Papers in Evolutionary Economic Geography

21.04

Better by design?
Collaboration and performance in the board-game industry

Nicola Cortinovis & Frank van der Wouden

Better by design? Collaboration and performance in the board-game industry

May 11, 2021

ABSTRACT

We study team-work among board-game designers to bring new insights on the effect of team composition on performance. Team-work in game design is an informal and unstructured process that strongly relies on creativity, imagination and out-of-the-box thinking. Current understandings on team composition and performance rely mainly on research in scientific, technological and corporate domains. In these fields team-work is organized in systematic, coordinated and formal processes. It is therefore unclear how the findings from these fields apply to team-work occurring outside corporate and academic boundaries and in less institutionalized and unregulated domains. Our work contributes by examining whether collaborating with someone with greater past performance increases the quality of output, both in the collaborated project and in subsequent single-authored output. Besides, we explore the role of cognitive distance in the team as a possible driver for performance improvement. We apply econometric and machine learning tools to a novel detail-rich database with information on 10,000 quality-rated board-games and their 5,167 designers. Our findings indicate that the quality of the output of a board-game designer significantly increases when (1) collaborating with a better performing designer and (2) having little or significant overlap in expertise with the collaborator. We conclude that the relation between team-work and performance in the board-game industry is different than in domains in which collaboration is coordinated in formal settings. We connect our results to other debates in the organization and innovation literatures and propose policy and managerial implications.

1. Introduction

Working in teams has become an increasingly common way to organize human activities in nearly every domain (42, 87, 93). Whether considering the technological innovation and science (2, 29), medical teams (1, 73) or entertainment (60, 75), team work is usually associated to greater levels of output, learning, problem solving and inventiveness. Such performance-enhancing effects are particularly evident at individual (7, 9, 14, 20, 30, 63, 72) and team-level (25, 40, 75, 84, 85), but effective team-work has also been shown to positively influence firms and organizations (22, 44, 64, 79).

To understand the reasons behind team effects, research has focused on the structure and composition of teams. Various team characteristics are associated with performance improvement. The most important and recurrent in empirical studies are cognitive diversity (71, 81, 82), network position (28, 56), status (75) and experience (38, 77, 81) of team members.

A feature of team composition that is likely to play a crucial role is the actual quality of team-members. According to Frabotta, a young team-member of the celebrated football player Cristiano Ronaldo of Juventus FC, being in the squad with such an exceptional player and observing Ronaldo train every day provides opportunity for asking advice and learning some of his secrets (Redazione Goal Italia) [1].

More generally and across a variety of industries, research findings suggest that the past performance of a team-member or collaborator is an important criteria in the organization and performance of the team (9, 30, 56). Recent contributions have used data on patents and scholarly articles to disentangle the effect of quality of team members, showing that working with someone of superior track record is indeed associated with better performance (4, 14, 72), especially when the quality of team members is relatively balanced (3). However, these findings relate to teams working in technological and scientific domains, often in capital-intensive industries. It is unclear how these apply to team-work in less institutionalized and informal settings such as in creative domains.

[1] In the article (in Italian), the young player Frabotta argues "Observing him (Ronaldo) everyday helps you understanding what "dedication" means, both on and out of the pitch. I ask him advice and try to steal some of his secrets"

Cultural and creative domains have been subject to numerous studies, some of which considering the quality of collaborators. For instance, Rossman et al. (75) show that past nominations for Oscars of coworkers on a movie generate positive spillovers and increases the number of Academy Awards nominations. In this context, the quality of collaborators is often linked to their status and success. This, however, has been proven to be not always correct. Salganik et al. (76) demonstrate how even weak social influence affect the perception of quality and increases the unpredictability of success of unknown songs. Piazza et al. (69) indicate that association with high-profile bands can ultimately have detrimental effects on new entrants in the music industry. While status and quality remain conceptually different (16, 75), their effects are difficult to disentangle and isolate empirically.

In this paper we contribute to the extant literature on teamwork and performance by studying the relation between quality of the collaborator and the performance of the focal designer in the board-game industry. Examining team-work and performance in this industry is different from similar research in corporate, technological or scientific fields. First, as in other creative industries, success in this highly dynamic and rapidly growing industry relies on knowledge and creativity rather than capital investments, hierarchy and institutionalized collaboration. Second, designers are not assigned to collaborate on a project together. Instead, designers have great amounts of agency to identify and select team-members. Third, board-game designing does not require any specific technical background to obtain the desired output from a creative idea. Yet, specific knowledge and skills are still essential. Fourth, the quality of an original idea is closely reflected in the end product that users play and evaluate. Fifth, the vast variety of existing board-games is classified along different dimensions (e.g. genres, mechanics of the game or complexity of the rules) that allow us to examine board-games in the same previous research has examined collaboration in corporate, technological and scientific domains. Finally, the long history of board-games, their growing popularity, the existence of communities of dedicated users and award-giving organizations provides detail-rich data for analysis.

Our main objective is to distinguish and empirically unravel the effect of a collaborator's quality on the performance of the focal designer, while controlling for effects from the collaborator's experience and status. We define the quality of a collaborator as the users' rating of the board-games the collaborator is associated with. Collaborators of teams producing highly rated board-games are considered of high quality. We argue, in line with previous studies on individuals (4, 9, 69) and organizations (39), that working together with a high-quality collaborator, a focal agent may be exposed to practices and routines, fostering learning opportunities and greater performance. Regardless of the quality of previous performance, we also consider the level of experience of the collaborator as a distinct driver for higher-level outcomes (14, 81). Collaborating with someone of longer tenure is likely to provide insights due to the breadth of experience accumulated by the coworker (73). Similarly, the status of individual coworker can prove to be a useful resource for the performance of team members (69, 75). On the one hand, high status signalled by social information may lead to greater success (76). On the other hand, social network studies suggest possible lower levels of creativity from centrally positioned team members (16), making learning and performance improvement less likely.

Our data originates from the website *www.boardgamegeek.com* and allows us to capture detail-rich information on designer quality, experience and status. This publicly available website hosts information on board-games, designers, publishers and other matters related to board-games (hereafter, we use games and board-games interchangeably). An interesting feature of the website is that registered users evaluate games[2] on their quality. We captured this data by web-scraping the website. We gathered information for the top 10,000 games and the corresponding 5,167 unique designers (snapshot of November 2018[3]). Each designer has a unique classifier, enabling us to follow the career of individual designers over time. This allows us to examine to what extent collaborating with a designer of higher average ratings of past games, knowledge of different game mechanisms and themes, award-related social status and greater experience in board game designing, impact the probability of producing a game with a higher rating than one's past average.

We examine this novel database and put forward four main contributions. First,

[2]This represent an important innovation compared to previous studies, mostly relaying on paper citations as a metric of quality

[3]This data-set is independently developed but similar to the one recently generated by Pollok et al. (71)

collaborating with a better performing designer significantly increases the probability of improving the quality of game design. This result is consistent with the idea of knowledge recombination and learning opportunities in teams. Exploiting the patterns of single-authored and collaborated games of designers, we are able to show that the effect persists even after the collaboration, thus excluding possible heavy-lifting effects from the better team member. Second, our results show no direct impact of either experience (captured by the number of games previously published) and status (proxied by number of awards and nomination obtained). However, we show that when working with someone of greater status *and* greater past performance, designers tend to improve more (10, 76). Third, having little or a lot of overlap in the collaborators' expertise improves the chances of the focal scholar to improve the quality of her output. This suggests there are benefits from collaborating with someone who is relatively similar in terms of experience or very dissimilar. This result is consistent with two recurrent arguments in the literature concerning the costs related to bridging differences in heterogeneous teams (similar) and variance-increasing effect of diversity on team performance (dissimilar) (81).

The remainder of the paper is organised as follows. Section 2 provides an overview on the chosen industry, discussing its evolution and features relevant to this study. Section 3 reviews the related literature, introduces the conceptual framework of our research and formulates the hypotheses to test in our analysis. Section 4 gives information on the construction of the database, variables and the methodological choices. Section 5 covers the results of the analysis and the robustness checks. We conclude and discuss our main findings in Section 6.

2. The board-game industry

Board-games in various shapes and forms have been a constant in human societies throughout history (59, 80). Titles like *Monopoly* and *Scrabble* are known to communities across the globe. In essence, board-games are bundles of pieces, rules and themes, which players use to try to achieve a given goal. In the example of *Monopoly*, players engage in negotiation and investments (*theme*) on the basis of specific mechanics (e.g.

dice rolling, skipping turns, etc.) in order to accumulate the most wealth (*goal*).

The mechanics (i.e. the rules upon which games are based) and the themes of the game are to some extent standard and relatively recurrent. For instance, *Monopoly* and *Game of the Goose*, while radically different in themes, rely on shared mechanics such as "roll and move" and "lose turn". Similarly, *Monopoly* and *Settlers of Catan*, based mostly on different mechanics (e.g. victory through elimination of opponents in one and through victory points in the other), share the thematic focus on negotiations and economic venturing. In these respects, the crucial component that characterizes each game is the specific combination of mechanics and thematic elements. As games are relatively often co-designed and the rules and theme components are essentially cognitive constructs derived mostly from knowledge and creativity, the creation of a board-game fits very well the purpose of our analysis: studying the relation between individual performance and team composition in creativity-intensive (but low capital-intensive) endeavours.

2.1. *Innovation in the board-game industry*

Since the 1970s board-game design has undergone a radical evolution. Different from classic products like *Monopoly* and *Cluedo*, games now are designed with a focus on the synergy between mechanics and theme and with a greater attention on interaction among players. Such a drastic change unfolded over various decades and was supported by changes in the cultural and technological domains.

Classic games produced before the 1970s were typically based on strict sets of rules, offered limited choices to players and did not require much or any social interaction. The emergence of a market for games for adults. outside the circuit of mass production since the 1970s led to the development of the first examples of "second generation" board-games. While still rather scant in terms of theme, these games were characterised by "relatively short and clear rule sets, manageable playing time, and a lack of player elimination" (89, p. 35)[4].

Social and cultural developments in society contributed to substantial innovation in

[4] "Acquire" (1962) is considered the predecessor of "second generation" games. "Crude: the oil game" (1974), "Hare and Tartoise" (1973) are other notable examples

the design characteristics of board-games (21, 89). The success of science-fiction and fantasy literature and the popularity of role-playing games paved the way to a greater attention to story and thematic elements in the game design. Games like "Cosmic Encounter" (1977) and "Dune" (1979) offered novel and more engaging science-fiction themes and highly innovative "variable player powers" mechanics [5]. Similarly, "Tales of Arabian Nights" (1985) assimilated role-playing features such as story-telling and flexibility with respect to the plot and victory conditions (89). These, among many other successful innovations, pushed board-game designers to increasingly focus their creative efforts on building a synergistic relation between theme and mechanics. A clear example of this is the "Settlers of Catan" (1995), which enormously contributed to the emergence of modern "eurogames", their popularity and the growth of the industry (21).

The emergence of magazines, conventions and awards (e.g. *Spiel des Jahres* in Germany) centered on board-games provided opportunities for the exchange of ideas and fostered the diffusion of knowledge in the growing community of board-game designers. Technological progress and the advent of the internet in the early 1990s enhanced information-sharing opportunities for both designers and users, revealing the growing potential customer base for the industry (59, 89).

The board-game industry has witnessed remarkable dynamism and growth in the last two decades, leading to what has been referred to as the "golden age" of board-games (59). On a global scale, the market for physical games grew by 12% in 2018 (59), while in the US board-games sales grew by 28 % in 2016 only (15). Socio-technological developments like the emergence of social media channels, online platforms for crowd-funding and smartphones have greatly fostered the inventiveness and dynamism in the industry (32, 65, 89). Crow-funding platform Kickstarter, for which physical games represent the biggest sector[6], has changed the production and sale of board-games, much in favour of independently designed ones (59). Physical games have also started integrating and blending with digital ones, fully exploiting technological innovations such as touch-screen, apps and virtual reality tools (59, 89). The astonishing growth

[5]Unlike classic games in which rules apply to all players equally, "variable player powers" imply that each player may, under certain circumstances, break a rule of the game.

[6]Livingstone and Wallis (59) report that, in 2018 alone, more than 2300 games were successfully launched on Kickstarter, raising more than 160 million USD.

Figure 1. Number of new board-games published per year

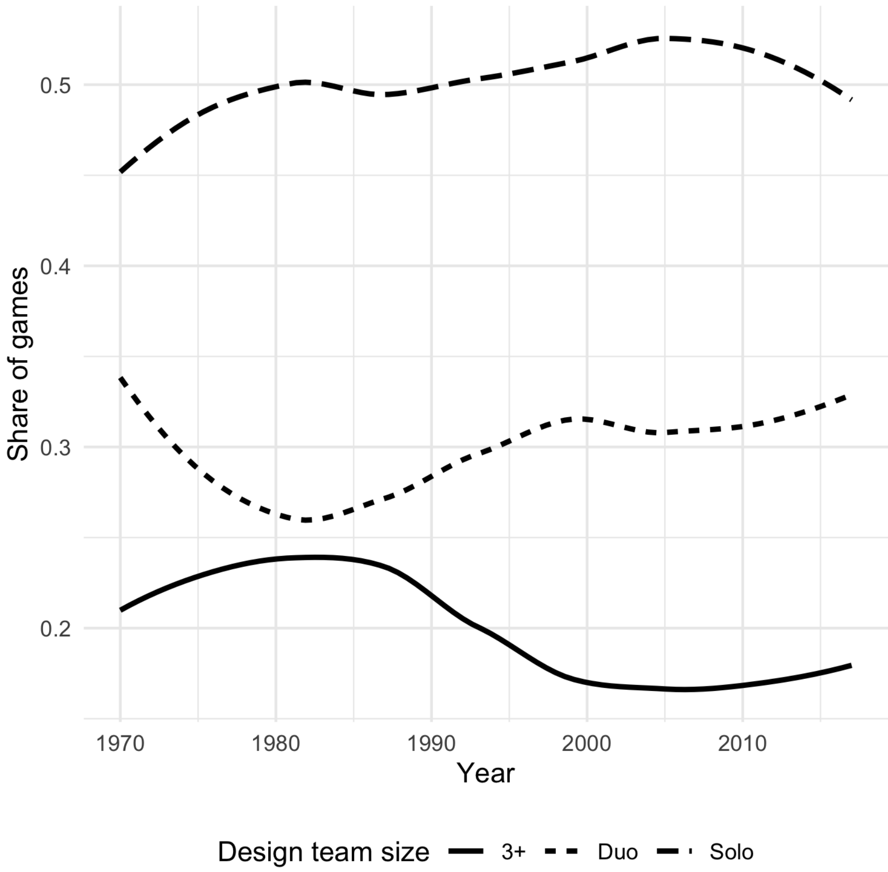
and vitality of the board-game industry is evident in Figure 1: since the early 2000s the number of published games increases super-linearly, reaching around 7000 new games published in 2019.

2.2. *Collaboration in board-game design*

In the board-game industry cognitive constructs (mechanics, themes) are combined by collaborating designers during the development of the game. As Livingstone and Wallis (59, p. 169), two successful game designers put it: "Games are magic and part of that magic is that anyone can make them. You don't need a qualification; all you need is an idea. Pick up some pebbles or draw few lines in the dirt, and see what happens when you try moving them in different ways". Collaboration allows individual designers to access knowledge they do not possess, to combine insights and produce more ideas more efficiently.

The innovative wave hitting the board-game industry since the 1970s is also reflected in the extent to which designers collaborate. Figure 2 is based on our board-game sample (see "Data development" section below), for which we have information on individual designers contributing to board-games. While games developed by single designers remain the most recurrent throughout the years, the share of games designed

Figure 2. Board-games published per year, by number of designers (sample)



by two authors has grown since the early 1990s. Unlike other fields, like scholarly papers and patents (3, 87, 93), teams larger than two are relatively uncommon in this industry. Considering the rapid and radical changes in the board-game industry, the relative importance of smaller teams is line with the empirical observation that small teams tend to be produce disruptive innovations (92).

The relative frequency of collaboration aside, do designers perceive working with others as important? Bruno Faidutti (27), an established board-game designer, indicates that developing a game with someone else "allows the authors (...) to work more quickly, effectively, and (...) with more pleasure". By working together, designers are able to exchange ideas, comments and insights and more easily and quickly overcome problems in the design. According to Faidutti, "these constant exchanges speed the process of the creation of a game much faster and more dynamically than when an author works alone and generally leads to better results". Interestingly, technological innovations such as the telephone, email and VoIP are mentioned as tools facilitating designers' collaborative efforts (27, 83). In another article, Faidutti (26) argues that collaboration fosters the combination of mechanic and thematic elements of games. Even-though anecdotal in nature, this suggests a crucial role of team work in both productivity and quality of output in board-game design.

3. Theory & hypotheses

The rising importance of team work and collaboration (53, 75, 84, 85, 87, 93) poses the question of the role of team composition in performance. The composition of collaborators in corporate, scientific and other professional and institutionalized teams is of critical importance and with long lasting consequences (52). While in certain settings team members are selected at random (11, 63), the selection of collaborators is often guided by preferences that originate from context-specific reasons and motivations for collaboration (14, 55, 86).

In the informal and unstructured settings of the board-game industry, collaborations are more likely to be driven by personal preferences (for example friendship or compatibility) than in corporate and formal settings. For instance, the designer Chad

Ellis (23) explicitly refers to three crucial factors for successfully co-designing a game: trust, shared values and vision, and complementary experience. At the same time, the selection of a specific collaborator may emerge in a rather serendipitous way, for instance through discussions and bouncing off ideas with other designers and play-testers during (conference) meetings (33, 68). Although the designer-specific nature of collaborative choices can be driven by countless preferences, the general underlying assumption is that the collaboration should be beneficial for the parties included, with a desired quality of output that can not be attained without collaboration.

3.1. *Quality of the collaborator*

Regardless of the industry or sector considered, research findings suggest that the past performance of a potential collaborator is an important criteria in selecting a collaborator (9, 56). Theoretically, a better performing team member is likely raise the team's performance and to generate learning opportunities (4). For instance, Rawlings and McFarland (72) find that having a grant-winning coauthor in a grant proposal increases the probability to obtain a grant. Bikard et al. (14) find that the quality an individual researcher's academic paper tends to improve after collaborating with a researcher from another department but that collaborating with a more senior researcher negatively impacts the quality. Examining patents and research articles, Ahmadpoor and Jones (3) find that, overall, there are benefits from collaboration on impact, but that these are largest in "balanced" teams in terms of previous performance.

In spite of the notable contributions offered by these studies, little is known to what extent these findings apply to highly creative industries in which collaboration is less formal, institutionalized and regulated. However, in the case of movie industry the probability of an actor being nominated for an Oscar award increases when working with high quality actors, writers and directors (75). This suggests there are benefits from working with people of superior performance in a creative industry. Based on these considerations, we put forward the following hypothesis:

Hypothesis 1. Collaborating with a designer of superior past performance increases the probability of improving the quality of the focal designer's output.

3.2. *Cognitive distance*

A crucial element when evaluating possible collaborators is the difference in terms of skills, knowledge and cognitive abilities of teams members. Innovation studies broadly refer to such heterogeneity in teams as cognitive distance (9, 66, 67). Scholars in management tend to distinguish between "job relevant" and "background" diversity (45, 46) [7]. Whereas conceptual differences between *cognitive distance* and *job relevant heterogeneity* exist, contributions from both fields come to similar expectations for the relation between team cognitive differences and performance. On this basis, we decide to adopt the concept of cognitive distance for this paper, given its broader and more encompassing definition.

In both areas of research the existence of a trade-off between cognitive heterogeneity in the team and its performance is recognized (40, 45, 67). When two agents are cognitively very similar, knowledge and ideas can easily be exchanged but are less likely to be informative or novel. Conversely, with highly heterogeneous teams, there is ample room for learning and novelty but exchanging knowledge is difficult because of the lack of a *common ground* and absorptive capacity (17). Empirical results have confirmed the "double-edged" nature of cognitive distance. At the team level, cognitive distance is associated to better performance (45), though not necessarily in a linear way (47, 67).

In developing our hypothesis on cognitive distance, we take into account two important aspects that distinguishes our analysis from previous studies. We focus on impacts on individual-level performance (rather than team output) and limit our analysis to teams of two designers (rather than larger teams). We assume designers aim to design a game of the highest possible quality. To do so, they have to select their collaborator carefully, because potential shortcomings or liabilities cannot be mitigated by attracting additional collaborators. Designers who chose to collaborate with a designer

[7] The former captures skill, education and knowledge related characteristics of the team members, while the latter is associated to bio-demographic attributes not related to the task at hand.

with a similar cognitive portfolio will face little coordination costs due to overlap in cognitive domains (large absorptive capacity), but also have little opportunities for learning during collaborating. Moving along the cognitive distance spectrum, designers will encounter greater coordination costs as overlap in cognitive domains decreases, but will be exposed to greater variety of ideas and knowledge, generating opportunities for learning, novelty and potential rewards (81). This low-risk/low-reward and high-risk/high-reward trade-off is generally accepted in innovation and management literature - for instance in the exploitation vs. exploration debate (61). We therefore derive the following two hypotheses:

Hypothesis 2. Increasing cognitive distance between the expertise of collaborators reduces the probability of improving the quality of the focal designer's output.

Hypothesis 3. Small or very large cognitive distance between collaborator and focal designer has positive effects on the probability of improving the quality of the focal designer's output.

3.3. *Status and experience*

Aside from quality of collaborator and similarity in cognitive portfolios, the status and experience of a collaborator has the potential to impact performance. Whereas we do not formulate specific hypotheses concerning status and experience, we account for their possible role in team-level dynamics both conceptually and empirically.

Theoretically (13, 35, 70), status essentially refers to the perception of the actual quality of an actor by other actors. In this sense, as quality is often difficult to observe and evaluate, a number of factors are likely to make the inherently connected quality and status diverge quite sharply (10, 76). For instance, publications of high-status scientist tend to attract more citations (10), signaling greater quality and thus may attract potential collaborators. In the context board-games, this would imply that teams with designers of varying status tend to design games of greater novelty and higher rating. However, status affects the way agents interact in team-settings due to informal hierarchies created through perceived differences in status and may result in ineffective communication and team functioning (41, 51). Teams of Wall Street research analysts with large number of high-status individuals are found to under-perform due to sub-optimal integration of capabilities (37). How individual-level status impacts team-level performance through collaboration is still rather unclear.

Similarly, the relation between experience and quality is also rather nuanced. Greater levels of experience are conceptually related to greater knowledge and understanding on how to successfully perform a certain task (34). At the same time, even though difference in experience can increase the resource base of the team (54), tenure disparity is likely to be associated with status and power, possibly hindering innovation and creativity (12). As a result, the effect of experience on performance is also ambiguous. Scholars with more experience tend not to produce publications of greater or smaller impact (57, 58, 78). Others have shown that successful teams of scientists have greater shares of experienced team-members (38). These may contribute specific skills and expertise from which the team benefits. Yet, studies on inventors report stark negative returns with experience (5, 24). In the comic-book industry, experience in the genre has a positive effect on variance but not ”mean-increasing” in performance (81). However, it is unclear how individual level experience in the board-game industry impacts performance, nor how experience benefits teams in this industry.

On these bases, in our analysis we control for the difference in experience and status of the team members. As additional contribution, we use interaction terms to explore possible impacts on performance of, one the one hand, quality and status and, on the other hand, experience and cognitive diversity.

4. Data methods

4.1. *Data source*

The board-game industry makes an interesting case for studying team work and individual performance in a highly creative and dynamic industry. However, data covering this sector has been lacking. We therefore developed a novel, detail-rich database. We identified the well-known online community of board-game enthusiasts on the website *www.boardgamegeek.com* (BGG) as our main data-source. BGG is the largest online community of board-game users and designers, and provides a reliable and constantly updated flow of information on board-games. The information provided is extremely rich, ranging from the title of the game and its designers, graphic artists and publishers, to details like the suggested range of age the game is appropriate to and the complexity of the rules. We web-scraped this information from the website (snapshot of November 2018) and constructed a data-base linking games to their designers.

Our novel data-set has three features that allows us to test our hypotheses. First, BGG users can rate (with a score between 1 and 10) the quality of any game. This rating is calculated based on the voters ratings and with an additional 100 votes with the score of 5 (18). This Bayesian-style average reduces the potential creation of outliers, for instance when novel games receive five times a score of 10 for their first ratings. Thus, it limits the impact of games with low voter counts in the rankings. This is important for our research, because we focus on the top 10,000 ranked games. Using this approach to calculate the average rating means our ratings are less likely to be impacted by games with only a few votes[8]. In the Supplementary Information we validate that these ratings capture the quality of the game by examining the correlation between number of awards won and average rating. We find a strong positive correlation.

Second, each game is classified into a number of categories based on their mechanisms (e.g. *dice rolling, worker placement, hand management*, etc.) and their themes (e.g. *zombies, pirates, Star Wars, novel-based*, etc.). We exploit the information on these features in order to obtain measures of designers' similarity in the game mechanisms and themes in their cognitive portfolio. This is done by finding the Jaccard distance between the mechanism and themes portfolio of designers, described in the next section.

Finally, games are linked to their designers for whom information is reported. This allows us to track disambiguated designers over time. The amount of information on the designers varies, but usually consists of the location where the designer resides, the number of fans the designer has on BGG, whether the designer has an account on BGG and a short description in the form of a biography.

We take two more important steps before the empirical hypotheses testing. First,

[8] If we compare the first 100 games and last 100 games (i.e. games between 9901 and 10000) of the ranking in our sample, the respective mean (median) number of voters are 20699.6 (16454) and 77.69 (54)

we subset our data and only focus on designers with at least four games. This allows us to obtain a measure of each individual designer's quality as a rolling average on the last three games she designed before the collaborated game (used for our dependent variable). To exclude possible consequences of looking at collaborated games only (e.g. heavy-lifting effect by the better designer), our analysis also considers single-designed games following a collaboration. This is the fifth game in the sequence of consecutive games by the focal designer. Not all sampled designers have a single-designed game after the collaboration and are excluded from the analysis using this data.

Second, we only select collaborated games designed by teams with two designers. We opt to sample only two-designer teams because these are the most prevalent and it allows us to clearly disentangle the effect of a collaborator on a focal designer. With increasing team-members, this relationship will be more difficult to disentangle analytically because knowledge diffusion and collaborative dynamics will be more complex. Focusing on teams with two designers allows us to obtain a clearer picture on the role of collaboration on individual performance in the board-game industry.

For each focal designer we now have a sample of four sequential games, of which the fourth is a collaboration with one collaborator. For a follow-up robustness check, we also include the post-collaboration game if the focal designer designed this game alone.

4.2. *Operationalization of Measures*

We model the probability of a designer to improve her performance as a function of team characteristics. Our dependent variable (*X Up*) is a binary variable which takes the value 1 when the rating of the collaborated game of focal designer X is higher than the average score of her previous three games, and 0 otherwise. As a robustness check we also use as dependent variable the actual difference between the collaborated game of focal designer X and the average score of her previous three games (*X Up cont.*).

Our main variables of interest capture the difference in quality and the cognitive distance between collaborators in the team. With respect to quality, we use two indicators both of them based on the difference between the average rating of last three games of designer Y (the collaborator) and designer X (the focal designer). One indicator, *Diff. Prev. Ratings*, captures the actual difference in the previous ratings of Y and X, and thus it is a continuous variable measuring the "intensity" of quality difference, as shown in Equation 1.

$$Diff.\ Prev.\ Ratings_{g,x,y,t} = \frac{\sum_{g=-3}^{-1} r_{g,y,t}}{3} - \frac{\sum_{g=-3}^{-1} r_{g,x,t}}{3} \tag{1}$$

In the equation above (as in the one that follow), the subscripts g, x, y and t refer respectively to the game (g), individual designers (x, y) and time (t) dimensions.

The other indicator, *Better collab.*, is a dummy variable, which takes value of 1 when designer X is collaborating with a designer Y of higher previous ratings, and 0 otherwise. In more formal terms:

$$Better\ collab._{g,x,y,t} = \begin{cases} 1 & \text{if } Diff.\ Prev.\ Ratings_{g,x,y,t} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

We use *Better collab.* to subset our sample into matched treatment and control groups. Focal designers who collaborate with someone of greater past performance receive the treatment (*Better collab.* equal to 1) while the other designers with similar characteristics (but with *Better collab.* equal to 0) end up the control. We hypothesize that *Diff. Prev. Ratings* and *Better collab.* increase, the board-game design quality of the focal designer X will increase. It is important to notice that both our measure of quality of the collaborator are based on a moving average based on previous games. This means that, as a designer may have more games published in the same year, *Diff. Prev. Ratings* and *Better collab.* vary at the game rather than year level.

With respect to cognitive distance, the variable *Jaccard dist. (M)* measures the difference in previous cognitive portfolio between designers X and Y in terms of game mechanics (M). Each designer has a cognitive portfolio for her experience with game mechanics (e.g. dice-rolling, lose-a-turn), which is built cumulatively and increases as the designer produces games with greater variety of mechanics. We compute the Jaccard distances between the portfolios of collaborating designers to capture the similarity of their cognitive capabilities. This is done as follows:

$$Jaccard\ dist._{x,y,t} = 1 - \frac{M_{x,t} \cap M_{y,t}}{M_{x,t} \cup M_{y,t}} \tag{3}$$

Intuitively, Equation 3 captures how many of the total number of mechanisms in the portfolio of designer X and Y (denominator of the second term in the equation) are *not* shared by designer X and Y (nominator of the second term in the equation). Thus, smaller distances indicate greater similarity between two designers, while large distances suggest the two designers have very different cognitive portfolios.

In addition, we include a number of control variables. Based on the discussion status, the *Diff. Status* variable records the difference in the number of awards received by designer Y and X in the last five years (71). Similarly with experience, our *Diff. Game #* measure records the difference in the number of running total of designed games between designer Y and X. A greater positive difference indicates that a relative inexperienced focal designer X collaborates with an experienced designer Y. Besides, *Prev. coll.* is a dummy variable taking the value 1 if the two designers have collaborated before. Designers who collaborated before have greater trust, which should boost knowledge sharing and, potentially, higher quality games. On the other hand, designers who collaborated before might have less novel ideas to share. We also include two game-specific control variables, namely the number of publishers who have produced and distributed the game (*Publishers #*) and the level of complexity of the game rules (*Game complx*). Including for the number of publishers helps us controlling for the possible greater visibility and popularity of the game of games published by numerous publishing houses. Controlling for the complexity of the game may be important as board-games based on various mechanics are likely to be more complex, making them either consistently more or less appealing to the public. Finally, time fixed effects and team fixed effects are included in all our regression specifications.

4.3. *Matching Strategy*

To isolate the impact of collaborating with someone of greater past performance on the quality of output (e.g. the rating of the collaborated game), we employ a matching strategy. We match designers who collaborated with a designer of greater past performance (treatment group), to similar designers who did not collaborate with a designer of greater past performance (control group). We use Coarsened Exact Matching (CEM) algorithms (49) to match cases from the treatment group to similar cases from the control group. The CEM algorithm prunes observations that dont have close matches on a series of covariates in both the treated and control groups. The covariates used for matching are *Number of Previous Games, Mean Rating of Previous Three, Year of Collaboration, Number of Fans* of the focal designer. Note that these are individual-level characteristics. Matching on dyad-level characteristics potentially interfere with our treatment variable and is therefore not possible.

This data pre-processing exercise using CEM has many advantages (43, 48, 50). The main advantage is that it allows estimation of the quasi-causal effect of treatment on quality of output, the main aim of this paper. Removing potential relationships between the treatment and control variables simulates the randomization effects of an experiment. In experimental settings the random assignment of participants to a treatment and control group makes sure that the covariates of these groups are balanced. CEM achieves this result by pruning observations from the treatment group that do not have a similar observations in the control group. The key consequence of this, is that results of the statistical models using matched data rely less on model specifications.

However, this approach has several limitations. Our sample size is reduced because unmatched observations are deleted. Thus, our analysis is representative only for designer with rather common covariates because those with outliers on the matching criteria are not matched. In addition, our matching exercise did not achieve perfect balance across the total set of covariates of interest in our analysis, primarily because our sample size is relative small compared to the variance in our covariates. This means we cannot estimate the impact of the treatment on the outcome directly, but we have to control for the confounding relationships of the covariates with the outcome variable through statistical modelling. First we describe the matched data.

From the data from *www.boardgamegeek.com*, we limit our data to 1114 collaborations on board-games by two designers yielding 557 board-games and 326 unique designers. During the subsequent matching exercise we prune 152 and 110 observations from the control and the treatment group, respectively. This is shown in Table 1. As a result, the matched data used for the empirical analysis contains 852 collaborations on 492 board-games by 295 unique designers.

The matching exercise has increased the balance in covariates between the treatment and control group. Table 2 shows the pre-matching balance across the four matching variables between the treatment and control group. Especially the variables *# Previous Games*, which is the number of games the designer has already produced, and *Number of Fans* are unbalanced. This is shown by the difference in mean statistic, differences at different points in the distribution and in the L1 score. This score can be thought of as the differences between the heights in the multi-dimensional histogram for covariates for the treatment group and the multi-dimensional histogram for covariates in the control group.

Table 1. Matched and Unmatched cases by Treatment and Control group

	Control	Treatment
All	728	386
Matched	576	276
Unmatched	152	110

Table 2. Balance before matching approach

	Statistic	Type	L1	Min	25%	50%	75%	Max
Year	1.18	(diff)	0.00	0.00	2.00	1.00	0.00	0.00
# Previous Games	-6.41	(diff)	0.11	0.00	-1.00	-5.00	-9.00	-57.00
Average Rating	0.25	(diff)	0.04	-0.17	0.20	0.28	0.31	0.53
Number of Fans	-46.36	(diff)	0.00	0.00	-1.00	-16.00	-92.00	753.00
Multivariate Imbalance Measure:			L1=	0.78				
Percentage of local common support:			LCS=	11.3%				

In Table 3 the same statistics are shown for the matched sample used for empirical analysis. The matching exercise reduced the unbalance across all matching variables, reflected in the difference in means (here, *Statistic*), uni-variate L1 score and differences in quantiles. The reduction in the difference in means is particularly remarkable: the gap between treated and control groups in the unbalanced sample is substantially reduced for *# Previous Games* (from a difference of 6 to 2 in previous games designed), *Number of Fans* (from a difference of 6 to less than 2) and *Average Rating* (from 0.25 to 0.02)[9]. After matching the multivariate L1 score is smaller, while the local common support improved. However, the matching did not achieve perfect balance. This means the effect of the treatment (collaborating with a game designer of greater previous performance) on the outcome variable (difference between quality of current collaborated game and the previous three designed games) can not be readily estimated in a model. However, we can alleviate the problem by controlling for covariates that both possibly impact the treatment and outcome variable. We do this in our empirical model discussed in the next section.

Table 3. Balance after Coarsened Exact Matching approach

	Statistic	Type	L1	Min	25%	50%	75%	Max
Year	-0.11	(diff)	0.00	2.00	0.00	0.00	0.00	0.00
# Previous Games	-1.15	(diff)	0.01	0.00	0.00	-2.00	-3.00	-1.00
Average Rating	0.02	(diff)	0.03	-0.19	0.00	0.00	0.02	0.02
Number of Fans	1.96	(diff)	0.00	0.00	1.00	4.00	-3.00	0.00
Multivariate Imbalance Measure:			L1=	0.71				
Percentage of local common support:			LCS=	17.0%				

Figure 3 shows the distribution of designers in the treatment and control group in the matched data sample based on the difference between the rating of the collaborated game and the focal designer's average rating of the previous three games. Comparing the distributions suggests that designers in the treatment group (who worked with a designer with a better track record than theirs), on average, see greater improvement in their rating than designers in the control group. The graph provides some *prima facie* evidence that a designer who worked with a collaborator of greater past performance, tend to design games who are rated higher than her average previous rating. To examine if this result is robust when controlling for various covariates, statistical models are designed and operationalized.

[9]In our "unbalanced" sample, the variable *Previous ratings* has a standard deviation of 0.45. Through the matching strategy, we are thus able to reduce the gap in previous ratings between treated and control groups by more than half a standard deviation

Figure 3. Rating by Treatment.

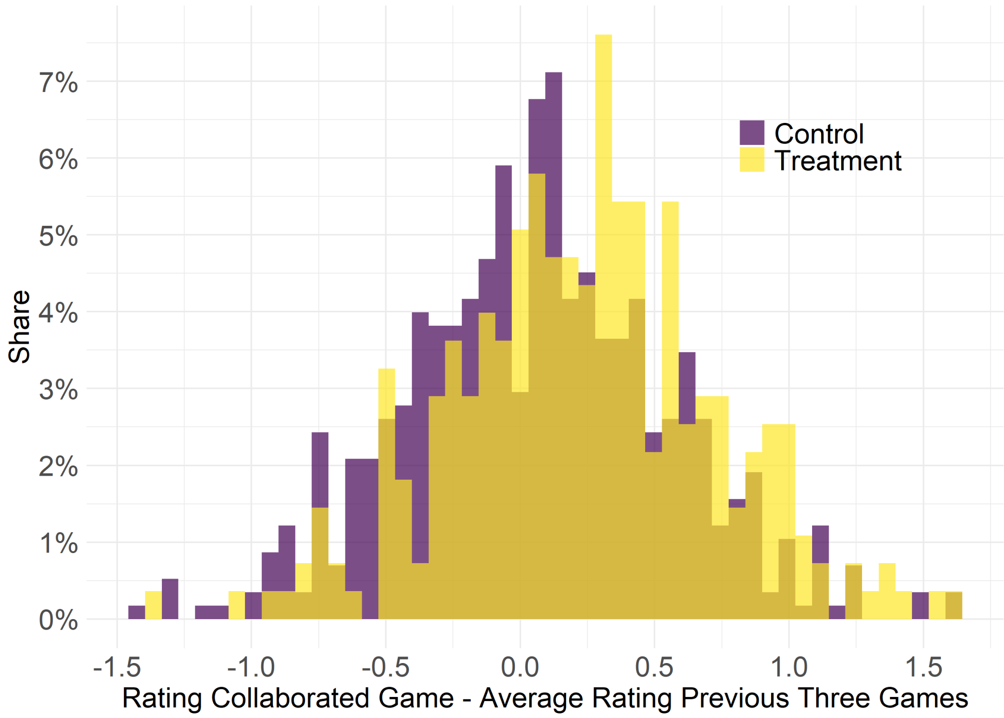


Table 4. Basic descriptive statistics

VARIABLES	(1) Variation	(2) N	(3) mean	(4) sd	(5) min	(6) max
X up	Game-level	852	0.603	0.490	0	1
X up cont.	Game-level	852	0.122	0.499	-1.423	1.617
Better collab.	Game-level	852	0.324	0.468	0	1
Diff. Prev. Ratings	Game-level	852	0.00855	0.280	-1.143	1.083
Jaccard dist. (M)	Year-level	757	0.338	0.304	0	1
Previous coll.	Game-level	876	0.541	0.499	0	1
Diff. Game #	Game-level	852	2.312	16.16	-56	152
Diff. Status	Year-level	852	0.108	1.922	-12	14
Publishers #	Game-level	852	2.825	2.846	1	21
Game complx	Game-level	852	2.256	0.841	0	4.640

4.4. ***Description of data***

Tables 4[10] and 5 report the basic statistics for the key variables in our research and their pairwise correlation scores. A few interesting aspects emerge from Table 4. Around 60% of the designers in our sample see an improvement in their rating (*X Up*). This means that improvement is prevalent, but not necessary. Some designers see their ratings drop after collaboration. In addition, the cognitive distance measures cover the whole spectrum of possible values. Teams are on average relatively close, with a mean value about 0.35 for both our distance measures. However, there is at least one team of designers with exactly the same (i.e. value of 0) and one team with completely different expertise (i.e. value of 1). On around 50% of the games the designers worked together on any of their previous games. Interestingly,Table 5 shows that the quality of the collaborator (rows 3 and 4) and the characteristics of the game (rows 9 and 10) are relatively highly correlated to our dependent variables (columns 1 and 2).

We further characterize and explore our data graphically. First, not all designers collaborated on a game of higher average rating than the average rating of their three previously designed games. In Figure 4 the average rating of the collaborated game

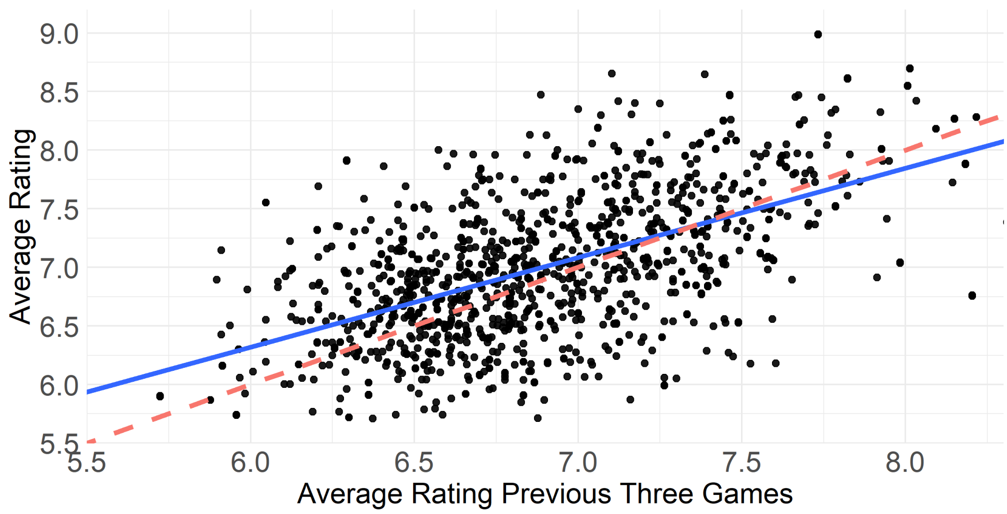
[10]For the sake of clarity, the second column of Table 4 reports at what level we observe the variation in each of our variables.

Table 5. Pairwise correlations

		1	2	3	4	5	6	7	8	9	10
X up	1	1									
X up cont.	2	0.78	1								
Better collab.	3	0.13	0.16	1							
Diff. Prev. Ratings	4	0.19	0.26	0.67	1						
Jaccard dist. (M)	5	-0.03	0.04	0.36	0.04	1					
Previous coll.	6	-0.06	-0.07	-0.17	-0.02	-0.45	1				
Diff. Game #	7	0.01	-0.02	0.02	-0.01	0.04	-0.01	1			
Diff. Status	8	0.02	0.04	0.09	0.08	0	0.05	0.35	1		
Publishers #	9	0.15	0.22	-0.02	0	0.06	-0.02	-0.04	0.03	1	
Game complx	10	0.2	0.28	0.08	0.03	0.04	-0.09	0	0.03	-0.03	1

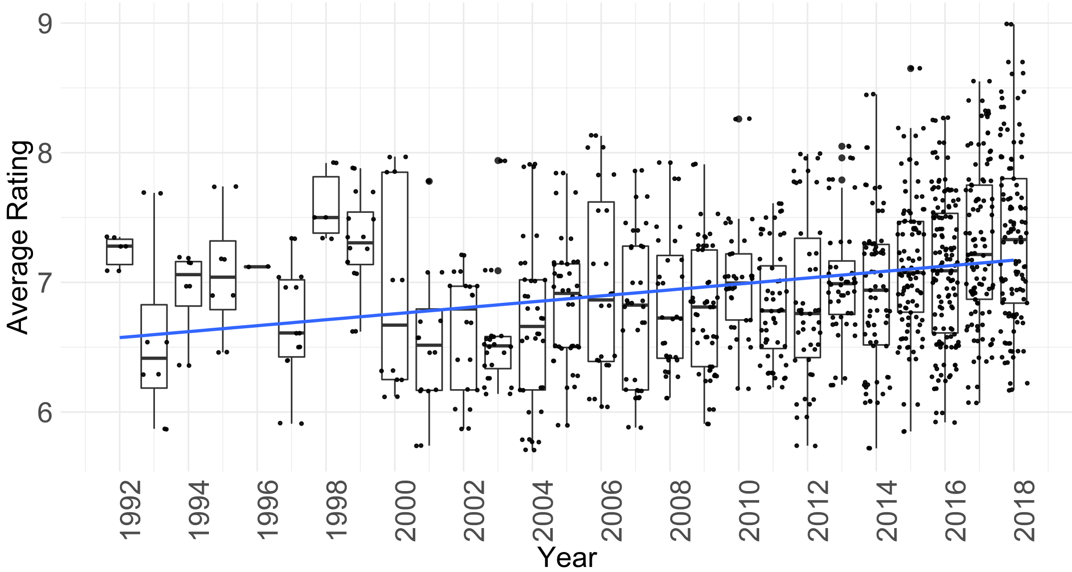
is plotted against the average rating of the previous three games of the designer. Observations above the red dotted line indicate that the rating of the collaborated game is greater than the rating of the previous three. Roughly 60% of the observations fall above the red dotted line, indicating improvement in rating. The difference between the red dotted line and the blue best fit line suggests that the improvement in rating is greatest for designers who had relative low average ratings for their previous three games. For designers with higher average ratings for their previous games it is more difficult to produce a game of higher rating.

Figure 4. Rating by Average Rating of Previous Three games.



Considering the rapid innovation that has characterized the board-game industry in the last decades, it is important to check the evolution of game ratings over time. In our data, newer board-games tend to be rated higher, on average, than older board-games. The blue fitted line in Figure 5 shows this trend in our sampled data for the period of 1991 to 2018 [11]. The average rating of games produced in 2018, as indicated by the horizontal line in the box-plots, is 7.31, while the ratings in the early 1990s hoovered around 6.8. Part of this trend might be explained with the rising popularity of board-games and the attention new board-games attract. Older games might not attract such attention and are less likely to be voted on by enthusiastic gamers. Regardless of the reasons why newer games tend to be rated higher on average, the evidence provide by Figure 5 indicates the need to control for time fixed effects [12].

Figure 5. Rating by Year of production.



4.5. *Modelling*

The main objective of our paper is to explore how team composition affects the performance of individual board-game designers and to offer an empirical test of the three

[11]This trend is also observed in our matched data-set and for the entire time period of our study (available upon request)

[12]Failing to control for this positive trend is likely to introduce some positive bias the results for more recent games. Unlike other recent papers using similar data (71), our regression models always include time fixed effects.

hypotheses we have put forward in our conceptual framework. To this end, we define the following empirical model to estimate through standard econometric tools.

$$up_{g,x,y,t} = \beta_1 \mathbf{Z1}_{g-1,x,y,t} + \beta_2 \mathbf{Z2}_{x,y,t-1} + \beta_3 Z3_{x,y} + \beta_4 5y_t + \phi_{x,y}, \qquad (4)$$

In the equation above, the dependent variable *up* captures the improvement of designer X compared to her previous performance. We empirically measure such improvement in two ways. First, we will use the variable *X Up*, which is binary and takes the value of 1 if designer X improves her rating compared to the average of her previous three games and 0 otherwise. In this case, our estimates captures the effect of game and team characteristics on the probability of improving designer X's rating. Second, we model the actual difference in ratings (X Up cont., i.e. by how much designer X improves) rather than whether she improves or not. When using X Up cont. as dependent variable, the estimated coefficients can be interpreted as the contribution of a specific variable to the improvement in X's scores. By considering both the extensive (whether X improves or not) and the intensive (by how much X improves) margins, we can thus explore both the probability and intensity of improvement as a function of team composition. [13].

In Equation 4, we model the performance of designer X as a function of various team characteristics, such as game-variant designers' features $\mathbf{Z1}_{x,y,g}$ (e.g. whether the previous game was also a collaboration between designer X and Y, or the the difference in previous scores between Y and X) and time-variant designers' features $\mathbf{Z2}_{x,y,t}$ (status, cognitive distance). In these respects, it is important to stress that all our models include team fixed effects and time fixed effects[14], in line with recent

[13]With our approach, we are also more confident that our results are not due to the specific way in which we measure performance. For instance, if designer A had a bad track-record of ratings (e.g. an average score of 5), he may be more likely to improve in the next game (since any value greater than 5 would grant him a 1 in *X Up*). On the other hand, using a continuous dependent variable X Up cont. implies a rather strict assumption, namely that there is a linear relation between team characteristics and difference in rating (e.g. a 0.1 increase in Jaccard dist. (M) would increase the rating of X by n points).

[14]In order to correctly estimate fixed effects, we decided to use dummies for 5-year periods rather than for each and every year between 1977 and 2019. The main reason for this choice is to ensure a suitable number of observations per period. Since in our matched sample games published before 1997 are rather scarce (only 4% of observations pertain games published between 1977 and 1996), we further modify our dummies so games in the period 1977-1996 are included in one dummy and the distribution of observations across time dummies is more balanced. It should be noticed that the problem of the number of observations per category is especially

contributions in the literature (14, 72). Including these helps controlling for sorting based on time-invariant characteristics (e.g. talent, intelligence, cultural traits, etc.) and for period-specific events. We cluster the standard errors at the team-level.

4.6. *Estimation and identification strategy*

From a methodological point of view, we perform our estimations using ordinary least squares in a panel setting. Given the binary nature of one of our dependent variable (*X up*), we could perform our analysis also using a probit or logit model. However, we opt for using a linear probability model since such specification is know to perform better when including fixed effects (90, 91)[15].

The objective of our analysis is to correctly identify the effect of having a "better" collaborator on the performance of the focal designer. To this aim, applying the matching procedure discussed above helps us reducing concerns for unobserved characteristics to affect our estimates. Whereas the matching considerably reduce the heterogeneity in our sample (cf. Table 3 and Table 2), treated and control groups are not perfectly balanced. For this reason, in addition to control variables, we include both time and team fixed effects in our estimation to capture the effect of time invariant characteristics (at team and, in the robustness checks (Table 9), at individual level) and time trends. The main results of this analysis are shown in Table 6

While control variables, matching and fixed effects limit the concern for our results being driven by unobservable characteristics, an important concern remains. Since our dependent variables (*X up* and (*X up cont.*) are constructed as the difference between individuals' previous ratings and the focal jointly produced game, the improvement in the rating of designer X may simply result from partnering up with a hard (and better) working designer Y (what we called "heavy-lifting" effects).

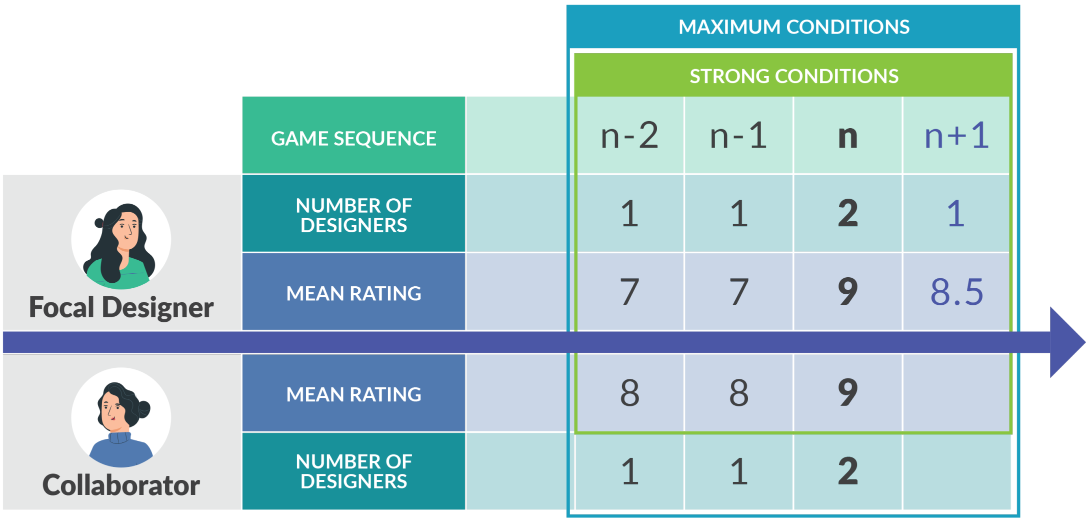
To test whether this is the case, we exploit the collaborative patterns of board-game designers in an additional analysis (see Table 8. Specifically, we identify designers (X)

relevant for our FE logit model (Table 10), while we are able to estimate a linear probability model even with yearly dummies. The results between LPM with 5-year or single year dummies do not change.

[15]Including fixed effects through dummy variables in a probit or logit model is likely to lead to a incidental parameter problem, possibly introducing a bias in our estimates (36). For completeness, our robustness checks report estimations based on a conditional/fixed effect logit regression (91), as shown in Table 10 in the Supplementary Inforation section.

who worked on their own producing at least two games before collaborating with someone else (Y), and who subsequently designed a game on her own after the collaboration, as shown in Figure 6. In other words, to rule out heavy-lifting, we focus on designers with 1-1-2-1 collaborative patterns (i.e. two single-designed games (1-1) followed by a collaboration (2), followed by another single-designed game (1)). We apply a similar logic to the Y designer [16] and proceed to subset our data. In spite of the very demanding conditions, we identify 30 observations in which the required collaborative patterns for X (1-1-2-1) and Y (1-1-2) meet these criteria. We define this as Subset 1 ("Maximum Conditions“ in Figure 6). In Subset 2 (("Strong Conditions“ in Figure 6)) we relax the conditions concerning designer Y, which results in 89 observations, almost a tripling with respect to Subset 1[17]. Due to the tough conditions we subject our data to, we are not able to include a vast variety of fixed effects. To control for possible time trends, we include 5-year period fixed effects.

Figure 6. Identification strategy for analyses on subsets



[16]We relax the condition that designer Y must have a single-designed game after the collaboration. Y's collaboration pattern should thus be: 1-1-2.
[17]These 30 games (as well as the 89 games of Subset 2) are the single-authored games represented by the last "1" figure in the pattern 1-1-2-1.

5. Econometric results and robustness checks

In our analysis, we aim at estimating the relation between the improvement of a designer's current rating and the characteristics of the team she belongs to. In Table 6 we report the main results concerning Hypotheses 1, 2 and 3.

With respect to Hypothesis 1, we find a rather strong positive relation between both *Better collab.* (in Column 1) and *Diff. Prev. Ratings* (in Column 2) and the probability of individual improvement. This suggests that when designer X collaborates with a designer Y of better previous performance, the probability of obtaining better ratings increases. This is in line with our expectations presented in the theoretical framework. For instance, if designer X collaborates with a better collaborator (i.e. *Better collab.*= 1), her probability to improve her previous rating increases by roughly 14 percentage points. Similarly, working with a collaborator Y with an average previous rating 0.1 point higher than X (i.e. *Diff. Prev. Ratings*= 0.1) increases the chances of improving for X by about 3%.

Hypothesis 2 and 3 focus on the cognitive distance between the collaborators on the team. The negative significant coefficients of *Jaccard dist. (M)* in Columns 1 and 2 indicate that a greater difference in the game mechanisms expertise tend to reduce the probability of designer X to improve her score. This confirms our second hypothesis. Venturing into collaboration with someone who has widely different expertise reduces, on average, the probability of improving one's rating. However, this effect appears to be non-linear, because the *Jaccard dist. (M) squared* term is positive, significant and roughly of the same size as the regular term (Column 3 and 4). In other words, the negative effect of cognitive distance on performance is not observed in teams which are either relatively similar (low difference in expertise) or relatively different (high difference in expertise). This confirms our third hypothesis.

We find no significant effects of our control variables on the probability of increasing one's rating, except for the number of publishers on the paper. Board-games that are published by more publishers are associated with an improvement in the rating of the focal designer. A possible explanation could be that these games tend to be better promoted through multiple channels, increasing visibility and, potentially popularity and rating.

Other interesting findings are the interaction effects between Difference in Status and Difference in Previous Rating or Better Collaborator (0/1). If designer X designs a game with a collaborator of better previous performance and greater status, her rating is more likely to improve (Column 5 and 6). This suggests that designing a game with a collaborator of greater status only positively impacts the probability of the focal designer to improve her rating if this collaborator also has greater previous performance.

Table 7 replicates these results using a continuous measure of improvement in the rating as the dependent variable. Our main findings reported earlier are qualitatively similar, except for the effect of cognitive distance. We do not observe the earlier reported significant negative relationships if the squared term is not included in the model. This finding indicates that the effect of cognitive distance between collaborators on the extend of improvement in the focal designer's rating doesn't follow a linear relationship, but rather a non-linear U-shape.

In the light of these results, working with someone of greater past performance not only increases the chances of a designer to improve, but also by how much one improves on average. This result has important implications since, pairing an average designer with slightly better one, as opposed to a much better one, is likely to lead only to a marginal improvement.

In Table 8 we address the issue of *heavy-lifting*. The increase in rating for the focal designer can also simply be the result of the contributions of the collaborator with superior past performance. In this case, the performance boost received by the focal designer is a one-time bonus and would undermine possible policy and managerial implications aimed at fostering long-term effects. To examine this we focus on Subset 1 and Subset 2, in which we select only observations for which the focal designer is the single designer on the post-collaboration game to rule out potential performance effects from new collaborators.

In this setup of the data we exploit the 1-1-2-1 pattern in the collaboration of X. Our dependent variable in this case is defined as the difference between the last single-authored game (the "1" following the "2" in the pattern) and the previous games (the

Table 6. Linear Probability Models with Team FE

	Dependent variable: *X up* (binary)							
Better collab.	0.140***		0.141***		0.131***		0.138***	
	(0.0310)		(0.0309)		(0.0315)		(0.0311)	
Diff. Prev. Ratings		0.327***		0.325***		0.329***		0.324***
		(0.0547)		(0.0552)		(0.0557)		(0.0550)
Jaccard dist. (M)	-0.813***	-0.792***	-2.283***	-2.221***	-0.816***	-0.790***	-0.826***	-0.798***
	(0.299)	(0.296)	(0.620)	(0.588)	(0.303)	(0.295)	(0.296)	(0.294)
Jaccard dist. (M) squared			1.603***	1.559***				
			(0.598)	(0.571)				
Previous coll.	-0.102	-0.101	-0.0810	-0.0803	-0.0960	-0.0991	-0.104	-0.102
	(0.0630)	(0.0625)	(0.0636)	(0.0631)	(0.0634)	(0.0629)	(0.0630)	(0.0625)
Diff. Game #	0.00124	0.00122	0.00115	0.00113	0.000954	0.00102	-0.000587	0.000291
	(0.00135)	(0.00123)	(0.00132)	(0.00122)	(0.00129)	(0.00122)	(0.00234)	(0.00222)
Diff. Status	0.00321	0.000134	0.00466	0.00161	-0.0155	0.00342	0.00343	0.000287
	(0.0135)	(0.0136)	(0.0138)	(0.0140)	(0.0190)	(0.0133)	(0.0135)	(0.0136)
Publishers #	0.0431***	0.0425***	0.0421***	0.0416***	0.0435***	0.0427***	0.0430***	0.0425***
	(0.00805)	(0.00790)	(0.00857)	(0.00842)	(0.00769)	(0.00761)	(0.00807)	(0.00791)
Game complx	0.0793	0.0840	0.0753	0.0800	0.0776	0.0835	0.0797	0.0842
	(0.0785)	(0.0760)	(0.0774)	(0.0749)	(0.0778)	(0.0748)	(0.0786)	(0.0762)
Better collab.*Diff. Status					0.0462**			
					(0.0222)			
Diff. Prev. Ratings*Diff. Status						0.0590*		
						(0.0331)		
Jaccard dist. (M)*Diff. Game #							0.00458	0.00232
							(0.00602)	(0.00590)
Constant	0.566***	0.591***	0.728***	0.750***	0.566***	0.587***	0.570***	0.593***
	(0.209)	(0.205)	(0.222)	(0.216)	(0.210)	(0.201)	(0.209)	(0.205)
Observations	719	719	719	719	719	719	719	719
R-squared	0.369	0.381	0.378	0.390	0.372	0.383	0.369	0.381
Team FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5Y FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors clustered at team-level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7. Models with Team FE

	Dependent variable: *X up cont.*							
Better collab.	0.192***		0.193***		0.186***		0.192***	
	(0.0216)		(0.0212)		(0.0225)		(0.0221)	
Diff. Prev. Ratings		0.459***		0.457***		0.461***		0.463***
		(0.0254)		(0.0248)		(0.0271)		(0.0290)
Jaccard dist. (M)	-0.343	-0.313	-1.975***	-1.889***	-0.345	-0.311	-0.344	-0.306
	(0.275)	(0.271)	(0.482)	(0.443)	(0.277)	(0.269)	(0.275)	(0.273)
Jaccard dist. (M) squared			1.779***	1.718***				
			(0.476)	(0.444)				
Previous Coll.	-0.00767	-0.00533	0.0161	0.0174	-0.00312	-0.00355	-0.00791	-0.00429
	(0.0618)	(0.0612)	(0.0628)	(0.0622)	(0.0625)	(0.0619)	(0.0621)	(0.0615)
Diff. Game #	0.000877	0.000855	0.000776	0.000755	0.000675	0.000668	0.000621	0.00191
	(0.000811)	(0.000800)	(0.000859)	(0.000877)	(0.000788)	(0.000827)	(0.00136)	(0.00131)
Diff. Status	0.00215	-0.00227	0.00376	-0.000645	-0.0111	0.000930	0.00218	-0.00245
	(0.00938)	(0.00936)	(0.00967)	(0.00975)	(0.0132)	(0.00933)	(0.00938)	(0.00943)
Publishers #	0.0585***	0.0578***	0.0575***	0.0568***	0.0589***	0.0580***	0.0585***	0.0578***
	(0.00944)	(0.00924)	(0.00964)	(0.00944)	(0.00903)	(0.00877)	(0.00945)	(0.00925)
Game complx	0.138	0.145*	0.133	0.140*	0.137	0.144*	0.138	0.144*
	(0.0893)	(0.0860)	(0.0876)	(0.0843)	(0.0884)	(0.0845)	(0.0893)	(0.0859)
Better collab.*Diff. Status					0.0329*			
					(0.0194)			
Diff. Prev. Ratings*Diff. Status						0.0574**		
						(0.0281)		
Jaccard dist. (M)*Diff. Game #							0.000642	-0.00265
							(0.00390)	(0.00377)
Constant	-0.324	-0.290	-0.143	-0.115	-0.323	-0.293	-0.323	-0.292
	(0.247)	(0.245)	(0.243)	(0.239)	(0.247)	(0.240)	(0.248)	(0.246)
Observations	719	719	719	719	719	719	719	719
R-squared	0.478	0.504	0.489	0.514	0.480	0.506	0.478	0.504
Team FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5Y FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors clustered at team-level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8. Models using Subsets

	Models using Subsets for identification							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Subset 1 (LPM)	Subset 1 (LPM)	Subset 1 (Cont.)	Subset 1 (Cont.)	Subset 2 (LPM)	Subset 2 (LPM)	Subset 2 (Cont.)	Subset 2 (Cont.)
Better collab.	0.360*		0.371**		0.154		0.239**	
	(0.179)		(0.158)		(0.102)		(0.107)	
Diff. Prev. Ratings		0.247**		0.331***		0.264***		0.346***
		(0.0928)		(0.0882)		(0.0577)		(0.0616)
Jaccard dist. (M)	1.859	1.047	-0.680	-1.640	-0.522	-0.444	-0.674	-0.621
	(1.631)	(1.409)	(1.258)	(0.989)	(0.724)	(0.706)	(0.713)	(0.700)
Jaccard dist. (M) squared	-1.645	-0.850	0.538	1.489	0.460	0.434	0.672	0.672
	(1.488)	(1.261)	(1.187)	(0.958)	(0.623)	(0.601)	(0.597)	(0.578)
Diff. Game #	0.00526**	0.00597***	0.00317	0.00386**	0.000959	0.00181	-0.000263	0.000708
	(0.00188)	(0.00175)	(0.00184)	(0.00175)	(0.00151)	(0.00132)	(0.00186)	(0.00162)
Diff. Status	-0.00867**	-0.0112***	-0.00638	-0.00901**	-0.00170	-0.00314	-0.00108	-0.00273
	(0.00403)	(0.00363)	(0.00401)	(0.00365)	(0.00361)	(0.00352)	(0.00395)	(0.00388)
Game complx	0.402***	0.369***	0.433***	0.385***	0.258***	0.233***	0.291***	0.262***
	(0.0799)	(0.0779)	(0.0874)	(0.0618)	(0.0700)	(0.0658)	(0.0700)	(0.0624)
Publishers #	-0.0521**	-0.0475**	-0.0283	-0.0160	-0.00817	-0.00398	0.00303	0.00817
	(0.0197)	(0.0214)	(0.0224)	(0.0185)	(0.0189)	(0.0182)	(0.0229)	(0.0207)
Constant	-0.743*	-0.388	-0.778**	-0.385*	0.0240	0.112	-0.598***	-0.454**
	(0.369)	(0.310)	(0.299)	(0.216)	(0.255)	(0.218)	(0.216)	(0.185)
Observations	30	30	30	30	89	89	89	89
R-squared	0.726	0.729	0.675	0.767	0.259	0.320	0.330	0.426
Team FE	No	No	No	No	No	No	No	No
5Y FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

"1-1" preceding the "2" in the pattern). A consequence of these harsh criteria result in a reduced sample does not allow for the inclusion of a team fixed effects. However, we still control for possible time trends using five year dummies.

Table 8 shows the results of this exercise. The columns of table 8 containing "LPM" in the header report estimates for the linear probability model (Up X) while those containing "Cont." report the estimates using our continuous depednent variable (*Up X cont.*). Regardless of the dependent variable used, our results indicate that the post-collaboration single-authored game of designer X tends to receive a higher rating compared to her pre-collaboration single-authored games if she collaborated with a designer of superior previous performance. The main difference with respect to the results from Table 6 and 7 is that cognitive distance has no significant effect on improvement. This is likely due to the reduction in the sample and the associated limited heterogeneity remaining in terms of cognitive distance. While this is a relevant limitation, the main purpose of this analysis is to examine whether the relationship between quality and performance detected in Table 6 and 7 is not due to *heavy-lifting* of the collaborator. Indeed, these results indicate that working with a superior collaborator has lasting effects. The probability of a designer to improve her rating increases greatly (between 24% and 37%) after working with a superior designer as compared to collaborating with someone without superior past performance.

5.1. *Robustness checks*

To check the validity of our approach, we perform a series of robustness checks. First, we test whether the inclusion of different fixed effects (or exclusion of thereof) has a major impact on our findings (see Table 9 in the Supplementary Information (SI)). Only the conclusions concerning our measure of cognitive distance is somehow affected in the case no fixed effects (other than the 5 year period dummies) are included. Second, we check whether using a conditional logit model rather than linear probability model has any effect on our results. Performing the analysis in logit settings, as shown in Table 10 provides a further confirmation to all our findings. These results are reported in subsection 7.3 in the SI. In subsection 7.4 of the SI, we step outside the regression frame-work and use supervised machine learning models to check the robustness of our findings. These models allow for complex interactions or additional non-linear relationships that may not have been accounted for in our econometric approach. These results confirm our main findings.

6. Conclusion

Does working with a collaborator of greater past performance increases the quality of an individual's output? The extant literature on relations between team-work and performance have mostly focused on capital-intensive (e.g. movie industry (75)), low creativity industries (e.g. professional sports (6)) or sectors in which collaboration has been institutionalized, formal and strongly (hierarchically) structured (e.g. patenting and scholarly collaborations (4)). In these studies, performance is often captured on the team or firm level. Insights and understandings on the relations between collaboration and performance sectors driven by creativity, experience and out-of-the-box thinking are lacking, as well as impacts of collaboration on the performance of individuals (notable exceptions are (9, 14)).

In this paper we addressed these two gaps. We gathered detail-rich data on board-games and their designers from the website *www.boardgamegeek.com*. We argued that in the board-game industry the crucial element for success is creativity, out-of-the-box thinking and the recombination of knowledge and expertise. We used the data on games and designers to estimate the relations between team composition and individual performance. Performance is inferred from the average rating of the game, given by thousands of board-game enthusiasts as registered users of the website. Linking designers to games we built portfolios of expertise and past performance. Based on the existing contributions, we focus on two main factors which may influence the performance of individuals: (1) quality of past performance and (2) experience in board-game mechanisms. We operationalized econometric models to estimate the impacts of quality of the collaborator and cognitive distance between team-members on the probability the focal designer increases her performance. Since previous performance is often conceptually linked to status and experience (10, 34), we explored possible interaction effects among these and our variables of interests.

Our key finding is that collaborating with a designer of greater past performance strongly increases the probability of the focal designer to improve her performance (hypothesis 1). This indicates that there are benefits for designers to team-up with a top performer. What is unclear is whether the increase in performance observed at the level of the focal-designer is related to the transfer of know-how, knowledge, skill or expertise from the top-performer to the focal designer. If this is the case, collaborating with a top-performer might have long-run impact on performance due to learning-effects. If not, than this increase in performance could be a one-time event. To test for the possible lasting effects, we exploited a specific pattern of collaborations among designers to capture the difference between the rating of the next post-collaboration single-designer game and the pre-collaboration mean rating for treated and untreated designers. For designers to be included in the sample, the last two pre-collaboration games and the post-collaboration game needs to be designed alone. The analyses of designers in this 1-1-2-1 sub-samples indicated that designers who collaborated with a designer of superior performance carried some of the benefits over to their next game, suggesting lasting effects.

While we are unable to pin down the specific mechanisms behind this finding, our specification helps us ruling out certain factors. First, since we are controlling for status and experience, these factors are less likely to drive our finding. Second, having included team (and individual) fixed effects, some of the aspects mentioned by designers (e.g. trust, complementarity and shared values and vision) (23) as key factor for success are also likely to be controlled for. Based on these considerations, learning effects remain a plausible but unobserved mechanism.

We captured team-level synergies by examining the diversity in the team's expertise in game mechanisms. Our results showed that increased diversity reduced the chances of improving performance (hypothesis 2) but in a non-linear U-shaped way (hypothesis 3). These findings suggest that during collaboration on board-games there are high-risk and high-return opportunities that are also observed in innovative and strategy domains (61, 81). In our research we found that chances of improving one's ratings are higher when the expertise of designers are rather similar or very much different. This is different from what recently has been observed in empirical studies on knowledge diffusion and cognitive distance (9, 67). In those studies, agents are found to benefit most from collaborating when there is some cognitive distance (otherwise no novelty) but not too much (do not understand each other). Perhaps the small size of the teams and the rather unstructured and informal setting of collaboration in creative industries allows for more exploring and enables team-members to bridge differences in expertise more easily or effectively.

We also explored the role of status and experience in the team. While neither of these variables is found to be significantly related to improvement in the focal designer, the difference in status between team members plays a moderating role. Specifically, when the focal designer collaborates with someone of greater quality *and* higher status, the probability and degree of improvement of the focal designer are significantly higher. While it is not possible to elaborate on the specific mechanism explaining this result, our finding is in line with previous contributions linking higher performance to higher collaborators' status (10).

These findings have several important managerial implications. First, the relation between team-work and performance in creative industry is different from that in capital and technology-intensive industries. Collaborating among teams with great diversity in expertise is associated with positive returns in the board-game industry, while benefits that arise from status and experience are not observed. These can be important features during decision-making when assigning individuals to teams. Second, stimulating collaboration of top-performers with employees lagging performance could be a tool to increase the performance of under-performers. In our models, the intensity of improvement in individual designers is found to depend on quality of the collaborator. In these respects,under-performers paired with a high-quality team member are likely to improve more than if they were paired with someone of only marginally higher quality. Our research provides indications that the increased performance possibly originates from learning-effects and lasts beyond the collaborated output.

Our results are also informative for policy-makers. The recent trends in automation have had important effects on job availability and income distribution. Unlike repetitive and easy to computerize jobs, creative tasks are unaffected by or even benefit from automation (8). As a result, scholars call for jobs to become *creative* (31, 62). Our contribution helps in these respects by highlighting factors associated with high performance in creative tasks. For instance, policy promoting connections between *new comers* and *super stars* may contribute to higher-quality creative outcomes. Similarly, policy that fosters collaboration among creative agents with different cognitive backgrounds can enhance creative outcomes.

Our findings are shown to be robust to several model specifications, modifications and using supervised machine learning approaches. Still, at least three limitations exist. First, while our data is detail-rich, the overall sample size is relatively limited compared to other studies on, for instance, patent data. While we are able to produce robust and reliable results, the smaller sample size constrains our analysis and the possibility to generalize our findings.Second, we were unable to specifically disentangle individual designers' contributions to each game - the information to do so does not exist in our data-set. Recent developments have offered novel methods for estimating individual impacts in joint productions (3, 30). Applying these methods to team-work and performance in the board-game industry is an exciting path for future research. Third, we have focused on a binary measure of improvement. This has limit us to only evaluate what factors contribute to the probability of obtaining a better rating rather than what factors lead to the greatest improvement. Using a continuous measure of individual performance would address this limitation.

Our research generates several insights that require further research. First, insights on the creative board-game industry can be helpful in widening our perspective on knowledge diffusion, collaboration, recombination and creativity. It offers a valuable complementing perspective to studies on corporate collaboration and team-work in science and technology. Second, our finding on the U-shaped effects of cognitive distance on performance requires further attention to examine whether our findings can be replicated for larger teams and in other creative industries.

References

[1] Agha, L., Ericson, K. M., Geissler, K., and Rebitzer, J. (2018). Team Formation and Performance: Evidence from Healthcare Referral Networks. *National Bureau of Economic Research*.

[2] Ahmadpoor, M. and Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science*, 357(6351):583 LP – 587.

[3] Ahmadpoor, M. and Jones, B. F. (2019). Decoding team and individual impact in science and invention. *Proceedings of the National Academy of Sciences*, 116(28):13885 LP – 13890.

[4] Akcigit, U., Caicedo, S., Miguelez, E., Stantcheva, S., and Sterzi, V. (2018). Dancing with the stars: Innovation through interactions. Working Paper 24466, National Bureau of Economic Research.

[5] Akcigit, U., Grigsby, J., and Nicholas, T. (2017). The Rise of American Ingenuity: Innovation and Inventors of the Golden Age.

[6] Archibugi, D. (2017). Blade Runner economics : Will innovation lead the economic recovery ? *Research Policy*, 46(3):535–543.

[7] Arcidiacono, P., Kinsler, J., and Price, J. (2017). Productivity spillovers in team production: Evidence from professional basketball. *Journal of Labor Economics*, 35(1):191–225.

[8] Autor, D. H. (2015). Why are there still so many jobs? the history and future of workplace automation. *Journal of Economic Perspectives*, 29(3):3–30.

[9] Ayoubi, C., Pezzoni, M., and Visentin, F. (2017). At the origins of learning: Absorbing knowledge flows from within the team. *Journal of Economic Behavior and Organization*, 134:374–387.

[10] Azoulay, P., Stuart, T., and Wang, Y. (2014). Matthew: Effect or fable? *Management Science*, 60(1):92–109.

[11] Bandiera, O., Barankay, I., and Rasul, I. (2009). Social Connections and Incentives in the Workplace: Evidence From Personnel Data. *Econometrica*, 77(4):1047–1094.

[12] Bell, S. T., Villado, A. J., Lukasik, M. A., Belau, L., and Briggs, A. L. (2011). Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *Journal of Management*, 37(3):709–743.

[13] Benjamin, B. A. and Podolny, J. M. (1999). Status , Quality , and Social Order in the California Wine Industry. *Administrative Science Quarterly*, 44(3):563–589.

[14] Bikard, M., Murray, F., and Gans, J. S. (2015). Exploring trade-offs in the organization of scientific work: Collaboration and scientific reward. *Management Science*, 61(7):1473–1495.

[15] Birkner, C. (Accessed: 2020-02-28). From Monopoly to Exploding Kittens, Board Games Are Making a Comeback.

[16] Cattani, G. and Ferriani, S. (2008). A core/periphery perspective on individual creative performance: Social networks and cinematic achievements in the Hollywood film industry. *Organization Science*, 19(6):824–844.

[17] Cohen, W. M. and Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1):128–152.

[18] Contributors, B. G. (Accessed: 2020-02-28). User ratinges.

[19] Cummings, J. N. and Kiesler, S. (2007). Coordination costs and project outcomes in multi-university collaborations. *Research Policy*, 36(10):1620–1634.

[20] Defazio, D., Lockett, A., and Wright, M. (2009). Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program. *Research Policy*, 38(2):293–305.

[21] Donovan, T. (2017). *Its All a Game: The History of Board Games from Monopoly to Settlers of Catan*. Thomas Dunne Books.

[22] Dul, J. and Ceylan, C. (2014). The impact of a creativity-supporting work environment on a firm's product innovation performance. *Journal of Product Innovation Management*, 31(6):1254–1267.

[23] Ellis, C. (2004). Living the dream 1.0. *The Games Journal*.

[24] Esposito, C. R. and van der Wouden, F. (2021). Learning, Fast and Slow: The Returns to Experience and Collaboration for High-Impact Innovation in the United States between 1836 and 1975.

[25] Faems, D. and Subramanian, A. M. (2013). R&D manpower and technological performance: The impact of demographic and task-related diversity. *Research Policy*,

42(9):1624–1633.

[26] Faidutti, B. (2005a). Themes mechanics 1.0. *The Games Journal*.

[27] Faidutti, B. (2005b). Themes mechanics 3.0. *The Games Journal*.

[28] Ferriani, S., Cattani, G., and Baden-Fuller, C. (2009). The relational antecedents of project-entrepreneurship: Network centrality, team composition and project performance. *Research Policy*, 38(10):1545–1558.

[29] Fleming, L., Mingo, S., and Chen, D. (2007). Collaborative Brokerage, Generative Creativity, and Creative Success. *Administrative Science Quarterly*, 52(3):443–475.

[30] Flores-Szwagrzak, K. and Treibich, R. (2020). Teamwork and individual productivity. *Management Science*, 66(6):2523–2544.

[31] Florida, R. (2014). The Creative Class and Economic Development. *Economic Development Quarterly*, 28(3):196–205.

[32] Garry, B. (Accessed: 2020-02-28). What's behind the success of tabletop gaming?

[33] Glenn, S. (2002). Protospiele. *The Games Journal*.

[34] Gonzalez-Mulé, E., S. Cockburn, B., W. McCormick, B., and Zhao, P. (2020). Team tenure and team performance: A meta-analysis and process model. *Personnel Psychology*, 73(1):151–198.

[35] Gould, R. V. (2002). The origins of status hierarchies: A formal theory and empirical test. *American Journal of Sociology*, 107(5):1143–1178+I.

[36] Greene, W. (2002). The Behavior of the Fixed Effects Estimator in Nonlinear Models,. *Unpublished Manuscript, Stern School of Business, NYU*, 29:1–31.

[37] Groysberg, B., Polzer, J. T., and Elfenbein, H. A. (2010). Too Many Cooks Spoil the Broth: How High-Status Individuals Decrease Group Effectiveness. *Organization Science*, 22(3):722–737.

[38] Guimera, R., Uzzi, B., Spiro, J., and Amaral, L. (2005). Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. *Science*, 308(5722):697–702.

[39] Gulati, R. (1999). Network location and learning: The influence of network resources and firm capabilities on alliance formation. *Strategic management journal*, 20(5):397–420.

[40] Harrison, D. A. and Klein, K. J. (2007). What's the Difference? Diversity constructs as separation, variety and disparity in organizations. *Academy of Management Review*, 32(4):1199–1228.

[41] He, J. and Huang, Z. (2011). Board Informal Hierarchy and Firm Financial Performance:

Exploring a Tacit Structure Guiding Boardroom Interactions. *Academy of Management Journal*, 54(6):1119–1139.

[42] Herbst, D. and Mas, A. (2015). Peer effects on worker output in the laboratory generalize to the field. *Science*, 350(6260):545–549.

[43] Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*, 15(3):199–236.

[44] Hoegl, M. and Gemuenden, H. G. (2001). Teamwork Quality and the Success of Innovative Projects: A Theoretical Concept and Empirical Evidence. *Organization Science*, 12(4):435–449.

[45] Horwitz, S. K. and Horwitz, I. B. (2007). The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography. *Journal of Management*, 33(6):987–1015.

[46] Hülsheger, U. R., Anderson, N., and Salgado, J. F. (2009). Team-Level Predictors of Innovation at Work: A Comprehensive Meta-Analysis Spanning Three Decades of Research. *Journal of Applied Psychology*, 94(5):1128–1145.

[47] Huo, D., Motohashi, K., and Gong, H. (2019). Team diversity as dissimilarity and variety in organizational innovation. *Research Policy*, 48(6):1564–1572.

[48] Iacus, S. M., King, G., and Porro, G. (2011). Multivariate Matching Methods That Are Monotonic Imbalance Bounding. *Journal of the American Statistical Association*, 106(493):345–361.

[49] Iacus, S. M., King, G., and Porro, G. (2012a). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 20(1):1–24.

[50] Iacus, S. M., King, G., and Porro, G. (2012b). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 20(1):1–24.

[51] Iosub, D., Laniado, D., Castillo, C., Fuster Morell, M., and Kaltenbrunner, A. (2014). Emotions under Discussion: Gender, Status and Communication in Online Collaboration. *PLOS ONE*, 9(8):e104880.

[52] Jaravel, X., Petkova, N., and Bell, A. (2018). Team-specific capital and innovation. *American Economic Review*, 108(4-5):1034–1073.

[53] Jones, B. F. (2009). The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? *The Review of Economic Studies*, 76(1):283–317.

[54] Joshi, A. and Roh, H. (2009). The role of context in work team diversity research: A

meta-analytic review. *Academy of Management Journal*, 52(3):599–627.

[55] Leahey, E. (2016). From Sole Investigator to Team Scientist: Trends in the Practice and Study of Research Collaboration. *Annual Review of Sociology*, 42(1):81–100.

[56] Lee, J. (2010). Heterogeneity, Brokerage, and Innovative Performance: Endogenous Formation of Collaborative Inventor Networks. *Organization Science*, 21(4):804–822.

[57] Liu, L., Dehmamy, N., Chown, J., Giles, C. L., and Wang, D. (2021). Understanding the onset of hot streaks across artistic, cultural, and scientific careers. *arXiv preprint arXiv:2103.01256*.

[58] Liu, L., Wang, Y., Sinatra, R., Giles, C. L., Song, C., and Wang, D. (2018). Hot streaks in artistic, cultural, and scientific careers. *Nature*, 559(7714):396–399.

[59] Livingstone, I. and Wallis, J. (2017). *Board Games in 100 Moves*. Dorling Kindersley.

[60] Mannucci, P. V. (2017). Drawing snow white and animating buzz lightyear: Technological toolkit characteristics and creativity in cross-disciplinary teams. *Organization Science*, 28(4):711–728.

[61] March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, 2(1):71–87.

[62] Martin, R. L. (2012). The u.s. needs to make more jobs more creative. *Harvard Business Review*.

[63] Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1):112–145.

[64] Montes, F. J. L., Ruiz Moreno, A., and García Morales, V. (2005). Influence of support leadership and teamwork cohesion on organizational learning, innovation and performance: An empirical examination. *Technovation*, 25(10):1159–1172.

[65] Newswire (Accessed: 2020-02-28). Detailed Research on Global Board Games Market 2017-2021.

[66] Nooteboom, B. (2000). Learning by interaction: Absorptive capacity, cognitive distance and governance. *Journal of Management and Governance*, 4(1-2):69–92.

[67] Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., and van den Oord, A. (2007). Optimal cognitive distance and absorptive capacity. *Research Policy*, 36(7):1016–1034.

[68] Petty, M. (2002). Protospiele 2002. *The Games Journal*.

[69] Piazza, A., Phillips, D. J., and Castellucci, F. (2020). High-status affiliations and the success of entrants: New bands and the market for live music performances, 2000-2012.

Organization Science, 31(5):1272–1291.

[70] Podolny, J. M. (1993). A status-based model of market. *American Journal of Sociology*, 98(4):829–872.

[71] Pollok, P., Amft, A., Diener, K., Lüttgens, D., and Piller, F. T. (2021). Knowledge diversity and team creativity: How hobbyists beat professional designers in creating novel board games. *Research Policy*, (December).

[72] Rawlings, C. M. and McFarland, D. A. (2011). Influence flows in the academy: Using affiliation networks to assess peer effects among researchers. *Social Science Research*, 40(3):1001–1017.

[73] Reagans, R., Argote, L., and Brooks, D. (2005). Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Science*, 51(6):869–881.

[Redazione Goal Italia] Redazione Goal Italia. Frabotta e la scalata fino alla juventus: "la mia storia solo all'inizio". Accessed: 2021-03-30.

[75] Rossman, G., Esparza, N., and Bonacich, P. (2010). I'd like to thank the academy, team spillovers, and network centrality. *American Sociological Review*, 75(1):31–51.

[76] Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856.

[77] Salter, A. and Gann, D. (2003). Sources of ideas for innovation in engineering design. *Research Policy*, 32(8):1309–1324.

[78] Sinatra, R., Wang, D., Deville, P., Song, C., and Barabási, A.-L. (2016). Quantifying the evolution of individual scientific impact. *Science*, 354(6312):aaf5239.

[79] Solheim, M. C., Boschma, R., and Herstad, S. J. (2020). Collected worker experiences and the novelty content of innovation. *Research Policy*.

[80] Solly, M. (2020). The best boardgames of the ancient worlds. *The Smithsonian Magazine*.

[81] Taylor, A. and Greve, H. R. (2006). Superman or the fantastic four? knowledge combination and experience in innovative teams. *Academy of Management Journal*, 49(4):723–740.

[82] Ter Wal, A. L., Criscuolo, P., and Salter, A. (2017). Making a marriage of materials: The role of gatekeepers and shepherds in the absorption of external knowledge and innovation performance. *Research Policy*, 46(5):1039–1054.

[83] Trzewiczek, I. (2014). *Boardgames that tell stories*. Portal Games.

[84] Uzzi, B., Mukherjee, S., Stringer, M., and Jones, B. (2013). Atypical combinations and scientific impact. *Science*, 342(6157):468–472.

[85] Uzzi, B. and Spiro, J. (2005). Collaboration and Creativity: The Small World Problem. *American Journal of Sociology*, 111(2):447–504.

[86] Vakili, K., Teodoridis, F., and Bikard, M. A. (2020). Detrimental Collaborations in Creative Work: Evidence from Economics. *Academy of Management Proceedings*, 2020(1):17719.

[87] van der Wouden, F. (2020). A history of collaboration in US invention: changing patterns of co-invention, complexity and geography. *Industrial and Corporate Change*, 29(3).

[88] van der Wouden, F. and Rigby, D. L. (2019). Co-inventor networks and knowledge production in specialized and diversified cities. *Papers in Regional Science*, 98(4).

[89] Woods, S. (2012). *Eurogames: the design, culture and play of modern European board games*. McFarland Company.

[90] Wooldridge, Jeffrey M., . ([2012]). *Introductory econometrics : a modern approach*. Fifth edition. Mason, Ohio : South-Western Cengage Learning, [2012] ©2012.

[91] Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, volume 1 of *MIT Press Books*. The MIT Press.

[92] Wu, L., Wang, D., and Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, 566(7744):378–382.

[93] Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–9.

7. Supplementary Information

7.1. *Validating average rating with award data*

Does the average rating given by the users of B*www.boardgamegeek.com* reflect the quality of the game? To examine this we use the award data listed on the website that links 12,351 awards to games. Out of these, 2,614 are awarded to winners and the remaining 9,737 are for runner-ups and other nominees. Most of these awards are assigned by experts, but some of them are awarded through contests and audience awards. We use these awards as a signal of quality.

In Figure 7 the relationship between the average rating of a game and the number of awards it has received is plotted. We find a significant positive relationship. An additional award is associated with an increase in the average rating of 0.04. This relationship is robust. In Figure 8 this relationship is plotted in which only 'Winner' awards are counted.

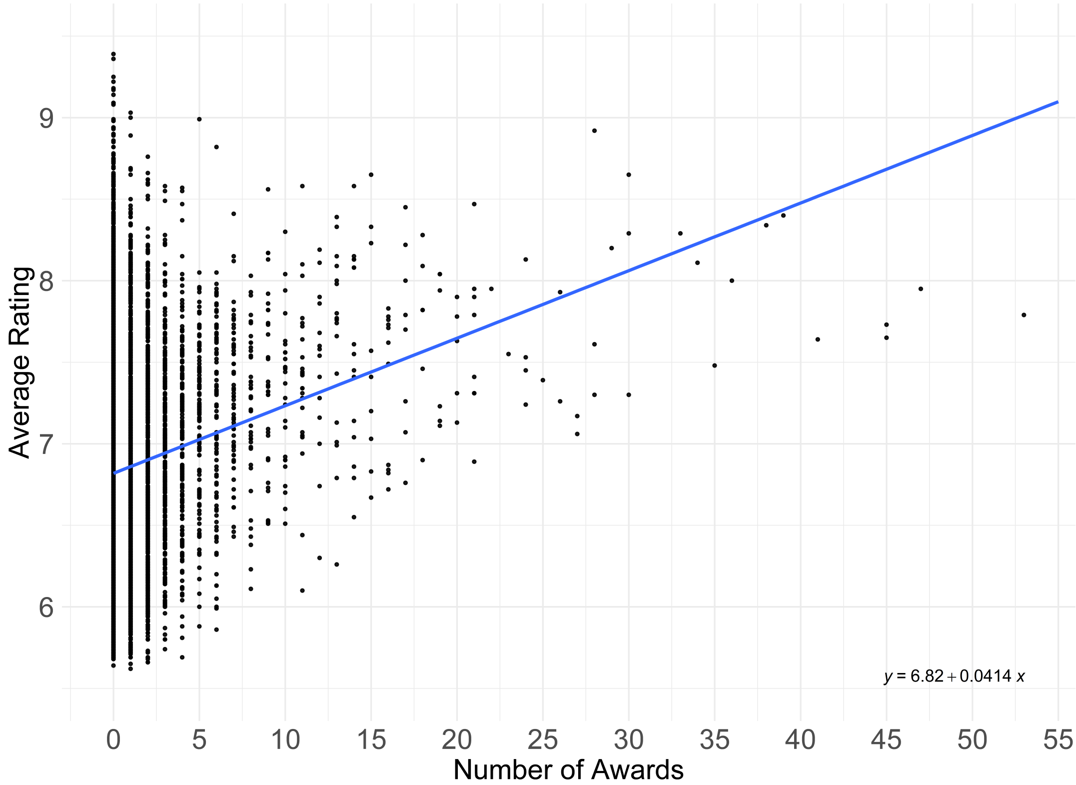


Figure 7. Relationship between average rating and number of awards received.

7.2. *Team-work and performance*

Is team-work associated with greater performance when designing board-games? In Figure 9 the relationship between average rating of a game is plotted against the number of designers on the team. The positive relationship indicates there are benefits of collaboration. However, when team-size gets too large (¿ 7) these benefits tend to disappear. This might have to do with the increasing coordination costs that scale super-linearly with growing team-size. These costs might not be off-set by the potential benefits. Insufficient coordination might also result in lower outcomes (19).

7.3. *Robustness checks using regression techniques*

To check the validity of our approach, we perform a series of robustness checks to test possible differences in the use of fixed effects and non-linear link functions. First,

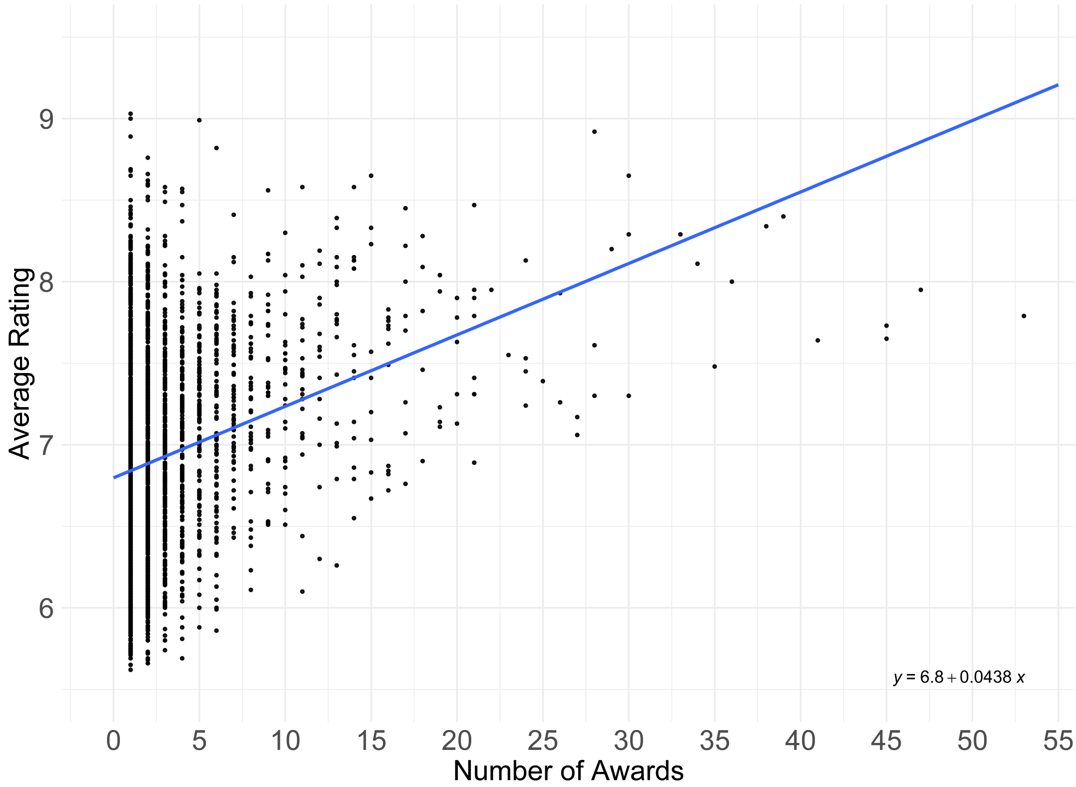


Figure 8. Relationship between average rating and number of 'Winner' awards received.

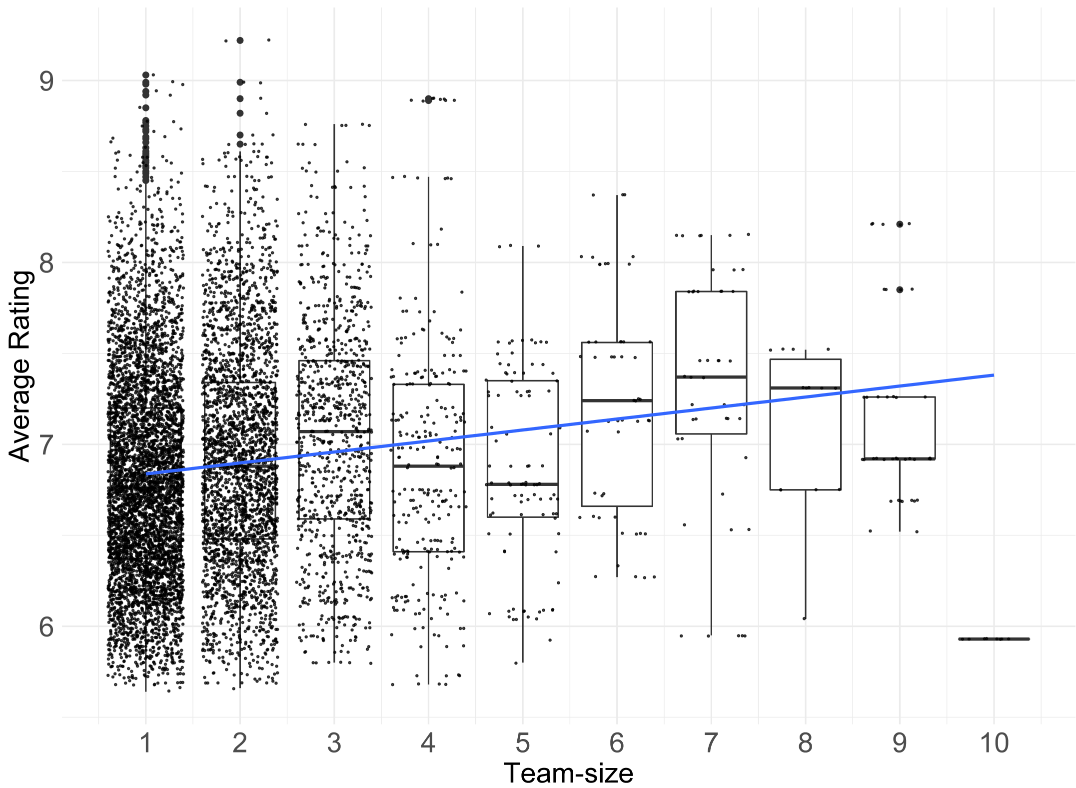


Figure 9. Relationship between average rating and number of designers on team.

Table 9. Robustness checks on fixed effects

	Robustness checks							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Main - No FE (LPM)	Main - No FE (LPM)	Main - No FE (LPM)	Main - No FE (LPM)	Main - X FE (LPM)	Main - X FE (LPM)	Main - X FE (LPM)	Main - X FE (LPM)
Better collab.	0.149***		0.158***		0.120**		0.123**	
	(0.0288)		(0.0286)		(0.0500)		(0.0498)	
Diff. Prev. Ratings		0.323***		0.321***		0.315***		0.302***
		(0.0472)		(0.0475)		(0.0773)		(0.0795)
Jaccard dist. (M)	-0.220***	-0.151*	-0.483**	-0.288	-0.605***	-0.590**	-1.991***	-1.889***
	(0.0787)	(0.0808)	(0.217)	(0.224)	(0.228)	(0.228)	(0.515)	(0.517)
Jaccard dist. (M) squared			0.330	0.174			1.431***	1.341***
			(0.235)	(0.241)			(0.481)	(0.477)
Previous collab	-0.0861**	-0.0888**	-0.0815**	-0.0865**	-0.106	-0.112*	-0.103	-0.110
	(0.0399)	(0.0400)	(0.0397)	(0.0399)	(0.0677)	(0.0666)	(0.0691)	(0.0682)
Diff. Game #	0.000541	0.000696	0.000615	0.000731	-0.00442*	-0.00399	-0.00376	-0.00340
	(0.000980)	(0.000912)	(0.000967)	(0.000906)	(0.00266)	(0.00262)	(0.00267)	(0.00264)
Diff. Status	0.00349	0.00122	0.00381	0.00154	0.0139	0.00928	0.0145	0.0103
	(0.00867)	(0.00846)	(0.00870)	(0.00854)	(0.0129)	(0.0132)	(0.0136)	(0.0138)
Publishers #	0.0281***	0.0268***	0.0271***	0.0262***	0.0449***	0.0447***	0.0453***	0.0450***
	(0.00514)	(0.00511)	(0.00534)	(0.00531)	(0.00835)	(0.00830)	(0.00901)	(0.00893)
Game complx	0.126***	0.129***	0.127***	0.130***	0.0664	0.0679	0.0593	0.0613
	(0.0286)	(0.0281)	(0.0285)	(0.0282)	(0.0727)	(0.0703)	(0.0720)	(0.0697)
Constant	0.295***	0.316***	0.312***	0.326***	0.517**	0.550***	0.700***	0.722***
	(0.0885)	(0.0878)	(0.0901)	(0.0901)	(0.198)	(0.191)	(0.208)	(0.201)
Observations	757	757	757	757	632	632	632	632
R-squared	0.107	0.121	0.110	0.122	0.251	0.261	0.264	0.273
X FE	No	No	No	No	Yes	Yes	Yes	Yes
Team FE	No	No	No	No	No	No	No	No
5Y FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

with respect to our choices in terms of fixed effects, Table 9 shows that the results are broadly consistent regardless of the set fixed effects included. The main difference compared to the previous findings is that, when *Diff. Prev. Ratings* is used without any team or individual fixed effect, the U-shaped relation between improvement and cognitive distance disappears. However, when using the dichotomous *Better collab.* and when dummies for designer X are included, the previous results are fully confirmed.

Second, in terms of functional form, we consider a possible specification that allows us to model a binary dependent variable while controlling for time-invariant characteristics of designers' teams. In this situation, according to Wooldridge (91), a number of options are possible with none of them being generally preferable the others. Among these options, we choose as alternative to our previous linear probability models, a fixed-effect logit. The results reported in 10 further confirm the validity of our previous findings [18].

7.4. *Robustness check using supervised machine learning algorithms*

The econometric models estimate the linear (and non-linear) effects of predictors on the outcome variables. However, possible complex interactions or additional non-linear relationships are not accounted for. To examine whether our results obtained from the econometric models hold under such conditions, we construct a supervised machine learning binary classification model.

[18] Compared to 6, some observations in the fixed-effect logit model are automatically dropped, due to the limited variation in the dependent variable (only positive or only negative outcomes within a team)

Table 10. Robustness checks on functional form

	FE logit models (Team FE)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Main	Main	Non-linear	Non-linear	Interaction	Interaction	Interaction	Interaction
Better collab.	0.786***		0.809***		0.751***		0.769***	
	(0.180)		(0.186)		(0.183)		(0.181)	
Diff. Prev. Ratings		1.705***		1.743***		1.715***		1.674***
		(0.319)		(0.341)		(0.306)		(0.315)
Jaccard dist. (M)	-4.101**	-4.055**	-12.67***	-12.41***	-4.070**	-4.022**	-4.197***	-4.124**
	(1.638)	(1.680)	(4.541)	(4.161)	(1.671)	(1.695)	(1.618)	(1.668)
Jaccard dist. (M) squared			9.482**	9.345**				
			(4.191)	(3.897)				
Previous Coll.	-0.373	-0.382	-0.270	-0.277	-0.330	-0.350	-0.386	-0.390
	(0.315)	(0.319)	(0.315)	(0.321)	(0.313)	(0.319)	(0.316)	(0.320)
Diff. Game #	0.00520	0.00589	0.00432	0.00459	0.00413	0.00494	-0.00491	-0.000633
	(0.00698)	(0.00777)	(0.00707)	(0.00807)	(0.00661)	(0.00785)	(0.0142)	(0.0143)
Diff. Status	0.0299	0.0155	0.0413	0.0299	-0.0682	0.0344	0.0303	0.0150
	(0.115)	(0.156)	(0.126)	(0.169)	(0.161)	(0.157)	(0.114)	(0.155)
Publishers #	0.230***	0.231***	0.230***	0.231***	0.237***	0.241***	0.230***	0.231***
	(0.0648)	(0.0646)	(0.0658)	(0.0652)	(0.0641)	(0.0628)	(0.0651)	(0.0646)
Game complx	0.376	0.400	0.357	0.382	0.368	0.396	0.379	0.404
	(0.373)	(0.365)	(0.368)	(0.358)	(0.372)	(0.361)	(0.374)	(0.366)
Better collab.*Diff. Status					0.235*			
					(0.140)			
Diff. Prev. Ratings*Diff. Status						0.377**		
						(0.189)		
Jaccard dist. (M)*Diff. Game #							0.0254	0.0166
							(0.0392)	(0.0391)
Observations	541	541	541	541	541	541	541	541
Team FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5Y FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors clustered at team-level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The observations from our matched sample are randomly assigned in a training set (80%), validation set (10%) and test set (10%). The training set holds labelled information on events (rating of designer goes up) and the corresponding covariates. Machine learning algorithms can be deployed to this data to *learn* which characteristics among the covariates can best classify observations. A validation data-set is used to tune the (hyper-)parameters of possible models and prevent the models from over-fitting to our training data. A test-set is used to evaluate the performance of the candidate models and select the model of best performance (88). This model is then used to predict which collaborations result in an increase of in the rating for designer X.

There are numerous machine learning algorithms that can be trained to predict and classify a binary outcome variable. As for most machine learning exercises, it is unclear what algorithm works best for our data. Therefore, we initially train models based on Support Vector Machine, Deep Learning, Random Forest and Gradient Boosting Machine algorithms. The latter algorithm produced predictions with greatest *Area Under the Curve* (AUC) scores on the test-set, so we select this algorithm and do extensive grid-searches to tune the hyper-parameters of this algorithm to improve predictions on the test-set.

This trained and validated a supervised gradient boosting machine (GBM) model is used to predict which observations (collaborations), based on the same set of predictors as in the statistical models, result in a game-rating that is higher than the average rating of the 3 previous games of designer X (0/1). Our trained model had an accuracy of 78% on the out-of-sample test-set. This informally translate to the fraction of predictions our model got right. Thus, based on the values of our predictor variables this trained model predicts correctly whether the rating of the collaborated game is higher or not than the average rating of the previous 3 games of designer X in 78% of the time.

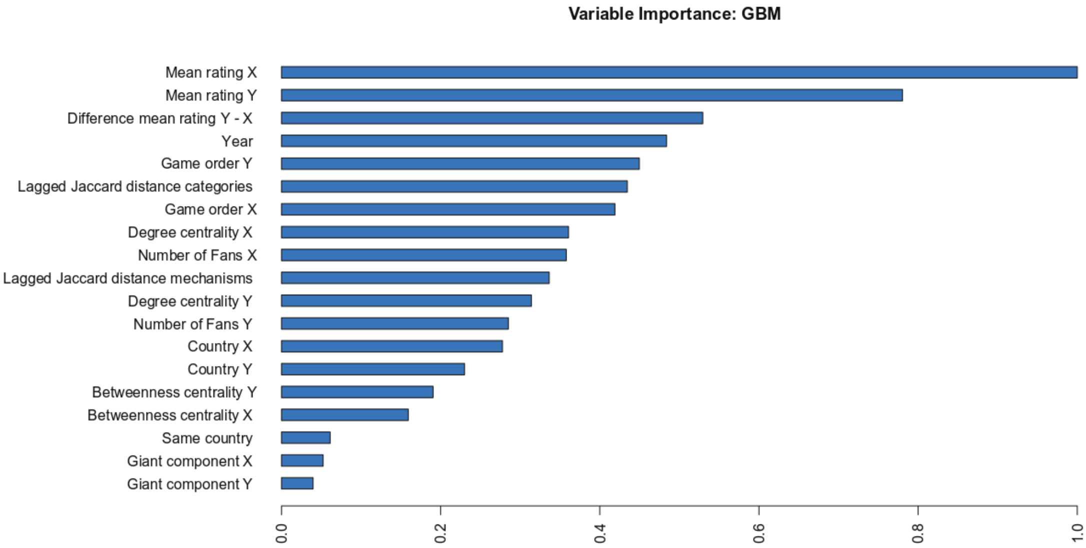


Figure 10. Variable Importance for Predicting Increase in Rating.

Figure 10 shows the relative importance of the variables of the trained Gradient Boosting Machine model that are used to make predictions whether the rating of the collaborated game is higher than the mean rating of the previous three games of the focal designer X. This means rating of designer X is the strongest predictor. This makes sense intuitively, because designers with relative low mean ratings have ample room for improvements.

The following two predictors are the key variables for this paper. The *Mean Rating Y* indicates that the past performance of a collaborator has relative strong influence

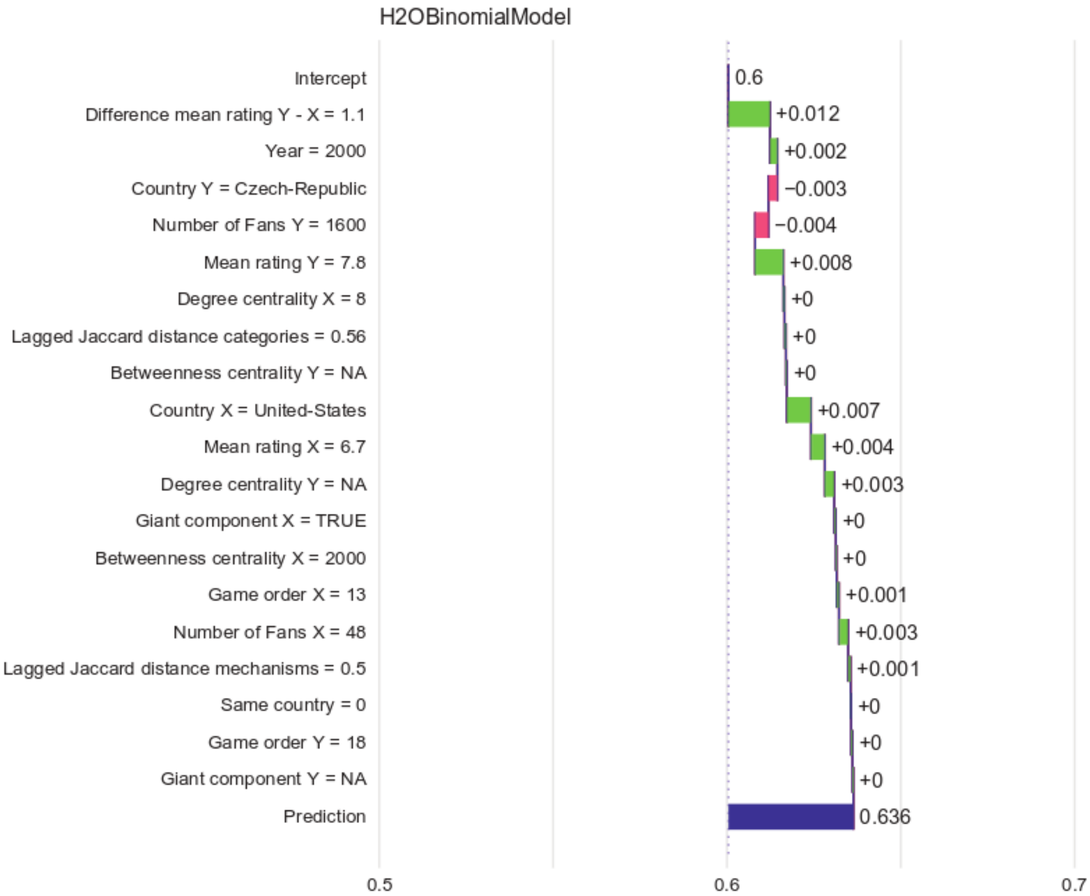


Figure 11. Example of Single Observation: How collaboration characteristics result in prediction

on whether the collaboration with designer X results in a game rated higher than the mean rating of designer X's previous three games. Moreover, the difference in the mean rating of the collaborator and the mean rating of focal designer X provides additional predictive power. This means that in addition to the mean rating of the focal designer and collaborator, the distance between these ratings provides valuable insights whether the mean rating of focal designer will increase or not.

Aside from quality of collaborator, the proximity between collaborators in terms of experience in categories and mechanisms helps to predict whether the focal designer improved its mean rating. Figure 10 indicates that the *Lagged Jaccard distance Categories* and *Lagged Jaccard distance Mechanisms* have relative strong predictive power. The proximity in experience in Categories is slightly more informative in predicting the outcome variable than proximity in experience in Mechanics. Cognitive proximity between collaborating designers impacts whether the rating of collaborating game increases the mean rating of the focal designer.

Network effects are important too, but seem to matter less than quality of collaborator and cognitive proximity. The degree centrality of the focal designer is a relative strong predictor of the outcome, whereas the degree centrality of the collaborator Y is less informative for prediction. Betweenness centrality and whether the focal designer and collaborator are in the giant component of the designer collaboration network has very limited predictive power in our trained Gradient Boosting Machine model.

Out of the four potential factors that influence how team-composition impacts performance, geography has the smallest predictive power in our machine learning model. The country of the focal designer and collaborator have predictive power, but considerably less than the other predictors in the model. In addition, whether the collaborating designers live in the same country or not holds very little predictive power.

Figure 11 provides the breakdown of *how* our trained model makes decisions for improvement of rating (1) or no improvement (0) for an individual case. The average rating of collaborator Y (7.8) is greater than the average rating of focal designer X (6.7). The difference between the ratings of Y and X (1.1) helps predict whether their collaboration results in game that has a higher rating than the average rating of designer X. In this case, the difference between Y and X is the variable that has the strongest prediction power (+0.012).

A couple things are important to note. First, the cut-off point in our GBM model was an optimized F1 score of 0.595 and not the usual 0.5 or 50%. This means that as soon as an observation hits a score of 0.595 or higher, the prediction will be classified with a 1 and 0 otherwise. Second, some variables (i.e. *Lagged Jaccard Distance Categories*) are not predictive in our chosen example observation. However, on the overall set of observations these covariates do have an impact on the prediction as shown in 10. This is because the actual score on these covariates in this particular observation is not predictive. In other cases these scores could very well be predictive.

Overall, the findings of our supervised machine learning model are in line with those obtained from our econometric analyses. Together with our matching approach in pre-processing our data, this suggest that the results for our analyses are robust and not subject or depending on model specifications.