Exploring Compatible Interaction Preferences with a Puzzle Video Game

Samuel Gomes[1,2], **Luis Costa**[1,3], **Marta Couto**[1], **Sandra Gama**[1,2], **Carlos Martinho**[1,2], **João Dias**[4,5], **Geraldo Xexéo**[3], **and Ana Moura Santos**[2]

[1]INESC-ID, Portugal
marta.couto@gaips.inesc-id.pt

[2]Instituto Superior Técnico, Universidade de Lisboa, Portugal
{samuel.gomes,sandra.gama,carlos.martinho,ana.moura.santos}@tecnico.ulisboa.pt

[3]Universidade Federal do Rio de Janeiro, Brazil
{luisfcosta,xexeo}@cos.ufrj.br

[4]CISCA, Portugal

[5]Faculdade de Ciências e Tecnologia, Universidade do Algarve, Portugal
jmdias@ualg.pt

ABSTRACT

As social beings, humans pursue happiness by craving to foment and maintain fulfilling social relations, a need that transcends from personal into professional life and that leveraged the study of psychological factors that can influence teamwork processes and outcomes. Following this tendency, the present work studies how the pairing of people with distinct interaction preferences influences their acquired ability and experience while training together. For this, a puzzle video game named *Alien Bar* was deployed and used to evaluate 31 pairs of players ($n = 62$). The results demonstrated teamwork benefits of including self-oriented and challenge-oriented subjects, and that care should be raised when joining others-oriented characters. Additionally, interpersonal closeness influenced subjects' experience at the perceived competence level, but not their enjoyment which may instead relate to task affinity. These findings may help develop instructor-driven and automatic group management, otherwise dependent on possibly inaccurate subjects' judgements alone.

Keywords

Interaction Preferences, Teamwork, Puzzle Video Game, Alien Bar

Introduction

Given their social nature, humans aim to achieve well-being and happiness by fomenting and maintaining gratifying social relations. This need transcends onto other animal communities such as primates, determining their societal ranking and reproductive success (Silk et al. 2013). In search for purpose and significance in their social connections, humans identify and assimilate practices and values from their environment. Such values are crucial for functional participation in various settings like sports, work, and school. In fact, according to the Self-Determination Theory (Deci and Ryan 1985; Ryan and Deci 2023), humans have basic needs for competence, relatedness, and autonomy, and these needs are satisfied in environments that promote well-being, contrasting with controlling environments that

Proceedings of DiGRA 2025

foment negative affect like apprehension and defensiveness towards others. As such, given that we spend a considerable portion of our lives working (Campbell 2017), it is assumed that our relationships at the work environment should be the most harmonious possible for us to routinely feel well and more willingly achieve settled goals.

Previous research has given particular attention to psychological factors that can influence the processes and outcomes underlying teamwork, notably habitual feeling states such as personality (Bradley et al. 2013). This research advocates that, beyond more direct task liking or interpersonal closeness (Gino and Galinsky 2012), the personality make-up of a team helps shape both the processes and outcomes that underline teamwork. Thus, we believe that more intrinsic aspects can be considered by instructors as well as game designers, when developing or tuning automatic systems for group management and game matchmaking.

Inspired by these thoughts, in this article, we study how personal preferences can combine to drive effective teamwork, i.e., if subjects with a disposition to interact in a certain way can fare better when working alongside subjects with other specific interaction preferences. For instance, we explore in which cases we can expect 'similarity attraction' (Insko et al. 1973), in the sense that subjects alike enjoy being paired; which cases can corroborate the 'complementary hypothesis' (Dryer and Horowitz 1997), in the sense that subjects' conduct can complement each other in completing a task; or even if certain profiles can have a more general appeal and work well alongside multiple profiles.

Our premise is that good team-driven learning outcomes and experience may emerge not only through task liking and interpersonal closeness (Gino and Galinsky 2012), but also when harmonious interaction preferences are present. In our work, we consider interaction preferences for completing a task in a certain manner, for instance, some people might prefer to complete a collective task by helping peers in need while others may prefer to compete with their peers. Thus, we explore the following research questions:

RQ1: When working together, how can the interaction preferences of subjects influence their and their peers' acquired ability and experience?

RQ2: Before working together, can subjects accurately estimate, solely based on their and their peers' interaction preferences, the most successful way in which the group should interact?

RQ3: Besides interaction preferences, how can other variables such as task liking and interpersonal closeness affect the development of ability and experience?

To respond to these questions, we start by presenting a simple model that profiles interaction according to two dimensions: *Focus* that connects to the tendency of subjects to pay attention to themselves or others while interacting; and *Challenge* that distinguishes embracing an easier or harder route for task completion. Afterwards, we present *Alien Bar* (Gomes et al. 2024), a puzzle video game in which a player combines fictitious ingredients to form recipes (see Figure 1). *Alien Bar* worked as a training task for the intents of this study.

(a) (b)

(c) (d)

Figure 1: Screenshots of the menu (Figure 1a) and gameplay (Figures 1b to 1d) of *Alien Bar*. The gameplay screenshots demonstrate different game modes. Figures 1b and 1c demonstrate the *Training* game mode from different in-game stations. Figure 1d is a screenshot of the game in its *Survival* mode.

Considering all the aforementioned aspects, this document is organised as follows. Firstly, we review studies on the psychology of work-driven social relations and describe the interaction model considered for our experiments. Afterwards, we describe the game *Alien Bar*. Then, we present our experimental process and evaluation. Later, we discuss the practical implications of our findings, and present some limitations and concluding remarks.

Background

To contextualise the current study, we start by reviewing research on social relations present in a work setting, and afterwards we discuss research on how to profile a subject's interaction preference.

The Psychology of Work-Driven Social Relations

Based on the sort of interaction developed among workers, research has divided the types of workplace relationships into three categories: work acquaintances, work friends, and social friends (Henderson and Argyle 1985). Work acquaintances consist of purely formal relations that are superficial and task-oriented, and neither have liking or disliking connotations; work friends are more intimate but still limited, in the sense that friendship social relationships exist, but are constrained to the work environment; and when there is an intimate enough relation between workers, they become social friends that meet at social events outside of the work setting. As such, to produce a harmonious work setting, work environments are often set up and laid out in a pleasant and comfortable way, and often promote a culture of inclusion that eases the proliferation of social relations. Even so, while perceiving closer relations, subjects may implicitly feel at ease to communicate negative information, motivated by the protection of their peers instead of self-enhancement (Dubois et al. 2016). Keeping this in mind, a question is raised: *is the pursuit of closer social relations enough to foment productive and fruitful teamwork, and can we further aid this by exploring the specifics of work affiliation, notably the presence of compatible interaction styles?* To approach such a question, we will consider psychology-driven research relating teamwork to individual differences.

Multiple models have been proposed to characterise personality, notably the Myers-Briggs Type Indicator (MBTI) (Myers et al. 1998) that characterizes subjects according to five dichotomies: Extraversion or Introversion (E–I), Sensing or Intuition (S–N), Thinking or Feeling (T–F), and Judging or Perceiving (J–P); and the Five Factor Model (FFM) (McCrae and Costa Jr 2008) that profiles subjects according to five dimensions: Conscientiousness, Agreeableness, Neuroticism, Extraversion, and Openness to Experience. These have served as a base for multiple studies regarding teamwork, which indicated that the composition of a team, that is, the psychological profiles of its members (Lewis and Smith 2008; Ahmed et al. 2010; Omar and Syed-Abdullah 2010; Shuto et al. 2017) or team coordinator (Rodríguez Montequín et al. 2013) have an impact on team performance and experience (conflict and satisfaction). For instance, some research defends that teams in which members have diverse profiles may succeed in challenging or more creative tasks as members can adjust to different roles, contrary to homogeneous teams that fare well in less challenging tasks (Omar and Syed-Abdullah 2010; Mazni et al. 2010). Furthermore, some research defends the consideration of personality diversity for group management over students' self-assigned groups or the creation of groups solely based on technical ability (Rutherfoord 2001).

The same line of research also insights how specific traits lead to better teamwork in certain environments, for instance, that the presence of members exhibiting higher extraversion and agreeableness, as profiled by FFM, can be beneficial in teams for computer science projects and that the positive effect of agreeableness prevails when teammates interact face-to-face (Omar and Syed-Abdullah 2010; Bradley et al. 2013); or that MBTI teams with intuition-profiled members will excel in creative tasks (Rodríguez Montequín et al. 2013).

Different FFM research also verified, among other aspects, that learning can be promoted in teams having members with similar levels of Neuroticism, or diverse levels of Extraversion or Openness to Experience (Shuto et al. 2017). Even so, the same research suggests that the influence of personality in learning may also vary with the teaching style of a course, namely with discussions, based on practice, or using exercises. Additionally, while defending the self-improvement benefits of students' insight about their and their team's FFM profile, Ogot and Okudan (Ogot and Okudan 2006) further review and relate an extended amount of literature on FFM traits in teamwork, assessing how the FFM makeup of a design team affects its dynamic. It is suggested that for some traits, homogeneity is preferred, e.g., low values of Neuroticism and high levels of Agreeableness, and in some other traits such as Extraversion and Conscientiousness, some heterogeneity may benefit team capability.

Furthermore, Rodrígez Montequín et al. (Rodríguez Montequín et al. 2013) measured the work quality produced by teams of a Project Management class implementing project-based learning. The authors proved that teams in which members had strong leadership MBTI profiles, e.g., ENTJ or ESTP, managed to achieve better results than teams in which members had MBTI profiles with weak leadership qualities and lower motivational skills, particularly

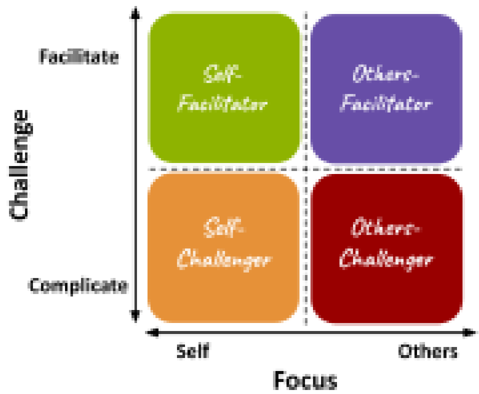


Figure 2: The space formed by *Focus* and *Challenge*.

the ones that imply the need for clear and strict organizations: ISTJ and ESTJ. Ahmed and colleagues (Ahmed et al. 2010) also revealed that fourth year undergraduate software engineering students who were MBTI thinkers and judges performed especially well in such a subject, as evaluated using paper-based exams, assignments, projects, and quizzes.

Although the previous studies give broad insight on what team compositions can be considered more or less beneficial, they focus on the habitual feeling states of subjects instead of their interaction preferences while completing a task. The latter is also relevant, considering that the influence of feeling states in teamwork can vary with the style of task (Shuto et al. 2017). Thus, we believe on the value of contributing to teamwork research by tackling how subjects' task-directed interaction preferences combine well in teams. *But how can we model task-directed interaction preferences?* We approach this aspect in the following section.

Modelling Social Interaction

To suit the needs of our study, we considered a simple model to represent subjects' interaction preferences as they perform a task, drawing its dimensions from previous research (Alves et al. 2020; Gomes et al. 2022). To create a holistic view of the interaction styles expressed in collective tasks, an interaction space was formed by joining two dimensions: *Focus* and *Challenge* (see Figure 2). Firstly, *Focus* distinguishes subjects' tendency to interact while paying attention to themselves or others. As an example, consider an academic scenario in which *Focus* is translated to the difference between studying alone or studying by solving exercises with others. Secondly, *Challenge* distinguishes the intention of the subject who interacts between 'Facilitate' and 'Complicate'. In other words, this dimension differentiates between embracing an easier or more challenging route for task completion. Consider, for example, that students can prepare for an exam by tackling the minimum required exercises (either alone or with others), or they can challenge themselves or others with hard assignments.

In Figure 2, we labelled four areas corresponding to different *Focus-Challenge* combinations, that can be described as:

- *Self-Facilitator*: the focus on facilitating own task progression, that is, opting for the most effortless route. This interaction may relate to a slower and highly granular learning process, and may be reflected by a student who prepares for an exam by studying easier exercises alone until reaching the knowledge required to solve the exam.
- *Others-Facilitator*: the consideration of an altruistic task completion as a way to learn, assisting others in following the path of least resistance. At an extreme, we can think of a peer helping other peer to study for an exam, even if that just means wasting time from the functional perspective of the helper. Nevertheless, the helping peer may inadvertently be learning and/or consolidating their knowledge, a form of the protégé effect (Chase et al. 2009).
- *Self-Challenger*: the completion of complicated self-oriented tasks as a way to learn. This interaction may relate to a faster and less granular learning process, led by the pursuit of difficulty and personal growth. As an example, a student can prepare for an exam alone by studying difficult exercises.
- *Others-Challenger*: encouraging and motivating others to challenge themselves. We can think of a student that prefers to learn while inciting other students to solve more demanding exercises.

Alien Bar

The present work applies the game *Alien Bar* (Gomes et al. 2024) as a training task. This game is similar to *Cook, Serve, Delicious!*[1] or *Overcooked!*[2], in the sense that it is a continuous game in which players use ingredients to form recipes. Yet, unlike those games, all ingredients and utensils present in *Alien Bar* have fictitious names (like *Thirpunasorec* and *Orgeine*), and its mechanics are built in such a way that in-game tasks become memory puzzles. The repository for the code of the game is available online, hosted on the *GitHub* platform[3].

The game was configured with a total of five difficulty levels. Yet, the levels considered by the game depend on the game mode. Each difficulty level comprehends a set of recipes, and relies on the steps and mechanics needed to generate them. During gameplay, three game modes can be selected in the main menu (see Figure 1a): *Tutorial*, *Training*, and *Survival*. *Tutorial* presents only recipes from the first level and has a fixed duration; *Training* also has a fixed duration, but throughout this phase, a player is free to start any level and play for as long as they want (this mode is the only one to display the top right button of Figures 1b and 1c that returns the game to the main menu); and in *Survival*, the game starts from the first level and progressively increases its level until either a maximum duration is reached or no more pending orders can be included.

Unlike in *Survival*, orders in both *Tutorial* and *Training* are generated every time the player delivers a pending order, thus maintaining a full pending orders area. Furthermore, in *Sur-*

1. https://www.cookservedelicious.com/yum/ (accessed 5th June 2025).
2. https://www.team17.com/games/overcooked/ (accessed 5th June 2025).
3. https://github.com/SamGomes/alien-bar (accessed 5th June 2025).

Figure 3: Participants playing *Alien Bar* during our pilot tests.

vival, an order is generated every time there are no pending orders left to deliver, keeping player focus and engagement (compare the gameplay screenshots of Figure 1).

Pilot Tests

We wanted to define an experimental procedure which was flexible enough to approach our research questions while brief enough so that it could be easily and systematically deployed. As such, before conducting our experiments, we executed a total of 9 pilot tests with player dyads ($n = 18$), observing how players interacted with the game and with each other (see Figure 3). Some gameplay metrics like player scores and the difficulty of delivered orders were collected, yet they were not used because of a less strict setting in which players could give researchers some feedback when facing an issue. At the start of these tests, we provided players with materials describing the game to inform their gameplay. The pilot tests allowed us to further perfect the game and gradually adjust our experimental process.

As a consequence, we identified and fixed several game issues that could affect our evaluation. Particularly, we simplified how a player interacted with ingredients, processors, and utensils (for instance, dragging instead of clicking to both grab and release an ingredient); we enhanced feedback animations and sounds; and, to avoid errors, we forced the game modes to be unlocked according to their order in our experiment.

We also gradually improved the experimental process over these tests. For instance, printed textual descriptions of the game and its levels were replaced by a video and by graphical representations integrated into a physical divider, an element that served to separate the participants and, at the same time, contained instructions on how to build the recipes. The final experimental process is described next.

Experimental Process

To address the questions of the present study, we prepared an experiment room according to the layout presented in Figure 4. We will use the areas signalled in the image to illustrate our explanations. A television to display a video explaining *Alien Bar* was positioned in a 'pre-game area' (area A), and two computers with *Alien Bar* installed (as well as two pens) were positioned approximately 1 meter apart at the B.1 and B.2 areas. An observation area was

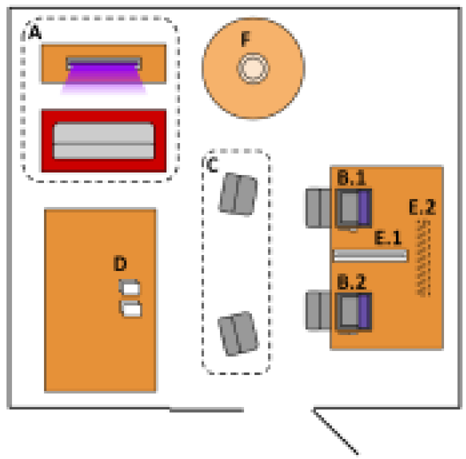


Figure 4: Layout of the experiment room.

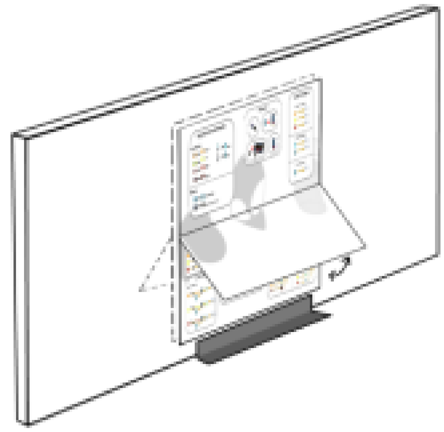


Figure 5: Representation of the divider with *Alien Bar* instructions.

also prepared for researchers, positioned behind both participants (area C). Hidden response sheets were added (elements labelled with D) to register participants' preference estimates. A crucial element for the execution of our experiment was a physical divider containing the game elements and recipes (Figure 5) that could be positioned in two different ways (E.1 or E.2 of Figure 4). Finally, we included some candy to give to the participants (area F). All the materials associated with the experiment, including the (anonymous) processed data, its analysis method, and forms, are available online[4].

Each experiment took around 35 minutes in total, going through four phases: *Experiment Introduction*, *Single-Player Tutorial*, *Team Training*, and *Single-Player Survival and Debriefing* (see Figure 6). While we measured the ability of a player by their final game score in the survival phase, a player's subjective experience was assessed resorting to their responses to questions of the Intrinsic Motivation Inventory (IMI) (Ryan 1982; Ryan et al. 1983; Center for Self-Determination Theory 2024) while remembering their experience in *Team Training*. Even though, in total, IMI comprehends seven experience dimensions, it is possible to consider a subset of these dimensions. As such, we deemed *Interest/Enjoyment* and *Perceived Competence* to be the most relevant dimensions for our study. All the collected independent and dependent variables are listed and described in Tables 1 and 2. Accounting for the layout presented in Figure 4, our experimental procedure is detailed below:

1 *Experiment Introduction (10 minutes)*

1.1 In the 'pre-game area' (area A), the researchers introduced themselves and the general procedure to participants, without going into details about the game. The video explaining *Alien Bar* was also presented in this step, and participants were given the opportunity to clarify any questions they had (this took around 5 minutes).

4. https://osf.io/2e54f/ (accessed 5th June 2025).

Table 1: Independent Variables

Variable	Description
$VideoGameFamiliarity$	A player's familiarity with video games, measured through the question: 'How often do you play video games?', answered either 'I make some time in my schedule to play video games'; 'I play video games occasionally when the opportunity presents itself'; or 'I do not play video games'.
$GameGenreEnjoyment$	A player's enjoyment of the target game genre, measured through the question: 'Do you enjoy games where you create recipes on demand (such as the "Cook, Serve, Delicious!" series)?', answered either 'I enjoy them and have played/watched others play them multiple times'; 'I played/watched others play them enough to understand I do not appreciate them'; or 'I am not familiar with these games and/or have no formed opinion on them'.
$TutorialScore$	The score obtained by a player in the *Tutorial* phase, automatically recorded by the game.
$Focus_{Rep}$	A player's self-reported pre-game *Focus*, measured as 'When given a task you prefer to...', responded through a 9-point Likert scale, ranging from '... be left alone to focus on the task.' to '...interact with the others, but also focus on the task.' (middle of the scale), finishing with '...interact with others disregarding the task.'.
$Challenge_{Rep}$	A player's self-reported pre-game *Challenge*, measured as 'When given a task you prefer that...', responded through a 9-point Likert scale, from '... the task provides an easy path for its completion.' to '...the task provides a difficult path for its completion.'.
$GroupFocus_{Rep}$	A player's self-reported pre-game estimate of what *Focus* should be promoted in the group for it to succeed (measured via the same options as $Focus_{Rep}$).
$GroupChallenge_{Rep}$	A player's self-reported pre-game estimate of what *Challenge* should be promoted in the group for it to succeed (measured via the same options as $Challenge_{Rep}$).
$Focus_{Obs}$	A player's focus while training, obtained via external observations made by the researchers. Similar to some behaviour observation protocols (Smith et al. 2013), the researchers annotated behavioural data, in this case the players' focus, for every 2-minute periods. The considered levels were: 'Alone focusing on own task' (level 1); 'Interact with others but also focus on own task' (level 2); or 'Interact with others disregarding own task' (level 3). The focus of a player was the average of the levels of all 2-minute periods.
$Challenge_{Obs}$	The (observed) challenge level of a player while training, measured through the in-game collection of the trained difficulty levels. Consider a set of trained levels T of size s: $\{T_1, ..., T_s\}$, a set containing the duration of trained levels D: $\{D_1, ..., D_s\}$, and the total duration td. A player's $Challenge_{Obs}$ value was computed as: $Challenge_{Obs} = \frac{[\sum_{i=1}^{s}(T_i \times \frac{D_i}{td})]-1}{4}$
$InterpersonalCloseness$	The interpersonal proximity between the participants (Gino and Galinsky 2012; Dubois et al. 2016), as rated by a player through the question: 'How well do you know your experiment partner?', answered via a 5-point Likert scale from 'Perfect Stranger' to 'Close Friend/Relative'.

Table 2: Dependent Variables

Variable	Description
$Interest/Enjoyment$	Experience dimension measured through the IMI questions for this component, ranked from 1 to 7.
$PerceivedCompetence$	Experience dimension measured through the IMI questions for this component, ranked from 1 to 7.
$SurvivalScore$	The final score of the survival phase, automatically recorded by the game.

Figure 6: Our experiment phases.

1.2 Each participant then sat next to the computers (areas B.1 and B.2), agreed and applied to an informed consent, and afterwards filled some demographic data (including $VideoGameFamiliarity$ and $GameGenreEnjoyment$). The current and next steps of this phase lasted for approximately 5 minutes.

1.3 In the same form, participants defined their $Focus_{Rep}$ and $Challenge_{Rep}$ (measured as described in Table 1) and then used their pen to copy these responses to a hidden preference sheet.

1.4 Then, a researcher gathered the hidden preference sheets and, without participants' notice, switched them (it was presented as if the switched preferences came from a random pool).

1.5 Given the preference of the other player, each participant estimated $GroupFocus_{Rep}$ and $GroupChallenge_{Rep}$.

2 *Single-Player Tutorial (3 minutes)*

2.1 Each subject individually played the *Tutorial* version of the game (only recipes of level 1) for 3 minutes, using the divider in position E.1 as a reference. At this point, the divider was set to hide the recipes of levels higher than 1.

3 *Team Training (15 minutes)*

3.1 A researcher placed the divider in position E.2 and uncovered the recipes of levels 2, 3, 4, and 5 to each participant (this took a few seconds).

3.2 Afterwards, the players collectively studied and played the *Training* version for 10 minutes (at this point, the participants were informed that they could freely interact with each other if they desired). At the same time, the two researchers annotated, for each 2-minute periods, the $Focus_{Obs}$ of each player on each other player, and the game recorded how many times and how long each training level was played (used to obtain $Challenge_{Obs}$).

3.3 After training, a researcher placed the divider in position E.1, and each player individually responded to the IMI questions regarding their $Interest/Enjoyment$ and $PerceivedCompetence$ while training (taking around 5 minutes).

4 *Single-Player Survival and Debriefing (7 minutes)*

4.1 In the last phase, each participant individually played the *Survival* version of the game for a maximum of 5 minutes. The $SurvivalScore$ achieved during this phase was stored by the game.

4.2 By the end of the experiment, each player was debriefed and rewarded with candy (element F) for their expended time. This took around 2 minutes.

Evaluation

The final evaluation focused on the use of *Alien Bar* to ascertain whether interaction preferences or other intrinsic factors can influence player ability and experience. Keeping this in mind, subjects were recruited in pairs, through standard convenience sampling procedures such as direct contact and through word of mouth. Most of the participants were invited during an international event held at one of our universities, and so our sample included subjects from 10 countries spread across Europe, America, Africa, and Asia. There were no potential risks and no anticipated benefits to individual participants. We conducted a total of 32 tests, comprising 64 participants. However, the participants of one test skipped an experiment step, and so we excluded them from the statistical analysis. Thus, our final data set comprises 31 tests with a total of 62 participants (42 males, 19 females, and 1 preferred not to say), between 16 and 44 years old (M = 25.29; SD = 6.56).

Regarding video game culture, 51.6% of our participants reported to make some time in their schedule to play video games, 21.0% reported to play video games when the opportunity presents itself, and 27.4% reported to not play video games. Additionally, 48.4% of our participants reported to not be familiar or have no formed opinion about games where the player creates recipes on demand, 40.3% reported to have played/watched others play them enough to understand they do not appreciate them, and only 11.3% reported to enjoy them and to have played/watch others play multiple times.

Data Transformation and Synthesis

We applied some transformations to our data in order to more easily analyse it[5]. Firstly, we transformed the preference variables (related to $Focus$ and $Challenge$) into three factors: $Preference_{Obs}$, $Preference_{Rep}$, and $GroupPreference_{Rep}$; each with four levels representing if the $Focus$ and/or $Challenge$ values were below or above their median values[6].

5. Because the exclusion of one test was due to skipping a step instead of any incidents during the completed steps, this initial transformation considered the data available from all tests to more accurately represent the population groups.

6. The equal values were considered in the above/high level.

This allowed us to build a model adjusted to our setting and that follows the quadrants of Figure 2: $\{$*Self-Facilitator*; *Self-Challenger*; *Others-Facilitator*; *Others-Challenger*$\}$ (normalised median values: $Mdn_{Focus_{Rep}} = 0.5$, $Mdn_{GroupFocus_{Rep}} = 0.5$, $Mdn_{Focus_{Obs}} = 0.4$, $Mdn_{Challenge_{Rep}} = 0.625$, $Mdn_{GroupChallenge_{Rep}} = 0.625$, $Mdn_{Challenge_{Obs}} = 0.453$). We also added peer preference variables, e.g., $OtherFocus_{Rep}$, $OtherChallenge_{Rep}$, and $OtherPreference_{Obs}$, so that we could measure the effects of the other player's preference in a given player's ability and experience. After logged and then compiled using the *R* statistical software version 4.4.2 (R Core Team 2024), the data collected throughout these experiments was analysed using the *IBM SPSS Statistics* software[7], version 28.

The other independent variables, $TutorialScore$ and $InterpersonalCloseness$, were considered factors with two levels: $\{$*Low*; *High*$\}$, depending on whether the values were below or above their median value[8] (median values: $Mdn_{TutorialScore} = 5250$, $Mdn_{InterpersonalCloseness} = 3$). Finally, $VideoGameFamiliarity$ was divided depending on whether the player plays or does not play video games: $\{$*Does Not Play*; *Plays*$\}$ and $EnjoysGameGenre$ was divided depending on whether the player likes, or either does not like or has no formed opinion about the game genre: $\{$*No/No Opinion*; *Yes*$\}$.

After processing the data, its normality was tested resorting to the Kolmogorov-Smirnov statistic with Lilliefors correction. Because the test revealed some non-normal data and due to the reduced sample size when dividing the data per condition, the influence of the different preference variables in the dependent measures: a player's ability ($SurvivalScore$) and experience ($Interest/Enjoyment$ and $PerceivedCompetence$), related to RQ1, were assessed by several (non-parametric) Kruskal-Wallis tests, one for each dependent measure. Each Kruskal-Wallis test was followed by a multiple comparison of the rank means, as presented by Marôco (Marôco 2021). The size of each Kruskal-Wallis effect was assessed using η^2_H. Following social sciences standards (Marôco 2021), an effect with: $\eta^2_H \leq 0.05$ was considered small; $0.05 < \eta^2_H \leq 0.25$ was considered moderate; $0.25 < \eta^2_H \leq 0.5$ was considered large; and $\eta^2_H > 0.5$ was considered very large.

To verify the subjects' estimation accuracy (RQ2), we directly compared the distributions of the estimation and training data. To complement the analyses, we also performed several Spearman's rank-order correlation tests (Bhandari 2023) relating the three performance-oriented variables: $TutorialScore$, $PerceivedCompetence$, and $SurvivalScore$. Following general interpretations for correlations coefficients (Bhandari 2023) and considering the Spearman's ρ, we assumed correlations with $0 < |\rho| \leq 0.3$ to be weak (none if 0); $0.3 < |\rho| \leq 0.5$ to be moderate; $0.5 < |\rho| \leq 0.7$ to be strong; and $0.7 < |\rho| < 1$ to be very strong (perfect if 1).

Finally, as also presented by Marôco (Marôco 2021), several (non-parametric) Wilcoxon-Mann-Whitney tests were executed to test RQ3, that is, the influence of the other variables: $VideoGameFamiliarity$, $GameGenreEnjoyment$, and $InterpersonalCloseness$; in a player's ability and experience. The size of each Wilcoxon-Mann-Whitney effect was assessed using Cohen's d. Following social sciences standards (Marôco 2021), an effect with: $d \leq 0.2$ was considered small; $0.2 < d \leq 0.5$ was considered moderate; $0.5 < d \leq 1$

7. https://www.ibm.com/analytics/spss-statistics-software (accessed 5th June 2025).
8. Like before, the equal values were considered in the above/high level.

was considered large; and $d > 1$ was considered very large.

The results of our analyses were divided into the aspects approached by our research questions. Firstly, we present the tests related to RQ1 and RQ2, that verified the influence of interaction preferences in a player's ability and experience. Then, we present some tests approaching RQ3, that is, the influence of the other variables, $VideoGameFamiliarity$, $GameGenreEnjoyment$, and $InterpersonalCloseness$, in the same metrics. In all the included box plots (e.g., Figure 7), distributions noted with different letters are significantly different according to the performed Kruskal-Wallis test and pairwise rank mean comparisons (Bonferroni-adjusted significant differences are highlighted in bold); and in the heatmaps (e.g., Figure 9), the numbers represent the frequencies for each combination, and darker cells contain higher frequencies. All analyses consider a significance level (α) of 0.05.

Effects of Interaction Preferences

Firstly, our results showed that the $Interest/Enjoyment$ measure did not vary significantly with $Preference_{Obs}$ ($H(3) = 5.352; p = 0.148$). $Interest/Enjoyment$ also did not vary significantly with $OtherPreference_{Obs}$ ($H(3) = 1.070; p = 0.784$). Thus, overall, enjoyment was not influenced by the way that subjects preferred to interact.

$PerceivedCompetence$ varied significantly with $Preference_{Obs}$, presenting a moderate effect ($H(3) = 9.963; p = 0.019; \eta^2_H = 0.120$). As indicated by the non-adjusted pairwise rank mean comparisons and by the data distributions (see Figure 7), the *Self-Facilitator* values were significantly lower than the *Self-Challenger* ones ($p = 0.022$), and lower than the *Others-Facilitator* ones ($p = 0.039$); and the *Others-Challenger* values were significantly lower than the *Self-Challenger* ones ($p = 0.017$) and lower than the *Others-Facilitator* ones ($p = 0.032$). These results, however, have to be considered carefully as the Bonferroni correction suggests that none of them reveals true significance ($p_{adj} > 0.05$). Even so, the fact that neither the *Self-Facilitator* differed from the *Others-Challenger* ($p = 0.906$), nor the *Self-Challenger* differed from the *Others-Facilitator* ($p = 0.733$) suggests a connection between the profiles of each of these pairs for perceived competence. Regarding the influence of players' partners, $PerceivedCompetence$ did not vary significantly with $OtherPreference_{Obs}$ ($H(3) = 4.758; p = 0.190$), suggesting that the ways of others' interaction may not be as important as the own profile for a subject's assessment of their competence.

$SurvivalScore$ varied significantly with $Preference_{Obs}$, presenting a moderate effect ($H(3) = 9.522; p = 0.023; \eta^2_H = 0.112$). According to the non-adjusted pairwise rank mean comparisons and as depicted by the data distributions (see Figure 8a), the *Self-Challenger* values were significantly higher than the *Others-Challenger* ($p = 0.015$), the *Self-Facilitator* ($p = 0.016$), and the *Others-Facilitator* ($p = 0.007$) ones. Even so, these results have to be considered carefully as the Bonferroni correction suggests only the latter difference (between *Self-Challenger* and *Others-Facilitator*) to reveal true significance ($p = 0.007, p_{adj} = 0.041$). The performance of the players in the survival game also varied significantly with $OtherPreference_{Obs}$, presenting a moderate effect ($H(3) = 9.020; p = 0.029; \eta^2_H = 0.104$), see Figure 8b. The non-adjusted pairwise rank mean comparisons present effects that go in-line with the ones discovered for self pref-

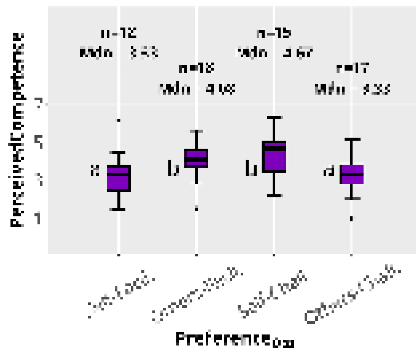
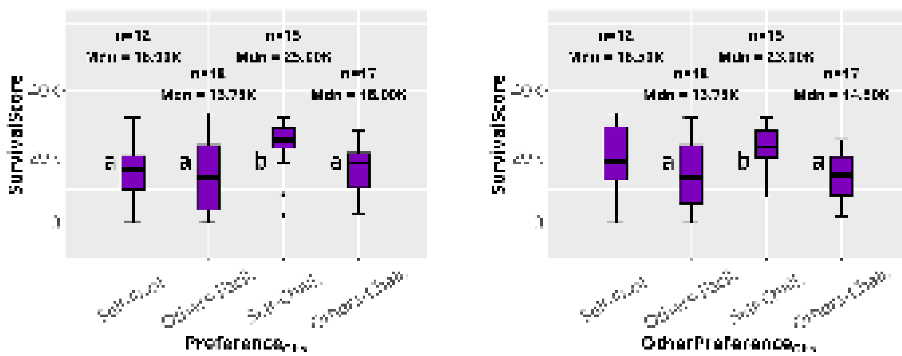


Figure 7: Distribution of the $PerceivedCompetence$ metric for each level of $Preference_{Obs}$.



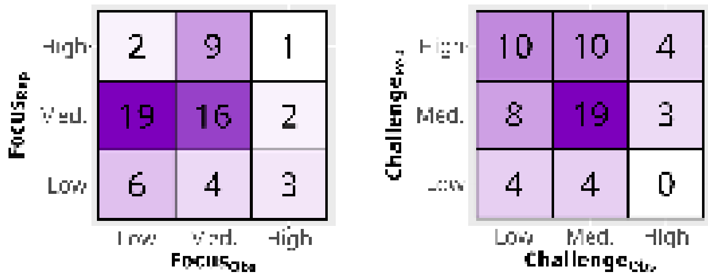
(a) (b)

Figure 8: Distribution of the $SurvivalScore$ metric for each level of $Preference_{Obs}$ (Figure 8a) and $OtherPreference_{Obs}$ (Figure 8b).

erence: the *Self-Challenger* values were significantly higher than the *Others-Facilitator* ($p = 0.014, p_{adj} > 0.05$), and the *Others-Challenger* ($p = 0.007, p_{adj} = 0.040$) ones. Thus, regarding $SurvivalScore$, the tests indicated that the *Self-Challenger* players had more positive outcomes, while also positively impacting the practical training outcomes of their partners.

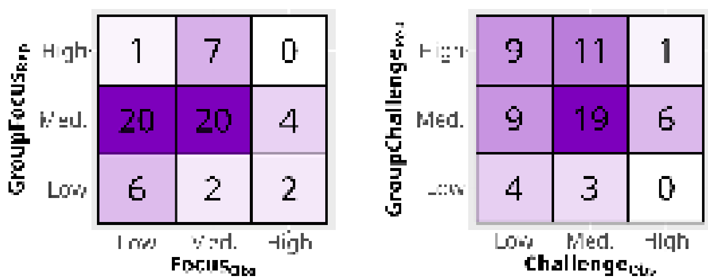
To verify the subjects' pre-game preference estimation accuracy (RQ2), we directly compared the distributions of the estimation and training data, while dividing the values of each preference dimension into three ordinal levels: {*Low*; *Med.*; *High*} (see Figures 9 to 12)[9]. Even though most of the data clustered around the middle of the scale, which may relate to the fallback subjects resort to when lacking task information, some interesting trends can be perceived. Only 37.1% of the $Focus_{Rep}$ and 43.5% of the $Challenge_{Rep}$ data matched $Focus_{Obs}$ and $Challenge_{Obs}$, respectively (Figure 9 plots the data distributions); and only 41.9% of the $GroupFocus_{Rep}$ and 38.7% of the $GroupChallenge_{Rep}$ data matched $Focus_{Obs}$ and $Challenge_{Obs}$, respectively (see Figure 10). Even so, higher

9. Assuming values between 0 and 1, each dataset was divided according to 3 buckets: {*Low*: [0,1/3[; *Med.*: [1/3,2/3[; *High*: [2/3,1]}.



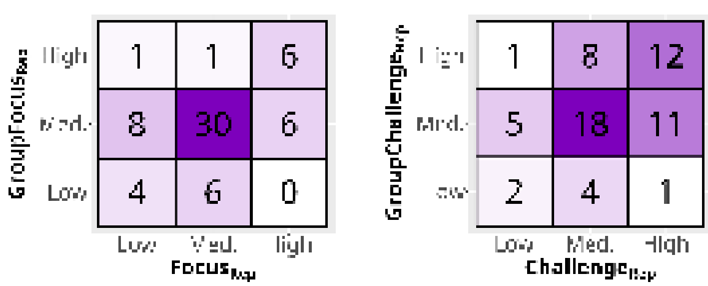
(a) (b)

Figure 9: Heatmaps obtained by crossing the results of $Focus_{Obs}$ with $Focus_{Rep}$ (Figure 9a), and $Challenge_{Obs}$ with $Challenge_{Rep}$ (Figure 9b).



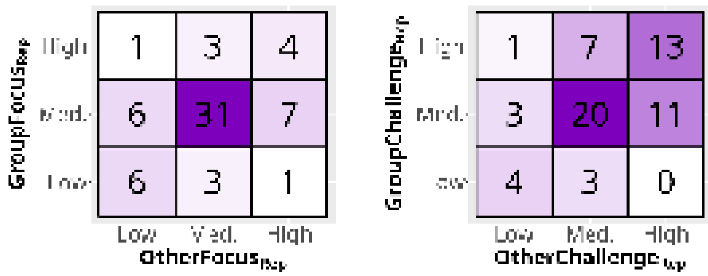
(a) (b)

Figure 10: Heatmaps obtained by crossing the results of $Focus_{Obs}$ with $GroupFocus_{Rep}$ (Figure 10a), and $Challenge_{Obs}$ with $GroupChallenge_{Rep}$ (Figure 10b).



(a) (b)

Figure 11: Heatmaps obtained by crossing the results of $Focus_{Rep}$ with $GroupFocus_{Rep}$ (Figure 11a), and $Challenge_{Rep}$ with $GroupChallenge_{Rep}$ (Figure 11b).



(a) (b)

Figure 12: Heatmaps obtained by crossing the results of $OtherFocus_{Rep}$ with $GroupFocus_{Rep}$ (Figure 12a), and $OtherChallenge_{Rep}$ with $GroupChallenge_{Rep}$ (Figure 12b).

data match rates were verified between $GroupFocus_{Rep}$ and either $Focus_{Rep}$ (64.5%) or $OtherFocus_{Rep}$ (66.1%); and between $GroupChallenge_{Rep}$ and either $Challenge_{Rep}$ (51.6%) or $OtherChallenge_{Rep}$ (59.7%). Indeed, in such cases, the values tended to proportionally approximate (see Figures 11 and 12). This reveals that subjects may have estimated the group preferences while considering their own and their peer's preferences, but at the same time, these estimations were not accurate indicators of training behaviour, which in turn suggests an effect of the task itself.

Aligned with the aforementioned results, we verified no significant effects of $Preference_{Rep}$ in: $Interest/Enjoyment$ $(H(3) = 0.708; p = 0.871)$, $PerceivedCompetence$ $(H(3) = 4.411; p = 0.220)$, or $SurvivalScore$ $(H(3) = 4.264; p = 0.234)$. Besides, a lack of significant effects was also verified for $GroupPreference_{Rep}$: $Interest/Enjoyment$ $(H(3) = 0.657; p = 0.883)$, $PerceivedCompetence$ $(H(3) = 6.415; p = 0.093)$, and $SurvivalScore$ $(H(3) = 5.548; p = 0.136)$. This indicates that experience and performance did not vary with the initial estimation that players made of their ways of interacting, relating instead to their behaviour while training.

Finally, even though the players' preference estimations were not relevant to predict ability, Spearman's Rank-Order Correlation tests indicated moderate significant correlations of $PerceivedCompetence$ with $TutorialScore$ $(\rho(60) = 0.328, p = 0.009)$ and $SurvivalScore$ $(\rho(60) = 0.434, p < 0.001)$, as well as a strong correlation between $TutorialScore$ and $SurvivalScore$ $(\rho(60) = 0.657, p < 0.001)$, indicating that subjects' initial skill and their perception of competence during training are both good outcome estimators.

Testing Combinations of Focus and Challenge

To further test the influence of players' preferences, we measured the effects of different combinations of $Focus_{Obs}$ (in Self: S or Others: O) and $Challenge_{Obs}$ (Facilitate: F or Complicate: C) in a player's ability and experience. For instance, we represented the data of a self-oriented player playing with an others-oriented one from the perspective of the self-oriented player as *S w/ O* and from the perspective of the others-oriented player as *O w/ S*. We derived the factors: $FocusCombination$, endowed with four levels: {*S w/ S*;

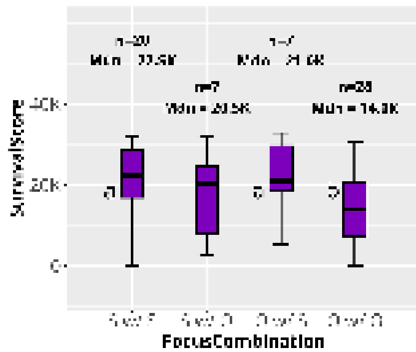


Figure 13: Distribution of the $SurvivalScore$ of a player divided by the combination of $Focus_{Obs}$ present in the player's group.

S w/ O; *O w/ S*; *O w/ O*$\}$; and $ChallengeCombination$, also endowed with four levels: $\{$*F w/ F*; *F w/ C*; *C w/ F*; *C w/ C*$\}$[10]. From all the results, we only verified a significant (moderate) effect of $FocusCombination$ ($H(3) = 8.592; p = 0.035; \eta^2_H = 0.096$) in the $SurvivalScore$ acquired by a given player. The pairwise rank mean comparisons and the data distribution (see Figure 13) suggest that the ability values when the two players were *others-oriented* (*O w/ O*) was significantly lower than when they both were *self-oriented* (*S w/ S*) ($p = 0.010$), and when the own player was *others-oriented* and the other was *self-oriented* (*O w/ S*) ($p = 0.038$). Even though these significance values have to be considered carefully as their adjusted values are non-significant according to the Bonferroni correction for the four groups, by observing the data distributions, it is clear to verify that *O w/ O* achieved the lowest score median and contrasted with the concentration of higher values in *S w/ S* and *O w/ S*.

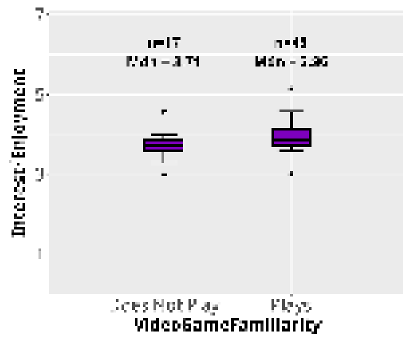
Impact of Video Game Familiarity, Game Genre Enjoyment, and Interpersonal Closeness

Regarding the metrics $VideoGameFamiliarity$, $GameGenreEnjoyment$, and $InterpersonalCloseness$, we also found multiple effects relating to teamwork quality. Given that we predicted the data distributions to shift to one of the sides, we considered one-tailed significance levels in our Wilcoxon-Mann-Whitney tests[11].

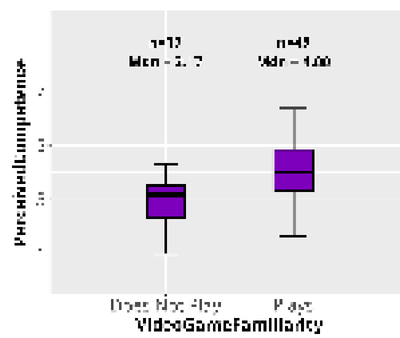
As expected, we verified a positive significant effect of $VideoGameFamiliarity$ in all the dependent variables (see Figure 14), notably a moderate effect with $Interest/Enjoyment$ ($U = 203.0; p = 0.002; d = 0.469$), a large effect with $PerceivedCompetence$ ($U = 173.5; p < 0.001; d = 0.546$), and a moderate effect with $SurvivalScore$ ($U = 241.0; p = 0.012; d = 0.370$), meaning that subjects who reported playing video games were more likely to achieve better results and to have a better experience while playing our game. Another expected result was the large significant effect of $GameGenreEnjoyment$ in $Interest/Enjoyment$ ($U = 93.5; p = 0.012; d = 0.514$), indicating that subjects enjoyed the game more when they reported to like its genre (see Figure 15).

10. Other tests were conducted by dividing these factors into two levels: $\{$*Similar*; *Different*$\}$, depending on whether the players presented similar (*S w/ S*, *O w/ O*) or distinct (*S w/ O*, *O w/ S*) $Focus_{Obs}$ or $Challenge_{Obs}$ levels. Even so, these tests are omitted due to their lack of significant results.

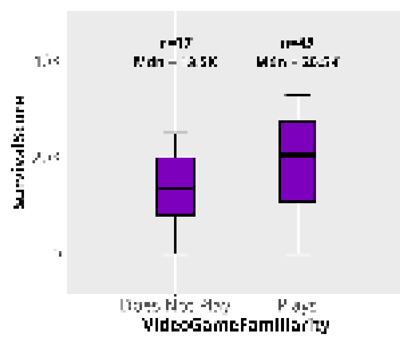
11. Nonetheless, as can be perceived, the two-tailed significance levels allow the same interpretations.



(a)



(b)



(c)

Figure 14: Distribution of $Interest/Enjoyment$ (Figure 14a), $PerceivedCompetence$ (Figure 14b), and $SurvivalScore$ (Figure 14c) by $VideoGameFamiliarity$.

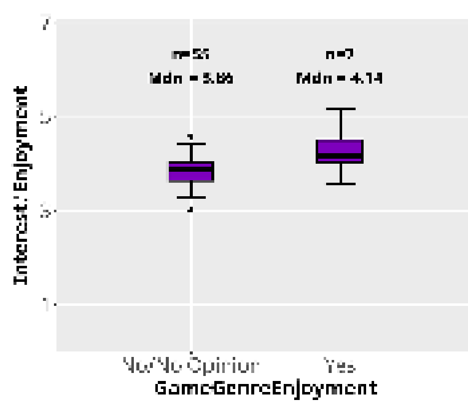


Figure 15: Distribution of $Interest/Enjoyment$ by $GameGenreEnjoyment$.

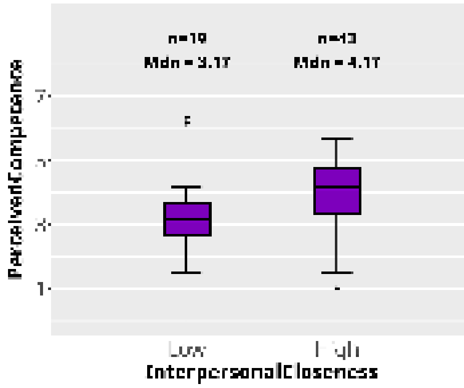


Figure 16: Distribution of $PerceivedCompetence$ by $InterpersonalCloseness$.

Interestingly, the only significant $InterpersonalCloseness$ effect that we found was a moderate effect of the metric with $PerceivedCompetence$ ($U = 243.0; p = 0.005; d = 0.405$), see Figure 16, meaning that the subjects' perception of their competence was the only factor affected by how well the subjects rated they knew each other. Curiously, enjoyment or demonstrated ability were not significantly affected by this factor, suggesting that players' interpersonal proximity perception did not influence their enjoyment or outcomes.

Discussion

In this study, we aimed to test how subjects with different interaction preferences fare when performing a task together. Thus, we deployed the puzzle game *Alien Bar* and used it as a training task to assess how acquired ability and experience varied when joining different profiles of players. Ability was measured resorting to the score obtained in a survival version of the game played at the end of the experiment, and a player's experience while training the game with a peer was measured via a questionnaire approaching two IMI dimensions: $Interest/Enjoyment$ and $PerceivedCompetence$. The most relevant results extracted from these tests are summarised in Figure 17.

Firstly, we concluded that enjoyment was not significantly influenced by the way that subjects interacted in our training setting. Even so, subjects' perception of competence and demonstrated ability were affected by their interaction preferences. We verified that, for $PerceivedCompetence$, *Self-Challenger* did not differ from *Others-Facilitator* subjects, and *Self-Facilitator* did not differ from *Others-Challenger* subjects, and that these two profile pairs diverged from each other, notably that the former pair achieved higher values on this metric (RQ1). Although these differences did not sustain after a Bonferroni correction, we still believe they reveal interesting tendencies. Particularly, we believe we can apply these findings by grouping work colleagues endowed with the profiles of the former pair. Said is reasonable in a similar (Insko et al. 1973) or complementary (Dryer and Horowitz 1997) fashion. On the one hand, while *Self-Challenger* subjects may feel competent when training together because of their non-disruptive nature towards others, or *Others-Facilitator* may feel competent when guiding each other; a *Self-Challenger* may also feel especially competent when provided the support of an *Others-Facilitator*, and vice-versa, an *Others-Facilitator* may feel especially competent when guiding a self-challenging subject.

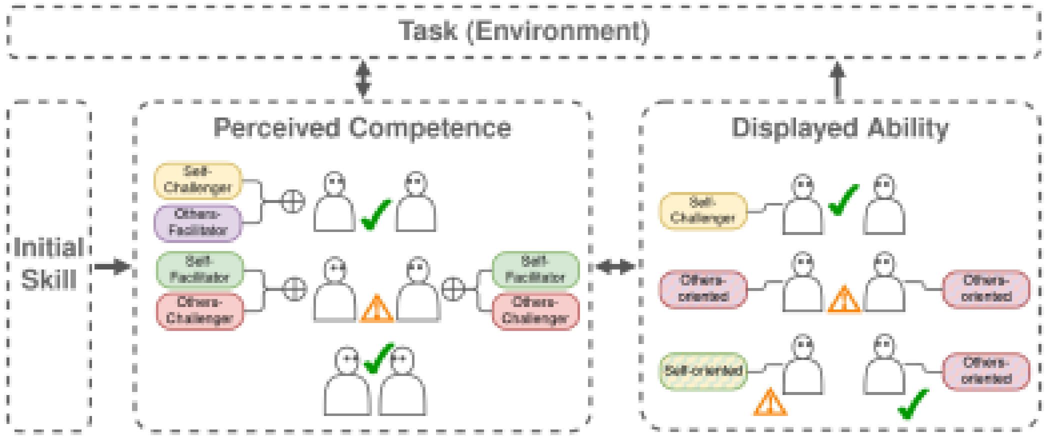


Figure 17: Scheme depicting our most important results. The labels next to the avatars represent subjects' preferences as acquired from the training phase (not their pre-game estimates). The two avatars positioned closer signify higher interpersonal closeness. The green 'check' signs were added to the profile pairings or subjects that were deemed to develop the respective metric; and the orange 'warning' signs were included for the ones that were deemed unable to develop the respective metric.

On the other hand, care should be taken when incorporating *Self-Facilitator* and *Others-Challenger* subjects in teams because of their tendency towards lower perceived competence: whenever possible, it may be preferable to distribute subjects with these profiles into different groups.

Besides the preference measures, $InterpersonalCloseness$ also had a singular effect, a positive one, on subjects' perception of their competence (RQ3). This is a particularly interesting result given that enjoyment may, at a first glance, be deemed to associate to interpersonal relations as well. Nonetheless, $Interest/Enjoyment$ was positively influenced by $VideoGameFamiliary$ and $GameGenreEnjoyment$. It may be that, while completing a gaming, training, or working task, interpersonal closeness comes into play in the form of making oneself or others feel fulfilled and competent, and that the enjoyment of individuals while performing a certain task is more predominantly associated with the players' affinity with the task and not as connected to how close peers feel to each other. In fact, aligned with related research (Rodríguez Montequín et al. 2013), we did not observe an effect of $InterpersonalCloseness$ in the survival game scores.

Regarding survival game performance, the results suggested that *Self-Challenger* players achieved higher $SurvivalScore$ while also contributing to their peers' $SurvivalScore$ success (RQ1). This may happen because their disposition for a faster, less progressive, learning behaviour may have an implicit leading and responsible connotation that motivates other players to succeed in a time-limited and individually-executed training task like in *Alien Bar*. Interestingly, these conclusions align with the remark that students who performed best were deemed more suited to solitary employment (Ahmed et al. 2010).

Further testing for effects between combinations of players' preferred focus and challenge levels in $SurvivalScore$ showed that a team joining *others-oriented* players developed lower ability than in all other pairing conditions (RQ1). This suggests that care should be taken when grouping *others-oriented* individuals because the groups may lack the application of actions and strategies of self-improvement, notably in a time-limited and individually executed training task such as ours. The results also informed that the particular case of having an *others-oriented* subject joined with a *self-oriented* partner benefited the *others-oriented* player more than the *self-oriented* one. Coupled with the revealed $PerceivedCompetence$ benefits for *Self-Challenger* and *Others-Facilitator* elements, such a particular case may signify that, after perceiving higher competence, there is an emergence of the protégé effect: learning by guiding a peer (Chase et al. 2009), and that this effect can overlap the insight coming from the *others-oriented* towards the *self-oriented* player. Nonetheless, further tests with a higher sample size may help better determine this aspect.

By directly comparing the pre-game player estimations with training data, we also posited that subjects may have estimated the group preferences while considering both their own and their peer's preferences, but at the same time, we also deemed that these estimations were not accurate indicators of training behaviour (RQ2). Additional tests identified no effects of the subjects' own preference estimations in their ability and experience, further corroborating the idea that the subjects' pre-task judgments of how they interact and how their group should successfully interact are shallow indicators of the potential of the group. This intuitively makes sense because the task and its surrounding environment may help shape the interaction and subjects cannot account for that beforehand. Nevertheless, tests correlating the three performance-oriented measures: $TutorialScore$, $PerceivedCompetence$, and $SurvivalScore$, demonstrated that initial skill and perceived competence were good indicators of self-ability potential, further aligning with our previous claims. Other tests also revealed that participants who reported playing video games were more likely to achieve better results and experience, and that participants who reported liking the game genre, enjoyed it more (RQ3). Thus, subjects may be able to estimate own task success based on the frequency and liking in doing similar tasks, their initial skill, or perception of competence during training, but although interaction preferences impact success, subjects may not accurately predict how such happens beforehand. We believe that this aspect connects to the drawbacks of self-picked teams (Rutherfoord 2001) as their elaboration usually does not take into account more intrinsic traits (personality diversity being identified by the reviewed research) and instead relies on more objective and easily comparable dimensions.

We believe that the presented findings can be considered by instructors as well as game designers, to help develop or tune automatic systems for group management and game matchmaking. The latter challenges usual game matchmaking approaches which tend to focus heavily on demonstrated player skill and not as much on how to proportionate a fulfilling experience using players' intrinsic preferences. For instance, an automated system may consider grouping subjects deemed as *Self-Challenger* and *Others-Facilitator* and avoid joining subjects deemed as *Self-Facilitator* and *Others-Challenger* to help create a more fulfilling environment that foments perceived competence. Additionally, joining subjects deemed as socially close, e.g., identified as friends, should also promote perceived competence. Even so, the previous aspects do not directly address actual team performance. To promptly foment positive outcomes, an automated system may propose groups that include *self-directed* subjects who prefer challenge, and avoid grouping too many *others-oriented* subjects because the group may lack the application of actions and strategies for self-improvement. Still, the system may acknowledge that the inclusion of an *others-oriented* subject next to a *self-oriented* one proved beneficial for the *others-oriented* subject.

Limitations

While we extracted multiple interesting results from our tests, we also recognise some limitations of the present work. Firstly, although our full sample size meets the usual requirements of social sciences, we acknowledge that dividing our sample according to the needed conditions inevitably decreased the power of our statistics. We also acknowledge that other data analysis criteria can affect the magnitude of some relations. Thus, future work may test if the same relations can be verified in other settings, namely while considering a larger sample. Another sample-related issue is the fact that, due to practicality, we neither controlled for nor enforced any kind of balance between pairs regarding their members' demographics or video game familiarity, skill, and liking. Also due to practicality, we did not control for the social relations between the members of each pair beforehand.

Another relevant limitation is the fact that, even though we tried our experimental setup to be as interaction-neutral as possible, i.e., not biasing players to perceive a certain training interaction style, the observed effects may still have been influenced by some characteristics of our training task. For instance, the fact that the game was played individually may have conditioned the perception of competence and performance by players oriented towards others. Also, the fact that training had a 10-minute time limit may have favoured more challenging profiles that selected higher difficulty levels from the start, although we believe that such a duration was enough for players to acquire all the required knowledge, even for those who were more reluctant to increase challenge. Furthermore, although observation was considered necessary and area C of Figure 4 was deemed distant enough to avoid obtrusion, observation may still have had some influence on how players' behaved, as a form of the Hawthorne effect (McCambridge et al. 2014).

Accounting for the aforementioned aspects, further testing with a larger and more controlled sample, with a refined experimental procedure, using other styles of task, or with less time-restricted training, may be conducted to corroborate, refute, or complement our findings.

Conclusions

Given our time-expensive work lives, it is crucial to understand different factors that help foment work success and a fulfilling experience. The present work embraces this premise by studying how the pairing of people endowed with distinct interaction preferences influences the level of their acquired ability as well as their overall experience. Inspired by how several psychological models were used to profile teams and study their effectiveness, a model for characterising a subject's interaction preference was derived to suit the needs of the present study. Then, the ability and experience acquired by different pairs of subjects were recorded while they played the puzzle game *Alien Bar*, where a player combines fictitious ingredients to form recipes.

Using this setting, we managed to extract several results related to the perception and display of competence and that may be considered by instructors, automatic group management processes, and game matchmaking. Our results revealed team outcome benefits when including self-oriented challenger subjects alongside other profiles, and guided us to raise care when joining others-oriented members. Even so, these results may be, to some extent, connected to our setting, considering that our evaluation used an individually-executed and time-limited game task as a basis for team training. An additional finding was that interpersonal closeness can have an impact on the perceived competence of subjects, but not in their demonstrated ability or enjoyment, as the latter two aspects may be more predominantly influenced by subjects' affinity with the task.

In the future, the knowledge acquired here may be applied in an automatic process to join players or work colleagues based on their preferences. Several studies can also be conducted with other games or applications, or even considering the use of a similar protocol for tasks that better approximate a work setting, in which individuals can use their previous knowledge more thoroughly. Along the way, the applied interaction model can be refined to better suit new deployments. Overall, we believe that the findings we provided can contribute to the improvement of teamwork methodologies, leveraging the proliferation of success and well-being in working or training environments.

Acknowledgements

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with references SFRH/BD/143460/2019 and UIDB/50021/2020 - DOI: 10.54499/UIDB/50021/2020, as well as project CRAI with reference C628696807-00454142 (IAPMEI/PRR).

BIBLIOGRAPHY

Ahmed, Faheem, Piers Campbell, Ahmad Jaffar, Shayma Alkobaisi and Julie Campbell. 2010. 'Learning & personality types: A case study of a software design course'. *Journal of Information Technology Education: Innovations in Practice* 9 (1): 237–252.

Alves, Tomás, Samuel Gomes, João Dias and Carlos Martinho. 2020. 'The Influence of Reward on the Social Valence of Interactions'. In *2020 IEEE Conference on Games (CoG),* 168–175. Osaka, Japan: IEEE.

Bhandari, Pritha. 2023. *Correlation Coefficient | Types, Formulas & Examples.* https://www.scribbr.com/statistics/correlation-coefficient/. Accessed: 02-11-2023.

Bradley, Bret H, John E Baur, Christopher G Banford and Bennett E Postlethwaite. 2013. 'Team players and collective performance: How agreeableness affects team performance over time'. *Small Group Research* 44 (6): 680–711.

Campbell, Leigh. 2017. *We've Broken Down Your Entire Life Into Years Spent Doing Tasks.* https://www.huffpost.com/entry/weve-broken-down-your-entire-life-into-years-spent-doing-tasks_n_61087617e4b0999d2084fec5/. Accessed: 01-02-2023.

Center for Self-Determination Theory. 2024. *Intrinsic Motivation Inventory (IMI).* https://selfdeterminationtheory.org/wp-content/uploads/2022/02/IMI_Complete.pdf. Accessed: 2024-03-05.

Chase, Catherine C, Doris B Chin, Marily A Oppezzo and Daniel L Schwartz. 2009. 'Teachable agents and the protégé effect: Increasing the effort towards learning'. *Journal of science education and technology* 18:334–352.

Deci, Edward L. and Richard M. Ryan. 1985. 'Conceptualizations of Intrinsic Motivation and Self-Determination'. Chap. 2 in *Intrinsic Motivation and Self-Determination in Human Behavior,* 11–40. Boston, MA: Springer US. ISBN: 978-1-4899-2271-7. https://doi.org/10.1007/978-1-4899-2271-7_2.

Dryer, D Christopher and Leonard M Horowitz. 1997. 'When do opposites attract? Interpersonal complementarity versus similarity.' *Journal of personality and social psychology* 72 (3): 592.

Dubois, David, Andrea Bonezzi and Matteo De Angelis. 2016. 'Sharing with friends versus strangers: How interpersonal closeness influences word-of-mouth valence'. *Journal of Marketing Research* 53 (5): 712–727.

Gino, Francesca and Adam D. Galinsky. 2012. 'Vicarious dishonesty: When psychological closeness creates distance from one's moral compass'. *Organizational Behavior and Human Decision Processes* 119 (1): 15–26. ISSN: 0749-5978. https://www.sciencedirect.com/science/article/pii/S0749597812000477.

Gomes, Samuel, Tomás Alves, João Dias and Carlos Martinho. 2022. 'Reward-Mediated Individual and Altruistic Behavior'. In *Videogame Sciences and Arts,* edited by Inês Barbedo, Bárbara Barroso, Beatriz Legerén, Licínio Roque and João Paulo Sousa, 84–97. Cham: Springer International Publishing. ISBN: 978-3-030-95305-8.

Gomes, Samuel, Luis Costa, Carlos Martinho and João Dias. 2024. 'The Alien Bar Game: Puzzling Social Interaction Between Baristas'. *Special Issue for GALA 2024 Games Competition & Exhibition Book of Abstracts,* 36.

Henderson, Monika and Michael Argyle. 1985. 'Social support by four categories of work colleagues: Relationships between activities, stress and satisfaction'. *Journal of Organizational Behavior* 6 (3): 229–239. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/job.4030060306. https://onlinelibrary.wiley.com/doi/abs/10.1002/job.4030060306.

Insko, Chester A, Vaida D Thompson, Wolfgang Stroebe, Karen F Shaud, Berna E Pinner and Bruce D Layton. 1973. 'Implied evaluation and the similarity-attraction effect'. *Journal of Personality and Social Psychology* 25 (3): 297.

Lewis, Tracy L. and Wanda J. Smith. 2008. 'Building software engineering teams that work: The impact of dominance on group conflict and Performance outcomes'. In *2008 38th Annual Frontiers in Education Conference,* S3H-1-S3H–6. Saratoga Springs, NY, USA: IEEE.

Marôco, João. 2021. *Análise Estatística com o SPSS Statistics.* 8ª. Pêro Pinheiro, Portugal: ReportNumber, Lda.

Mazni, Omar, Sharifah-Lailee Syed-Abdullah and Naimah Mohd Hussin. 2010. 'Analyzing personality types to predict team performance'. In *2010 International Conference on Science and Social Research (CSSR 2010),* 624–628. Kuala Lumpur, Malaysia: IEEE.

McCambridge, Jim, John Witton and Diana R Elbourne. 2014. 'Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects'. *Journal of clinical epidemiology* 67 (3): 267–277.

McCrae, Robert R and Paul T Costa Jr. 2008. 'The Five-Factor Theory of Personality'. Chap. 5, 159–181. New York, USA: The Guilford Press.

Myers, Isabel Briggs, Mary H McCaulley, Naomi L Quenk and Allen L Hammer. 1998. *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator.* Vol. 3. California, USA: Consulting Psychologists Press Palo Alto, CA.

Ogot, Madara and Gül E. Okudan. 2006. 'The Five-Factor Model personality assessment for improved student design team performance'. *European Journal of Engineering Education* 31 (5): 517–529. eprint: https://doi.org/10.1080/03043790600797335. https://doi.org/10.1080/03043790600797335.

Omar, Mazni and Sharifah-Lailee Syed-Abdullah. 2010. 'Identifying effective software engineering (SE) team personality types composition using rough set approach'. In *2010 International Symposium on Information Technology,* 3:1499–1503. Kuala Lumpur, Malaysia: IEEE.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rodríguez Montequín, Vicente, José Manuel Mesa Fernández, J Villanueva Balsera and A García Nieto. 2013. 'Using MBTI for the success assessment of engineering teams in project-based learning'. *International journal of technology and design education* 23:1127–1146.

Rutherfoord, Rebecca H. 2001. 'Using Personality Inventories to Help Form Teams for Software Engineering Class Projects'. In *Proceedings of the 6th Annual Conference on Innovation and Technology in Computer Science Education,* 73–76. ITiCSE '01. Canterbury, United Kingdom: Association for Computing Machinery. ISBN: 1581133308. https://doi.org/10.1145/377435.377486.

Ryan, Richard M. 1982. 'Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory.' *Journal of personality and social psychology* 43 (3): 450.

Ryan, Richard M, Valerie Mims and Richard Koestner. 1983. 'Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory.' *Journal of personality and Social Psychology* 45 (4): 736.

Ryan, Richard M. and Edward L. Deci. 2023. 'Acting as One: Self-Determination Theory's Scientific and Existential Import'. In *The Oxford Handbook of Self-Determination Theory,* 1173–1176. Oxford, UK: Oxford University Press, February. ISBN: 9780197600047. eprint: https://academic.oup.com/book/0/chapter/396161622/chapter-ag-pdf/50147256/book_45638_section_396161622.ag.pdf. https://doi.org/10.1093/oxfordhb/9780197600047.013.59.

Shuto, Masashi, Hironori Washizaki, Katsuhiko Kakehi, Yoshiaki Fukazawa, Shoso Yamato, Masashi Okubo and Bastian Tenbergen. 2017. 'Relationship between the five factor model personality and learning effectiveness of teams in three information systems education courses'. In *2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD),* 167–174. Kanazawa, Japan: IEEE, June.

Silk, Joan, Dorothy Cheney and Robert Seyfarth. 2013. 'A practical guide to the study of social relationships'. *Evolutionary Anthropology: Issues, News, and Reviews* 22 (5): 213–225. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/evan.21367. https://onlinelibrary.wiley.com/doi/abs/10.1002/evan.21367.

Smith, Michelle K., Francis H. M. Jones, Sarah L. Gilbert and Carl E. Wieman. 2013. 'The Classroom Observation Protocol for Undergraduate STEM (COPUS): A New Instrument to Characterize University STEM Classroom Practices'. _Eprint: https://doi.org/10.1187/cbe.13-08-0154, *CBE—Life Sciences Education* 12 (4): 618–627. https://doi.org/10.1187/cbe.13-08-0154.