BETA
Bureau
d'économie
théorique
et appliquée

Documents de travail

«Disentangling the role of surface and deep-level variables on individuals' and groups' creative performance: A cross-level experimental evidence»

Auteurs

Anne-Gaëlle Maltese, Sara Gil-Gallen and Patrick Llerena

Document de Travail n° 2023 – 19

Juin 2023

Bureau d'Économie Théorique et Appliquée
BETA

www.beta-umr7522.fr

@beta_economics

Contact :
jaoulgrammare@beta-cnrs.unistra.fr

Université de Strasbourg · cnrs · Université de Lorraine · INRAE · AgroParisTech

Disentangling the role of surface and deep-level variables on individuals' and groups' creative performance: A cross-level experimental evidence.

Anne-Gaëlle Maltese[†*], Sara Gil-Gallen[†], and Patrick Llerena[†]

[†]*BETA, University of Strasbourg*

Abstract

Our societies are based on the principle of heterogeneity of individuals who possess diverse multidimensional characteristics. Social interactions among those profiles significantly impact collective activities, including creative outcomes. There exists a growing literature studying the variables influencing both individual and collective creative performance, but due to the complexity of the phenomenon, the literature does not find a consensus on their impact. The novelty introduced by this paper is first to capture and disentangle the role of surface and deep-level diversity variables in individual and collective creative performance. Secondly, we run a collective experiment involving real social interactions, which is a dimension rarely captured in the experimental economics literature, to measure groups' creative performance and the creative process behind it. Finally, to the best of our knowledge, we are the first experimental paper disentangling the role of such variables within individual creative performance, considering both convergent and divergent thinking by introducing three different types of tasks: open, open with constraints, and closed. The results of our analysis concluded that exists a mixed pattern of the impact of surface and deep-level variables on the individual creative performance, knowing that it will differ according to the degree of openness and the criteria of creativity. The only factor that arose persistently across degrees of openness was the self-evaluation of the performance in the task, which positively relates to creative performance (open and closed), while open with constraints is detrimental for subjects who self-evaluate their performance better than others. At the collective level, we observe different types of results depending on the evaluation criteria by means of feasibility fostered by homogeneous female groups and instead originality by half male and half female. Moreover, we also observe the implications of individual training, driven by subjects' from programs with the formation in creativity being detrimental for feasibility but instead increasing originality in collective creative performance. To be noted, further improvements in this work have to be expected, and we invite the reader to refer to the last section of this paper for more information.

Keywords: Creativity, Diversity, Collective, Experiment.
JEL Codes: C91, C92, O31.

[*]Contact: ag.maltese@unistra.fr

1 Introduction

The process of creativity is the stepping stone for innovation development, and the idea-generation process, which is central to creativity, strongly relies on the individuals involved in it. Capturing individuals' profiles becomes essential to understand fully how individuals can be creative and, more precisely, to understand the interplay between their characteristics and the creative process. Moreover, we observe a growing collective nature of creative activities. Those interactions between individuals to produce a common output induce the critical role of group composition in the creative process. This group composition has different impacts on the creative process and the creative performance through various channels such as group atmosphere, decisions making, or even conflict management. In this work, to understand the impact of group composition on creativity at the collective level, we chose to focus on the concept of diversity, which will be defined later. In fact, the study of diversity has become an important aspect of the literature on creativity because it has emerged as a source of competitive advantage and is established as a multilevel phenomenon that may boost creativity under certain conditions.

This paper's main contribution lies in identifying participants' profiles, their related individual creative performance, and the interactions of those in the collective performance. In fact, this paper aims to analyze the cross-level relationship between creativity and *surface* and *deep-level* diversity variables at the individual and collective levels. Individually, to the best of our knowledge, we are the first paper disentangling the role of *surface* and *deep-level* variables within individual creative performance, considering the three different levels of openness [see Attanasi et al., 2021]: *open*, *open with constraints*, and *close*. In addition, collectively, we focus on the diversity-creativity relationship by focusing on the group gender composition at the *surface-level* as well as the perceived group experience at the *deep-level*. Finally, we implemented a collective experiment (defined later in Section 2.4) involving face-to-face interactions, which is rarely addressed in the literature on experimental economics and allows us to study the collective process.

Thus, we aim to address the following research questions: how do *surface* and *deep-level* variables impact individuals' creative performance regarding the degree of openness of the task? And, to what extent will groups' diversity, in terms of *surface* and *deep-level* variables, and group members' individual creative performance impact the group's creative performance as a whole?

The results of our analysis concluded that exists a mixed pattern of the impact of *surface* and *deep-level* variables on the individual creative performance, knowing that it will differ according to the degree of openness of the task and even the criteria of creativity selected. The only factor that arose persistently across degrees of openness and criteria of creativity was the self-evaluation of the performance in the task, which promotes individual creative performance for an open and closed task, but self-perception of performing better than others is decreasing creativity in an open with constraints task. At the collective level, we find different results according to the creativity criteria we focus on (originality or feasibility) for the gender composition of the groups. Indeed, heterogeneity in the gender of group members (half split) fosters originality, whereas feasibility increases in women's homogeneous groups. However, programs with the training of creativity

foster originality but are detrimental to the feasibility.

The remainder of the paper is organized as follows: Section 2 outlines the state of the art, introducing the relevant literature and key concepts presented in this work. Subsequently, Section 3 describes the experimental design and procedure. Section 4 presents the measure of creativity (our dependent variable), and illustrates the results for both the individual and collective analysis. While, Section 5 displays the discussion, and finally, Section 6 depicts the conclusion and the future improvements.

2 State of art

2.1 Individual and collective creativity

Numerous disciplines, such as psychology, management, sociology, and neuroscience, have focused on the question of creativity. By dint of this multidisciplinary, extensive literature exists that sets the bases of the topic of creativity and aims to define it. Economics made no exception, and various authors have tried to define creativity for many decades [Poincaré, 1908, Mednick, 1962, Osche, 1990]. The definition we selected comes from Teresa Amabile, who is considered the reference author on the topic of creativity across disciplines, and is the following: "*the production of novel and useful ideas by an individual or small group working together*" [Amabile and Pratt, 2016, p.158]. Beyond the first part of this definition, where creativity relies on novelty and new combinations of knowledge, the second part highlights that creativity can be both an individual and collective process, as in our study. As illustrated by Fischer et al. [2005] as well, "*most scientific and artistic innovations emerge from joint thinking, passionate conversations and shared struggles among different people, emphasizing the importance of the social dimension of creativity.*" [p. 483]. Including this social and collective dimension in the conceptualization of creativity implies taking into account the bidirectional dynamic between the individuals and their social environment and how some factors can act positively or negatively on creativity [Amabile, 1982, 2012, Amabile and Pratt, 2016]; the former influencing the latter and conversely. As well, Csíkszentmihályi [1996] defined collective creativity as the product of the interaction of individuals and not only the summation of individuals' ideas. In this study, we will not only focus on individual or collective creativity but study both of them as well as their relationship. Moreover, as a consequence of creativity's broadness, we chose to focus on the notion of *surface* and *deep-level* diversity variables and the interaction of different profiles based on the previous factors to understand the mechanisms that will impact respectively individual and collective creativity. The following paragraph will focus on the notion of diversity, its definition and implications, and how we implement the same categorization for individual variables.

2.2 Diversity as a key factor of creativity

Heterogeneity, multiplicity, and variety are different terms to discuss diversity in the literature. If everybody has a clear idea of what diversity is, finding its unique and consensual definition in the

literature is quite difficult. In this paper, we rely on the definition provided by van Knippenberg et al. [2004], who sees diversity as "*differences between individuals on any attribute that may lead to the perception that another person is different from self*" (p. 1008). When someone needs to collaborate with other individuals on a specific idea, task, or project, she or he needs 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 defended the benefits of diverse groups [Jackson et al., 1991, Amabile et al., 1994, Phillips and O'Reilly, 1998, Kavadias and Sommer, 2009, Dutcher and Rodet, 2022]. 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 has a positive or negative impact on groups' performance and functioning because such results differ as long as we modify the type 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 chose to distinguish the *surface* from *deep-level diversity*. On the one hand, at the *surface-level* of diversity, differences respect three principles: *immutability*, *immediate observation*, and *simple and valid measurement* [Jackson and Wolinsky, 1996, Milliken and Martins, 1996]. The variables that we could consider as *surface* variables can be: 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]. 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]. Those characteristics can include skills, personality traits, moods, or even life experiences. If so, *deep-level diversity* can not verify the three concepts of *immutability*, *immediate observation*, or *simple and valid measurement*, and its measures become more challenging due to the complexity of collecting such *deep-level* information on individuals. However, even if we distinguish the *surface* from *deep-level diversity*, it does not mean that these two categories should be analyzed separately. In fact, both are interrelated, and their conjoint analysis should not suffer from the fact that they need to be measured through different channels. For us, studying one without the other would always imply losing one side of the story each time. On the individual side of creativity, we will use the same intuition and terminology of *surface* and *deep-level* characteristics and categorize them as *surface* and *deep-level* variables relative to subjects' profiles.

As mentioned above, our work is part of already dense literature and previous studies focusing non-exhaustively on age, gender, personality traits, or cognitive style effects on creativity at the individual level. According to the survey by Attanasi et al. [2021], there are no conclusive results about the impact of these variables on creative performance in the literature. While in the review by

Hundschell et al. [2021] on the relationships between diversity and creativity, the authors consider a multilevel effect at individual, collective, and organizational levels. The authors conclude, among other caveats, the need to advance the study of the interplay between different levels of diversity, such as individual and collective, on creative performance. Our contribution aims at filling this gap by using the experimental methodology and, more specifically, with the introduction of a collective experiment.

2.3 Measure and evaluation of creativity

To measure and evaluate creativity, we need to answer three main questions; (1) What type of creative output are we considering? (2) How do we elicit the creativity assessment? (3) Who is evaluating the creative output?

Creative outputs might differ first regarding the type of thinking process activated in a specific task. We can distinguish two main thinking processes that are the convergent thinking process and the divergent thinking process. The first can be defined as a form of problem-solving based on the generation of one specific logical and unique answer, whereas the second can be defined as a form of problem-solving based on the generation of logical and various solutions [Guilford, 2017]. Thus, one creative task will not always activate the same thinking process, and this difference between tasks allows us to categorize them. Based on the literature, we assume that creative tasks differ in terms of their degree of openness. Indeed, the spectrum of openness of tasks is characterized by the level of autonomy left to individuals to solve the tasks. Beyond the traditional dichotomy between *closed* and *open* tasks [Charness and Grieco, 2013, 2019] exists a third type of task entitled *open with constraints* task [Attanasi et al., 2021]. The final categorization that we support in this work is the following: a *closed* task that is defined as having only two possible scenarios, solving the problem or not, and this is activating a convergent thinking process. Then, an *open* task activates a divergent thinking process and does not limit the number of possible answers. Finally, an *open with constraints* task still activates a divergent thinking process while including some constraints inducing creative tasks closer to real-life problems that one could encounter.

Then, creativity is multidimensional, and one can not simply define an output as creative with no further consideration. The first thing to consider is that the creativity assessment is based on an objective or subjective evaluation. To do so, we have to consider the degree of openness of tasks as stated above. On the one hand, a *closed* task that activates a convergent thinking process and thus has only one possible true solution is based on an objective evaluation that does not require specific evaluation guidelines. On the other hand, *open* and *open with constraints* tasks that activate a divergent thinking process are based on a subjective evaluation and require either a taxonomy or score approach. To define what should be chosen between a taxonomy or a score, we have to decide if the creativity assessment is based on more than one creativity criterion. Guilford [1950] stated that creativity can be assessed according to three main criteria: originality (*how infrequent a particular solution is*), fluidity (*how many ideas were generated to solve a specific problem*), and flexibility (*how many themes cover the set of ideas generated*). Those three criteria, widely used in

the literature, will be key in this work but will be completed by two other criteria that decompose the *originality* criteria: *expansivity*, which is defined as "*the ability to provide solutions outside the fixation effect*" [Camarda et al., 2017] in comparison with all ideas generated by the entire sample and *persistency* which we can define as the ability to provide solutions outside the participant's own fixation effect[1]. The taxonomy allows isolating of different aspects of creativity. Nevertheless, we can instead be interested to follow a general evaluation of creativity by introducing a score (usually from 1 to 10), which follows *Amabile's consensual assessment technique*, according to which raters should be capable of independently agreeing upon a creativity judgment, without defining ultimate objective criteria for creativity. Therefore, the choice between a score and a taxonomy mainly relies on the type of creative task.

Finally, we had the define the type of task and the different criteria of creativity, but it is missing a critical point, who/how is in charge of asses creativity. As mentioned above, an objective creativity assessment based on a convergent thinking process does not require specific guidelines and, thus, does not require specific judges. On the contrary, a subjective creativity assessment will require the implication of judges. Here we have two main possibilities in the literature; an evaluation assessed by peers (participants in the same experiment) or by external judges. In this work, we based the subjective creativity assessment of our tasks on external judges that, we assume, provide more accurate evaluations being blind to the treatment.

2.4 Collective experiments

Doing collective experiments is unusual. If some experiments do involve interactions between participants, it is mainly through a computer interface and, most of the time, anonymously. This corresponds to the established literature on the good practices of experimental economics as in Jacquemet et al. [2019], who promoted that the application of the following four basic principles ensures the proper conduct (to have a controlled environment) of the experimental method: (1) minimize the likelihood that participants know each other (2) prevent communication (3) isolate participants and (4) prevent identification. Part of our contribution in this paper is to demonstrate that it is possible to run collective experiments with real social interactions in a controlled environment (by controlling all those factors suggested by Jacquemet et al. [2019]) and provide significant accurate results at the same time.

In light of those considerations, we define collective experiment in our work as *"an experiment enabling to observe group processes by means of the social interactions of subjects without systematic intermediary"*. Thus, we can already warn readers that in our protocol, we bypass the four principles presented above, but we control all of them. Indeed, as it will be presented in the experimental design (section 3), we try to control for any weakness that could appear due to the non-minimization of the likelihood that participants know each other, their interactions, their non-isolation, and non-anonymization. Because working on collective processes and, in our case, collective creativity implies integrating social interactions and wanting to encourage them. Thanks to this, we are able

[1] *Persistency* is computed as the ratio between *fluidity* and *flexibility*

to study collective processes in a scenario closer to real life, which is slightly being explored by the literature and has enormous potential.

3 Experimental design

3.1 Timeline of the experiments

Our experimental protocol is composed of three different parts extended in time[2]. To be noticed, when one participant could not attend one of the sessions, she or he was removed from our sample. The following paragraphs describe in detail those three different sessions, and Figure1 schematically represents the overall protocol.

3.1.1 Preliminary Questionnaire

The first session was a pre-questionnaire on the computer that took place in the lab, which started with written instructions also read by the experimenter. The in-lab version was chosen to make sure that participants can not communicate with each other and are only focusing on the questionnaire.

The aim of this first broad pre-questionnaire is to collect various information on our participants in terms of *surface* and *deep-level* variables to provide us with a complete profile for each one. The pre-questionnaire consists of 66 questions that can be broken down into four main themes. First, we collect sociodemographic factors. Second, adapted and translated from Attanasi et al. [2019], questions on creativity and social habits are asked. Third, we use the "Diversity of Life Experiences" questionnaire [Douthitt et al., 1999] and adapt it to the French/European context. Finally, we ask participants about their domains of expertise, as presented in Carson et al. [2005] and Dutcher and Rodet [2022].

3.1.2 Individual experiment

The individual experiment is a lab experiment performed individually and composed of three different tasks aimed at obtaining the reference point in terms of the creative performance of each participant.

Those three distinct tasks differ in their degree of openness [Attanasi et al., 2021] and are the following[3] :

1. An *open task* where participants are asked to draw an alien animal from a planet different from Earth [Ward, 1994] (from now on, Draw task).

[2]To avoid any order effect, participants did not always follow the same experimental scheme. Half of the participants did the individual experiment before the collective one, and conversely. Only the first part, the preliminary questionnaire, stayed at the same spot for everyone because we needed this first collection of data to build our two populations, thus, the individual profiles.

[3]To avoid any order effect, half of the participants completed the tasks from the less opened to the most and the others from the most to the less.

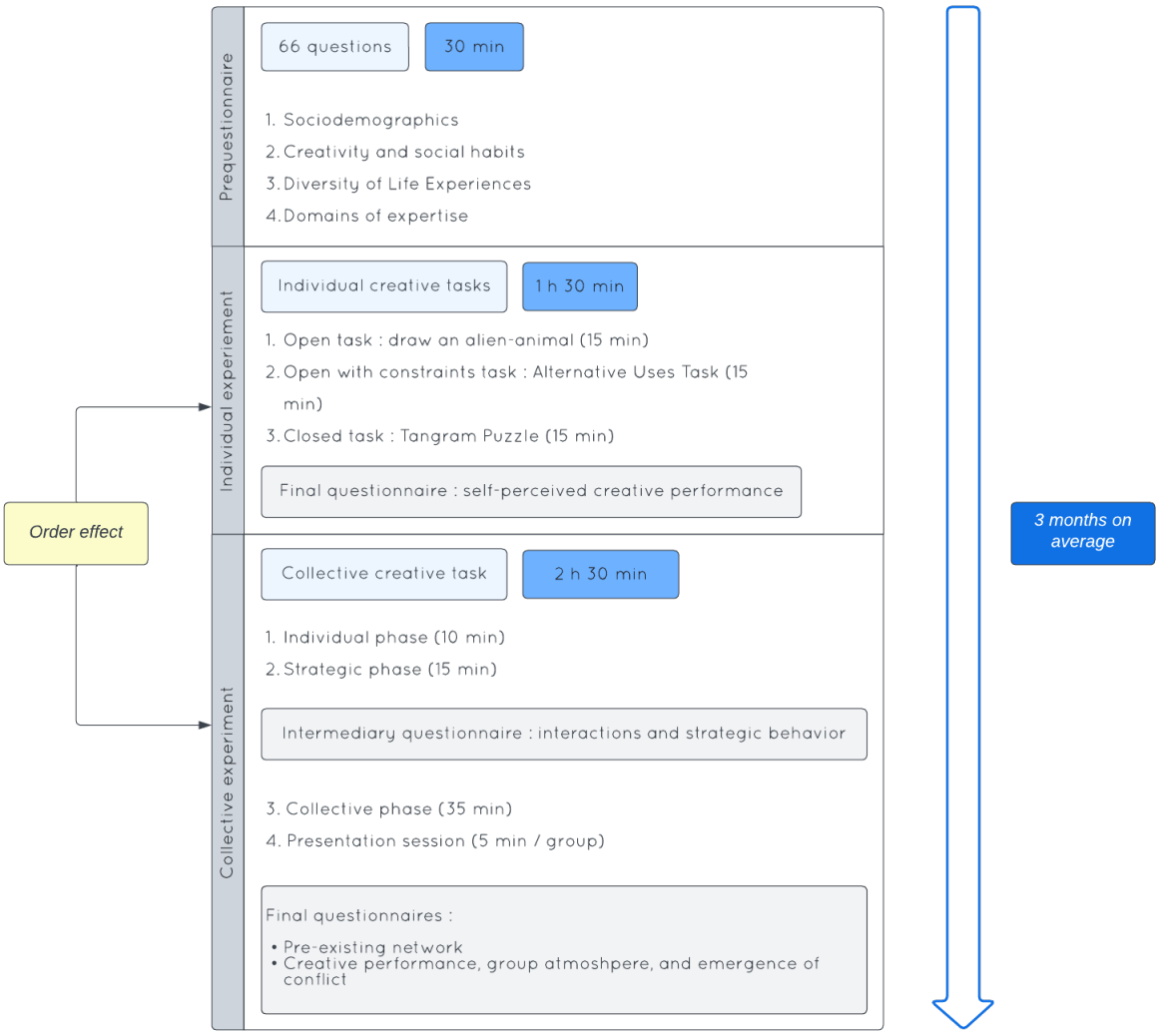

Figure 1: Protocol timeline

2. An *open with constraints task* is known as the *Alternative Uses Task* (from now on, AUT task), where participants have to find unusual uses for everyday objects [Guilford, 1967, Dutcher and Rodet, 2022].

3. A *closed task* where participants have to solve a Tangram puzzle [Ariely et al., 2009] (from now on, Tangram task).

Once the subjects have performed the three tasks, they are asked to fill out a final questionnaire about their perceived own creative performance. First, we ask them about their general self-perceived creative score for on a scale from 1 to 5. Then, for each task, we ask them (1) what difficulties they encountered during this task, (2) their self-perceived creative score for this task on a scale from 1 to 5, and (3) if they think they performed better than others on average.

3.1.3 Collective experiment

The collective experiment aims to evaluate the collective creative performance of groups of three to five participants in an open creativity task. We ask participants to elaborate and present an idea or a project that could improve their daily university life [Baruah and Paulus, 2011, 2016, Kohn and Smith, 2011, Kohn et al., 2011, Harvey, 2013]. The only instruction they received was that, in the end, the most creative idea would be the one that is the furthest from what is already done in their university but also furthest from other groups' ideas. The experiment is composed of four steps:

1. An *individual phase* during which participants are asked to think about the problem by themselves. We provide participants with scrap paper to keep track of their ideas, and all participants sit on their own with no possibility of talking to each other. This phase lasted 10 minutes, and the configuration of the room was as displayed in *Layout A* of Figure 2.

2. A *strategic phase* where participants are allowed to talk to other participants under certain conditions. Each participant receives 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, participants can choose to refuse to receive a token (as a refusal to give information) and/or to spend only a part or all of their tokens or none. This part lasted 15 minutes, and the configuration of the room was as displayed in *Layout B* of Figure 2.

3. A *collective phase* with groups of three to five participants (depending on the size of the session[4]). During this step, each group has its own area with a table, chairs, and a whiteboard. They are allowed to move and organize themselves as they want in this area. The aim of this step is to allow group members to exchange their ideas and converge toward one. It is also mentioned in the instructions that each group has to prepare a short presentation after converging on a specific idea. This part lasted 35 minutes, and the configuration of the room was as displayed in *Layout C* of Figure 2.

4. A *presentation session* as a 5-minutes speech exercise. Each group is completely free on the presentation besides the time constraint. This part lasts 5 minutes per group, and the configuration of the room is as displayed in *Layout C* of Figure 2.

During this experiment, participants receive three different questionnaires. The first one is an intermediary questionnaire given at the end of the *strategic phase* as a control of the tokens they received and/or kept and an explanation of their strategy, asking to whom they talked and why. Then, at the end of this experiment, the second questionnaire aims to explore, through 27 questions, participants' self-evaluation of creativity, work-group experience, conflicts, and daily reactions when

[4]We have a total of 3 groups of 3 group-members; 11 groups of 4 group-members, and 5 groups of 5 group-members.

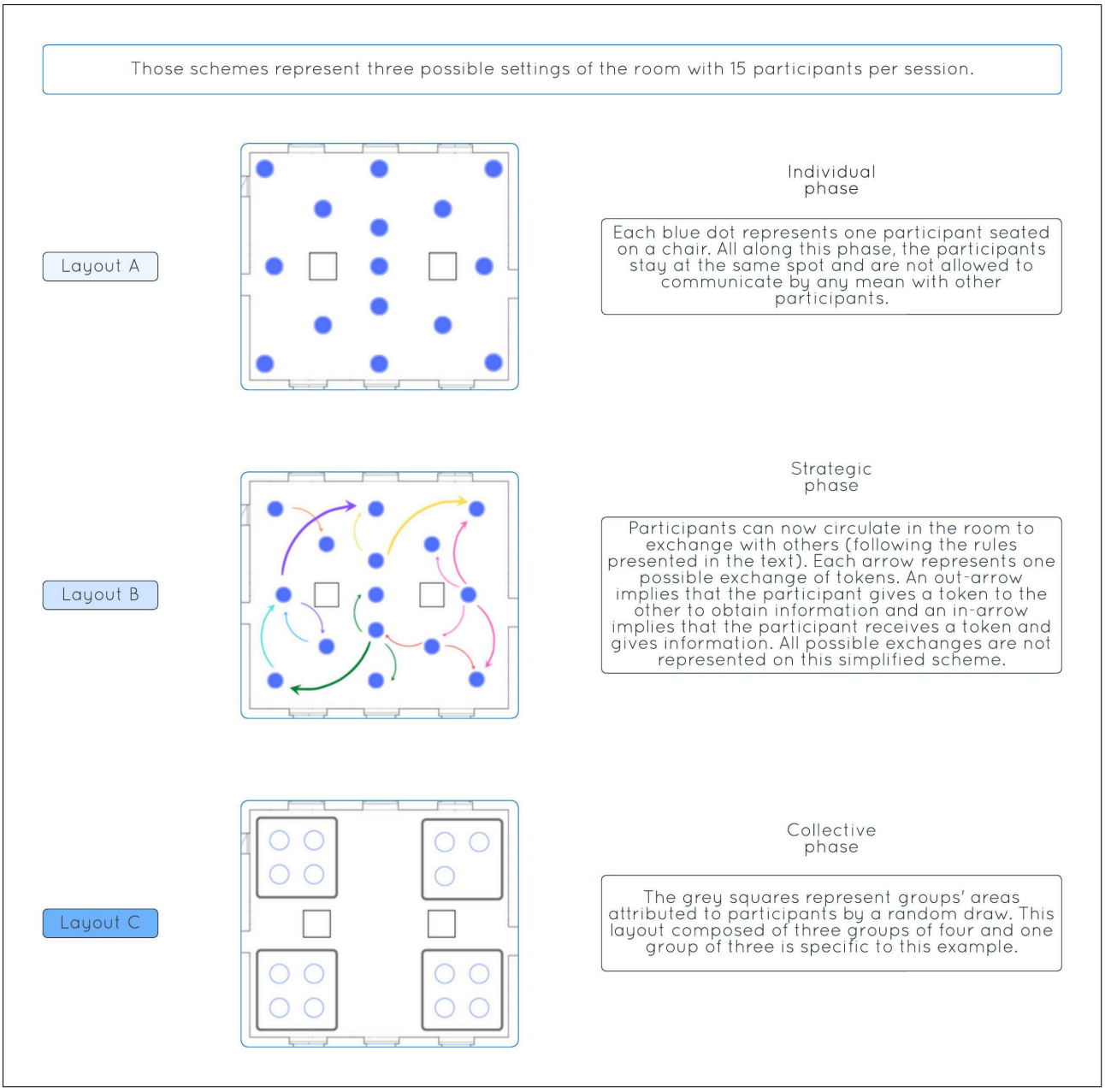

Figure 2: Experimental layouts

facing negative or uncomfortable events. Finally, the third *pre-existing network* questionnaire allows us to understand what are the pre-existing links and relationships between participants in their respective experimental sessions. We put the name and photo of each participant in the session, and for each one of them, they answered whether they knew this person before this experiment and, if so, what was the type of their relationship (friendship, professional, or both and to score it

from 1 to 5 each time) and how frequently they saw or talked to each other.[5]

To conclude on the timeline of this protocol, it has to be noted that three months on average separated the first stages to the third to diminish as much as possible the effect that one can have on the other.

3.2 Experimental procedures

We recruited 74 participants from the University of Strasbourg (France). We chose three different categories of students depending on their background specificities in terms of creativity. First, we selected a large sample of 55 master students from *economics* because they do not have any specific creativity skills acquired, at least within their program. Our second sample of 12 students was constructed by merging two groups of master students in *innovation management* similar in terms of their training in creativity[6]. Finally, we recruited 7 participants from another program designed for *students-entrepreneurs*. Table 1 presents the characteristics of the programs from which the participants come, and as a matter of population similarities, we do not find any significant difference in terms of gender (p=0.5107; Two-sample Wilcoxon rank-sum (Mann-Whitney) test).

Table 1: Participants distribution and specific background

Programs	*N*	*Creativity*	*Entrepreneur*	*Economics*
Economics	55			X
Innovation management	12	X		
Student-entrepreneurs	7		X	
Percentage of women	48.6	41.7	42.9	50.9

The experimental sessions took place at the University of Strasbourg between April and December 2021. All the experiments have been carried out in French[7] and implemented in the *Creativ'lab*, a creativity room of nearly 118m^2 conceived to observe collective processes thanks to a Tacking system which allows us to capture the movement of the participants.

Moreover, the number of sessions differs depending on the stage of the experiment. There are 3 sessions for the pre-questionnaire divided by the three programs. While for the individual experiments, we have 4 sessions with 20 participants, three of them and 14 in the other one. Finally, for the collective experiment, we ran 7 sessions, which total distribution in groups is the following: 3 groups of 3 group members; 11 groups of 4 group members, and 5 groups of 5 group members.

[5]This final questionnaire allowed us to control for the deviation from the principle of non-minimization of the likelihood that participants know each other we mentioned earlier.

[6]To be precise, the training in creativity refers to training in creativity techniques and not in the sense of artistic training.

[7]Because not all participants are native French speakers, we controlled for their native language.

The payment implemented is a flat payment by means of paying a fixed amount instead of based on the creative outcome by a performance-based or competition. We select this incentive scheme taking into account the different types of tasks and the literature about the effect of incentives depending on the degree of openness, which results are mixed. For the *closed* tasks, the performance-based payment can be beneficial, while for *open* tasks present no effect or even a negative effect [Charness and Grieco, 2019]. Moreover, Amabile et al. [1996] mentions a *crowding out effect* of monetary incentives that diminish the intrinsic motivation of individuals and, thus, negatively impact their creative performance. Supporting our design, Eckartz et al. [2012] concludes that there is no significant effect of performance-based payment in confront to flat payment on participants' creative outcomes. For the sake of consistency, we have chosen to pay for all our tasks in the same way, avoiding any monetary effect on the performance of our participants. The payment scheme for this experimental protocol is stated in Table 2.

Table 2: Payment scheme

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

3.3 Hypotheses

In this section, we present the four hypotheses formulated as follows. First, as was mentioned earlier, the degree of openness of a task indicates what type of thinking process is activated [Attanasi et al., 2021]. We assume that having differences in terms of the thinking process will imply different performances from participants. No participant would ultimately succeed in each task or completely fail in each task. Depending on each participant's profile and characteristics at both the surface and deep-level of analysis, his or her performance in a specific task will vary. This leads us to the following hypothesis :

H1: *The individual performance and the surface and deep-level variables impacting it will differ according to the level of openness of the task.*

By establishing each participant's profile, we obtain a set of information as complete as possible, including their proximity to creative activities. We consider that having been trained in creativity or practicing creative activities will have a positive impact on participants' creative performance at both the individual and collective tasks. Then, our second hypothesis is the following :

H2: *Participants with creativity and entrepreneurial training background will perform better*

individually and as a group.

Then, the literature on diversity and its impact on creative performance at the collective level presents mixed results[Attanasi et al., 2021, Hundschell et al., 2021]. If diversity can be seen as beneficial, it seems that diversity can also increase the emergence of problems that would brake the production of creative outputs. Hence we are able to state the third hypothesis:

H3: *A higher level of diversity in a group, either at the surface or deep-level, will imply a greater performance in the collective creative task.*

Finally, we assume that a group's creative performance also relies on the creative performance of each participant individually. Thus, the individual level of creativity of each group member fosters group creativity. Therefore, we conclude with the fourth hypothesis:

H4: *The greater the level of individual creative performance of each one of the group members, the greater the group's creative performance as a whole will be.*

4 Results

In this section, we present the results of our analysis, which is organized as follows. First, in section 4.1, we define our dependent variables, hence the measure of creativity differentiated by both the type of task and the creativity criteria selected. Then, we test the role of *surface-level* diversity variables in Section 4.2 and the *deep-level* variables in Section 4.3 on individual creativity. We decomposed the study of deep-level diversity variables into two dimensions: "Diversity of life experiences" in section 4.3.1 and personality traits in section 4.3.2. While in section 4.4, we address the effect of the interplay of diversity on collective creative performance. Moreover, we investigate if the profiles from individual creative performances, group atmosphere, and group conflict influence the collective creative outcome. Finally, in Section 4.5, we provide an exhaustive regression analysis to test our hypotheses and confirm the results arising from our descriptive analysis.

4.1 Evaluation of creativity

This section aims to define our dependent variables, hence, we will measure creativity for all three individual tasks (Tangram, AUT, and Draw tasks) and the collective task. First, the *Tangram task* is a *closed* task that allows eliciting convergent thinking given that the subjects overcome the task only by solving the puzzle. However, we can also compare the performance by considering the time spent to solve the task. In this line, we consider when subjects solve the task in less time than average (39.19%), equal or over average time to solve the task(32.43%), and fail to solve the task (28.38%).

While in the *AUT task*, we are dealing with an *open with constraints* task, where divergent thinking with significant restrictions is activated. Subjective assessment is implemented by "blind" external judges whose characteristics are displayed in Table 3. Then, the judges had to select the number of valid ideas for every five items of the task of each participant. From those valid ideas, they had to determine the themes related to these ideas. We did not provide them with an exhaustive list of themes, and they had to determine them by themselves[8] . Finally, they evaluated on a scale from 0 to 10 the level of elaboration of each participant. To be noted, the results were presented to the judges in different orderings to control for potential order effects.

Table 3: External judges' profiles

Judge	*Gender*	*Position*	*Domain*
1	M	PhD Student	Environmental economics, Growth theory
2	F	Associate Professor	International innovation management, Values-based innovation management
3	M	CEO	Research and development, Health
4	F	PhD Student	Labour market, Poverty, Unemployment benefits

From the judges' evaluation, we obtain our multi-dimensional measure of creativity based on the AUT task literature [Bradler, 2015, Bradler et al., 2016, Dutcher and Rodet, 2022]. The five dimensions of creativity construct in the literature are the following: *fluidity* (number of ideas), *flexibility* (number of themes), *expansivity* (frequencies)[9], *persistency* (Number of ideas/Number of themes), and *elaboration* (the details introduced in the explanation of the unusual use). We selected to describe subjects' creativity in AUT task through expansivity and persistency, knowing that their Cronbach's alphas are acceptable and reliable [an alpha reliable from 0.6-0.7 and over 0.7 acceptable and reliable, see Ursachi et al., 2015], as we report in Table 4 and Table 5, respectively[10].

Table 4: Cronbach's alpha test and related statistics for Expansivity in the AUT task

Judges	*Obs*	*item-test correlation*	*item-rest correlation*	*average inter-item covariance*	*alpha*
Judge 1	355	0.8806	0.7493	0.0011	0.7854
Judge 2	350	0.8180	0.6456	0.0012	0.8310
Judge 3	347	0.8429	0.7222	0.0012	0.8034
Judge 4	356	0.8237	0.7020	0.0013	0.8128
Test scale				0.0012	**0.8493**

[8]The final elicitation of themes and cleaning of judges' scoring is presented in the Appendix 7.1.

[9]Expansivity is "the ability to provide solution outside the fixation effect" (Camarada et al., 2017). To measure the degree of expansivity of participants' creative solutions, we use the frequency of their ideas in the entire sample.

[10]Fluidity and flexibility are integrated into the computation of the persistency indicator, and elaboration was not acceptable and reliable measure with a Cronbach's alpha of 0.4963.

Table 5: Cronbach's alpha test and related statistics for Persistency in the AUT task

Judges	*Obs*	*item-test correlation*	*item-rest correlation*	*average inter-item covariance*	*alpha*
Judge 1	370	0.8315	0.7035	0.2164	0.8488
Judge 2	370	0.8526	0.7151	0.1975	0.8466
Judge 3	370	0.8652	0.7437	0.1957	0.8332
Judge 4	370	0.8668	0.7705	0.2109	0.8263
Test scale				0.2051	**0.8739**

For the *Draw task*, we rely on Amabile [1982] *consensual assessment technique*, according to which participants are considered creative (or not) when all external judges rate them independently and agree upon their evaluation. The judges are the same ones who evaluated the AUT task (see Table 3). 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. Judges are supposed to have an acceptable experience in the target domain. They should use their own subjective judgments of creativity to rate the creative outputs independently, and from the evaluation of the four judges, we need to obtain acceptable inter-judge reliability as before, measured by Cronbach's alpha [Ursachi et al., 2015]. In this line, the drawings of our experiment were evaluated by the four judges on a scale from 0 (minimum) to 10 (maximum). Therefore, creativity was evaluated in ascending order, i.e., the higher the level of creativity expressed in the judge's opinion, the higher the grade. Moreover, they were presented to the judges in different orderings to control for potential order effects. In Table 6, we present the results of our Cronbach's alpha test, where the alpha was equal to 0.6567; hence we conclude that it is a reliable measure of creativity [Ursachi et al., 2015].

Table 6: Cronbach's alpha test and related statistics for Draw task

Judges	*Obs*	*item-test correlation*	*item-rest correlation*	*average interitem covariance*	*alpha*
Judge 1	74	0.6977	0.4760	1.6106	0.5701
Judge 2	74	0.7578	0.4961	1.3125	0.5458
Judge 3	74	0.8288	0.6143	0.9580	0.4475
Judge 4	74	0.5041	0.1983	2.4108	0.7292
Test scale				1.5730	**0.6567**

Finally, regarding the *Collective task*, which is an open task, the creativity assessment implemented is a subjective assessment following a taxonomy attributed to the three following criteria: originality, elaboration, and feasibility[11]. Accordingly, judges indicated a score from 0 to 10 (where 0 is the minimum value, and 10 is the maximum) to evaluate the creative outcome, in this case, the

[11]It refers to how implementable those ideas are.

videos of the presentations of the groups. The Cronbach's alpha differs depending on the measure as continuation: a reliable value equal to 0.6774 for feasibility, a non-reliable value equal to 0.5376 for elaboration, and a reliable and acceptable value equal to 0.8022 for originality. Therefore, we consider the analysis to measure collective creativity feasibility and originality.

4.2 Surface variables and individual creativity

The sociodemographics variables are *surface-level* variables that, as put forward in the literature [Attanasi et al., 2019], can have various influences on creative performance. In this line, this section depicts the descriptive analysis of the role of *surface-level* diversity variables on individual creativity[12] for the three tasks: *Tangram* (closed), *AUT* (open with constraints), and *Draw* (open) tasks.

Firstly, in the *Tangram task*, we do not find any significant effect from the *surface* variables. While for the *AUT task*, the effect differs between the two measures of creativity. On one side, we do not find a significant effect of any surface diversity variable on expansivity. On the other side, we observe that French speakers (evidence) and French citizens (mild evidence) demonstrate lower levels of persistency. We suggest is driven by the fact that the natives have a greater vocabulary to interrelate ideas, which arise the internal fixation effect. Finally, we devoted our attention to the *Draw task*. We discover mild evidence about subjects who speak fewer languages (apart from French) and those whose domain is technology have lower scores on their drawing.

Thus, we can conclude that there is no significant systemic pattern of socio-demographic variables influencing individual creative performance. Actually, the significant variables are totally different among tasks, by means of the degree of openness, and even among the different criteria to assess creativity for the same task. This leads us to the first result:

R1: *Surface level variables' effect on individual creativity differs depending on the degree of openness of the task and the criteria of evaluation.*

- *Tangram task: We do not observe any significant effect of surface variables on creative performance in the closed task.*
- *AUT task: It is a matter only of persistency, which is decreased by being French and inline, a French speaker.*
- *Draw task: We find a negative effect of having an academic background in Science and Technology and a positive one by the number of spoken languages.*

[12] To run the descriptive analysis along sections 4.2, 4.3, and 4.4, we implement a non-parametric Kruskal-Wallis equality-of-population rank tests. We apply the terminology indicated by Moffatt [2020], which is the following: if $p<0.10$, there is *mild evidence* of an effect; if $p<0.05$, there is *evidence*; if $p<0.01$, there is *strong evidence*.

4.3 Deep-level variables and individual creativity

Moreover, as we have already pointed out in section 2, variables at the *deep-level* can be one of the key drivers of creativity. As previously, we run a descriptive analysis for the three tasks and the variables that define *deep-level* diversity. The *deep-level* variables are separated into two dimensions: diversity of life experiences (subsection 4.3.1) and personality traits (subsection 4.3.2).

4.3.1 Diversity of life experiences and individual creativity

Firstly, we focus on the "Diversity of life experiences," borrowing the questionnaire from [Douthitt et al., 1999], which is a validated measure. From this comprehensive questionnaire, the author derived five different categories of life experiences: (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 [GRWOF]. Following those categories, we aim to study the role of diversity of life experiences on individual creativity by tasks and criteria to assess creativity.

In the *Tangram task*, we do not uncover any significant effect from our data. While for the *AUT task*, none of the diversity of life experiences items influences expansivity. Nevertheless, this differs for persistency. Indeed, we discover evidence of a negative effect by the fact of experiencing different interests, likes, and attitudes on persistency. We hypothesize that this result is led by the fact that having too much information can create a sort of confusion, leading to an increase in the internal fixation effect. On the contrary, in the *Draw task*, we present evidence that subjects with a better relationship with their parents/family seem to perform better. This leads us to our second result:

R2: *Deep-level variables', measured by diversity of life experiences, effect on individual creativity differs depending on the degree of openness of the task and also on the criteria of creativity we focus on.*

- *Tangram task: We do not observe any significant effect of the diversity of life experiences on creative performance in the closed task.*
- *AUT task: Persistency is negatively impacted by the DILA component, while we do not observe an effect on expansivity.*
- *Draw task: Only the RPFE component has a positive impact on subjects' performance.*

4.3.2 Personality traits

The second dimension of *deep-level* variables is personality traits. To address the relationship between personality traits and individual creativity, we can break it down into four categories based on the information collected in the several questionnaires: *self-reported creativity*, *open-mindedness*

and extroversion, *trust and risk preferences*, and *domain-related skills*. In this section, we aim to study each category's influence on the three tasks and criteria to assess creativity.

Beginning with the *Tangram task*, in the category of self-reported creativity, we find strong evidence of a positive effect of greater self-evaluation in general creativity and inside the Tangram task and thinking to perform better on average than others in the task (evidence). While in the category of open-mindedness and extroversion, we discover evidence that subjects who are more extroverted do not only tend to solve the task, even though under the average time. In contrast, we uncover mild evidence of negative effects on subjects who often do new things and that prefer to work in groups. Finally, in the category of domain-related skills, we find evidence of a positive effect of subjects having a hobby or/and talent on humor, inventions_technologies, and a bit weaker effect for visual arts (mild evidence).

Afterwards, we continue with the *AUT task*, which seems to be a matter mainly of self-evaluation of creativity and the category of domain-related skills, but the effect differs between the criteria to assess creativity. On the one hand, we find mild evidence of lower levels of expansivity for subjects who often go to the cinema/theatres and out in the evening. In contrast, we encounter a greater expansivity for subjects' tolerance of homosexuals. Furthermore, we discover mild evidence of a positive effect on expansivity from subjects' self-evaluation of creativity. On the other hand, we discover mild evidence of a negative effect on persistency from often going to bars or restaurants. Moreover, we have evidence of a greater level of persistency in subjects who have a hobby or/and talent in individual sports. Moreover, we depict a negative effect on persistency of the thinking that their creative performance was better than others.

Finally, in the *Draw task*, we obtain mild evidence of a positive effect on the creative performance from subjects who have a hobby or/and talent in visual arts, theatre/films, inventions and technologies, and an even stronger effect by finding evidence for the case of culinary art. Therefore, we can report our third result:

R3: *Deep-level variables', measured by personality traits, show that domain-related skills foster individual creativity. Even though the relevant variable differs by task and the criteria to assess creativity.*

- *Tangram task: Subjects' performance is positively impacted by self-evaluation, extroversion, and, considering domains-related skills, humor, inventions and technologies, and visual arts trigger convergent thinking.*
- *AUT task: On one side, expansivity is reduced by the social habit of going to the cinema and theatre or going out in the evening but is positively impacted by self-evaluation. On the other side, persistency decreased from going to bars/restaurants and thinking they perform better than others, but is promoted by domain-related skills such as individual sports.*
- *Draw task: The performance in the open task is elicited by domain-related skills in terms of visual arts, theatres and films, and invention and technologies.*

4.4 Diversity and Collective creativity

In this section, we depict the descriptive analysis aiming to capture the effect of individual diversity on collective creativity by also considering two different criteria to assess creativity: feasibility and originality. As in the previous sections, we differentiate between *surface* and *deep-level* diversity. For the collective experiment and *deep-level* diversity, we consider the group experience that is captured in the post-questionnaire by the following categories: *self-evaluation of creativity*, *workgroup experience, conflict*, and *daily reactions when facing negative or uncomfortable events*.

First, we study the role of *surface-level* diversity in the group on collective creative performance by considering the gender composition of the groups and the program of studies: *economics*, *innovation management*, and *student-entrepreneurs* (for more information, see Table 1). We began by devoting our attention to feasibility, and we found evidence of the effect of gender on collective creative performance by means of groups homogeneously of females performing better than heterogeneous and, lastly, homogeneous males. Although we find strong evidence for the program, *innovation management*[13] performs better. Nevertheless, the dynamics change in the case of originality. We have strong evidence of heterogeneous groups having greater levels of creativity than homogeneous, even though the homogeneous male groups perform better than the homogeneous female. At the same time, we encounter evidence that students from the *innovation management*[14] program have higher levels of originality. Thus, we can point out our fourth result:

R4: *Surface diversity has an essential role in collective creativity, which differs by criteria to assess creativity. For gender, feasibility is fostered by homogeneous female groups, and originality is triggered by the heterogeneity of gender in the groups. In comparison, training in creativity promotes both measures of collective creative performance.*

Secondly, we analyze the role of *deep-level* diversity on collective creativity. On the one side, we discover mild evidence that subjects who expect the problems to solve themselves are negatively related to feasibility. On the other side, we uncover mild evidence of a positive effect on originality driven by satisfaction to have worked with their group, to feel that your team has group spirit, and even stronger by finding evidence for the perception to form part of a good group to work on. Hence, we can state our fifth result:

R5: *The deep-level diversity directly influences collective creativity, but differently depending on the criteria to assess creativity. Feasibility is altered by individuals who indicate as a behavioural trait to prefer when problems solve themselves. On the contrary, originality is fostered by subjects' satisfaction with their group experience.*

[13] The order in average for feasibility is the following: Innovation management>Economics >Student-entrepreneurs.

[14] The order in average for originality is the following: Innovation management>Student-entrepreneurs>Economics.

4.5 Regression analysis

In this section, we report results from regression analysis to gain more insight into the dynamics of creative performance and check previous results with a more accurate analysis[15]. Thanks to this analysis, we are able to study the causal between creativity and several relevant explanatory variables. We began by analyzing each task from the individual experiment controlling for *surface* and *deep-level* diversity variables. In addition, we investigated the creative performance in the collective experiment controlling for the gender composition of the group and program as a measure of *surface-level* diversity. At the same time, we consider *deep-level* diversity by means of task self-evaluation of creativity, workgroup experience, conflict, and daily reactions when facing negative or uncomfortable events. The details of every regression used in this analysis are depicted in Appendix 7.2.

In Table 7[16], we present the summary of our results for the individual experiment. First, we discuss the results of the *Tangram task* (for more details, see Appendix Table 9). Regarding the *surface-level* diversity variables, no significant effect arises from the data, in line with the descriptive analysis. Alternatively, for the *deep-level* diversity variables, the results differ between the diversity of life experience and personality traits. For the first, we do not observe any statistically significant effect. However, for personality traits, we encounter evidence for subjects who positively evaluate their performance in the tangram task having an effective better performance by solving the puzzle and even also under the average time. These results reinforce the conclusions from the descriptive analysis and confirm that the key determinant of creativity for our *closed task* is the self-evaluation of the task.

Then, we looked at the *AUT task* and disentangling the creative measure between expansivity and persistency in Table 7 (for more details, see in Appendix Table 10 for expansivity and Table 11 for persistency). In the case of *expansivity*, none of the variables of *surface-level* diversity is significant. In regard to *deep-level* variables exist mild evidence of the negative effect on expansivity of diversity of interests, generalized trust (evidence), frequency of theatres, and self-evaluation of the own creative performance better than others (evidence). Looking at persistency, the *surface-level* diversity variables damaging are younger age (evidence), *innovation management* program (mild evidence) and *student-entrepreneur* program (evidence). Then, we now look at the role of *deep-level* diversity variables, and we observe the detrimental effects on persistency from diver-

[15]The analysis was carried through the software STATA.

[16]As we already presented in the subsection 4.1, the variable we aim to study in this work is creativity, but the measure and codification of the variable differ depending on the type of task. For the Tangram task, the dependent variable takes 1 if the subject fails to solve the puzzle, 2 if the subjects solve it but above or in the average timing, and 3 if the puzzle was solved below the average timing. For this reason, we decided to implement in the regression an ordered logit model, because it is designed for ordinal dependent variables. While for the AUT task, we follow the same logic for the two measures of creativity, expansivity and persistency. The dependent variable, in this case, is a dummy variable that takes 1 when the expansivity/persistency is below the mean value, and 0 if the expansivity/persistency is equal to or above the mean. We selected a Probit model because our dependent variables are binary, and the same results were found with a Logit model. Finally, for the drawing task, the evaluation of creativity was captured by giving a score from 0 (minimum) to 10 (maximum). Thus, we used a Tobit model, which allows to censored data in the extremes, hence minimum and maximum evaluation. We obtained the same results by OLS (Ordinary Least Square), but we chose the Tobit because it had a higher coefficient of determination.

sity of life experiences (mild evidence), self-evaluate own performance better than others (strong evidence), and frequent often bars/restaurants (evidence). Even though some of the variables significant in the descriptive analysis persist and order vanishes, the logic is still the same, by means of the drivers of creativity in our *open with constraints* task difference between criteria of creativity; hence, expansivity and persistency.

Table 7: Summary of the results of the regressions for the Individual experiment.

Trangram task			*Ordered logit*
Age			n.s.
DOLE			n.s.
Self-evaluation puzzle			+++
	AUT task		
Expansivity	*Probit model*	Persistency	*Probit model*
Age	n.s.	Age	++
Female	n.s.	Female	n.s.
French nationality	n.s.	French nationality	n.s.
French speaker	n.s.	French speaker	n.s.
DILA	−	Program==2	−
General self-evaluation creativity	n.s.	Program==3	−−
Better-in-average word	−−	DILA	−
Trust	−−	Better-in-average word	− − −
theatres	−	Trust	n.s.
Music	n.s.	Theatres	n.s.
Individual sport	n.s.	Music	n.s.
Bars	n.s.	Individual sport	n.s.
		Bars	−−
Draw task			*Tobit model*
Program==2			n.s.
Program==3			n.s.
DILA			n.s.
Self-evaluation Draw			++
Theather & Flim			+

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.

Lastly, we focus on creative performance for the *Draw task*. As previously, we look at the summary of the results for the individual experiment in Table 7 (see for more details in the Appendix in Table 12). Concerning the *surface-level* variables, we do not report any significant effect. By contrast, for *deep-level* variables, we find evidence of greater self-evaluation of the performance in the task leading to better effective performance, and mild evidence of subjects who attend more cinema/theatre presents a greater level of creativity in this task. Thus, the effect in the descriptive analysis of *surface* variables and life of experience disappears with more accurate analysis, and the personality traits remain with some modification, but they still have the frequency of visiting theatres/cinemas remains and arise as new for the analysis the key role on the self-evaluations.

Therefore, we can confirm results 1, 2 and 3 (with some modification on the variables, but the same logic) and adds the last, the only persistent effect found in the data across degrees of openness in the tasks, by stating self-evaluation of the performance in the task as a trigger of creative performance.

Furthermore, the summary of the outcome of the collective task is depicted in Table 8[17] (see for more detail Appendix Table 13), where creative collective performance is broken down into feasibility and originality. In relation to *surface-level* diversity, we will consider, as before, the gender composition of the groups and the programs. On the one hand, for feasibility, we discover strong evidence that homogeneous female groups perform better in comparison with heterogeneous with equal gender division. On the other hand, for originality, homogeneous groups of females perform more poorly, followed by the heterogeneous with a greater number of females, playing down always with equal distribution of gender. Therefore, gender has an essential role in collective creativity and differs between criteria to asses creativity because groups with only females foster feasibility and have a detrimental effect on originality. While if we devote our attention to the role of the program, for feasibility, we have strong evidence of the detrimental effect of the entrepreneurial training, while for originality, we have a positive role (strong evidence) as creativity training but with a smaller measure (mild evidence). Nevertheless, we investigated the role of *deep-level* diversity by studying the group experience, and no significant result emerged from the data. Therefore, we are able to support result 4 and deny result 5 with a more accurate analysis.

Table 8: Summary results of the Tobit regressions for Collective task.

	Feasibility	*Originality*
	Model 1	Model 2
Homogeneous Female	+++	− − −
Homogeneous Male	n.s.	n.s.
Heterogeneous Female	n.s.	− − −
Heterogeneous Male	n.s.	− − −
program==2	n.s.	+
program==3	− − −	+ + +

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. The variable program==2 corresponds to participants in and program==3 refers to *Student-entrepreneurs* in relativize to economics.

[17]The dependents variables is always referring to creativity, and for the collective experiment is divided in two components feasibility and originality (for further information see subsection 4.1). Both of them take values from 0 (minimum) to 10 (maximum). Therefore, we decided to implement a Tobit model which allows as to censored the extremes, in this case the minimal and maximal score for each measure of creativity. We have run the same regression by OLS and the results do not differ, whereas, we selected the Tobit because it had a greater coefficient of determination.

5 Discussion

This paper aims to analyze the role of surface and deep-level diversity variables on individual and collective creativity. First, at the individual level, we showed that neither *surface* nor *deep-level* variables have a constant role in our creativity assessment. The variables impacting creative performance change as long as we change the nature of the task in terms of the degree of openness and the criteria chosen to define creativity. These mixed results confirm our first hypothesis and corroborate the necessity for future research to dig into this research question of how different profiles of individuals will perform differently according to the thinking elicited. Despite the mixed results, we discover a variable that systematically influences creativity for all the tasks, regardless of the degrees of openness, which is the self-evaluation of the performance in the task. But still, it is positive for open and closed tasks, whereas perceiving a better performance of the other is negatively impacting creative performance for open with constraints tasks. Thus we suggest that more research should direct the attention to self-evaluation of performance in the study of individual creativity.

Second, at the collective level, we selected two variables of interest for *surface* diversity: the first one, group gender distribution, and the second one, the curriculum of participants (by means of training). On the one hand, we confirmed that perfectly heterogeneous groups, in terms of gender, would perform better than homogeneous groups for the criteria of originality, which confirms our hypothesis 3 in the sense that diversity of gender has a generally positive impact. Besides, for the criteria of feasibility, homogeneous female groups performed better than heterogeneous and male homogeneous groups. Therefore, even though we confirm hypothesis 3 for originality, instead, it is denied for feasibility because females foster it, and males have a detrimental effect. Hence we can suggest that further research is needed to decompose the role of gender on group performance.

While, the group performance depends on the programs that subjects are involved in. We conclude that the *innovation management* program promotes creativity for both measures of feasibility[18] and originality. The difference between criteria to asses creativity arises if we look at the program in which students perform the most poorly, that is, *students-entrepreneurs* for feasibility and *economics* for originality. Therefore, we can partially confirm hypothesis 2 because the principal foster of creativity is the program of *innovation management* by means of the training of creativity. Thus, universities and enterprises should consider this program of study as a requirement or training to offer with the potential to foster creativity. Even though there are still a few things to say on this topic, and is necessary to measure the dimension of this effect.

By contrast, we do not find significant evidence for the role of *deep-level* diversity in the regression analysis. We hypothesize that this is driven by the small number of groups in our sample, and this might will change after the extension of our sample. Therefore, we remark on the importance of continuing to investigate the role of group experience has on collective creativity. Moreover, we had neither found evidence to support hypothesis 4, we suggest that the lack of a number of

[18] Even if it is not significant for feasibility is the highest creative outcome in comparison to the other two programs.

groups also is a keen factor in improving and properly testing the hypothesis, which has enormous potential.

The main contribution of this paper to the existing literature is the attempt to capture the role of surface and deep-level variables on creative performance. However, they seem to be highly sensitive to task specificity in terms of their degree of openness and criteria of creativity. Much work remains to be done to disentangle the black box that still triggers creative performance.

6 Conclusion and further improvements

In this paper, we have a three-fold objective to address. First, we aim to study the role of variables at the surface and deep-level on individual creative performance. Second, we want to moderate the effect of those variables according to the degree of openness of the task; by considering open, open with constraints, and closed tasks. Third, we target to study the role of diversity (*surface* and *deep-level*) on collective creativity. Indeed, to the best of our knowledge, we are the first to consider the interplay between diversity levels with regard to both the degree of openness and individual and collective creativity.

For individual creativity, the results of this study conclude that diversity variables, regardless if we refer to surface or deep-level, have different effects depending on the degree of openness and even between criteria to assess creativity. The only factor that arose persistently was the self-evaluation of the performance in the task, which positively relates to creative performance (open and closed), whereas negatively when subjects self-evaluate their performance better than others for an open with constraint task.

In comparison, for collective creativity, gender and program of studies have a keen role in fostering creativity at the surface-level of diversity. But the effect differs by criteria to assess creativity for gender composition to the group, by being a matter of the heterogeneous composition of the group for originality, and females performing better for feasibility. Nevertheless, in the program, we observe a positive systemic effect of the program with the training of creativity fostering collective creative performance.

We cannot deny that the main limitation of this paper is the number of observations. For this reason, it is important to note that a second set of experimental sessions was carried out from September to December 2022. Thanks to this, we raised our first sample size of 74 subjects and 18 groups to 146 subjects and 35 groups. This increase in the sample allows us to replicate the previous analysis, leading to more robust results. Additionally, we will also include improvements in the data analysis. The following paragraphs aim to present further extra analysis that will be carried out.

First, the interrelation between the individual and collective performance of subjects. In this work, we highlight the role of surface and deep-level variables on both individual and collective creative performance. However, we still need to investigate the role of individual performance on collective creative outcomes. The idea is to issue subjects' individual performance in the three

different tasks, hence by the degree of openness, as independent variables, in other to settle more light of drivers' of collective creative performance, our dependent variable. The aim is to investigate the group's diversity in terms of the individual performance of all group members.

Second, the expansion of the diversity analysis with the implementation of the *Blau Index* that can be defined as follows: "*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, 2012, p.7]. Thus, we aim to study diversity in the group for any variable at both surface and deep-level dimensions to measure asymmetry. Scoring from 0 (no diversity) to 1 (complete diversity), the Blau Index allows us to have a measure of diversity to compare each group and understand more precisely how this diversity impacts collective creative performance.

Third, the implementation of a social network analysis. This additional layer of analysis would help us, first, to control for the interactions during the experiment from the perspective of real social interactions where anonymity and isolation are waived. Then, it would give a broader understanding of the dynamics occurring during the experiment and their direct influence on the collective creative performance. In the end, we collected information on five networks linking our subjects. We can categorize them into two parts: (1) the experimental network (*the exchanges of tokens during the strategic phase*) and (2) the pre-existing network (*friendship, knowing previously, professional, and meeting-frequency networks*).

References

Teresa M. Amabile. Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43:997–1013, 1982.

Teresa M. Amabile. Componential theory of creativity. 2012.

Teresa M. Amabile and Michael G. Pratt. The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research in Organizational Behavior*, 36:157–183, 2016.

Teresa M. Amabile, Karl G. Hill, Beth A. Hennessey, and Elizabeth Tighe. The work preference inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology*, 66:950–967, 1994.

Teresa M. Amabile, Regina Conti, Heather M. Coon, J. Lazenby, and M. Herron. Assessing the work environment for creativity. *Academy of Management Journal*, 39:1154–1184, 1996.

Dan Ariely, Uri Gneezy, George Loewenstein, and Nina Mazar. Large stakes and big mistakes. *The Review of Economic Studies*, 76:451–469, 2009.

Giuseppe Attanasi, Ylenia Curci, Patrick Llerena, Adriana Carolina Pinate, María del Pino Ramos-Sosa, and Giulia Urso. Looking at creativity from east to west: Risk taking and intrinsic motivation in socially and culturally diverse countries. 2019.

Giuseppe Attanasi, Michela Chessa, Sara Gil-Gallen, and Patrick Llerena. A survey on experimental elicitation of creativity in economics. *Revue d'économie industrielle*, 2021.

Jonali Baruah and Paul B. Paulus. Category assignment and relatedness in the group ideation process. *Journal of Experimental Social Psychology*, 47:1070–1077, 2011.

Jonali Baruah and Paul B. Paulus. The role of time and category relatedness in electronic brainstorming. *Small Group Research*, 47:333 – 342, 2016.

Christiane Bradler. How creative are you? – an experimental study on self-selection in a competitive incentive scheme for creative performance. *ERN: Experimental Economics (Topic)*, 2015.

Christiane Bradler, Susanne Neckermann, and Arne Jonas Warnke. Incentivizing creativity: A large-scale experiment with tournaments and gifts. *Behavioral & Experimental Economics eJournal*, 2016.

Anaëlle Camarda, Gregoire Borst, Marine Agogué, Marianne Habib, Benoit Weil, Olivier Houdé, and Mathieu Cassotti. Do we need inhibitory control to be creative? evidence from a dual-task paradigm. *Psychology of Aesthetics, Creativity and the Arts*, 10 2017. doi: 10.1037/aca0000140.

Shelley H. Carson, Jordan B. Peterson, and Daniel M. Higgins. Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal*, 17:37 – 50, 2005.

Gary Charness and Daniela 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, February 2013. URL `https://ideas.repec.org/p/cdl/ucsbec/qt4mr6p1d5.html`.

Gary Charness and Daniela Grieco. Creativity and Incentives. *Journal of the European Economic Association*, 17(2):454–496, 2019. URL `https://ideas.repec.org/a/oup/jeurec/v17y2019i2p454-496..html`.

Mihály Csíkszentmihályi. Creativity: Flow and the psychology of discovery and invention. 1996.

Shane S. Douthitt, Lillian T. Eby, and Shana 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.

Glenn Dutcher and Cortney S. Rodet. Which two heads are better than one? uncovering the positive effects of diversity in creative teams. *Behavioral & Experimental Economics eJournal*, 2022.

Katharina Eckartz, Oliver Kirchkamp, and Daniel W. Schunk. How do incentives affect creativity? *Labor: Personnel Economics eJournal*, 2012.

Gerhard Fischer, Elisa Giaccardi, Hal Eden, Masanori Sugimoto, and Yunwen Ye. Beyond binary choices: Integrating individual and social creativity. *Int. J. Hum. Comput. Stud.*, 63:482–512, 2005.

Joy P. Guilford. Creativity. *American Psychologist*, 5:444–454, 1950.

Joy P. Guilford. *The nature of human intelligence*. New York: McGraw-Hill, 1967.

Joy Peter Guilford. Creativity: A quarter century of progress. In *Perspectives in creativity*, pages 37–59. Routledge, 2017.

David A. Harrison, Kenneth H. Price, and Myrtle 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.

Sarah Harvey. A different perspective: The multiple effects of deep level diversity on group creativity. *Journal of Experimental Social Psychology*, 49:822–832, 2013.

Andreas Stefan Hundschell, Stefan Razinskas, Julia Backmann, and Martin Hoegl. The effects of diversity on creativity: A literature review and synthesis. *Applied Psychology*, 2021.

Matthew O. Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71:44–74, 1996.

Susan E. Jackson, Joan F. Brett, Valerie I. Sessa, Dawn M. Cooper, Johan A. Julin, and Karl Peyronnin. Some differences make a difference: Individual dissimilarity and group heterogeneity as correlates of recruitment, promotions, and turnover. *Journal of Applied Psychology*, 76:675–689, 1991.

Nicolas Jacquemet, Fabrice Le Lec, and Olivier L'haridon. Précis d'économie expérimentale. 2019.

Stylianos Kavadias and Svenja C. Sommer. The effects of problem structure and team diversity on brainstorming effectiveness. *Manag. Sci.*, 55:1899–1913, 2009.

Nicholas W. Kohn and Steven M. Smith. Collaborative fixation: effects of others' ideas on brainstorming. *Applied Cognitive Psychology*, 25:359–371, 2011.

Nicholas W. Kohn, Paul B. Paulus, and Yunhee Choi. Building on the ideas of others: An examination of the idea combination process. *Journal of Experimental Social Psychology*, 47:554–561, 2011.

Sarnoff A. Mednick. The associative basis of the creative process. *Psychological review*, 69:220–32, 1962.

Frances J. Milliken and Luis L. Martins. Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of Management Review*, 21:402–433, 1996.

Peter G. Moffatt. *Experimetrics: Econometrics for Experimental Economics*. New York: McGraw-Hill, 2020.

R. Osche. Before the gates of excellence: the determinants of creative genius. 1990.

Katherine Phillips and Charles O'Reilly. *Demography and Diversity in Organizations: A Review of 40 Years of Research*, volume 20, pages 77–140. 01 1998.

Henri Poincaré. *A description of his own creativity*. Trans. G.B. Halstead, 1908.

Alexander Spickermann, Martin Zimmermann, and Heiko 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.

Ivan D. Steiner. Group process and productivity. 1972.

George Ursachi, Ioana Alexandra Horodnic, and Adriana Zait. How reliable are measurement scales? external factors with indirect influence on reliability estimators. *Procedia. Economics and finance*, 20:679–686, 2015.

Daan van Knippenberg, C. D. De Dreu, and Astrid C. Homan. Work group diversity and group performance: an integrative model and research agenda. *The Journal of applied psychology*, 89 6:1008–22, 2004.

Thomas B. Ward. Structured imagination: the role of category structure in exemplar generation. *Cognitive Psychology*, 27:1–40, 1994.

7 Appendix

7.1 Elicitation of themes

After receiving their list of themes, we had to clean it and regroup them. We followed the process hereafter :

1. We selected all themes from all judges and classified them by categories. From their subjective themes, we finished with 28 final themes.
2. For each evaluator, we corrected the themes according to the classification and counted for unique themes.

From the number of valid ideas and the number of unique themes for each word of each participant, we could compute our measures :

1. Expansivity is the originality of the participants' answers, represented by the frequency of their answers in the entire sample.
2. Persistency which the indicator calculated from the ratio of fluidity and flexibility (the number of valid ideas divided by the number of themes).

7.2 Regressions

Table 9: Ordered logit regression on Tangram task

	Model 1	Model 2	Model 3	Model 4	Model 5
Age	-0.1389*	-0.1629**	-0.1207	-0.1142	-0.1311
	(0.0723)	(0.0740)	(0.0789)	(0.0783)	(0.0813)
Gender (female=1)	0.0501	-0.1609	0.3668	0.2565	0.4310
	(0.4403)	(0.4622)	(0.5496)	(0.5844)	(0.6151)
Program=2	1.5348	1.7246*	1.6299	1.1057	0.5756
	(0.9749)	(1.0137)	(1.1838)	(1.2521)	(1.2416)
Program=3	0.4389	0.6999	1.4708*	1.2821	1.1472
	(0.6866)	(0.7141)	(0.8353)	(0.8839)	(0.8987)
DOLE		0.0662*	0.0134	-0.0008	0.0016
		(0.0385)	(0.0490)	(0.0499)	(0.0519)
General self-evaluation creativity			0.0651	-0.1524	-0.3606
			(0.3206)	(0.3746)	(0.3893)
Self-evaluation puzzle			1.1353***	1.1189***	1.1163***
			(0.3059)	(0.3144)	(0.3194)
Better-in-average puzzle			0.6210	0.6993	0.8199
			(0.7615)	(0.7774)	(0.8091)
Do new things				0.4829	0.5550
				(0.4936)	(0.5133)
Introvert				0.4701	0.4630
				(0.3145)	(0.3197)
Prefer to work in groups				-0.4485	-0.1941
				(0.5708)	(0.6100)
Visual arts					0.1741
					(0.3051)
Humor					-0.0903
					(0.3384)
Inventions and technologies					0.5245
					(0.3346)
cut1	-4.0528	-3.6825	-0.2163	1.0069	0.7351
	(1.6875)	(1.6894)	(2.0841)	(2.2738)	(2.2751)
cut2	-2.5814	-2.1581	2.2364	3.6283	3.4738
	(1.6404)	(1.6440)	(2.0950)	(2.3139)	(2.3119)
Observations	74	74	74	74	74
Pseudo R^2	0.0419	0.0607	0.3007	0.3340	0.3535

Robust standard errors in parenthesis ***p<0.01, **p<0.05, *p<0.1.

Table 10: Probit regression of Expansivity for AUT task.

	Model 1	Model 2	Model 3	Model 4	Model 5
Gender (female=1)	-0.1538	-0.1090	-0.0999	0.0399	0.3144
	(0.3098)	(0.3164)	(0.3439)	(0.3813)	(0.4243)
Age	0.0102	0.0164	0.0280	0.0054	0.0289
	(0.0367)	(0.0369)	(0.0388)	(0.0457)	(0.0555)
French speaker	0.1901	0.3370	0.3853	-0.2252	-0.1485
	(0.4525)	(0.4586)	(0.4696)	(0.5731)	(0.6811)
French nationality	-0.1834	-0.2032	-0.2280	0.1141	-0.0743
	(0.4667)	(0.4684)	(0.4889)	(0.5752)	(0.6623)
DILA		-0.1236**	-0.1469**	-0.1150	-0.1874**
		(0.0597)	(0.0646)	(0.0728)	(0.0840)
Self-evaluation creativity			0.2517	0.1935	0.0717
			(0.2248)	(0.2623)	(0.2893)
Self-evaluation creativity			0.1598	0.1625	0.2211
(Individual experiment)			(0.1799)	(0.1929)	(0.2042)
Better-in-average words			-1.2285**	-1.5995***	-1.2820*
			(0.5474)	(0.6064)	(0.6634)
Generalized trust			-0.2145**	-0.3046**	-0.2997**
			(0.1042)	(0.1237)	(0.1344)
Theatres				-0.4303*	-0.5486*
				(0.2524)	(0.2846)
Meet people				-0.2308	-0.2009
				(0.2404)	(0.2551)
Bars				0.0367	0.1213
				(0.2713)	(0.3166)
Association				0.7818	0.6595
				(0.5106)	(0.5410)
Night out				-0.2912	-0.4009
				(0.2537)	(0.2922)
Music					0.2257
					(0.2210)
Individual sports					0.2856
					(0.2384)
Humor					0.3634
					(0.2331)
Sciences					0.1543
					(0.2083)
Constant	-0.1703	0.2490	0.1540	3.1725*	2.0554
	(0.9448)	(0.9723)	(1.2859)	(1.9050)	(2.1529)
Observations	74	74	74	74	74
Pseudo R^2	0.0047	0.0481	0.1463	0.2442	0.3104

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

Table 11: Probit regression of Persistency for AUT task.

	Model 1	Model 2	Model 3	Model 4	Model 5
Gender (female=1)	-0.1901	-0.1117	-0.0324	0.1763	0.2250
	(0.3318)	(0.3428)	(0.3721)	(0.4169)	(0.4270)
Age	0.1293**	0.1621**	0.1545*	0.1656*	0.1875*
	(0.0642)	(0.0701)	(0.0758)	(0.0889)	(0.1003)
French speaker	-0.3671	-0.17208	-0.1230	-0.4075	-0.4185
	(0.4942)	(0.4979)	(0.5213)	(0.6559)	(0.6949)
French nationality	-0.3041	-0.3271	-0.2551	-0.0719	-0.2274
	(0.4884)	(0.4859)	(0.5023)	(0.6144)	(0.6785)
Program=2	-0.8019	-0.9653	-1.2395*	-1.8431**	-2.0525**
	(0.5879)	(0.6296)	(0.7331)	(0.8875)	(0.9241)
Program=3	-1.2274**	-1.5532**	-1.4437**	-1.6192**	-1.6732**
	(0.5640)	(0.6048)	(0.6451)	(0.6907)	(0.7553)
DILA		-0.1521	-0.1617**	-0.0897	-0.1350
		(0.0720)	(0.0773)	(0.0890)	(0.0990)
Self-evaluation creativity			0.1091	-0.1057	-0.0989
			(0.2411)	(0.2786)	(0.3197)
Self-evaluation creativity			0.1135	0.1387	0.1773
(Individual experiment)			(0.2056)	(0.2252)	(0.2306)
Better-in-average words			-1.6643***	-2.0583***	-1.9164**
			(0.6340)	(0.7371)	(0.7743)
Generalized trust			-0.0421	-0.0874	-0.0526
			(0.1081)	(0.1235)	(0.1296)
theatres				-0.1521	-0.1382
				(0.2585)	(0.2629)
Meet people				0.1330	0.1231
				(0.2381)	(0.2555)
Bars				-0.6890**	-0.6809**
				(0.2861)	(0.3042)
Association				0.0705	-0.0314
				(0.5128)	(0.5284)
Music					0.1988
					(0.2482)
Individual sports					0.2453
					(0.2448)
Humor					-0.0821
					(0.2217)
Sciences					0.1225
					(0.2378)
Constant	-1.9142	-1.9176	-2.0280	0.0235	-1.1137
	(1.4797)	(1.5594)	(2.0709)	(2.5934)	(2.869)
Observations	74	74	74	74	74
Pseudo R^2	0.1324	0.1824	0.2679	0.3499	0.3741

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

Table 12: Tobit regression for Draw task.

	Model 1	Model 2	Model 3
Spoken languages	0.1200	0.0779	0.0191
	(0.1793)	(0.1805)	(0.1731)
Domain=2	0.5257	0.6285	0.6880
	(0.7215)	(0.7172)	(0.6952)
Domain=3	-2.1566	-1.8874	-2.4883**
	(1.2995)	(1.2829)	(1.2256)
Program=2	1.1501	1.0003	-0.6569
	(0.8226)	(0.8094)	(0.7693)
Program=3	0.4068	0.4579	0.2973
	(0.6272)	(0.6213)	(0.5851)
DILA		0.0699	-0.0275
		(0.0654)	(0.0675)
DGLOHR		-0.1198	-0.1061
		(0.0746)	(0.0760)
RWPFE		0.0236	-0.0012
		(0.0491)	(0.0471)
Self-evaluation draw			0.3936**
			(0.1612)
Visual arts			0.0711
			(0.1844)
Inventions and technologies			0.1818
			(0.1942)
Theatre & film			0.4428*
			(0.2411)
Culinary arts			-0.1762
			(0.1875)
Constant	3.7665***	3.7392***	2.7464***
	(0.4733)	(0.5780)	(0.6649)
Observations	74	74	74
Pseudo R^2	0.0273	0.0399	0.0818

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

Table 13: Tobit regression of creativity components for Collective task.

	Feasibility		*Originality*	
	Model 1	Model 2	Model 1	Model 2
Homogeneous Female	0.7655**	0.8065**	-2.1998***	-2.3008***
	(0.3270)	(0.3270)	(0.4089)	(0.4220)
Homogeneous Male	-0.5226	-0.4528	-0.6361	-0.5898
	(0.3327)	(0.3369)	(0.4161)	(0.4380)
Heterogeneous Female	0.0833	0.0757	-1.5956***	-1.6019**
	(0.4525)	(0.4493)	(0.5659)	(0.6328)
Heterogeneous Male	0.4860	0.5155	-1.3917***	-1.4736***
	(0.3378)	(0.3365)	(0.4224)	(0.4541)
Program=2	0.5972	0.6868	1.0461*	1.005*
	(0.4207)	(0.4263)	(0.5262)	(0.5521)
Program=3	-1.0347***	-1.0350***	2.2321***	2.2549***
	(0.3052)	(0.3030)	(0.3817)	(0.3927)
Satisfaction group work				0.1505
				(0.2179)
Group spirit				0.1111
				(0.2159)
Good group to work in				-0.1969
				(0.2168)
Problem solve itself		-0.1012		
		(0.0966)		
Constant	5.125***	5.3402***	5.7188***	5.1517***
	(0.2290)	(0.3063)	(0.2864)	(1.3437)
Observations	74	74	74	74
Pseudo R^2	0.1293	0.1341	0.1599	0.1643

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