessions	211
4.2	External Evaluators' Profiles	215
4.3	Summary of the Descriptive Statistics	218
4.4	Balance Tests	219
4.5	Tobit Regression for the Effect of Treatment on Feasibility	224
4.5	Tobit Regression for the Effect of Treatment on Feasibility (continued)	225
4.5	Tobit Regression for the Effect of Treatment on Feasibility (continued)	226
4.6	Tobit Regression for the Effect of Treatment on Originality	227
4.6	Tobit Regression for the Effect of Treatment on Originality (continued)	228
4.6	Tobit Regression for the Effect of Treatment on Originality (continued)	229
4.7	Tobit Regression for Feasibility and Originality (Experience of Surprise)	230
5.1	Tasks' Characteristics	257
5.2	Number of Creative Outputs Generated and Evaluated	259
5.3	Creativity Criteria Assessed by Evaluators	260
5.4	Evaluators' Socio-Demographic Characteristics per Task	261
5.5	Comparison of Evaluators' Socio-Demographic Characteristics Across Tasks	261
5.6	Comparison of Agents' Performance for the Text task	268
5.7	Polynomial Logit Regression for the Text task	268
5.8	Comparison of Agents' Performance for the Alternative Uses Task . .	271
5.9	Polynomial Logit Regression for the Alternative Uses Task	271
5.10	Comparison of Agents' Performance for the Draw Task	273
5.11	Polynomial Logit Regression for the Draw Task	273
5.12	Summary of Regression Results	275
5.13	Text Task Descriptive Statistics	295
5.14	Alternative Uses Task Descriptive Statistics	295
5.15	Draw Task Descriptive Statistics	296

5.16	Comparison of Agents' Performance for the Text Task (non-parametric)	296
5.17	Comparison of Agents' Performance for the Alternative Uses Task (non-parametric)	297
5.18	Comparison of Agents' Performance for the Draw Task (non-parametric)	297
5.19	Text Task PCA – All Creativity Criteria	298
5.20	Alternative Uses Task PCA – All Creativity Criteria	299
5.21	Draw Task PCA – All Creativity Criteria	300
5.22	Socio-Demographic Variables Descriptive Statistics	301
5.23	Comparison of Evaluators' Socio-Demographic Characteristics Across Tasks	302
5.24	OLS – Average Score by Evaluators for the Text Task	303
5.25	Polynomial Logit – Evaluators Profiles for the Text Task	304
5.26	OLS – Average Score by Evaluators for the Alternative Uses Task . .	305
5.27	Polynomial Logit – Evaluators Profiles for the Alternative Uses Task	306
5.28	OLS – Average Score by Evaluators for the Draw Task	307
5.29	Polynomial Logit – Evaluators Profiles for the Draw Task	308

Anne-Gaëlle MALTESE

ESSAYS ON INDIVIDUAL, TEAM, AND ARTIFICIAL CREATIVITY:
AN EXPERIMENTAL APPROACH

RÉSUMÉ

Cette thèse étudie les déterminants de la créativité à travers les niveaux d'analyse individuel, d'équipe et artificiel. Le chapitre 1 souligne la nécessité pour l'économie de repenser les méthodologies expérimentales afin de mieux capturer les processus collectifs, en mettant l'accent sur l'importance des interactions directes entre les sujets. Le chapitre 2 explore la créativité individuelle, en analysant comment les profils des individus et le degré d'ouverture des tâches influencent les résultats créatifs. Le chapitre 3 examine l'interaction entre la créativité individuelle et celle d'équipe, en se concentrant sur l'impact de la diversité et des réseaux au sein des équipes, sur leur performance. Le chapitre 4 approfondit l'étude de la créativité d'équipe, notamment face à des événements inattendus et la manière dont les équipes s'adaptent aux surprises pour performer. Enfin, le chapitre 5 compare la créativité humaine et celle de l'IA, en évaluant leurs performances sur des tâches de différents degrés d'ouverture, tout en contrôlant les différentes stratégies de génération de requêtes.

Mots clés: Créativité individuelle; Créativité d'équipe; Créativité artificielle; Economie expérimentale; Expériences collectives

RÉSUMÉ EN ANGLAIS

This thesis investigates the determinants of creativity across individual, team, and artificial levels. Chapter 1 underlines the need for economics to rethink experimental methodologies to better capture collective processes, emphasizing the importance of direct interaction between subjects. Chapter 2 investigates individual creativity, analyzing how individuals' profiles and task openness shape creative outputs. Chapter 3 explores the interplay between individual and team creativity, focusing on how team diversity and networks influence performance. Chapter 4 further examines team creativity, particularly in response to unexpected events, and how teams adapt to surprise to perform. Finally, Chapter 5 compares human and AI creativity, evaluating their performance across tasks of varying openness and controlling for different prompting strategies.

Keywords: Individual creativity; Team creativity; Artificial creativity; Experimental economics; Collective experiments