# Behavioral Insights Into Knowledge Work

## Information Sourcing, Peer Dynamics, and Gender Disparities in Ideation

**Svenja Friess**

Dissertation

Munich 2025

Max Planck Institute for Innovation and Competition
LMU Munich, Department of Economics

# Behavioral Insights Into Knowledge Work

Information Sourcing, Peer Dynamics, and
Gender Disparities in Ideation

Inaugural-Dissertation
zur Erlangung des Grades

**Doctor oeconomiae publicae**
**(Dr. oec. publ.)**

an der Volkswirtschaftlichen Fakultät der
Ludwig–Maximilians–Universität München
2025

vorgelegt von
**Svenja Frieß**

Referent: Prof. Dietmar Harhoff, Ph.D.
Korreferent: Prof. Dr. Florian Englmaier
Promotionsabschlussberatung: 29. Januar 2025

# Acknowledgements

This dissertation could not have been achieved without the support and encouragement of many fantastic individuals.

First and foremost, I am profoundly grateful to my doctoral advisor, Dietmar Harhoff, for his initial trust, guidance, and generous support of my research endeavors over the past years. He gave me the freedom to pursue topics I was passionate about within the productive and supportive department he had established. I also sincerely thank Florian Englmaier for including me in his organizational economics chair at LMU Munich, my second academic family. The fruitful discussions and his loyal support have been invaluable. Finally, I am thankful to Tobias Kretschmer for his input on one chapter, for serving on my committee, and for believing in my future research potential.

A large portion of this dissertation is the result of teamwork, and I am deeply thankful to my co-authors for our fruitful collaborations. I have learned many valuable lessons from each of them. I thank Marina Chugunova for her calmness in challenging situations, her fresh perspectives, and for drawing me into my first project with an intriguing topic and ice cream. I am also grateful to Lea Heursen for the academic rigor and ambition she showed me and for teaching me patience throughout the scientific process. Finally, I sincerely thank Laura Rosendahl Huber for her trust, guidance, and encouragement to explore my own research ideas. I also gratefully appreciated the research assistants who supported my dissertation, especially Buse N. Caba.

I extend my gratitude to Karim Lakhani and the Laboratory for Innovation Science at Harvard (LISH) group for their hospitality, inspiring environment, and scholarly exchange. I also gratefully acknowledge financial support from Fulbright Germany for my research stay at Harvard University in spring 2022.

Igor, Yves, Ingrid, and Valeria have been inspiring academic role models. Their integrity, work ethic, and perspectives on the academic profession have greatly motivated me throughout my academic journey.

The Max Planck Institute for Innovation and Competition provided an exceptional research environment. I am especially grateful to Erik for his invaluable coding assistance and zen moments in yoga and to Gisela for her unwavering support. My thanks also go to Myriam for enhancing my research both linguistically and visually, as well as for her encouragement. I also acknowledge the support of the MPI's IT team and administration. Karen, thank you for sharing your fresh perspectives on academia and your knowledge of writing techniques with me.

My MPI colleagues Dennis, Nadine, Ann-Christin, Timm, Felix, Alexander, and Rainer (and Ginnie) have impacted my doctoral journey with their insights, support, and inspiring personalities — it truly takes a village. I will fondly remember our research seminars (*DokSems*) at unique locations like Ringberg Castle and Schneefernerhaus, as well as the cakes on *the bridge*. Special thanks to my stellar friends from the law departments, Carolina, Rebeca, and Miriam, for their encouragement, countless laughter, and coffee chats that always brighten my days.

I am deeply grateful to LMU Munich's Micro Group for including me and exchanging ideas over the years. The friendships I've formed — my true "top 5's" (plus) — with Silvia, Jae, Carolin, Lena, Dominik, Michael, Hoa, Tim, Friederike, Helene, Matylda, and Katharina are invaluable. I'll always cherish the Eisbach dips, "research walks" in the Englischer Garten, and conversations over lunch and (Venetian) coffee. Our shared highs and lows have genuinely shaped my doctoral journey.

To my family, thank you for being supportive and understanding during my periods of silence. In particular, I am beyond grateful to my mom for always believing in me and for her love and unconditional support in letting me pursue my own path without reservation.

Svenja Friess

*Munich, September 2024*

# Contents

# Preface

Understanding people's choices and behavior is crucial for designing effective organizational policies. Traditionally, economics has been rooted in the core assumption of neoclassical theory: the *homo oeconomicus*, a model of a perfectly rational decision-maker. Over the past decades, *behavioral economics* has grown into a well-established subfield. It challenges this rationality assumption by incorporating concepts from human psychology, such as cognitive biases, heuristics, or social concerns, into its models of decision-making (Rabin, 2002). The foundations of this field originate from seminal works by Simon (1955), Kahneman and Tversky (1979), and Thaler (1980) that documented systematic patterns in how human behavior diverges from rational economic predictions.

These early works paved the way for applying the modeling of "behavioral anomalies" to other subdisciplines, such as *organizational economics* (Camerer & Malmendier, 2007). Scholars began to acknowledge that workers were not perfectly rational agents motivated solely by wages and incorporated behavioral concepts to understand how to incentivize workers' performance optimally. To this end, organizational economists also consider how non-monetary factors, such as job satisfaction, intrinsic motivation, and organizational culture, influence worker behavior and performance.

As organizations have evolved over the past few decades, so has the nature of work. Previously dominating manual, routine labor has increasingly been replaced by non-routine, analytical tasks to sustain organizations' productivity (Autor, Levy, & Murnane, 2003; Price & Price, 2013). This transition has led to the emergence of a new class of employees – the so-called *knowledge workers*. This concept describes employees who primarily think for a living and generate value for an organization by applying their high-level knowledge to their work (Drucker, 1999). Often, knowledge workers have acquired their specialized knowledge through formal (university) education or professional certification programs.[1]

With knowledge-intensive labor becoming increasingly central to the modern economy, the importance of *knowledge flows* for organizational success has grown accordingly (Sandvik et al., 2020). Specifically, smooth knowledge flows within organizations can reduce information search time, prevent redundant work, and foster collaborative networks among members. However, knowledge dissemination mechanisms within organizational structures remain understudied. Both knowledge seekers and providers face various frictions that may hamper efficient information exchange, whether in physical workplaces or remote environments. For example, employees may hesitate to source information out of fear that it could make them appear less competent, which can impair their learning opportunities and, thus, performance. Understanding these frictions is essential for organizations aiming to enhance productivity and stay competitive in a rapidly evolving economy.

Innovative professionals such as scientists, inventors, and entrepreneurs stand out as an important subgroup among knowledge workers. *Innovations* create new markets, increase organizational productivity, and drive economic growth (Agarwal, Audretsch, & Sarkar, 2007; Akcigit, Grigsby, & Nicholas, 2017). Schumpeter's 'creative destruction' highlights the pivotal role of innovation in reshaping industries by replacing outdated technologies with breakthroughs (Schumpeter, 1942). Similarly, endogenous growth theory links economic expansion to the arrival of new ideas (Romer, 1990; Aghion & Howitt, 1992). Innovation is naturally linked to *creativity*, relying on creative ideation at its onset (Amabile, 1996; M. Baer & Frese, 2003). Consequently, fostering creative thinking and problem-solving is essential not only for organizations' success but for sustaining economic growth altogether, making it a topic of interest to scholars and practitioners alike.

Despite overall trends of gender convergence in employment and earnings, women continue to be underrepresented in innovative knowledge work (Bechthold et al., 2021; Albanesi, Olivetti, & Petrongolo, 2023). And even among those who entered innovative fields, women tend to engage less frequently in tangible innovative activities compared to their male counterparts. For example, women represented 28% of the US science and engineering workforce in 2021 (National Science Foundation, 2023). In contrast, the share of female inventors on granted USPTO patents was 10.5% in the same year (World Economic Forum, 2024). In the start-up world, female founders

[1] Source: https://www.forbes.com/councils/forbestechcouncil/2020/12/10/the-year-of-the-knowledge-worker/, last retrieved on Sept 9, 2024.

represented 23% of all US entrepreneurs in 2019 (European Investment Bank, 2022). During the same year, the share of venture capital (VC) funding acquired by founding teams, including at least one woman, was 14.5%, and by all-female US startups was only 2.6% (PitchBook, 2024). The negative economic impact is significant: these 'lost Marie Curies,' as described by Hoisl, Kongsted, and Mariani (2023), represent untapped innovation potential that could boost firms' capabilities, productivity, and economic growth. Therefore, understanding and removing the barriers that hinder women's full participation in these fields is essential to unlocking their innovative contributions.

The overarching objective of this dissertation is to investigate factors that influence the transmission and generation of knowledge within organizations. Specifically, this thesis aims to understand core trade-offs in knowledge-sourcing behavior (Chapter 1), peer dynamics in engaging and communicating on digital platforms (Chapter 2), and gender disparities in innovative processes (Chapter 3). The first essay examines how reputational concerns influence advice-seeking behavior at work, revealing how misperceptions about its reputational consequences can inhibit knowledge sourcing. The second study provides insights into how early, positive interactions can enhance future participation on a knowledge exchange platform, offering platform designers guidance on sustaining user engagement and improving platform value. Finally, the third essay investigates how competition incentives and competitor gender composition impact gender disparities in generating and selecting innovative ideas.

In order to comprehend the phenomena mentioned above, all chapters of this dissertation regard broad and heterogeneous samples of knowledge workers from the United States and the United Kingdom. This dissertation utilizes a modern mix of quasi-experimental, experimental, and text data methods. This dissertation exploits a quasi-experimental setting in Chapter 2 and uses two artefactual online field experiments in Chapters 1 and 3. For the questions studied in these two experimental chapters, the availability of real field settings that would allow for a consistent comparison of tasks and decisions without potentially significant career concerns for participants was limited. The employed methodology is described in the next paragraphs.

Quasi-experimental and experimental methods compare "treated" individuals with untreated "counterfactuals" to draw causal inference but differ in how this comparison group is chosen. First, quasi-experiments take advantage of naturally occurring variation in circumstances if a sufficient degree of randomness in individuals' allocation to being treated can be guaranteed. Prominent examples include natural disasters, unanticipated policy changes, or “policy lotteries” (Angrist & Pischke, 2009). Second, researchers can apply experimental methods to collect data themselves whenever no suitable natural experiment is available. In doing so, individuals are randomly assigned to treatment groups by the researcher. This approach ensures internal validity due to outcome differences attributable to the intervention rather than (unobserved) pre-existing factors. Experimental methodology broadly distinguishes laboratory from field experiments, where the latter are further categorized into natural, framed, and artefactual field experiments, according to Harrison and List (2004).

Natural field experiments have been considered a methodological benchmark within experimental methods because they combine high internal and external validity. As natural field experiments occur in real-world settings where participants are often unaware of their participation in an experiment, the risk of behavioral changes simply due to knowing to be observed — as in laboratory experiments — is reduced. The real-world context also enhances external validity, making findings more generalizable across populations. These features make natural field experiments particularly insightful for public policy and business decisions (Duflo, Glennerster, & Kremer, 2007; Angrist & Pischke, 2009). In contrast to laboratory experiments with "standard" student samples, framed and artefactual field experiments still use "non-standard" participant samples but with more realistic versus more abstract framing, tasks, information structures, or contexts, respectively (Levitt & List, 2009).

In recent years, social scientists have started to use text data as novel input to their research because digital text records such as social media postings, news articles, or patent texts make up an ever-increasing share of human interactions, communication, and codified knowledge. The information contained in the text, thus, offers a rich complement to more traditional data sources. Text data is, however, quite different from other structured data sources frequently used in economic analysis in that it is high dimensional. Consequently, the statistical methods used for analyzing text data must account for its high-dimensionality. This includes Natural Language Processing (NLP) methods that apply machine learning models to extract meaningful patterns, sentiments, or common themes from unstructured text data (Athey & Imbens, 2019; Gentzkow, Kelly, & Taddy, 2019).

Within the context of economic experiments, NLP methods allow researchers a more advanced analysis of qualitative responses from participants, e.g., by categorizing open-ended survey responses or examining linguistic patterns from communication between participants when studying group decision-making. Hence, these tools offer two advantages: potentially generating novel insights into human behavior and the possibility to classify and quantify large volumes of text data more efficiently than manual coding would allow.

In the following, I provide a non-technical description of each chapter included in this dissertation.

**Chapter I** researches the decision-making processes in workplace environments, focusing on how reputational concerns affect knowledge-sourcing behavior. Sourcing information or advice from colleagues is an important way to learn and enhance performance. However, individuals often fail to do so, potentially due to concerns about how it may affect others' perceptions of their competence — so-called reputational concerns. The central research question is whether such reputational concerns lead employees to strategically avoid seeking advice to appear more competent in the eyes of others, even when doing so may hinder their productivity.

My co-authors, Lea Heursen, Marina Chugunova, and I conduct an online artefactual field experiment with a large sample of 2,521 white-collar professionals and managers to establish a causal relationship between the propensity to seek advice and potential reputational concerns. Participants ("Employees") completed a general knowledge quiz and were given the option to seek advice on the quiz, which could improve their performance. The experiment varied whether the decision to seek advice was visible to a bonus-awarding "Manager" (i.e., switching reputational concerns on or off) and the quiz topic. The latter dimension manipulates gender-stereotypical beliefs about competence (i.e., high or low), which likely affects trading off the information benefit against the expected reputation costs of seeking advice. We gather participants' detailed explanations about their decisions regarding the main outcomes and categorize these text responses to gain deeper insights into their motives.

We find that the rate of seeking advice decreases by 16% when the advice-seeking decision was visible to a "Manager", i.e., when reputation concerns are present, even though it could improve performance. There was little evidence that stereotypical beliefs about competence played a strong role in influencing behavior, although we document substantial heterogeneity in expected reputational costs to seeking advice across participants. "Managers"' data shows that seeking advice has no economically meaningful (negative or positive) impact on the bonuses they assign based on their performance estimates when visible to them in brief "Employee" profiles.

This research contributes to several strands of literature. First, it extends work on advice-seeking behavior by providing the first causal evidence on the strategic trade-off between the benefit of advice and reputational concerns in workplace settings. It also adds to the literature on stereotypes and competence, offering new insights into how higher-order beliefs about stereotypes may influence knowledge-sourcing behavior. Finally, it builds on research about reputational concerns in the workplace, showing that such concerns can lead to inefficient behaviors, in particular, forgoing beneficial information.

In summary, these findings document professionals' widespread misperceptions and heterogeneity about the reputational consequences of sourcing knowledge from others. This has important implications for organizational policies aimed at promoting effective knowledge flows among members. For instance, organizations may focus on creating environments that address and correct these misperceptions while also considering the roles of knowledge providers and leveraging technological solutions to enhance knowledge flows.

**Chapter II** examines how early peer interactions influence long-term user engagement and persistence on digital knowledge exchange platforms. Digital advancements are reshaping knowledge exchange, enabling information sharing in various organizational and private contexts. Despite the growing market for such platforms, many struggle with declining user engagement over time, a critical issue as sustained engagement is necessary for generating valuable knowledge flows.

My co-author Laura Rosendahl Huber and I use novel interaction data from over 12,000 working professionals participating in online business courses over a five-year period. We causally analyze how initial activity influences a user's long-term engagement and persistence on the platform by leveraging quasi-random variation in peer behavior during the first period. We differentiate between two types of user activity: *general platform activity* and *direct interactions* targeted at specific users. This distinction helps us to shed light on how different types of engagement norms — broad, top-down norms versus personalized, bottom-up nudges from peers impact user behavior.

Our data reveals significant variation in user engagement across cohorts and time periods. Our findings show that users who receive early directed comments or likes are more likely to contribute and stay active. Engagement and persistence are particularly high for those receiving 'elaborating and agreeing'-comments. Interestingly, users in cohorts with fewer peers sharing likes in the first period are 3% more likely to persist compared to those with many such peers.

This study makes a contribution to several lines of research. First, it extends the literature on user-generated content platforms by exploring how the timing, type, and content of peer interactions affect user behavior early on. This area has been understudied due to challenges related to peer interactions' timing and endogenous nature. Second, we contribute to the broader literature on peer effects by adding evidence from fine-grained interaction data over a longer-term engagement period in an online setting. Lastly, our findings have implications for labor force training, as it focuses on upskilling working professionals and executives. Overall, our study offers insights into how digital platforms can be designed to foster continued user engagement and, thereby, smooth knowledge exchange among members. Its results highlight the relative benefits of integrating individual members versus establishing norms for cohort interactions early on.

**Chapter III** studies whether the often competitive and male-dominated nature of innovative environments discourages women from innovating. Although innovation is a key driver of economic growth, women participate less in measurable, innovative activities. Understanding the barriers women in such environments face is a necessary first step to overcoming them and unlocking their untapped innovative potential.

I conduct an artefactual field experiment in an online labor market to investigate gender differences in creative ideation and idea selection — both critical early stages of the innovation process. The experiment exogenously varies the disclosure of gender, gender composition of competitors, and the incentive structure to assess their impact on creative output. This approach allows for causal analysis, overcoming limitations related to endogenous selection, e.g., into specific firms, of observational settings, and permits the consistent measurement and comparison of creative output across incentive structures and individuals. Moreover, I employ NLP techniques for robustness analyses of text-based outcome measures and classify participants' text responses about their main experimental choice for additional insights into their behaviors.

The results indicate that women outperform men in creative ideation, even in competitive settings. Women consistently select more original and higher-quality ideas than men. However, the gender composition of the competitive environment plays a significant role, with men improving their idea selection when gender is disclosed and their performance in gender-balanced competitions. Additionally, women are less overconfident in anonymous competitions but become more confident in male-dominated settings. They also see gender-balanced environments as more competitive than men do.

My findings have important implications for knowledge work in organizations, particularly for those relying on teamwork and creative problem-solving. Women excel in creative ideation, suggesting that organizations should better leverage their strengths. The gender composition of competitive environments affects performance differently. Thus, organizations should recognize that one-size-fits-all competition may not yield the best results. By ensuring equal opportunities, fair evaluations, and tailored feedback to calibrate confidence about creative tasks well, organizations can help both men and women to realize their full creative potential and harness their best innovative ideas.

This research contributes to the literature strands on gender differences in mixed- versus single-gender environments, on incentives for creativity, and, more broadly, on gender gaps in innovative knowledge work. I provide novel micro-level evidence on gender gaps in creative idea selection and isolate the effects of the competitors' gender composition from a pure competition effect on creative processes and decision-making.

In summary, this dissertation offers new behavioral insights for academics and practitioners alike. Its findings highlight the importance of psychological and social dynamics in knowledge management. Individuals in organizations do not always act as perfectly rational *homines oeconomici*; their knowledge exchange and creative ideation behaviors are influenced by factors such as misconceptions, reputational concerns, and environmental conditions like gender composition and peers' activities. While this dissertation provides a starting point for understanding these intricate dynamics, further research — especially through (natural) field experiments in collaboration with firms — could offer deeper insights into how organizations can better harness and disseminate innovative ideas, which are critical for their long-term success.

# 1

# Reputational Concerns and Advice-Seeking at Work

## 1.1 Introduction

In the workplace, individuals engage in various activities to manage their reputation by signaling attributes like motivation, dedication, or competence (e.g., Anger, 2008; Campbell & Hahl, 2022). This study focuses on the decision to forego an individually beneficial action to potentially enhance one's perceived competence in the eyes of others, specifically, the decision to seek advice. Seeking advice at work is an important way to learn and to improve. While it facilitates knowledge flows crucial for firm productivity (Garicano, 2000; Sandvik et al., 2020), people often fail to seek advice (Lee, 1997, 2002). We investigate whether reputational concerns contribute to this and ask: do individuals strategically forego the information benefit of advice to appear more competent?

When trading off the information benefit of advice against its expected reputational cost, others' beliefs about competence could amplify or mute perceptions of the cost. Beliefs about competence are often rooted in stereotypes (Coffman, 2014; Bordalo et al., 2019) with documented impacts on important economic choices such as hiring and performance evaluations (Reuben, Sapienza, & Zingales, 2014; Bohren, Imas, & Rosenberg, 2019; Sarsons, 2019; Coffman, Exley, & Niederle, 2021; Barron et al., 2024; Campos-Mercade & Mengel, 2024).[1] People rely on stereotypes the most when competence is uncertain, for example, at a new job. This coincides with a time when the information value of advice is particularly large. Hence, we investigate whether others' beliefs about competence affect workplace behavior that may be interpreted as a negative signal of competence. How individuals respond to others' stereotypi-

*This chapter is based on joint work with Lea Heursen and Marina Chugunova. We are grateful to Valeria Burdea, Thomas Buser, Erina Itsma, Karin Hoisl, Nathalie Roemer, Roberto Weber, seminar participants at the Max Planck Institute for Innovation & Competition LMU Munich, HU Berlin, CREST, University of Zurich, Center for European Economic Research and conference participants at the NBER, CESifo Behavioral Economics Conference, OESS, COPE, SIOE, and ESA for valuable comments. We thank the Diligentia Foundation for Empirical Research for the research grant that funded data collection for this project. Chugunova and Heursen also gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119). The experimental design and empirical analysis were pre-registered with the Open Science Framework (https://osf.io/24zwt/). We thank Erik Buunk for his help in programming the experiment. We thank Nishan Lin, Margarita Gatsou, and Buse Caba for their excellent research assistance. We used the generative AI tool ChatGPT for coding and the AI-based writing assistance tool Grammarly to edit the text for spelling, grammar, and style. We reviewed and edited this content as needed and take full responsibility for the content of this study.

[1] While all these studies document stereotypical beliefs about competence connected to gender, they can be linked to other salient group characteristics such as age, ethnicity or socioeconomic status.

cal beliefs about competence is the next step to understanding how such perceptions influence economic behavior and outcomes. We are the first to consider if anticipating others' stereotypical beliefs about competence changes behavior in the workplace when identity is known.

We conducted a large-sample (N=2,521) artefactual field experiment with white-collar professionals ("Employees") and professionals with managerial experience ("Managers") to establish a causal relationship between the propensity to seek advice and potential reputational concerns. Employees answer a general knowledge quiz for a piece-rate. After that, they have the option to seek computerized advice, which simplifies but does not solve the quiz, preserving the Employees' agency over the final answer.[2] The experiment manipulates two dimensions. First, we vary whether the decision to seek advice is visible to a Manager, who estimates the Employee's independent performance in the quiz based on a short profile. This estimate determines the size of a substantial bonus that the Employee obtains. Second, we vary the topic of the quiz — "Science & Technology" or "Psychology & Linguistics" — to probe how others' stereotypical beliefs about competence affect the decision to visibly seek advice.

Our results demonstrate that reputational concerns hinder knowledge flows, as the rate of advice-seeking decreases by 16 percent when the choice is visible to the Manager. Employees refrain from seeking advice when it is visible, despite the potential for performance improvement and increased payment. Interestingly, we find little evidence that others' stereotypical beliefs about competence influence how individuals weigh the information benefit of advice against reputational concerns on average. If anything, the estimates, though insignificant, point in the direction that professionals may be more reluctant to visibly seek advice when it conflicts with an advantageous competence stereotype. We document significant heterogeneity in the expected reputational cost of seeking advice, ranging from perceiving substantial costs to substantial benefits. However, our analysis of managers' data reveals no actual, economically meaningful reputational cost or benefit of seeking advice.

Taken together, our findings highlight widespread misperceptions concerning the reputational consequences of advice-seeking among professionals. Correcting them could foster knowledge exchange and learning. Our empirical strategy offers several key advantages to address our research questions. First, the artefactual field experiment allows us to uncover whether individuals are willing to forego the benefit of advice

[2] Agency over decisions has been described as the key feature that distinguishes advice from other forms of help (see Brooks, Gino, & Schweitzer, 2015).

due to reputational concerns. To achieve this, we simplify the complex interaction of advice-seeking and -giving, while initially holding constant other mediating factors like advisor characteristics, advice quality, or type.

Moreover, we focus on the information benefit of seeking advice, excluding other purposes such as relationship-building or networking and initially limiting, by design, the potentially positive signaling value that seeking can have.[3] This streamlined decision environment can serve as a foundation for further research on the micro-determinants of knowledge flows in organizations. Second, we can manipulate others' beliefs about an individual's competence on a task without altering the task or work environment. This allows us to investigate the causal role of (higher-order) stereotypical beliefs in influencing behavior that can be interpreted as a negative signal of competence. Third, our experimental design enables us to shed light on mechanisms since we can precisely measure the information benefit of advice, the expected reputational cost of seeking advice, and quantify beliefs about others' stereotypical beliefs about competence. Finally, by running our experiment with a sample of white-collar professionals, we consider a population that likely regularly faces a trade-off between the information benefit of advice and perceptions of competence, due to the nature of knowledge-intensive jobs, where remuneration and career progression also depend on subjective evaluations from superiors and peers (e.g., Benson, Li, & Shue, 2024).

Our study contributes to three distinct literatures. First, it extends a growing body of evidence in economics and related disciplines on what determines the decision to seek advice and how it is perceived.[4] Previous empirical work has isolated the role of social costs in the form of shame or stigma for advice-seeking (Chandrasekhar, Golub, & Yang, 2018) and studied whether anonymity can encourage knowledge-seeking on organizational platforms (Mickeler et al., 2023). Other recent studies have focused on the relationship between adviser and advisee, investigating homophily (Heikensten & Isaksson, 2019) and the fear of rejection when asking for help in general (Bénabou, Jaroszewicz, & Loewenstein, 2022). Our findings highlight strategic advice-seeking, where professionals weigh the information benefit of advice against possible reputa-

[3] Advice-seeking at work could also be perceived positively, as a sign of self-awareness or self-assuredness. Given our research questions, we decided to abstract from this additional potential benefit to seeking advice. Introducing it would be a very relevant and interesting extension of this study. In principle, this experiment's measure of the expected reputational cost of seeking advice also elicits an expected reputational benefit by allowing for negative values. See Section 4.3. for a description of this measure and the corresponding results.

[4] Highlighting the importance and complexity of understanding advice at work, recent studies examine the content of advice that seekers receive (Gallen & Wasserman, 2021; Kessel, Mollerstrom, & Van Veldhuizen, 2021).

tional costs. Further, our results provide the first evidence that higher-order beliefs about competence cannot encourage individuals to seek advice.

Turning to perceptions of advice-seeking, initial evidence is mixed as to whether it is perceived negatively or positively by others (Brooks, Gino, & Schweitzer, 2015; Rosette, Mueller, & Lebel, 2015; Blunden et al., 2019). By jointly studying beliefs about how seeking is perceived and how it is actually perceived, our research design enables us to document widespread misperceptions among professionals across a variety of organizations. Our results reveal a potentially significant obstacle to knowledge flows within organizations, and they also suggest a potential solution by addressing misperceptions.

Additionally, we contribute to the literature on the behavioral implications of stereotypical perceptions of competence. Previous research has extensively documented the presence of gender-based stereotypes in beliefs about competence (e.g., Coffman, 2014; Bordalo et al., 2019) and their impacts on important economic decisions such as hiring or performance evaluations (Reuben, Sapienza, & Zingales, 2014; Bohren, Imas, & Rosenberg, 2019; Sarsons, 2019; Coffman, Exley, & Niederle, 2021; Barron et al., 2024; Campos-Mercade & Mengel, 2024). Understanding how individuals respond to others' stereotypical beliefs is an important area for further investigation. Initial evidence suggests that people behave strategically and attempt to act upon others' stereotypical beliefs: they conceal or misreport their gender at the hiring stage (Alston, 2019; Charness et al., 2020) and hide their ethnic minority status or affinity to LGBTQ+ to encourage others' prosocial behaviors towards them (Kudashvili & Lergetporer, 2022; B. Aksoy, Chadd, & Koh, 2023). We advance this nascent literature by providing the first evidence of whether anticipating others' stereotypical beliefs changes behavior when identity is known.

Third, we contribute to the literature investigating how reputational concerns influence workplace behavior when the ability is not observable.[5] Reputational concerns can incentivize agents to engage in workplace activities—effort or (excessive) risk-taking—that can improve the performance measure from which ability is inferred (see, e.g., the models by Gibbons & Murphy, 1992; Holmström, 1999). Theoretically, it has been argued that reputational concerns can further lead to undesirable behaviors, such as an inefficient use of new information (Prendergast & Stole, 1996), unwillingness

[5]The term "reputational concerns" is also used in theoretical and empirical work on relational contracting, referring to a mechanism to build trust and credibility between parties. This differs from its use in the above-cited literature, where reputational concerns denote an agent's desire to enhance their reputation in the labor market.

to convey a true judgment (Morris, 2001; Ottaviani & Sørensen, 2006) or a reduced willingness to help colleagues (Auriol, Friebel, & Pechlivanos, 2002).
Causal empirical evidence on how reputational concerns change workplace behavior is relatively scant. Early laboratory experiments tested the incentive effect of reputational concerns (Irlenbusch & Sliwka, 2006; Koch, Morgenstern, & Raab, 2009). Beyond that, experiments find that reputational concerns cause individuals to opt for unnecessarily complex and risky solutions to a problem (Katok & Siemsen, 2011) or to spend too much effort on activities that can influence a superior's tenure decision (de Janvry et al., 2023). We expand this literature by empirically showing the causal impact of reputational concerns on workplace behavior that could be interpreted as a negative signal of competence.[6]

The remainder of the study is structured as follows. In section 1.2, we present the research design. Section 1.3 introduces a conceptual framework and develops hypotheses. In section 1.4, we show our results, and section 1.5 discusses them and concludes.

## 1.2 Design

To answer our research questions, we designed an online field experiment and a survey. We present them in turn.

### 1.2.1 Online Experiment

Participants took part either as an "Employee" or a "Manager". In a simulated work environment, Employees worked on a knowledge task for a piece-rate and could decide whether to seek advice to improve their performance and pay. They were randomly paired with a Manager who estimated their independent task performance based on seeing the Employee's profile. Employees' pay increased in this estimate. The experiment has a 2x2 between-subjects design. First, we randomly varied if the Employees'

[6]In management, there is a large literature studying impression management. Impression management manifests in explicit or implicit attempts to manage how one is perceived by others. This literature largely focuses on self-presentation and its impact on desired outcomes such as hiring decisions, performance evaluations, and career success (Al-Shatti & Ohana, 2021). Liljenquist (2010) finds that self-promotion disguised as seeking advice—allows to increase others' perceptions of warmth without affecting perceptions of competence. Feedback-seeking, a behavior that does not signal a lack of competence, was frequently studied as a tool of impression management (e.g., Ashford, Blatt, & VandeWalle, 2003).

advice-seeking decision was visible to the Manager (*Private* or *Visible* conditions). Second, we relied on gendered stereotypes about domains of knowledge to manipulate stereotypical perceptions of competence with the topic of the knowledge task: "Psychology & Linguistics" or "Science & Technology". Third, we stratified recruitment by sex to ensure a balanced sample of women and men across experimental conditions.

### Employee Version

Before starting with the main study, Employees answered a brief survey on demographics, education, and their labor market status (see Table A.1 for descriptive statistics). Some of their answers were used as input later (see Figure 1.1 Panel 1.1a for a general outline of this version of the experiment).

#### - Part 1 -

**Knowledge Quiz** In Part 1, participants answered a general knowledge quiz of 10 multiple-choice questions with 5 answer options either on "Psychology & Linguistics" or "Science & Technology". Correct answers paid £0.10. Each question referred to a picture that was instrumental for answering correctly (see Figure A.1 Panel A.1a for an example). Participants had 30 seconds to answer each question. The combination of pictures and the time limit made it practically infeasible to search online for the correct answer.[7] This way, we ensured that Employees had to rely on their knowledge to answer the quiz. The quiz score in Part 1 is our measure of Employees' task competence. We opted for a knowledge quiz to highlight the role of competence in task performance and limit the role of effort. After finishing the quiz, Employees reported how many correct answers they thought they had provided, which served as a measure of their performance. They received £0.25 if their report was correct. Employees did not learn their actual quiz scores until they had completed the study.

#### - Part 2 -

**Advice-seeking Decision** In Part 2, Employees were offered the option to *seek advice* and revisit the simplified version of the same quiz. If Employees decided to seek advice,

[7] The time limit of 30 seconds was calibrated with pre-tests. It allowed participants to meaningfully consider the question.

each question had 2 instead of the initial 5 answer options.[8] Therefore, advice could help to improve Employees' performance and pay since the chance of simply guessing correctly increased from 20% to 50%. A computer randomly removed the same 3 incorrect answer options for all the advice-seekers. If their initial answer from Part 1 was among the 2 remaining answer options, it was highlighted (see Figure A.1 Panel A.1b for an example). Employees had 15 seconds to revise each simplified question. The advice came at a small one-time fee of £0.08. Employees received £0.10 per correct answer, regardless of whether it was obtained in Part 1 or after seeking advice in Part 2. Hence, a perfect quiz score with or without advice yielded a "performance pay" of £1.00.

If Employees decided not to seek advice, they proceeded to the final questionnaire without the opportunity to revise their answers. Employees were informed that they could learn the correct answers to the quiz at the end of the experiment, regardless of their decision to seek advice. Before making the decision to seek advice, as a reminder, they saw for 15 seconds the list of questions they encountered in Part 1 without the corresponding pictures.

Our design aims to create a stylized decision-seeking environment that mimics some important characteristics of advice-seeking in the workplace. Seeking advice on the entire quiz rather than individual questions parallels advice-seeking on complex tasks. For example, seeking advice on a project report as a whole instead of its individual sentences would provide more targeted insights, ensure coherence, and promote strategic improvement for the overall task. By computerizing advice, we eliminated the social component of seeking and reduced its non-monetary costs, for example, due to shame or fear of rejection. At the same time, we gained control over the quality of advice and, importantly, *beliefs* about its quality. By simplifying the quiz with advice instead of solving it correctly for the advice-seeker, we preserved the agency of Employees in how to respond to advice, which is the key feature that distinguishes advice from other forms of help (see, e.g., Brooks, Gino, & Schweitzer, 2015). By introducing a small advice cost, we modeled the seekers' and the advice-giver's opportunity cost of time. We announced that participants could learn the correct answers to all the questions at the end of the experiment. We reminded them of the questions they had just encountered before making the advice decision to limit the role that curiosity or memory constraints may play in the decision to seek advice.

[8] During Part 1, participants did not know that they would have the possibility to revisit the quiz later. Therefore, Part 1 measured their true knowledge of the quiz topic.

**Manager's Reward** Employees were randomly paired with a Manager and, in addition to the £0.10 piece rate for correct answers, received a "Manager's reward". This reward increased linearly in the Manager's estimate of their independent quiz score without advice, which had been recorded in Part 1. The Manager's estimate could be any number between 0 and 10, and the Employee received £0.30 times this estimate. Hence, Employees could earn a "Manager's reward" of up to £3.00. To estimate the Employee's initial quiz score, Managers saw a short profile of their Employee. This profile was common knowledge between the Employee and the Manager (see Panel 1.1b of Figure 1.1 for example profiles). Employees knew that their Manager would be a professional with reported experience in a managerial role who would complete the Manager's version of the study. They were blind to their Manager's gender.
The large 3:1 ratio of "Manager's reward" to "performance pay" intends to mirror a promotion in a real organizational setting, where assessments of competence can play an important role, for example, for promotion decisions or allocation of important tasks that are associated with high rewards. We focused on the estimates of the Employee's prior competence, as opposed to the performance with advice, as in knowledge-intensive jobs, ability or competence is typically valued in and of itself.

**Treatments** The profiles shown to the managers to base their estimates on introduced both dimensions of the treatment variation. The profiles showed the Employees' sex, age range, country of residency, education level, and quiz topic.

The first treatment dimension manipulated the presence of reputational concerns around the decision to seek advice. We varied experimentally whether the Managers would observe the Employees' decision to seek advice before reporting their estimate of the Employees' initial quiz score. In the *Visible* condition, the information on advice-seeking was included in the profile on which the managers based their decision, for example, "She did not seek advice on the quiz." or "He sought advice on the quiz." In the *Private* condition, this information was absent.

The second treatment dimension randomly varied the topic of the general knowledge quiz. Employees either took a quiz on "Science & Technology" or "Psychology & Linguistics".[9] With the combination of a reported group characteristic, in our case sex, and a quiz topic, we manipulated the Managers' beliefs about the Employees' knowledge of a topic and, more importantly, the Employees' beliefs about Managers' beliefs, following the method of Coffman (2014). Figure 1 Panel B displays examples of profiles by condition. The profiles were common knowledge between the Employees and the Managers (see Panel 1.1b Figure 1.1 for example profiles): Employees saw the profiles that their Managers would also see before making their decision of whether to seek advice. Employees in $Visible$ saw two possible versions of their profiles next to each other in a randomized order. The two profiles differed in the bullet about their advice-seeking behavior, which would be determined by their upcoming decision.

Our recruitment strategy ensured that all aspects other than sex, quiz topic, and visibility of advice remained constant across experimental conditions.[10] To make sex less salient, we opted for natural filler characteristics in the context of this study, such as the minimum education level or current country of residence.

The second treatment dimension randomly varied the topic of the general knowledge quiz. Employees either took a quiz on "Science & Technology" or "Psychology & Linguistics".[11] With the combination of a reported group characteristic, in our case sex, and a quiz topic, we seek to manipulate the Manager's beliefs about the Employee's knowledge on a topic and, more importantly, the *Employee's* belief about this belief, following the method of Coffman (2014).

We calibrated the two quizzes to be of comparable difficulty for men and women (see Section A.2.2 of the Online Appendix for details on the calibration of the two quizzes). We pre-tested several dozen questions per topic in the same subject pool. Based on the knowledge of women and men in our pre-test sample, we selected 10 questions

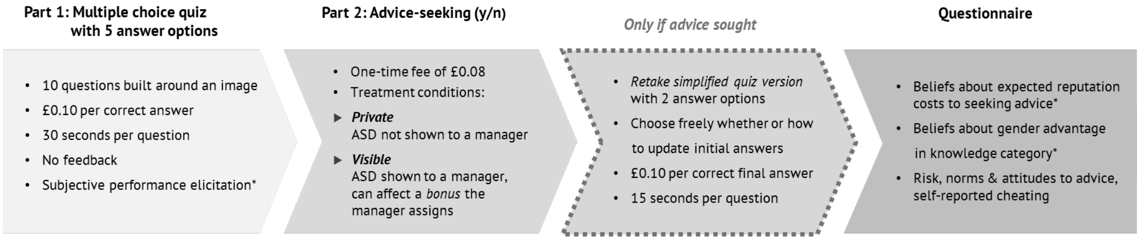
[9] In pre-tests in the same subject pool, we established that these two topics are highly gender-stereotyped and that women are seen as knowing more, on average, on "Psychology & Linguistics" and men are seen as knowing more, on average, on "Science & Technology". We calibrated the two quizzes for comparable difficulty for men and women (see Section A.2.2 of the Online Appendix for details). Based on the knowledge of women and men in our pre-test sample, we selected 10 questions per topic that, on average, yielded 6 correct answers, with 7 being the modal number of correct answers. With this calibration, we wanted to ensure that the information benefit of advice and the difficulty-induced misestimation of own knowledge (Bordalo et al., 2019) would be comparable across experimental conditions. Further, we aimed for a final quiz that was neither too difficult nor too easy. We targeted a unique mode since Managers were incentivized to report the mode of their believed distribution of knowledge in the sample that they were evaluating (see details in Section ).

[10] Data from the short survey at the beginning of the study ensured that these profiles were factually correct.

[11] In pre-tests in the same subject pool, we established that these two topics are highly gender-stereotyped and that women are seen as knowing more, on average, on "Psychology & Linguistics" and men are seen as knowing more, on average, on "Science & Technology".

**Figure 1.1:** Outline of the Experiment and Examples of Employees' Profiles

**(a)** *General Outline of the Experiment (Employee Version)*




**(b)** *Examples of Employees' Profiles as Shown to the Manager*


**Notes.** In Panel A: ASD stands for Advice Seeking Decision of the Employee,* indicates incentivized beliefs. Panel B reproduces examples of profiles shown to Managers by experimental condition (left: *Private*, right: *Visible*). Sex, quiz topic, and (if applicable) advice-seeking varied between participants. All other characteristics remained unchanged, and their factual accuracy was ensured through recruitment filters and confirmed in participants' initial survey responses.

per topic so that participants would, on average, answer 6 questions correctly, with 7 being the modal number of correct answers. With this calibration, we wanted to ensure that the information benefit of advice and the difficulty-induced misestimation of own knowledge (Bordalo et al., 2019) would be comparable across experimental conditions. Further, we aimed for a final quiz that was neither too difficult nor too easy. We targeted a unique mode since Managers were incentivized to report the mode of their believed distribution of knowledge in the sample that they were evaluating.

**Questionnaire** First, in an open-form field, participants were asked to explain what drove their advice-seeking decision. They then indicated how useful they thought it was to seek advice in this study. Afterward, we elicited beliefs about the Manager's quiz score estimate for Employees with two profiles. First, participants guessed a Manager's estimate for another Employee with the exact same profile as theirs. The characteristics of the second profile varied according to the experimental condition. In *Visible*, participants saw a profile identical to theirs in all dimensions but counterfactual advice-seeking. In $Private$, participants saw a profile that was identical to theirs in all dimensions but the Employee's sex. We incentivized these beliefs with an additional £0.25 if the guess equaled the estimate of a randomly selected Manager. One of the two guesses was randomly chosen to be evaluated for payment. The difference between the two guesses measures the belief about either the reputational cost of advice-seeking ($Visible$) or others' stereotypical beliefs about competence ($Private$).

The questionnaire proceeded with eliciting *beliefs about others' stereotypical beliefs about competence* for 6 domains of knowledge with a modified version of the slider measure introduced by Coffman (2014). For each domain, participants reported whether they think that *most people* think that men or women, on average, know more about it. This was done by positioning a slider anywhere between −1 (*most people* think there is a female advantage in knowledge) to 0 (no gender difference) to 1 (*most people* think there is a male advantage in knowledge). This was not incentivized and administered to all participants, regardless of the treatment. Further, participants reported beliefs about the quartile in which they place their independent quiz performance relative to others with the same profile as theirs (i.e., same age range, sex, country of residence, and education).

We elicited risk preferences with two unincentivized measures (Falk et al., 2023). Eight items elicited views on advice-seeking and social norms pertaining to it on a 7-point Likert scale (see the Appendix Table A.5 for a list of all items). To assess whether participants perceived the general knowledge quiz as a measure of competence, we asked about the relative role of luck versus knowledge in performing well on this quiz type, using a scale ranging from 0% ('no luck') to 100% ('only luck'). Participants also reported the gender composition of their workplace, the prevalence of teamwork, and several questions about how they experienced the experiment. They also reported their gender and gender identity (Brenøe et al., 2022). After the questionnaire, participants received feedback on their experimental earnings and had the option to learn the correct quiz answers. A general outline of this version of the experiment is displayed in Panel 1.1b of Figure 1.1.

**Manager Version**

The Manager version of the study also consisted of two parts. Part 1 was identical to the Employees' version. Managers answered either the general knowledge quiz on "Science & Technology" or "Psychology & Linguistics". The topic of the quiz was randomly assigned. Managers received £0.10 per correct answer. Thereafter, they guessed how many of their answers were correct and received £0.25 for a correct guess. Managers took the quiz to experience the knowledge task themselves before estimating their matched Employee's performance.

In Part 2, the Manager's main task was to *estimate* how well a matched Employee performed on the general knowledge quiz *without* advice. To make this estimate, the Manager saw an Employee's profile as explained in Section and shown in Figure 1.1 Panel 1.1b. The Manager received a bonus of £3.00 for a correct estimate. They also knew how this estimate would influence the experimental earnings of their matched Employee. These incentives ensured that the Manager attempted to estimate the Employee's initial quiz score correctly and that the Manager's estimate mattered for the Employee.[12]

Managers were randomly assigned to one of 12 profiles that differed in the sex of the matched Employee (2), the quiz topic (2), whether advice was visible (2), and when it was visible, the profile either showed that advice was sought or that it was not sought (x2 |visible). Each Manager saw only a single profile and reported a single estimate. This way, Managers were not aware that many filler characteristics were held constants in all profiles.

After the incentivized estimation task, Managers proceeded to the questionnaire. First, they were asked to briefly describe how they arrived at their estimate in an open-form field. Then, they stated their beliefs about the Employee's advice-seeking strategy. Specifically, they reported a threshold defined as the number of answers that someone with that profile must *not* know to decide to seek advice on the quiz. Lower numbers indicate a believed higher willingness to seek. The rest of the questionnaire was similar to the one presented to Employees. It included questions on demographics, attitudes towards advice-seeking, gender stereotypes about competence for several categories of knowledge, their views on the role of luck versus knowledge in the quiz, and their beliefs about the likelihood of cheating in the Employee version of the study.

[12] While social preferences of Managers might affect their estimate, it does not pose a challenge for our identification. These parameters are kept constant in all treatments and, therefore, can not explain treatment differences.

**Procedures**

We run our experiment with a sample of white-collar professionals. This population might be more likely to regularly face a trade-off between the information benefit of advice and perceptions of competence due to the measurability issue typical for knowledge-intensive jobs. In this type of jobs, remuneration and career progression more strongly depend on subjective evaluations from superiors and peers (e.g., Lev, 2000; Gibbs et al., 2004; Benson, Li, & Shue, 2024).

The experiment was conducted online in May 2022. Through Prolific Academic, we recruited participants who: (1) were residents of the UK or Ireland[13], (2) were between 25 and 60 years old, (3) had at least a Bachelor's degree, and (4) had an approval rate of over 95% on Prolific. As Managers, we recruited participants who, in addition to these criteria, reported having experience in management positions.[14] Each participant completed the study only once in the role of either Employee or Manager. Table A.2 shows the number of participants per experimental condition in both versions of the study.

The Employee version was implemented in oTree (D. L. Chen, Schonger, & Wickens, 2016) and the Manager version in Qualtrics. Employees took, on average, 17 minutes to complete the study, Managers 12 minutes, and they, on average, earned £4.01 and £2.14, respectively. Final earnings included a participation fee of £1.50 for Employees and £1.00 for Managers. Employees and Managers were randomly matched ex-post to calculate their payoffs. As the focus of the study is Employees' behavior, we recruited more Employees (1,800) than Managers (721). Managers were informed that, with some probability, they would be matched with several Employees with identical profiles. In this case, their estimate counted for all these Employees, and their estimation bonus was calculated based on one randomly selected match. We implemented a random matching procedure such that 20% of Managers were matched with a single Employee and the rest with several Employees with identical profiles.

Mandatory comprehension questions throughout the study ensured attention to and comprehension of experimental instructions.

[13]The experiment uses general knowledge questions as the main task. What is considered "general knowledge" is, however, specific to a certain cultural and geographical space. To obtain a degree of control over the knowledge space when constructing the knowledge task, we limit recruitment to participants from the UK or Ireland. See Online Appendix Section A.2.2 for further details on the calibration of the knowledge task.

[14]The question to determine relevant participants read as follows: "Do you have any experience being in a management position?".

### 1.2.2 Advice-Seeking at Work Survey: Design and Procedure

We surveyed 500 working professionals about knowledge-seeking at work and their workplace characteristics. The survey shed light on the demand side of advice and served two research objectives: First, to provide empirical support to the anecdotal evidence that people underutilize advice. Second, to offer a perspective on how important different barriers to seeking are, therefore putting experimental results into perspective.

In the survey, we ask white-collar professionals about their behavior and motives in a structured way. As for our experimental study, we recruited these professionals through Prolific and targeted the same population in terms of education, age, and reliability as participants. We recruited 400 U.S. residents and 100 residents of the UK or Ireland. The survey was conducted in December 2023. It took an average of 5 minutes and 47 seconds to complete, and participants were paid £1.00 (around $1.27[15]) if they passed at least one of two attention checks.

The following primary outcomes were collected. First, we asked about typical sources of information at work, asking the respondent to allocate 100 percent among six potential sources of information to indicate how often each one is used. Second, we asked whether the respondent had ever chosen to solve a work-related challenge independently, even though the respondent knew that asking someone for advice would have been quicker. If this question was answered positively, we also asked about the typical reasons for not seeking advice even though this could yield a quick solution, prompting the respondent to select all that apply out of seven. Third, we asked why professionals seek advice on a work-related challenge, ranking six reasons from most to least important. Lastly, we asked participants to reflect on whether they are satisfied with how much advice they seek at work and to indicate their satisfaction on a three-point scale. All answer options to these questions were presented in random order (randomization at the respondent level) to mitigate any concerns of order effects.

In addition, we asked demographic questions related to current or most recent employment (sector, industry, duration of employment, sizes of organization as a whole, and own organizational unit). We also asked about specific job characteristics (managerial responsibility, prevalence of teamwork) and perceptions of psychological safety at work. Finally, we asked the same set of questions on attitudes about advice-seeking and social image concerns that we also elicited in the experimental study.

[15]Exchange Rate £ to $ retrieved on Aug 10, 2024 here:https://g.co/kgs/wdq3mnf.

## 1.3 Conceptual Framework and Hypotheses

The Employee's decision to seek advice: $s \in \{0,1\}$ trades off the benefit of advice against its cost. The Employee has knowledge on a topic $t$ and knows $a_i \in \{0,1,2,\ldots,8,9,10\}$ answers. The Employee (she) observes her knowledge $a_i$, but the Manager (he) does not. Her quiz performance $p(a_i,s) \in \{0,1,2,\ldots,8,9,10\}$ is her knowledge when she does not seek $p(a_i,0) = a_i$, and it can increase with advice, $a_i \leq p(a_i,1) \leq 10 \; \forall \; a_i$. Advice costs a one-time fee of $c$.

Further, the Employee may experience a non-monetary cost to seeking advice, $\gamma_i$, for example, because she feels bad if she cannot accomplish this task independently. The Employee receives performance pay $b$ for correct answers. In addition, she receives a bonus $r\hat{a}(g,t,s)$ from her Manager, which increases linearly in the Manager's *estimate* of her knowledge $a_i$. This estimate is denoted $\hat{a}(g,t,s) \in \{0,1,\ldots,9,10\}$. The variable $g$ stands for the Employee's characteristics, her sex, education, etc., that the Manager observes. By design, Employees' characteristics only differ in their sex. The Manager's estimate also depends on the quiz topic $t$ that — together with the Employee's observable sex — induces beliefs about the Employee's knowledge. The Manager can interpret the Employee's behavior when he sees it. Specifically, he can condition his estimate of the Employee's knowledge on her decision to ask for advice $s$. When she decides whether to seek advice, the Employee does not know what the Manager's estimate of her knowledge will be. But she has a belief $\psi(\hat{a}) \in \{0,1,\ldots,9,10\}$ about it. In *Private*, the Employee's utility from seeking is $u(a_i,s=1,\psi) = bp(a_i,1) + r\psi(\hat{a}(g,t) - c - \gamma_i$, which she compares to her utility from not-seeking $u(a_i,s=0,\psi) = ba_i + r\psi(\hat{a}(g,t))$. Thus, the Employee will seek whenever the information value of advice exceeds the advice fee and any non-monetary cost to seeking advice, weighted by the piece rate for correct answers:

$$p(a_i,1) - a_i \geq \frac{(c+\gamma_i)}{b} \tag{1.1}$$

The piece rate for correct answers can be interpreted as the opportunity cost of renouncing on advice.

In *Visible*, the Employee's choice is more involved since her Manager observes her decision to seek advice. The Employee now also considers how her advice choice $s$ may influence the Manager's estimate of her knowledge. Her utility from seeking is $u(a_i,s=1,\psi) = b(p(a_i,1) + r\psi(\hat{a}(g,t,s=1)) - c - \gamma_i$, which she compares to her utility from not-seeking $u(a_i,s=0,\psi) = ba_i + r\psi(\hat{a}(g,t,s=0))$. In $Visible$, the

Employee seeks whenever the information value of advice exceeds the advice fee, any non-monetary cost to seeking advice and her expected reputational cost, weighted by the piece-rate for correct answers:

$$p(a_i, 1) - a_i \geq \frac{c + \gamma_i}{b} + \frac{r}{b}(\psi(\hat{a}(g, t, s = 0)) - \psi(\hat{a}(g, t, s = 1))) \qquad (1.2)$$

This expected reputational cost $r(\psi(\hat{a}(g, t, s = 0)) - \psi(\hat{a}(g, t, s = 1)))$ is the Employee's belief about how her decision to seek advice will change the Manager's estimate of her knowledge and, through that, her bonus.
The two thresholds for seeking advice, (3.1) and (1.2), differ by this expected reputational cost, weighted by the piece rate for correct answers. Whenever the Employee believes that the Manager would interpret her decision to seek advice negatively, that is, $(\hat{a}(g, t, s = 0)) > \psi(\hat{a}(g, t, s = 1))$, this expected reputational cost is positive. A higher threshold for advice-seeking in $Visible$ implies that an Employee seeking in $Private$ may not seek when randomly assigned to $Visible$.

Comparing the rate at which advice is sought in $Private$ to $Visible$ estimates of the causal effect of reputational concerns on the willingness to seek advice, given the random treatment assignment. We pre-registered the following hypotheses:

**Hypothesis 1.** *The rate at which advice is sought is lower in $Visible$ compared to $Private$.*
Moreover, we test whether reputational concerns change with others' stereotypical perceptions of competence. By design, female and male participants in our study face different perceptions of competence for a given quiz topic. Therefore, we pre-registered that we would test Hypothesis 2 separately for women and men:

**Hypothesis 2.** *The change in the rate at which advice is sought in $Visible$, when compared to $Private$, differs when stereotypes about competence assign participants an advantage in knowledge compared to a disadvantage.*

Others' beliefs about competence may amplify or mute the expected reputational cost of seeking advice or not affect it at all. Theoretically, the direction of the effect is ambiguous, and therefore, it is an empirical question. Our study seeks to uncover empirically—with actual choices and a direct measure of beliefs—whether others' stereotypical beliefs about competence change the expected reputational cost of seeking advice, on average. This is an important question to ask and answer: if professionals facing high or low beliefs about their competence perceived, all else equal, the reputational consequences of seeking advice at work systematically differently, ramifications for learning and productivity could prove substantial.

## 1.4 Results

### 1.4.1 Online Experiment

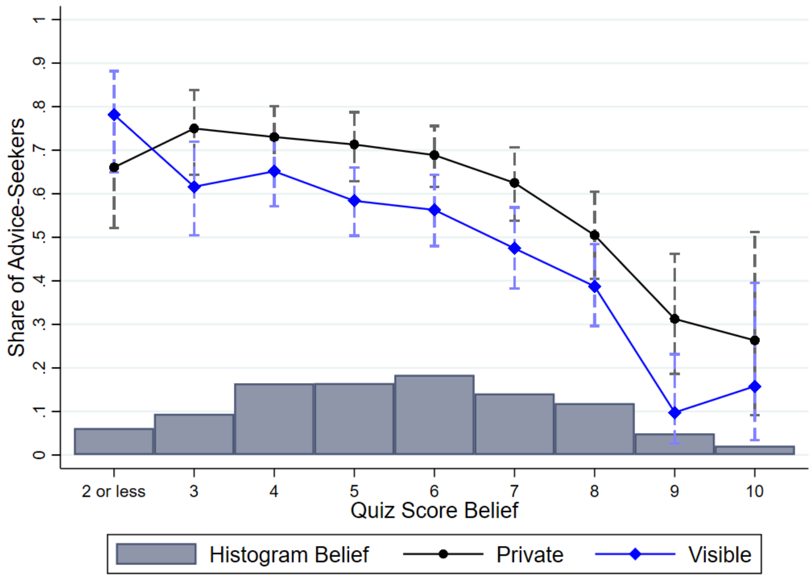
**How does visibility affect advice-seeking?**

To begin, we report results from analyzing pooled data (both quiz topics, women and men). The rate at which advice is sought is 64% when this decision is *Private*. Revealing it to the Manager leads to a decrease of 10 ppts [6.1,15.1] ppts[16] to a rate of about 54% ($p < 0.0001$, two-sided test of proportions), corresponding to a 16% reduction. This is consistent with the interpretation that participants trade off the information benefit of advice against a reputational cost when this choice is visible.
The rate of advice-seeking decreases nearly monotonically with participants' subjective quiz performance in both *Private* and *Visible* (see Figure 1.2). This strongly suggests that the information benefit of advice plays a role in the decision to seek. The gap between *Private* and *Visible* is remarkably constant for all subjective performance levels except the very lowest, for which we cannot reject that the share of advice-seekers is the same in *Visible* and *Private*.
Estimates of a linear probability model in which we control for beliefs about own performance, quiz topic, and Employee's sex confirm that visibility causes a decrease in the propensity to seek (see column 1 of Table 1.1). Conditional on these controls, the propensity to seek advice decreases by 11 ppts, on average, when it is visible to the Manager (p<0.001) with a 95%-CI of [-15.5, -6.8] ppts. Estimating the model separately for women and men, we observe, respectively, a 13 ppts and 10 ppts drop in the propensity to seek advice when it is visible ($p < 0.01$) with 95%-CI of [-19.0,-6.7] ppts and [-16.3,-3.9] ppts, respectively (see columns 2 and 3 of Table 1.1). This analysis leads to the first main result of the study:

**Result 1:** Visibility causes a large decline in the propensity to seek advice.

[16] 95%-confidence interval (CI).

**Figure 1.2:** Share of Employees Who Sought Advice by Quiz Performance Belief and Visibility




**Notes.** The *Quiz Performance Beliefs* is the Employees' subjective performance belief elicited after completing the quiz in Part 1 and before visibility treatment by answering the question: "Guess, how many of your answers are correct?". Whiskers show 95%-CI. The bars show the frequency of these reported beliefs in the sample (pooled across visibility conditions).

From the perspective of individual performance, this large reduction in the willingness to seek is inefficient since the expected information benefit of advice is positive for most participants. Whenever subjective (actual) quiz performance is lower than or equal to eight correct answers, an Employee's expected increase (actual expected increase) in performance and pay exceeds the advice fee.[17] In *Private* (*Visible*), 92.6% (93.3%) of participants believe to have an initial quiz score lower than or equal to eight, and 80.95% (80.16%) have such a score. Figure 1.2 shows that the rate at which advice is sought in *Private* is well below 100% at every level of subjective quiz performance. In fact, it does not exceed 74%. The rate of advice-seeking is also well below 100% at every actual performance level (see Figure A.3 in the Appendix). In sum, we find that a substantial share of Employees refrain from seeking advice, even when it has a positive expected net monetary value based on what they believe or how they performed. Visibility exacerbates this inefficiency. The low level of advice-seeking, even in *Private*, points to the fact that factors other than the expected net

[17]After advice, a participant who has no clue about the correct answer has a 50% chance of guessing correctly, compared to a 20% chance before advice.

benefit of advice influence the propensity to seek. We explore such additional factors in Section 1.4.1.

Regarding the realized gain from advice-seeking, 84% of seekers improved their quiz score after advice. On average, the pay increased by £0.10 net the advice fee. This corresponds to an average pay increase of 22% relative to what seekers would have received for their independent quiz score from Part 1.

**Table 1.1:** Linear Probability Models Predicting Employees' Propensity to Seek Advice

| DV: *Advice* (1/0) | (1) Pooled | (2) Female Employee | (3) Male Employee | (4) Female Employee | (5) Male Employee | (6) Pooled |
|---|---|---|---|---|---|---|
| *Visible* (1/0) | -.112*** | -.128*** | -.101*** | -.169*** | -.083* | -.082*** |
| | (.022) | (.031) | (.032) | (.044) | (.045) | (.032) |
| *Science & Tech* (1/0) | -.030 | -.021 | -.032 | -.061 | -.014 | |
| | (.023) | (.031) | (.034) | (.043) | (.045) | |
| *Visible x Science & Tech* | | | | .08 | -.036 | |
| | | | | [-.043, .203] | [-.160, .088] | |
| | | | | (.063) | (.063) | |
| *Male* (1/0) | -.035 | | | | | -.034 |
| | (.023) | | | | | (.023) |
| *Advantageous Competence Stereotype* (1/0) | | | | | | .023 |
| | | | | | | (.031) |
| *Visible x Advantageous Competence Stereotype* | | | | | | -.060 |
| | | | | | | (.045) |
| *Private* mean *Advice* | .643 | .700 | .588 | .700 | .588 | .643 |
| Subj. performance-level-dummies | yes | yes | yes | yes | yes | yes |
| Adjusted $R^2$ | .078 | .051 | .094 | .052 | .093 | .078 |
| # of Employees | 1800 | 900 | 900 | 900 | 900 | 1800 |

**Notes.** The dependent variable in all specifications is *Advice* that equals 1 if the Employee sought advice and 0 otherwise. *Visible* indicates that the Employee was in the treatment condition in which the advice-decision was revealed to the Manager. *Science & Tech* indicates that the Employee took the Science & Technology quiz. Male indicates that the Employee's sex is male. *Advantageous Competence Stereotype* indicates whether a stereotype about competence and an Employee's sex are congruent. For women, it takes the value of 1 in the "Psychology & Linguistics" quiz and for men in the "Science & Technology" quiz. *Private mean Advice* is the mean of *advice* for the (sub-) sample of Employees as described in the column header. *Subj. performance-level-dummies* bin Employees' incentivized beliefs about their independent quiz score in Part 1 into five levels: 0-2, 3-4, 5-6, 7-8, and 9-10, with 5-6 as the omitted category. Robust standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

### Do others' stereotypical beliefs about one's competence change how professionals trade off the information benefit of advice against reputational concerns?

Our second treatment dimension randomly varied the topic of the knowledge task. This way, we successfully manipulated participants' beliefs about the Manager's stereotypical beliefs about their competence. We confirm this with two measures of higher-order beliefs, one incentivized elicited between-subjects and one unincentivized elicited within-subject. These manipulation checks strongly suggest that the average woman who took the “Psychology & Linguistics” quiz believed that her matched Manager would believe that she was competent on the topic. In contrast, the average woman who took the “Science & Technology” quiz believed that her Manager had a relatively low perception of her competence. The reverse is true for men. Section A.1.2 of the Appendix gives a detailed account of these manipulation checks.

Because women and men held, by design, different higher-order beliefs about their competence, we split the sample for an analysis of the interaction between the decision to visibly seek advice and the quiz topic (in line with our pre-registration). In a linear probability model, we add the interaction of the treatment indicators Visibility and Science & Technology and estimate the model separately for women and men (Table 1.1, columns 4 and 5). For women, this interaction term estimates whether the visibility gap in the propensity to seek advice increases or closes, on average, when the person who interprets this choice has a relatively low perception of her competence compared to a high one, holding constant subjective performance beliefs. For men, the interpretation is reversed.

Turning first to our sample of women, we estimate a positive coefficient on the interaction term, 8 ppts, with the 95%-CI [-4.32, 20.27] ppts that includes zero. For men, we estimate an interaction term of -4 ppts with the 95%-CI [-16.04, 8.82] ppts that also includes zero. These different signs of the estimated coefficients point to the same average behavior. Directionally, these estimates suggest that individuals may be more reluctant to visibly seek advice when it conflicts with an advantageous competence stereotype [18], compared to a disadvantageous one.

However, despite our substantial sample size of 900 women and 900 men, we cannot reject the null hypothesis that higher-order beliefs about competence do not mediate the visibility gap in the propensity to seek advice. Indeed, even when pooling the sample of women and men and estimating the interaction between visibility and an indicator for an advantageous competence stereotype, the 95%-CI around the estimated coefficient includes zero [-14.74, 2.75] ppts (see Table 1.1, column 6). This leads to the second main result of the study:

**Result 2:** We find little evidence that others’ stereotypical perceptions of competence

[18]The indicator takes the value of 1 for women who took the “Psychology & Linguistics"-quiz and 1 for men who took the “Science & Technology"-quiz.

change the way that professionals, on average, trade off the information benefit of advice against reputational concerns.

We consider the robustness of our two main results by replicating Table 1.1 with the actual quiz scores from Part 1 (see Table A.4 in the Appendix). The estimated coefficients and the statistical inference on them are very similar.

### Expected Reputational Cost of Advice-Seeking

By asking participants to report incentivized beliefs, we attempted to quantify their expected reputational cost of advice-seeking. In $Visible$, each Employee reported two beliefs about a Manager's quiz score estimate: one for another participant with a profile identical to theirs (which includes identical advice-seeking decision) and the other one for another participant with a profile identical to theirs except for the counterfactual advice-seeking decision. We interpret the difference between the belief reported for someone who did not seek advice and someone who sought advice, all else equal, as the *expected reputational cost* of seeking. Whenever it is positive, an Employee perceived that the Manager would interpret the signal "advice was sought" negatively and lower their quiz score estimate. Whenever it is negative, an Employee perceived that the Manager would interpret the signal "advice was sought" positively and increase their quiz score estimate.
According to this measure, the *expected reputational cost* to seeking advice differs substantially (see Figure 1.3 for a histogram), ranging from large expected benefit (-6) to large expected cost (+6). The average expected cost is close to zero (-0.08), and the median expected cost is at 0. Interestingly, about half of the participants reported beliefs consistent with an expected reputational *benefit* to advice-seeking in this setting. For women and men, the distributions of expected reputational cost do not differ systematically between the quiz topics (p of rank sum tests >0.64). This is in line with our Result 2.
Overall, the rate at which advice is sought when visible varies for different levels of expected reputational cost and benefit (see Figure A.4). After an initial and significant increase[19] in the rate of visibly seeking advice moving from an expected reputational cost of zero to one (i.e., a difference in one point in the Manager's performance estimate), the rate of seeking declines as the expected reputational cost increases. The

[19]Two-sided test of proportions p=0.017.

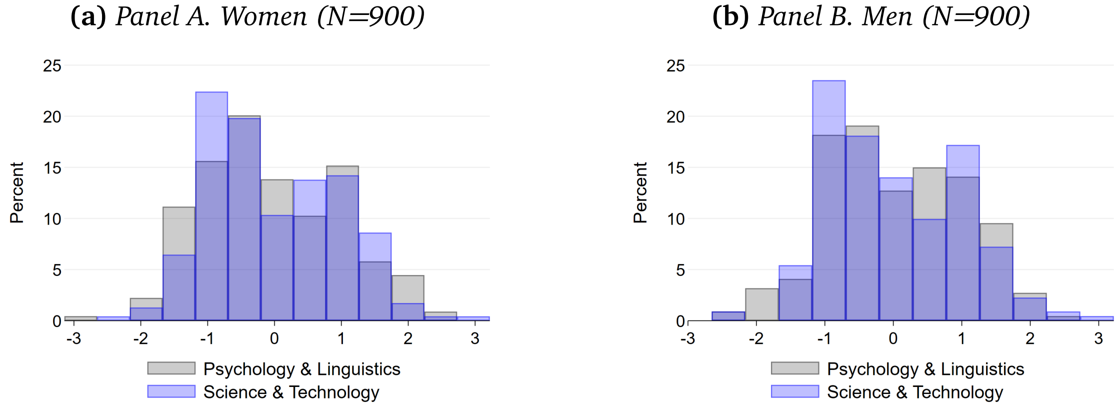
pattern for an expected reputational benefit is similar: the rate of visibly seeking advice nearly monotonically decreases in the size of the expected reputational benefit.
Given this inverted V-pattern, we correlate actual advice-seeking behavior with second-order beliefs, conditional on expecting a reputational benefit or a reputational cost when compared to no reputational consequence. We add our usual controls for subjective performance levels in a linear probability model that predicts the propensity to seek advice. The estimated correlation coefficients are 0.055[20] for an expected reputational benefit and -0.030 for an expected reputational cost. This analysis suggests that the slopes illustrated in Figure A.4 are (weakly) statistically significant ($p$-values$<0.08$). It suggests, counterintuitively, that a reported larger expected reputational benefit correlates with less advice-seeking. We elicited these second-order beliefs after measuring behavior.
Further, we elicited beliefs about two other Employees and not about self when compared to another Employee. This was done to limit the extent to which a self-other gap in judgment could influence the second-order belief measure (e.g., Möbius et al., 2022). Both factors can, however, contribute to the fact that these beliefs correlate weakly and inconsistently with observed advice-seeking behavior.
However, other direct evidence speaks to the reputational mechanism through which visibility lowers the rate at which advice is sought. To gain insights into how the professionals in our study reasoned about advice-seeking, we asked about their motives behind this choice in an open-form field. The free-from responses were classified by 3 raters blind to the research question into ten pre-defined categories (see Table A.6).[21]
Of the 10 motives, there are only two that are mentioned significantly more often in $Visible$ compared to $Private$, and both pertain to the Manager.

In $Visible$, 16% of those who did not seek state that this choice was driven by an expected reputational cost; they gave the negative impact of advice-seeking on the manager's quiz score estimate as a reason. In $Private$, this share is merely 2% of those who did not seek. Turning to those who did seek advice, 1% of them in $Visible$ explained that this was driven by an expected reputational benefit, compared to 0% of them in $Private$. While we observe a substantial share of participants in $Visible$ who expect a reputational benefit to seeking advice according to our second-order beliefs, measure, negligibly few of them state that it motivated them to visibly seek advice.

[20] The expected reputational benefit is a negative number, such that a positive correlation coefficient indicates that larger absolute values are associated with a lower propensity to seek.

[21] There is generally high to very high agreement in the classification of the free-form responses among the raters (see Krippendorff's alphas for each category in Table A.13).

**Figure 1.3:** Expected Reputational Cost to Seeking Advice by Topic and Gender




**Notes.** *Expected Reputational Cost* is the belief of the Manager's quiz score estimate for a *non*-seeker [0,10] minus the belief of the Manager's quiz score estimate for a seeker [0,10]. Positive numbers indicate an expected reputational cost and negative numbers an expected reputational benefit to advice with respect to a Manager's performance estimate. These incentivized beliefs were elicited in *Visible*.

### What explains further heterogeneity in advice-seeking?

Of all the 10 motives for the seeking decision, confidence in their own performance is mentioned most frequently: seekers stated that they did not feel confident in their performance and therefore sought advice (64% in *Private* and 68% in *Visible*) and non-seekers stated that they felt confident and therefore did not seek advice (44% in *Private* and 42% in *Visible*). The information benefit of advice (or perceived lack of it) was the second most frequent explanation for behavior (63% in *Private* and 60% in *Visible* among advice-seekers and 23% and 27% among non-seekers). Among those who did not seek, around 17% mentioned that they preferred solving the quiz on their own without external input, and this rate is the same in *Private* and *Visible*. Less than 1% of the responses suggest that the participant had not properly understood the incentives.

These insights into the professionals' motives for (not) seeking advice are corroborated with estimates of a linear probability model in which we correlate the decision to seek advice with items from our questionnaire controlling for subjective performance beliefs, separately for *Visible* and *Private*. Given the exploratory nature of this analysis and the multitude of correlations we are testing, we use the 0.5% level as the threshold for statistical significance (Benjamin et al., 2018). These results (see Table A.5) indicate that lower perceived usefulness of advice, as well as a generally negative attitude towards not accomplishing tasks independently, significantly correlate with lower seeking in both $Private$ and $Visible$. We find no evidence that factors such as, for example, attitude toward risk, the belief that reputation matters for career advancement, or social image concerns in general correlate with advice-seeking in our study. [22]

Contrary to the stereotype that men, on average, have a lower propensity to seek advice than women, we find little evidence that women and men differ in their propensity to seek when we condition on performance beliefs (see Table 1.1, column 1). The estimated coefficient of the male indicator is -3.5 ppts with 95%-CI [-8.0, 1.1]. If we, instead, control for actual quiz performance, the estimated coefficient of the male indicator is -7 ppts with 95-% CI [-11.6,-2.5] ppts (Appendix Table A.4, column 1).

Once we condition on actual rather than subjective performance levels, we can additionally consider whether self-stereotyping might play a role in advice-seeking (Coffman, 2014). Since self-stereotyping has been found to operate largely through confidence, we control for it in our main specifications. The estimates presented in columns 3 and 4 of Table A.4 suggest that self-stereotyping may influence men's propensity to seek advice but not that of women. For men, the average propensity to seek advice on the "Psychology & Linguistics" quiz is significantly higher compared to the "Science & Technology" quiz, conditional on actual performance levels and an indicator for the Visibility treatment.

**How do Managers interpret the decision to seek advice?**

Turning to the Managers' side, we present results on how they interpret the decision to seek advice. Each Manager evaluated a single profile that was randomly assigned. We analyze how characteristics conveyed by the profiles (gender, quiz topic[23] and—in $Visible$—the decision to seek advice) affected the Manager's estimates of the Employee's competence. While the profiles included more information (e.g., age, education, and country of residence), only these characteristics varied experimentally. In linear regressions, we can estimate the average causal effect of a specific profile characteristic by comparing estimates for profiles that differ in this characteristic, condi-

[22] Since we do not find systematic differences for women and men, we only report pooled specifications in Table A.5.

[23] A manipulation check confirms that the quiz topics induced stereotypical perception of competence in this sample of professionals with reported managerial experience, as measured with a slider ranging from -1 "women know more, on average" to 1 "men know more, on average". The average slider position is -0.17 for "Psychology & Linguistics" and 0.25 for "Science & Technology". Both averages are significantly different from 0 "no gender difference" (t-test p<0.001).

tional on the other randomly varying characteristics. We include the Manager's own subjective quiz performance as a control variable.

**Table 1.2:** OLS Regressions Predicting Managers' Quiz Score Estimate (std.)

| DV: *Manager's Estimate (std.)* | (1) Visible Pooled | (2) Visible Female Employee | (3) Visible Male Employee | (4) Private & Visible Female Employee | (5) Private & Visible Male Employee |
|---|---|---|---|---|---|
| *Advice sought* (1/0) | -.060 | -.103 | -.052 | | |
| | [-.222, .101] | [-.331, .126] | [-.282, .178] | | |
| | (.082) | (.116) | (.117) | | |
| *Visible* (1/0) | | | | -.194** | .036 |
| | | | | (.090) | (.091) |
| *Science & Tech* (1/0) | .323*** | .115 | .552*** | .064 | .426*** |
| | (.083) | (.118) | (.117) | (.091) | (.092) |
| *Female Employee* (1/0) | -.131 | | | | |
| | (.083) | | | | |
| *Own subj. quiz performance* (#) | .237*** | .261*** | .214*** | .249*** | .200*** |
| | (.023) | (.032) | (.030) | (.026) | (.023) |
| Mean *Estimate* | 5.554 | 5.325 | 5.779 | 5.661 | 5.750 |
| Adjusted $R^2$ | .254 | .262 | .253 | .258 | .231 |
| # of Managers | 480 | 241 | 239 | 362 | 359 |

**Notes.**. The dependent variable in all specifications is the Manager's *Estimate* of a matched Employee's quiz score. This variable is standardized to have a mean of zero and a standard deviation of 1 in the sample specified in the column header. *Advice sought* indicates that the matched Employee sought advice on the quiz. *Visible* indicates that Managers observed the matched Employee's advice-seeking decision. *Science & Tech* indicates that Manager and the matched Employee took the Science & Technology quiz. *Female Employee* indicates that the matched Employee is a woman. *Own subj. quiz performance* is the Manager's subjective belief of their own quiz performance and ranges from 0 to 10. *Mean Estimate* is the overall mean of the Managers' estimate for the sample specified in the column header. Results presented in Columns (1)-(3) are restricted to Managers who were randomly assigned to $Visible$, while columns (4) and (5) include all Managers who were matched with female and male Employees, respectively. 95% confidence intervals in squared brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

First, we consider the sample of professionals with reported managerial experience who saw their matched Employee's advice-seeking decision (Managers in $Visible$). We estimate that the decision to seek advice lowers the Manager's quiz score estimate by about -0.060 $\sigma$ (6 percent of a standard deviation), on average, with a 95%-CI [-0.222 $\sigma$, 0.1011 $\sigma$] that includes zero (N=480) (see column 1 of Table 1.2). The overall takeaway is the same if we split the sample by Employee sex, considering separately managers in $Visible$ who evaluated women and those who evaluated men

(columns 2 and 3 of Table 1.2). The estimated coefficients of the indicator "Advice" are small and negative for women (-0.103 $\sigma$) and men (-0.052 $\sigma$) with a 95%-CI of [-0.331 $\sigma$, 0.126 $\sigma$] and [-0.282 $\sigma$,0.178 $\sigma$], respectively. Though insignificant, the estimated effect size is double for women than for men. Yet, we cannot reject that the coefficients are the same when estimating the interaction of "advice sought" and "female employee" in the pooled sample of Managers who observed the advice decision(estimated coefficient -0.054, with a 95%-CI [-0.379,0.270] and an associated p=0.74, N=480) Further, we find no evidence that the manager's sex impacts their estimate in any way (see Online Appendix Table A.15).
We collected additional data to better understand how Managers reason about their quiz score estimate. This also confirms that they attended to the information provided in the profiles. In the final questionnaire, Managers were prompted to briefly explain in their own words how they arrived at their estimate.[24] Their free-form answers were coded independently by 3 raters who were blind to the research question (Table A.7).[25] These results give confidence that a large majority of Managers attended to and used the profile information when providing their estimates.

Overall, the Managers mentioned the profile information on education (41.3%), quiz topic (25.2%), and age (16.9%) most often. Further, 24.8% of responses show that the Managers were aware of the Employee's sex, although only 3.8% mention it as a reason for providing a certain estimate. In addition to profile information, 34% of Managers reported that they had compared the Employee's profile to themselves when forming a belief about the Employee's quiz score.

The reasons are largely the same in $Private$ and $Visible$ with the notable exception of the mention of advice. Specifically, 18.5% of Managers in $Visible$ explicitly mention the Employee's advice-seeking behavior when explaining how they arrived at their performance estimate, compared to 2.4% in $Private$. Overall, about 18.3% of Managers state that they have guessed their estimate. Significantly more Managers report having guessed the quiz score of their Employees in $Private$ (24%) than in $Visible$ (15%). This may suggest that observing the choice to seek advice helps the Managers form a belief about the Employee's competence. However, looking at actual correct estimates, there is no indication that Managers in $Visible$ are better at estimating the quiz score:12.9% of Managers correctly estimated their matched Employee's quiz

[24] The wording of the question was: "We would like to understand how you arrived at your estimate of the quiz-taker's quiz performance without advice. Please briefly describe your thought process."

[25] There is a high agreement in the classification of the free-form responses among the raters (see Krippendorff's alphas for each category in Online Appendix Table A.19).

score in $Private$ and 13.3% in $Visible$ (test of proportions, p=0.86).

Why does the decision to seek advice not affect the score estimates, on average? If Managers believed that all or most employees sought advice, regardless of their competence, then their score estimates should not change when observing the decision to seek advice. Direct evidence on Managers' beliefs about their matched Employee's advice-seeking strategy that we elicited in the questionnaire speaks against this interpretation.[26] The average Manager believes that someone with the profile they evaluated would decide to seek advice if at least 4.5 answers were unknown to them. This threshold is virtually the same in $Private$ (4.5) and $Visible$ (4.5) and, overall, for female (4.6) and male (4.5) Employees. Importantly, the average believed threshold is well above zero or one unknown answer. This zero or one unknown answer threshold would be consistent with the belief that all participants in the role of Employee would always seek advice.

### 1.4.2 Advice-Seeking at Work Survey

**Sample Description** We recruited on employment status, and 99 percent of respondents have a main job outside of taking surveys, and 85 percent report being employed full-time. Appendix Table A.8 reports the job characteristics of the surveyed professionals separately for the UK and the U.S. subsamples.

Since our main experimental study was conducted in the UK, we recruited some professionals residing there. Generally, these two subsamples do not differ greatly in their job characteristics, such as length at current employer, size of organization or organizational unit, and the prevalence of teamwork at their current job. More UK respondents work in the public sector (35 percent vs. 29 percent in the U.S.), and fewer report being self-employed (12 percent vs. 16 percent).

We also report employment characteristics for a reference sample in the U.S. These are U.S. residents who participated in the American Community Survey (ACS) in 2022, a large and nationally representative survey. The ACS data were filtered to match the demographic and socioeconomic recruitment filters from Prolific (age 25-60, educational level of at least a Bachelor's degree, currently employed, N=509,530).

There are some differences between our U.S. sample of professionals recruited through

[26]The question's wording was "In your opinion, how many answers must a quiz-taker with this profile not know to decide to seek advice on the "Science & Technology"-/"Psychology & Linguistics"- quiz?". The topic of the quiz varied, depending on which one the respondent was randomly assigned.

Prolific and the broader ACS sample. For example, the share of public sector workers and the share of self-employed are higher in the Prolific sample (public sector: 29 percent vs. 23 percent, self-employed: 16 percent vs. 9 percent). Aside from these, there are no stark differences in important job characteristics such as sector and industry of employment, considering that only 400 U.S. residents answered the detailed industry question (with 21 answer options) in our survey. These comparisons indicate that our sample of white-collar professionals recruited through Prolific generally, though not perfectly, resembles the broader U.S. population of white-collar professionals.

**Demand for Advice** Colleagues and superiors are a typical source of information for professionals when they require help on work-related challenges (see Panel a) of Appendix Figure A.5). On average, colleagues are asked 29 percent of the time, and superiors are asked 17 percent when respondents allocate 100 percent among six types of sources. By contrast, online resources are recurred to only 22 percent of the time, on average. Given the prevalence of interpersonal information sourcing at work, it is important to understand what can cause friction on the demand side, such as reputational concerns.

Further, we directly asked professionals whether they have ever refrained from seeking advice from others, even though it would have provided a faster solution to a work-related challenge. Most respondents (84 percent) answered the question affirmatively. We asked those participants to reflect on typical motives for this type of behavior, selecting all that apply from a list of seven. This list was exhaustive since only 5 percent of respondents wrote down an additional motive that was not listed. A desire for self-reliance is the most important motive for not seeking advice at work—67 percent of respondents selected it as a typical reason—followed by skill development and learning, which 56 percent selected (see Panel b) of Appendix Figure A.5). Every third professional (33 percent) reported that a fear of judgment was a typical motive for not seeking advice at work.

Turning to the supply side, only 10 percent reported that a fear of rejection typically hinders them from seeking advice at work. This contrast indicates a potential mismatch between the demand and supply of advice at work. Our question about general satisfaction with advice-seeking at work aims to reveal subjective inefficiencies in professionals' advice-seeking at work. About 25 percent of the survey participants are not content with their current level of advice-seeking at work, and most of them—23 percent in total—report that they could seek more advice. Asked about why they seek advice at work, most respondents listed access to knowledge or experience as the most important reason (28 percent). As the second-most important reason for seeking advice at work, 20 percent of participants gave “learning and development” and 19 percent of participants gave “seeking feedback and reassurance”. Motives like “showing engagement” or “building relationships” were less frequently selected among the most important reasons for seeking advice at work.

Overall, our survey evidence shows that advice from colleagues and superiors remains an important and valued source of information and knowledge at work. However, 33 percent of respondents who report foregoing at times the information value of advice at work report that a fear of judgment is a typical reason for this behavior. Psychological safety has been suggested to play a key role in group dynamics and team productivity (Edmondson & Lei, 2014; Castro, Englmaier, & Guadalupe, 2022), and we measured perceived psychological safety at current employers on a 7-point scale, where higher numbers mean higher levels of psychological safety. With an average of 5.25, these surveyed professionals report a relatively high level of psychological safety on average. Those selecting a ‘fear of judgment’ as a reason for not seeking advice reported significantly lower perceptions of psychological safety (t-test $p<0.001$) — by about 48 percent of a standard deviation—than those who selected other reasons for not seeking advice.

## 1.5 Discussion and Conclusion

In this study, we present evidence from an artefactual field experiment investigating whether white-collar professionals seek advice strategically and whether their strategies depend on others’ stereotypical beliefs about their competence. While asking for advice is an important way to access knowledge and improve performance and decision-making, professionals may be reluctant to seek advice if they believe it signals incompetence. Our research aims at uncovering this potentially important barrier to knowledge flows in organizations.

Our experimental design simplifies a complex interaction between advice-seeker and -giver to causally study the role of reputational concerns. We experimentally vary whether the decision to seek advice is visible to a Manager who estimates the Employee’s independent performance in a knowledge quiz, our measure of the Employee’s task competence. The Employee’s pay increases in her quiz performance— which can improve with advice—and in the Manager’s belief about her task competence. The randomly assigned topics of the knowledge task varied (higher-order) beliefs about competence, relying on gendered stereotypes about domains of knowledge.

In this decision environment, we document strategic advice-seeking: the rate of seeking advice on a knowledge task decreases by about 16% when it is visible to a Manager compared to when it is $Private$, despite its potential to increase task performance and earnings. Moreover, we find no evidence that Managers interpret the decision to seek advice negatively when estimating task competence, on average. However, professionals who play the role of Employees hold heterogeneous beliefs about how the decision to seek would affect the Manager's perception of their competence, ranging from a large negative to a large positive effect. Taken together, these results point to a widespread misperception about the reputational consequences of advice-seeking among professionals.

We find little evidence that others' stereotypical beliefs about competence cause a change in the visibility gap of advice-seeking, on average. This finding is noteworthy, considering that the literature has extensively documented stereotypical perceptions of competence to matter for important labor market outcomes. At first glance, this finding suggests that professionals may, on average, not cater to the signals about competence they send to others' stereotypical beliefs about their competence. An alternative interpretation is that, on the one hand, the desire to disconfirm a disadvantageous stereotype and, on the other hand, the desire to confirm an advantageous one are equally strong, on average. Both channels would imply that the willingness to seek is lower compared to a situation when others have no preconception of competence, which would be an interesting extension of our research.

The results further suggest that seeking may be subject to other internalized barriers since, even in $Private$, a considerable share of professionals does not tap into the increased earnings potential from advice. The desire to perform tasks independently is, according to this study's findings, one such barrier. Defying conventional wisdom, our results also show that in this study's setting, internalized barriers to seeking advice are not systematically different for women and men once we control for confidence.

The experimental decision environment we designed to answer our research questions deliberately abstracted from factors that could additionally affect the willingness to seek advice at work, some of which have been studied in related work. Objective difficulty of the task on which advice is sought is likely to also play a role in the expected reputational consequences of seeking advice. Looking at other costs, the seeker can experience psychological costs in the form of stigma and shame (Chandrasekhar, Golub, & Yang, 2018) or simply fear that her request for help may be rejected (Bénabou, Jaroszewicz, & Loewenstein, 2022). The opportunity cost of the advice-givers' time can be non-negligible (Espinosa & Stanton, 2022), something that the seeker may also consider when deciding whether to ask.

Regarding the benefits of advice-seeking at work, the wish to build a relationship with the advice-giver or to show self-assuredness and self-awareness are likely motivators of the willingness to seek. Our novel experimental design can be flexibly adapted to isolate some of these or other potential determinants of the willingness to seek or study the compound effect of several ones.

In our stylized environment, the evaluating Manager does not see the Employee's actual task performance. In practice, however, Managers can observe work outcomes—which they care about—in addition to competence. This makes a decision-maker's trade-off between the information benefit of advice and an expected reputational cost even more complex. The benefit of seeking advice is larger when a work outcome is visible. Yet, the Manager will have to *judge* how to attribute the outcome to inherent competence versus external input. This way, advice-seeking can lead to ambiguity regarding the source of performance outcomes. Previous work has shown that in the face of ambiguous performance outcomes, people may use simple heuristics to attribute credit, such as teamwork based on seniority or gender (Jin et al., 2019; Sarsons et al., 2021).

Just as managers typically have more information to base their evaluation on, professionals outside of our stylized decision environment can resort to sourcing knowledge without seeking advice, for example, online. We believe that our results also bear relevance for such environments. In these cases, reputational concerns linked to advice-seeking could manifest as preferring other sources of knowledge over asking others for advice. With multiple sources of knowledge, not seeking advice does not necessarily compromise a final work result but can make the process longer. On the one hand, searching independently may be inefficient while, on the other hand, excessive asking can also lower the productivity of others (Espinosa & Stanton, 2022). Whom, when, and about what to ask for advice is a skill in and of itself in the knowledge economy, with a fast-moving knowledge frontier.

This study uncovers a potentially important barrier to efficient knowledge flows: reputational concerns and widespread misperceptions concerning the reputational consequences of seeking advice among professionals. Correcting such misperceptions could foster knowledge exchange and learning from others at work.

# 2

# Breaking the Ice: Can Initial Peer Activity Enhance Platform Engagement and Persistence?

## 2.1 Introduction

Digital advancements have revolutionized the exchange of knowledge. Ubiquitous knowledge exchange platforms facilitate information dissemination beyond traditional geographical, social, or temporal boundaries (Faraj et al., 2016; Rietveld & Schilling, 2021). In 2023, knowledge exchange platform software reached a market valuation of $ 17.43 billion (Straint Research, 2024).[1] Within this large market, organizations increasingly rely on digital knowledge management tools to facilitate efficient information flow among members. Platforms like "Slack," "Microsoft SharePoint," or "Starmind" exemplify this trend. This type of information exchange is critical in the knowledge economy, where complex tasks prevail (Autor, Levy, & Murnane, 2003).

The value of any knowledge exchange platform for stakeholders (e.g., firms) and users depends heavily on the contributions made by other users, both in terms of quality and quantity. Contributors, thus, not only source and contribute knowledge in direct interactions but also create positive externalities for their peers on the platform. Not surprisingly, underinvestment is common: Too few potential users contribute, and many users could contribute more. In particular, while users may be willing to contribute initially, engagement tends to drop over time, making declining user engagement and response rates a common challenge across knowledge exchange platforms (Ren et al., 2012; Gallus & Frey, 2016; J. Baek & Shore, 2020; Mickeler et al., 2023). Thus, it remains an ongoing effort to understand barriers to and facilitators of efficient online knowledge exchange and how digital interactions can be designed and governed to create value.

We argue that the initial phase of the user platform lifecycle is pivotal for determining follow-on engagement behaviors, yet it is understudied. This gap in evidence is partly due to the empirical challenges of studying how individuals select into and the

*This chapter is based on joint work with Laura Rosendahl Huber. For valuable comments and suggestions, we thank Alessandra Allocca, Sofie Cairo, Florian Englmaier, Dietmar Harhoff, Tobias Kretschmer, Ambre Nicolle, Timm Opitz, Helene Strandt, Rainer Widmann, and participants of DRUID 2024, the seminars at CBS, LMU Munich and the Max Planck Institute for Innovation and Competition. We thank the platform for data access and generous support in understanding the platform. We also thank Haylee Ham and Yilun Xu for their invaluable support in extracting and processing the data. We used the generative AI tool ChatGPT for coding and the AI-based writing assistance tool Grammarly to edit the text for spelling, grammar, and style. We reviewed and edited this content as needed and take full responsibility for the content of this study.

[1]Source: https://straitsresearch.com/report/knowledge-management-software-market. Last retrieved on July 14, 2024.

endogenous timing, nature, and formation of social ties on knowledge exchange platforms.

This lack of evidence is partly rooted in empirical challenges related to selection into and the endogenous timing, nature, and tie formation of social interactions on knowledge exchange platforms (Manski, 1993).

In this paper, we use interaction data from an online educational platform to describe the dynamic impact of initially active peers on long-term user engagement and persistence. Data stems from over 12,000 working professionals in online business skills training courses. We capture 36 cohorts over five years, from April 2017 to June 2021, observing the overall cohort activities and interactions directed at the pivotal user. Exploiting quasi-random variation in the share of active peers per cohort in the first period allows us to estimate the *impact of having a higher share of active peers and users' early social engagement on users' future engagement and persistence*. We employ cutting-edge natural language processing to further investigate the nature of the users' comments.

All knowledge exchange platforms share standard interaction features. Users generate content and interact on knowledge exchange platforms in standardized modes - posts, comments, and reactions ("Likes"). Statements or questions are posted on a timeline or a discussion board to add new content. In this paper, we specifically distinguish between general activity on the platform and activity directed to a focal user from their peers. This approach allows us to pin down empirically which "engagement norm" - broad, top-down norms versus personalized, bottom-up nudges from peers impact user behavior.

We find that only directed interactions affect users' future persistence and engagement. Specifically, users who receive direct comments and like endorsements early on are significantly more likely to persist. These positive effects are strongest for users who have received 'affirming and elaborating' comments in the first period. Yet, being merely surrounded by an initial high share of actively commenting peers does not lead to more engagement or persistence on the platform. Indeed, there seems to be a crowding-out effect: A higher share of peers giving likes significantly decreases platform persistence by about 3% on average and slightly diminishes later commenting probability.

Existing literature on online knowledge exchange platforms has primarily focused on studying user interactions revolving around individuals' *own* asking and responding *behavior* (Wasko & Faraj, 2005; Hwang, Singh, & Argote, 2015). These studies investigate, for instance, various determinants of knowledge sharing (Faraj et al., 2016; Xu, Nian, & Cabral, 2020) and knowledge sourcing (Borgatti & Cross, 2003; Mickeler et al., 2023). Only recently, studies have begun to explore the broader influence of *peer activities* on user outcomes. For instance, in a virtual college education context, Bettinger, Liu, and Loeb (2016) find suggestive evidence that direct peer referrals improve learning outcomes of inactive users. Furthermore, investigating user interactions on a 3D printing platform, Claussen, Halbinger, and Hermida Carrillo (2022) show that users reciprocate tool-related assistance to others and that effects are strongest early in users' community life cycle. Beyond that, which activities and how the activities of one's peers may facilitate continuing engagement and persistence on online platforms has received little scholarly attention.

Building on this emerging research, we focus on the dynamic impact of *early peer activities* on other users' future engagement and persistence on an online knowledge exchange platform. An extensive literature in economics has convincingly demonstrated that peers can have a positive impact in various settings such as education, technology adoption, or entrepreneurship (e.g., Zimmerman, 2003; Bandiera & Rasul, 2006; Lerner & Malmendier, 2013; Feld & Zölitz, 2017), there is limited research on peer effects in online settings. Furthermore, in offline settings, the precise mechanisms of how and under which conditions interactions with active peers are beneficial remain unobserved. We capture the notion of peer activities and apply that to online user interaction behavior at the crucial stage in users' platform life cycle — the beginning.

This paper sits at the intersection of the literature bodies on digital platforms, peer effects, and text data methods in the social sciences and offers several contributions to the literature. First, we contribute to the scholarship on user-generated content platforms (Wasko & Faraj, 2005; Bettinger, Liu, & Loeb, 2016; D. J. Zhang, Allon, & Van Mieghem, 2017; Claussen & Halbinger, 2021; Claussen, Halbinger, & Hermida Carrillo, 2022; Kretschmer et al., 2022; Kreyer & Wang, 2022; Loh, 2022; Loh & Kretschmer, 2023) by investigating the timing, type, and content of online peer interactions to deepen the understanding of how early general and directed peer activities influence subsequent user behavior.

Second, we also contribute to the literature on peer interactions (Sacerdote, 2001; Ammermueller & Pischke, 2009; Feld & Zölitz, 2017; Bostwick & Weinberg, 2022; Feld & Zölitz, 2022) by adding evidence from an online setting where our fine-grained interaction data and cohort setting allows us to measure how individuals may benefit from early peer activities and input over a more extended period. Furthermore, our paper has implications for the labor force training literature (Bidwell & Briscoe, 2010; Cappelli, 2015) due to our novel focus on a diverse, global sample of working professionals and executives. To our knowledge, this paper is among the first to advance the understanding of peer effects among working professionals in a digital learning environment. The MBA training context used in this study allows us to assess platform behaviors in a highly relevant labor market setting. At the same time, its structure and core features are sufficiently similar to bear applicability to a broader set of knowledge exchange platforms.

Lastly, we add to a recent stream of works in social science using text data to generate measures for empirical analyses by introducing a novel scheme of comment classification (Athey & Imbens, 2019; Gentzkow, Kelly, & Taddy, 2019; Ash & Hansen, 2023).

The remainder of this paper is structured as follows. In Section 2.2, we outline the theoretical considerations and formulate our hypotheses. Section 2.3 describes the empirical platform setting, estimation strategy, and data. Section 2.4 presents the results. Finally, Section 2.5 offers a discussion and conclusive remarks.

## 2.2 Theory

### 2.2.1 The Role of Early Peer Interactions for Future Engagement and Platform Persistence

Interaction and knowledge exchange constitute the fundamental purposes of numerous online communities (Q. Jones, Ravid, & Rafaeli, 2004). Similarly, many online education platforms are intentionally equipped with features such as internal discussion boards to facilitate interactions. All these knowledge exchange platforms comprise similar communication modes, such as posts, comments, and likes intended to encourage public posts or other forms of user interaction. The underlying premise is that the information shared among peers, and feedback given as well as received is meaningful (D. J. Zhang, Allon, & Van Mieghem, 2017; Wang, Zhang, & Hann, 2018). In reality, however, many knowledge exchange platforms encounter challenges related to low response rates and sustaining user engagement (Ren et al., 2012; Gallus & Frey, 2016; J. Baek & Shore, 2020; Mickeler et al., 2023). Hence, the value creation of a platform can only reach its full potential with a sufficiently high number of active contributors (Faraj, Jarvenpaa, & Majchrzak, 2011; Loh & Kretschmer, 2023).

A pivotal phase in users' platform life cycle is the initial period following their entry into the platform. This is the time when users are likely to be most motivated and active. Hence, what happens in the early stages of a platform life cycle can have an imprinting effect on users in terms of norms around interactions and knowledge exchange on the platform (Claussen, Halbinger, & Hermida Carrillo, 2022). There are two competing theories on how best to encourage online users to engage and how to establish such a beneficial norm. First, a cohort might adopt the norm if it is modeled on the platform (i.e., top-down) and observed by them. Second, individual platform members may receive individual feedback and comments on the content they have provided that could encourage them to adhere to the norm through a mechanism of reciprocity (i.e., bottom-up or individual nudges).

To understand how these different types of initial peer activity influence users' future engagement and persistence, we distinguish general activity from their peers on the platform from activity directed to a focal user in the first period. We define *General Peer Activity* as exposure to undirected peer activities, i.e., observing others' commenting and liking activities. Whereas *Directed Peer Activity* is defined as receiving comments or likes from one's peers on one's own posts or comments. Thus, we conceptually disentangle which "interaction norm" — overall top-down versus bottom-up individual nudges — can foster future user engagement and platform persistence. We will start by discussing how these different types of interaction norms may influence future engagement, followed by a discussion on their impact on platform persistence.

**Future Engagement**

Starting with *General Peer Activity*, it could be argued that observing a high share of peers being active on the platform could be seen as setting the norm around (social) interactions on the platform. That is, observing that many peers distribute likes or comments in the initial phase on the platform could establish a *general social norm of high engagement* for future periods, encouraging the focal user to become active in their interactions on the platform themselves.

On the other hand, information overload and heterogeneity in contribution quality is a common challenge in large online forums (Q. Jones, Ravid, & Rafaeli, 2004; Makos et al., 2013). This issue is particularly pertinent in settings where posting is mandatory or online communities comprise many (active) users. The resulting abundance of content makes it challenging for users to discern valuable insights from repetitive or low-quality contributions. Additionally, in many settings, users are professionals with full- or part-time employment, limiting their time on any platform.

As such, the critical question of where users should focus their scarce attention arises. We propose that the activity, particularly *likes* endorsing posts or comments, could help identify promising content. Users may like the content they consider high quality, directing attention toward valuable contributions (Makos et al., 2013). Consequently, seeing that many peers are active early on, provide useful content, and "like" endorsements can motivate users to interact and contribute content to the platform in the future. Such motivators could be exchanging ideas, soliciting feedback, or aiming at "collecting likes" for one's posted content (J. Baek & Shore, 2020). Hence, we hypothesize:

**Hypothesis 1.** *A high share of initially active peers (with respect to distributing likes and comments) positively affects the focal user's future engagement (i.e., commenting and liking activity).*

Moving from the "passive" observation of peers' activity on the platform to peer activity *directed* at the focal user, these directed interactions typically consist of *receiving likes* and/or *comments*. It has been established in the literature that reciprocity can be a key motivator for user contributions online (Ardichvili, Page, & Wentling, 2003; Wasko & Faraj, 2005). For instance, L. Chen, Baird, and Straub (2019) use panel data from a technical Q&A-forum to find that "positive votes" by peers have a motivational effect on future engagement, particularly knowledge contributions. Hence, receiving likes on one's posts early in a user's life cycle on a platform may create a positive feedback loop encouraging their future engagement. Specifically, hypothesize:

**Hypothesis 2a.** *Receiving like(s) in the initial phase of the platform's life cycle positively affects users' likelihood of distributing likes to other users in future periods on the platform.*

Moving beyond likes, we consider the role of receiving comments from peers early on. There are several facets to receiving comments. First, similar to receiving a like, a received comment can trigger a direct reaction or impact future engagement positively via the specific or general reciprocity channel (Surma, 2016). Previous research has used data from an open-source software development community to analyze the role of community responses and member roles on platform persistence and found a positive impact of community member responses, particularly on "users" continued participation (C. Zhang, Hahn, & De, 2013). Hence, we hypothesize:

**Hypothesis 2b.** *Receiving comment(s) in the initial phase of the platform life cycle positively affects users' likelihood of distributing comments to other users in future periods on the platform.*

**Platform Persistence**

Besides having an impact on future engagement, one could also imagine that (observing) early peer interactions could also influence user persistence on the platform. Starting again with *General Peer Activity*, we argue that observing interactions between other peers, in particular when it comes to likes, could signal a supportive environment, which could motivate the focal user to do their best to collect likes from their peers as well. Furthermore, as mentioned above, observing likes being given to a certain post could help direct attention to useful content, making knowledge accumulation and time spent on the platform more effective and motivating for users to continue and persist. With regards to general *commenting* activity, we believe this could have the same effect as observing general liking activity, i.e., helping to direct attention, but it could also serve another purpose. While likes only provide a positive evaluation of the content provided, comments to such content could contain additional information that could be beneficial for the success and fruitful knowledge production on the platform. That is, by observing and reading other users' comments, users could gain additional knowledge that could encourage them to return and persist on the platform. Taking these arguments together, we hypothesize that:

**Hypothesis 3.** *A high share of initially active peers (with respect to distributing likes and comments) positively affects the focal user's platform persistence.*

Moving from general peer activity to *directed* peer activity, we postulate that receiving likes from peers on one's posted content can boost motivation to continue (K. Baek et al., 2011), and it may positively impact persistence on the platform through the channel of a sense of appreciation or reassurance in one's posts. Furthermore, when it comes to directed comments on one's own post, this could provide valuable insights for the knowledge accumulation process by either validating or correcting the content provided in the respective post. The value of online knowledge exchange is, to a large extent, determined by the value created by the knowledge exchange experience.

Hence, we argue that the more useful the knowledge exchange on the platform is, i.e., by receiving positive reactions or valuable comments, the longer users will persist on the platform:

**Hypothesis 4.** *Receiving like(s) or comment(s) in the initial phase of the platform life cycle positively affects the focal user's platform persistence.*

### 2.2.2 Differential Effects of Comment Type on Future Platform Behavior

To delve deeper into the mechanisms driving the positive impact of peer activity on prospective engagement and platform persistence, we consider various types of comments as a moderating channel. The type of user contributions on a platform is naturally heterogeneous in terms of the number of contributors, the quality, and the type of contributions made. Within the scope of this paper, we distinguish the following types of comments: Comments might either *agree* or *disagree* with an original post or previous comments. If users intend to solely state their (dis-)agreement with peers' content, they provide a *valuation*. Conversely, users may wish to provide a valuation of their peers' content and *elaborate* on why they (dis-)agree. First, considering comments that provide a *valuation* are either *purely agreeing* or not. Such purely agreeing comments in the initial phase on the platform may provide positive feedback and reassurance to the recipient. Similar to receiving a like, this early positive experience can have a motivating effect to reciprocate by engaging with peers' content in the future and persist longer on the platform. Using Facebook brand page data, Khobzi, Lau, and Cheung (2019) indeed find that the sentiment of messages matters for users' engagement behavior. They find that more positively and negatively framed comments trigger user engagement. In our setting, positive comments correspond to "agreement". Hence, we hypothesize:

**Hypothesis 5a.** *Receiving purely agreeing comments in the initial phase of the platform life cycle positively affects the focal user's future engagement (i.e., commenting and liking activity).*

**Hypothesis 5b.** *Receiving purely agreeing comments in the initial phase of the platform life cycle positively affects the focal user's platform persistence.*

Second, users may wish to provide a valuation of their peers' content and additionally *elaborate* on why they (dis-)agree. Looking at these longer, *elaborating comments* might be particularly insightful as users might share new insights or explanations. Such insightful, "knowledge-exchanging" comments might open up new threads or trigger the continuation of existing discussions among users.

Previous work by Ziegele et al. (2018) investigates factors influencing the impact of the civility of comments on news websites on readers' willingness to comment by experimentally manipulating comment nature. Their findings suggest that the type of comments can dynamically shape user engagement in online discussions. Particularly, "deliberative" reader comments containing discussion features such as questions or additional information, increased participants' willingness to reply to these comments primarily via cognitive involvement. Similarly, Kwon et al. (2019) study the effects of different types of instructor comments on engagement in an online discussion in an educational setting. Their descriptive content analysis suggests that elaborating comments are positively associated with interactivity among learners. Hence, we hypothesize:

**Hypothesis 6a.** *Receiving elaborating and agreeing comments in the initial phase of the platform life cycle positively affects the focal user's future commenting behavior.*

**Hypothesis 6b.** *Receiving elaborating and agreeing comments in the initial phase of the platform life cycle positively affects the focal user's platform persistence.*

## 2.3 Data and Empirical Strategy

The empirical context of this study comprises the course progression and user interactions of a large, diverse population of working professionals of an elite U.S. business school's upskill training courses. We capture cohorts over a five-year period from April 2017 to June 2021, with most course offerings once per quarter. Unlike Massive Open and Online Courses (or MOOCs), these online business courses require prospective users to apply to the program, get accepted, and pay non-trivial tuition to enroll alongside a cohort of virtual peers.

### 2.3.1 The Empirical Setting

#### Platform, Course & Cohort Structure

In particular, we assess the business school's online flagship program, which consists of three business and economics courses. These three courses are taken simultaneously over 10-12 weeks, starting staggered one after another. This program's tuition is above $ 2,000, and approximately one-third of enrolled users receive at least some corporate tuition reimbursement. Through our interviews with the program administrators, past users have indicated their motivations for taking the online courses, including having helped to bolster their CVs, improved their job performance, and enabled them to join a network of like-minded peers.

The learning model for each course is designed around three core components that are all delivered asynchronously, which include video lectures, case-based learning where users discuss and debate solutions to real-world business cases with their cohort peers, as well as social learning, where cohort members exchange ideas, offer input, seek out different viewpoints, and learn from one another's experiences and perspectives. This paper focuses primarily on the latter element of the courses. Upon course completion, users receive an online certificate from the business school. To earn a certificate of completion, users must complete each week's lessons by the weekly deadlines and earn an average quiz score at the end of each lesson of at least 50%.

#### Platform Objectives

Generally speaking, knowledge exchange platforms facilitate digital interactions and value creation between users (Rochet & Tirole, 2003). These users can be grouped typically into knowledge providers and knowledge seekers. The value created on knowledge exchange platforms depends on the quality and quantity of the provided content (Faraj, Jarvenpaa, & Majchrzak, 2011).

The setting of our study is an online learning platform, which we consider to be a special type of knowledge exchange platform. Apart from the objectives online learning platforms share with a broader set of knowledge exchange platforms, they further aim to maximize users' learning outcomes and their "customer" satisfaction (Alavi & Leidner, 2001).

Within the learning interface, social interaction features typically include discussion boards, built-in Q&A-forums, or, as regarded in this study's context, (mandatory) public postings (see paragraph below for a detailed description of the interaction features in place). These features are conducive to maximizing users' learning outcomes if more users make more high-quality postings and comments. Furthermore, the possibility and display of reactions (i.e., "Likes"), are a way to "curate" high-quality content by steering users' attention toward it. Such "curation" may benefit both active and passive users by guiding them to high-quality content generated by their peers.

Given that a large subset of users take the course in preparation for a full-time MBA or upskilling for professional reorientation, a degree of career concerns and networking incentives are also likely to be present on the platform. We assume that engaging these inactive users would increase the amount of high-quality content, benefiting the exchange of new information and learning experiences of themselves and their course peers. High customer satisfaction, in terms of engagement and persistence on the platform, increases the likelihood of future enrollment, recommendations, and positive reviews, which benefits the platform's reputation and may attract future users.

**Suitability of the Platform and User Incentives**

Our study context is uniquely suited to examine peer activity in online engagement for several reasons. First, our setting circumvents concerns of selection and common shocks, which may create identification problems in other peer effects studies (Manski, 1993). Course syllabi are publicly posted, and video lectures are pre-recorded. These features remove heterogeneity in the delivery of course content that may arise due to changing instructors or spontaneous rearrangement of modules. More importantly, the set of enrolled users applied independently and cannot select which (virtual) cohort "classroom" they are in, as cohorts are constructed based on the individuals who decided to apply to the same offering of the course, i.e., for the same "course wave" (Rosendahl Huber, N Lane, & Lakhani, 2020).

Second, interviews with the program administrators suggest they aim to keep the cohorts in a target size of 400 users (Mean = 384, SD = 73.6, Median = 376, N = 36 cohorts) to ensure a critical mass of users on the platform for interaction purposes. A cohort of several hundred users is robust to time zone and work habit differences. It limits the likelihood that a user will not have any peers engaging with their contributions and ideas on the platform without becoming too large either. Users can only enroll in the course if they have passed the program's admission criteria, and about

50% of admitted users choose to enroll. Upon enrollment, users are assigned a cohort of peers who simultaneously start the course.

Due to the popularity of the flagship program, many course waves were larger than the above-mentioned target size. If a course offering has more than 600 users, the users are randomly split into two separate but simultaneous cohorts. The random assignment of program users strengthens our empirical approach (see below) by alleviating concerns related to selection into a cohort with already-known peers such as friends or co-workers. Hence, we observe social engagement among (mostly) unknown peers, which is representative of many online environments such as MOOCs, large companies' knowledge-sharing platforms, and interest-based online communities.

Finally, there are significant opportunities for peer-to-peer interaction, as the course content is structured around interactive learning via case studies and weekly discussions, where users are encouraged to discuss case and homework prompts with their cohort peers. The opportunities for peer-to-peer interaction are particularly useful for working professionals due to the knowledge and skills experienced users bring to the course (Littlejohn et al., 2016). Moreover, active participation in the social course elements is an explicit learning requirement within this program. It is *incentivized* such that active participation contributes to improving the final grade, conditional on having passed the program based on the module quiz and exam performance.

Finally, the platform actively advertises the network that users can build during their program participation on their website. These networking opportunities likely increase users' active engagement. Since the platform offers business education for professionals, one can credibly assume that some non-monetary incentives, such as status or career concerns (Xu, Nian, & Cabral, 2020), as well as social comparison (Gerber, Wheeler, & Suls, 2018), might additionally be at play.

**Peer Interactions on the Platform**

While the courses are asynchronous in general, i.e., users can log in whenever they want—there is a certain degree of synchrony imposed by common (bi-)weekly deadlines and locked content. Each course module is unlocked at a fixed time, determining when users can start working on a module. Within a module, some content is locked/not visible until the user completes the previous content, meaning that one can look back at previous work but cannot skip ahead. There is also a shared module deadline by which all work of a particular module must be completed. Most modules are available for two weeks. This ensures that all users in a cohort work through the asynchronous material for a module in the same two to three-week period.

In addition to completing the quiz for each module, users are also expected to periodically post and respond to “shared reflections” on a joint message board. These shared reflections represent a key element of social interaction in this knowledge exchange platform. They are distributed consistently throughout each course, providing a consistent stream of interaction data that can be examined at multiple points within any given course. Therefore, shared reflections are the focus of measurement around users’ relative levels of on-platform social activity. See Figure B.1 for an exemplary screenshot of a module reflection.

These reflections typically ask users to reflect on some questions about the course material and offer their thoughts and opinions on the discussion board. Other users can then engage with the reflection by “liking” it or responding to it with their thoughts. In some cases, this results in a back-and-forth discussion between the original poster and other users within the comments of a reflection. While sharing "reflections" is a compulsory element of the program, both commenting and liking are optional; a user could hypothetically complete the courses without submitting a single comment or liking a single response.

The platform also offers an activity feed that presents a news-feed-like presentation of course-related activities, which is displayed on the user on their program landing page. These feeds include an indication with a *link* if a shared reflection written by the user has been commented on or liked, a *pointer* to a peer help response given by the user that has received a response, platform announcements, and more. These discussion boards include familiar communication modes, such as *posts, comments*, and *likes* intended to encourage public posts or other forms of interaction with other users.

### 2.3.2 Sample Descriptives

The sample contains 12,687 users coming from diverse backgrounds and experiences with 52 industries (e.g., consulting, education, energy, and healthcare), 35 fields of study (e.g., accounting, computer science, engineering, psychology, sociology), and 129 countries (e.g., US, China, India, Australia, Brazil). Overall, 40% of users identified as women, and the age ranged between 20 to 76 years with a median age of 34 years (see Table B.1).

Unfortunately, their work experience and study field information are only available for a considerably smaller sample subset. Hence, the following proportions should be considered only a crude approximation. The largest two groups of study fields are STEM and business-related fields, with 35% and 18% of users, respectively. The share of users with a background in any Social Science is about 13%. The residual share of users has studied a subject other than the three groups.

A large overall heterogeneity in user activity can be observed in terms of their interaction behaviors on the platform. Table B.1 displays summary statistics for the peer interactions on the platform. These statistics show that the median user receives 12 comments over their life cycle on the platform[2], conversely, the median of comments given is at 6. Likes, i.e., the faster and arguably cheaper type of peer interaction, are considerably more frequent, with a median of 56 given and 66 received likes, respectively. Hence, the aggregate interaction statistics suggest that the median user receives more interaction instances than they contribute.

### 2.3.3 Engagement and Persistence Measures

The outcome variables of interest are future engagement and persistence on the platform. We measure users' future engagement with two distinct variables, i.e., indicating whether they gave (1) any *Comment(s)* or (2) any *Like(s)*, to their peers in any of the later periods (all except for the first one). Descriptive statistics in Table B.1 indicate that 66% of users give at least one comment, and 85% of users distribute at least one like in any of the periods following the first one. Users *Persistence* on the platform is measured by the number of completed modules ranging from 1 to 17 (see Table B.1). The average platform persistence is 15.6 modules, with a standard deviation of 3.6 modules.

Due to the nature of the data, specifically, the considerable share of inactive users not making any comments or likes initially, i.e., zeros (see Figure 2.1) and overall decaying engagement over time (see Figure B.2), our main analysis focuses on the extensive margin using indicator (dummy) variables for the engagement outcomes and main explanatory variables.

In terms of explanatory variables, the interactions at the initial stage of users' life cycle on the platform are of particular interest to this paper. In each cohort, a share of users actively engage by giving comments or likes, while the rest remain inactive (see

[2]In our setting, a life cycle is the same in every cohort of the course program.

**Figure 2.1:** Cohort Shares without Engagement in Module 1

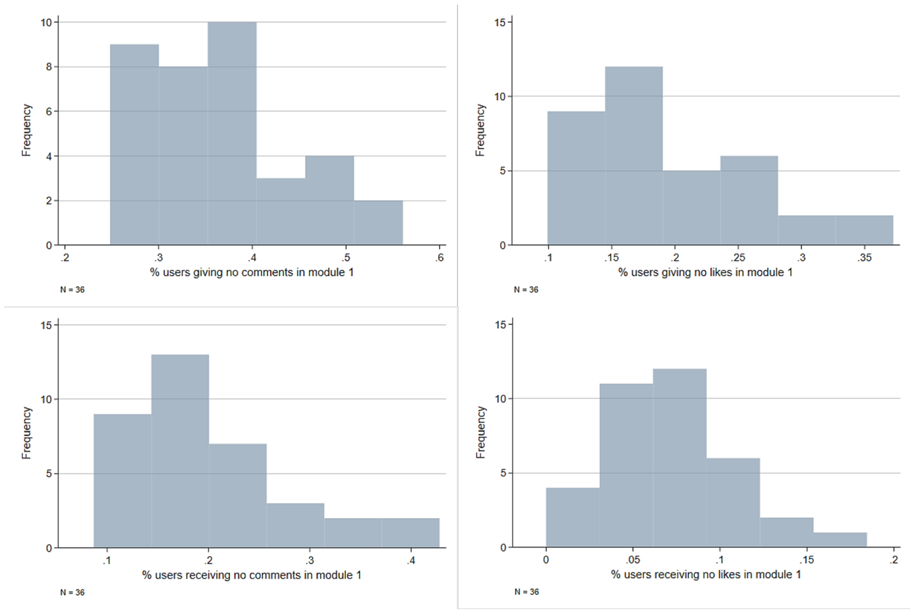
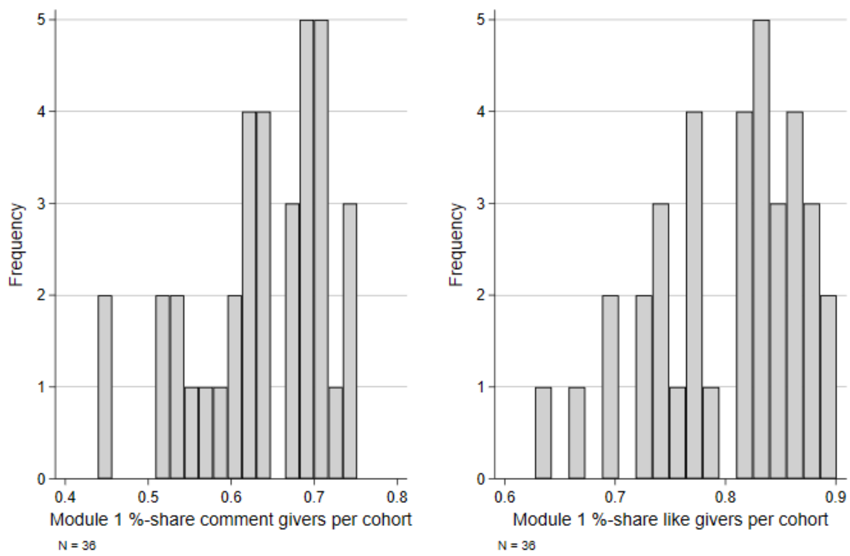



Figure 2.1). To test our hypotheses, we distinguish between *general peer activity* and *directed* peer activities toward the focal user. The former type of peer activity refers to users' exposure to the general *undirected* peer activities on the platform. In the first phase (called "Module 1"), the share of users giving at least one comment is 62% on average. There is, however, considerable variation between the different cohorts in these shares, ranging from 44% to 75%.

Conversely, the share of users giving likes to their peers in the initial phase of their platform participation is higher, with an average of 81% giving at least one like and a variation in this share ranging from 63% to 90% between cohorts (see Figure 2.2). We will use this variation in initial peer activity in our estimation strategy (see Section 2.3.5). That is, we measure the share of active users per cohort in the first period and then calculate the median activity shares across all cohorts with respect to the number of peers giving comments and likes. We define "high" activity cohorts in terms of *commenting* as those with a share larger than 65% of users who give at least one comment and for *liking* as those with a share larger than 83 % of users who distribute at least one like (see Table B.2). The *directed* peer activities toward a user are indicated by whether a user received any $Comment(s)$ (83% or users) or $Like(s)$ (94% of users) during the initial period on the platform.

**Figure 2.2:** Histograms of the Share of Active Peers in Module 1 by Interaction Type




### 2.3.4 Natural Language Processing for Measure Construction

To test the differential effect of comment type on platform engagement and persistence, we engineered computational measures of the content of users' comments, classifying comments as agreeing or disagreeing with the original post and classifying comments as either elaborating on a point or making a simple valuation statement. As the value created on knowledge exchange platforms depends on the quality and quantity of the provided content (Faraj, Jarvenpaa, & Majchrzak, 2011), our measure of "elaborating" comments serves as a proxy capturing the "knowledge exchange" dimension of user interactions.

#### Comment Classification Matrix

We classified comments as *agreeing* or *disagreeing* with the original post by using a DistilBERT model as the base.[3] We then fine-tuned the sequence classification model on a subset of labeled data from the dataset. For training data, we labeled 82,000 comments as either agreeing or disagreeing with the original post, i.e., such that there are no neutral comments. After fine-tuning the model, we achieved an accuracy of 94%. The trained model was then used to classify all comments as either agreeing or disagreeing with the original post.

[3]Source: https://huggingface.co/docs/transformers/model_doc/distilbert, accessed on Sept 15, 2024.

We developed a heuristic to classify a comment as either containing elaboration or not. We found that a simple heuristic of comment word count could be used to make the classification. We hand-labeled 906 comments as either elaborating on a new point or simply agreeing or disagreeing with previously established points. Using our heuristic, a comment would be classified as containing elaboration if it contained more than 10 words and not containing elaboration if it contained less than or equal to 10 words. With this heuristic applied to the labeled data, we were able to achieve a 91.5% accuracy rate.

**Figure 2.3:** Given Comment Type Shares by Course

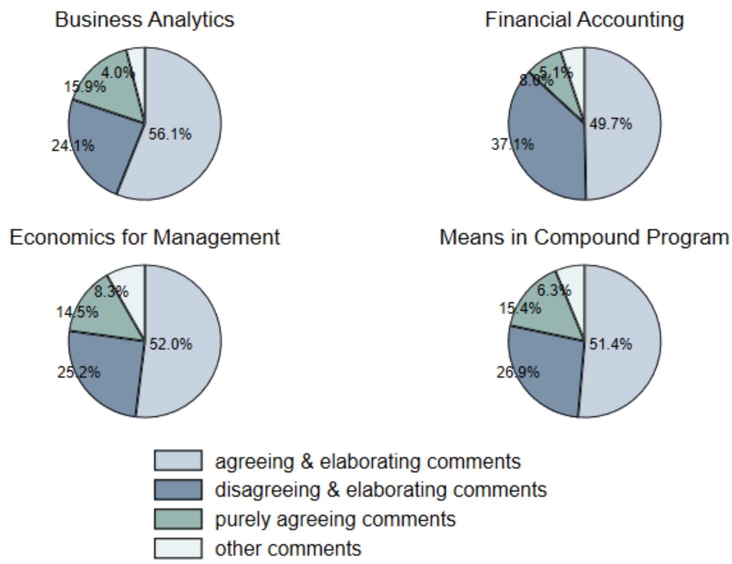



Figure 2.3 depicts descriptive information on the aforementioned text classification measures. The overall share of agreeing comments among all comments is 66.8%, while the share of purely agreeing comments without elaboration (e.g., “Well said. ") is 15.4%. In particular, elaborating comments are the ones in which participants can share knowledge with their peers. We observe 78.3% of comments being classified as elaborating and a subset of 26.9% of comments as elaborating and disagreeing. See Appendix B.4 for examples of every comment type.

### 2.3.5 Empirical Strategy

We aim to estimate the effect of general and directed peer activity on users' subsequent engagement and persistence on the platform. To test our hypotheses regarding the effect of *general peer activity*, the key variation we exploit is the share of active peers in the initial period that differs from cohort to cohort, which we argue is quasi-random (see Figure 2.2. Specifically, there is *uncertainty*, both on the part of the admissions office and on the part of potential users about the realized set of active peers in each cohort. For this quasi-random variation to have a *causal* interpretation, we have to assume that this share is also uncorrelated with other unobservable factors impacting the outcomes of interest at the cohort level (Ammermueller & Pischke, 2009; Bostwick & Weinberg, 2022).

Since there may be some endogeneity in terms of the timing when users decide to join the platform, e.g., users joining in winter may be different from users joining in spring, we only include so-called user "waves" where subscriptions were sufficiently high so that these users were *randomized* into two (or more) cohorts of similar size. This randomization further strengthens our identification strategy because the assignment of more or less active peers across these cohorts is arguably exogenous. Together, we estimate the following Ordinary Least Squares (OLS) regression model:

$$Y_{ic} = \beta_0 + \beta_1 HighCommentGivers_c + \beta_2 HighLikeGivers_c + \beta 3 CommentsReceived_i + \beta 4 LikesReceived_i + \beta_5 \boldsymbol{X}_i + \beta_6 YearFE_c + \epsilon_{ic} \tag{2.1}$$

Here, $Y_{ic}$ is the dependent variable of user $i$ in cohort $c$. It is either their future engagement on the platform, precisely, a dummy indicating whether they gave any *Comment(s)* or *Like(s)* to their peers in the later periods on the platform or their *Persistence* on the platform measured by the number of completed modules ranging from 1 to 17 (see Table B.1). We include four main explanatory variables on the right-hand side of the estimation equation. The first two explanatory variables capture the *general peer activity*, i.e., the exposure to the general *undirected* peer interaction activities on the platform (i.e., Hypotheses 1 and 3). $HighCommentGivers_c$ and $HighLikeGivers_c$ are dummies indicating cohorts with an above-median number of users giving comments (>65% of users) and likes (>83 % of users) in the initial phase on the platform, respectively (see Table B.2). The latter two explanatory variables capture *directed* peer activities toward a user, and hence provide a test for Hypotheses 2a and 2b, as well as Hypothesis 4. These are dummy indicators for any $CommentsReceived$ and $LikesReceived$ in the initial platform period.

$\boldsymbol{X}_{ic}$ is a vector of user controls, including a dummy for female sex (yes/no), age (in years), living in the US (yes/no), English as an official language at their location (yes/no), metrics of demographically similar peers at the cohort level (# of same gender, # similar age (+/- 2 years), # same country, and # same citizenship peers), procrastination (submission hours to quiz deadline), cohort size (N), and quiz score in the first submitted module as a proxy of their platform-related ability. We also include year-fixed effects (FE) to eliminate omitted variable bias caused by unobserved factors that evolve but are constant across users on the platform in a given year. Finally, $\epsilon_{ic}$ is the residual error term. Table B.1 contains the summary statistics for the variables used in the main analyses.

Because of the highly endogenous nature of later-period interactions among users, we use a cross-sectional analysis to isolate the impact of early peer activities on users' later engagement and persistence.

## 2.4 Results

### 2.4.1 Main Findings

#### Future Engagement

This paper asks how the presence of initially active peers impacts users' future engagement and persistence on a digital platform. Table 2.1 presents linear probability estimates of how the cohort composition of more or less active peers affects users' future engagement on the platform. Future engagement is measured by two dummy variables capturing *any comments given m2-7* (see Columns 1 and 2) and *any likes given m2-7* (see Columns 3 and 4) in later periods.

To begin, we consider the effect of *general peer activities* on users' prospective behavior on the platform. First, we test if a high share of peers who initially engage in commenting or liking has a positive effect on users' future engagement (Hypothesis 1). The results show that being in a cohort of users with an above the median share of comment-givers increases an individual's likelihood of giving comments in the future by around 2-3 percentage points (pp) (Column 1). However, this result does not hold once we include year fixed effects (Column 2). Furthermore, there seems to be no effect of being in a cohort with a high share of comment-givers on the propensity to distribute likes in the future. The effect is negligible with less than 0.5 pp, and the switch of the coefficient sign from Column 3 to Column 4, i.e., without and with year-fixed effects included in the model, hints towards further instability of any directional effect.

With regards to general peer activity in terms of liking, we find that users in a cohort with an above-median share of peers distributing likes in the initial period are less likely to give comments in future periods (see Columns 1 and 2 of Table 2.1). This effect is 4.5 pp large and statistically significant at the 5% level, including a large set of individual user control variables. However, once year-FEs are controlled for, the effect decreases to 3 pp and lacks statistical significance at conventional levels. The high share of peers distributing likes during the initial period on the platform also does not seem to increase users' probability of giving likes in future periods. The estimated coefficients in Columns 3 and 4 are small and not statistically significant at any conventional level. Hence, overall, we do not find support for our Hypothesis 1 that a high share of general *undirected* peer activity affects users' future engagement.

Next, we turn our attention to the early peer activities *directed* toward the users (see rows 3 and 4 in Table 2.1). The estimates indicate that receiving a comment in the initial period increases the propensity to give comments by 20 pp and likes by 12 pp, respectively. Conversely, receiving likes initially is associated with an increased propensity for commenting and liking in future periods of 10 pp and 14 pp, respectively. Hence, the effects are strongest by interaction type, i.e., comments in the first period are most strongly related to future comments, and vice versa for likes. All estimates of the early peer activity directed to users on their future platform engagement are statistically significant at the 1% level when including user-level controls and year FE. Thus, we find that *directed* user engagement at an early stage is positively associated with individual users' future platform engagement in terms of providing comments and giving likes to other users. We thus find support for Hypotheses 2a and 2b.

**Platform Persistence**

Table 2.2 presents the estimates for platform persistence as the second outcome of interest.[4] We start again by looking at undirected *general peer activity* as measured by the share of active comment- and like-givers in each cohort (Hypothesis 3). We find no clear indication that early actively commenting peers have any effect on platform persistence. While the estimated coefficient is positive in both models, it is not statistically significant in any of the specifications. Furthermore, our findings indicate

[4]For consistency and comparability of main outcomes, we use an OLS model throughout. See Appendix A, Table B.13 for an alternative estimation using a Poisson count model.

**Table 2.1:** Linear Probability Models Predicting Future Engagement

| | (1) *Comments given m2-7* (1/0) | (2) *Comments given m2-7* (1/0) | (3) *Likes given m2-7* (1/0) | (4) *Likes given m2-7* (1/0) |
|---|---|---|---|---|
| *High # comment givers m1* (1/0) | .028* | .019 | .004 | -.002 |
| | [-.003,.058] | [-.012,.050] | [-.019,.027] | [-.025,.021] |
| | (.016) | (.016) | (.012) | (.012) |
| *High # like givers m1* (1/0) | -.045** | -.030 | .013 | .023 |
| | [-.087,-.004] | [-.073,.012] | [-.019,.045] | [-.009,.055] |
| | (.021) | (.022) | (.016) | (.016) |
| *Comment(s) received m1* (1/0) | .205*** | .204*** | .122*** | .121*** |
| | [.180,.229] | [.179,.228] | [.101,.143] | [.100,.143] |
| | (.013) | (.013) | (.011) | (.011) |
| *Like(s) received m1* (1/0) | .107*** | .104*** | .147*** | .146*** |
| | [.067,.146] | [.065,.143] | [.109,.185] | [.108,.184] |
| | (.020) | (.020) | (.019) | (.019) |
| Dep. Variable Mean | .658 | .658 | .847 | .847 |
| Controls | Yes | Yes | Yes | Yes |
| Year FE | No | Yes | No | Yes |
| Adjusted $R^2$ | .097 | .101 | .086 | .089 |
| N | 12,687 | 12,687 | 12,687 | 12,687 |

**Notes.** The dependent variable in columns (1) and (2) *Comments given m2-7* equals 1 if any comments were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in columns (3) and (4) *Likes given m2-m7* equals 1 if any likes were given in later modules 2 to 5/7 in the program and 0 otherwise. *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

that an initial high share of active peers giving likes significantly decreases platform persistence by half a module or about 3% of overall persistence. Thus, we cannot confirm Hypothesis 3 as we find the opposite empirical result for a high share of peers distributing likes early and a null result for the share of comment givers.

Turning to the *directed* peer activity (see rows 3 and 4 in Table 2.2), we find that receiving early comments and likes correlates with a 1 to 1.5 module increase in (or 7% and 9% higher) platform persistence, respectively. These results suggest that while having a high share of general, undirected peer activity has no (or even a small negative effect) on platform persistence, early-stage *directed* peer activity can significantly boost platform persistence. Hence, we find support for Hypothesis 4.

**Table 2.2:** OLS Regressions Predicting Platform Persistence

| | (1) *Persistence* (# modules) | (2) *Persistence* (# modules) |
|---|---|---|
| *High # comment givers m1* (1/0) | .089 | .084 |
| | [-.135,.312] | [-.139,.308] |
| | (.114) | (.114) |
| *High # like givers m1* (1/0) | -.559*** | -.518*** |
| | [-.899,-.219] | [-.862,-.175] |
| | (.174) | (.175) |
| *Comment(s) received m1* (1/0) | 1.058*** | 1.058*** |
| | [.840,1.275] | [.840,1.275] |
| | (.111) | (.111) |
| *Like(s) received m1* (1/0) | 1.455*** | 1.460*** |
| | [1.039,1.872] | [1.043,1.876] |
| | (.212) | (.212) |
| Dep. Variable Mean | 15.643 | 15.643 |
| Adjusted $R^2$ | .137 | .137 |
| Controls | Yes | Yes |
| Year FE | No | Yes |
| N | 12,687 | 12,687 |

**Notes.** The dependent variable in all columns *Persistence* is the number of completed modules of the entire program (0-17). *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### 2.4.2 Mechanism: Comment Type

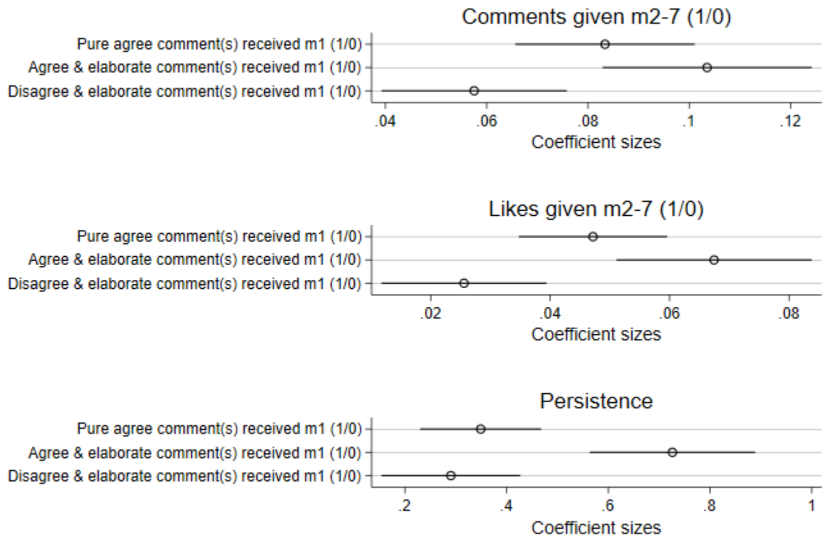
The main analysis suggests that receiving directed comments from peers at an early stage significantly increases users' future engagement and persistence on the platform. Against the backdrop of these empirical results, we investigate whether different types of comments matter differently for fostering engagement and persistence.

As detailed in Subsection 2.3.4, we employ computational measures of users' comments' text, classifying each comment as *agreeing* or *disagreeing* with the original post and as either *elaborating* on a point or making a simple *valuation* statement. Consequently, we obtain four types of comments: (1) elaborating & agreeing comments, (2) elaborating & disagreeing comments, (3) purely agreeing comments, and (4) other comments. Note that the latter category contains all comments that could not fit the categories mentioned above. These are, however, not "purely disagreeing" in nature, as such comments have not been found much on the platform.

To analyze the impact of the distinct comment types, we replaced the explanatory indicator variable $Comment(s)Received - m1$ in Equation 3.1 with dummies for each comment type (1) to (3) in joint regressions. Figure 2.4 presents coefficient plots derived from regressing the receipt of diverse comment types from peers during the initial platform period (m1) on the likelihood of engaging in future commenting and liking activities across subsequent periods (m2-7) as well as persistence depicted along the horizontal axis. In aggregate, the results presented in Figure 2.4 indicate that the nature of comments received from peers during the initial stage of users' platform presence matters for their subsequent behavior on the platform. Specifically, while all three comment types show a positive and significant association with future engagement and persistence compared to the benchmark category of 'other comments', the estimates are largest for *elaborating & agreeing* comments for all three studied behavioral outcomes.

The full estimation results of comment type on future engagement and platform persistence are shown in Table B.12. In line with Hypotheses 5a and 5b, we find a positive and significant association between receiving a purely agreeing comment in the initial stage of the platform life cycle and the focal user's future engagement and platform persistence. Specifically, receiving purely agreeing comments, i.e., without elaboration, yields a 8.3 pp increase in the probability of subsequent commenting and a 4.7 pp increase in the probability of subsequent liking (Hypothesis 5a), and in increases platform persistence by a 0.34 of a module (Hypothesis 5b).

**Figure 2.4:** Coefficient Plots by Dependent Variable and Comment Types




To test Hypotheses 6a and 6b, we look at the estimates for receiving elaborating& agreeing comments during the initial period on the platform. These estimates show that receiving such comments leads to a significant increase of 10.4 pp in the likelihood of future commenting behavior (providing support for Hypothesis 6a) and an increase of 0.73 modules in terms of platform persistence (Hypothesis 6b). Both estimates are significant at the 1 percent level. Moreover, receiving agreeing & elaborating comments during one's initial platform life-cycle also exhibits the strongest correlation with the propensity to offer likes to other users, with a coefficient estimate of 6.7 pp.

Interestingly, the results for agreeing & elaborating comments are significantly larger (more positive) than the estimated coefficients for disagreeing & elaborating comments. That is, the effect of receiving elaborating but *dis*agreeing comments leads to a 5.8 pp increase in future commenting behavior and an increase in the probability of distributing likes of 2.6 pp. This suggests that the combination of positive reinforcement and elaborating on a point is more positively associated with future engagement than a negative valuation combined with elaboration.

The mechanisms analysis shows that receiving early peer comments significantly boosts future engagement and persistence on the platform, with the type of comment playing a crucial role. Specifically, comments that both agree and elaborate on the original post have the strongest positive impact on subsequent user behavior, namely future commenting, liking, and platform persistence. This suggests that the quality and nature of early feedback are key drivers of sustained user interaction on the platform, with agreeing and elaborating comments being the most effective in fostering longer-term engagement.

### 2.4.3 Robustness Checks

To confirm the robustness of our findings, we perform several robustness checks. First, in our main specification, we did not include focal users' own engagement in Module 1 as controls due to concerns about them being subject to a bad control problem. Specifically, the interactions might occur post-"treatment" or are endogenous, i.e., related to or influenced by observing and receiving peers' interactions. Hence, they might bias our estimates and distort any causal interpretation of regression coefficients, leading to underestimating or overestimating the treatment effect. For completeness, Table B.3 reports the main Table 2.1 and Table 2.2 results of the most rigorous specification, including these controls. These results are qualitatively the same but quantitatively smaller, suggesting a downward bias.

Second, Table B.13 contains the results of an alternative specification using a Poisson count model instead of OLS. The estimates are qualitatively very similar, but effect sizes are slightly larger for the general peer activity variables and smaller for the directed peer activity variables. For instance, users in cohorts with a high number of peers distributing likes early persist 0.97 modules less instead of 0.52 according to the estimates column (2), which includes controls and year-fixed effects. In the same specification, receiving comments and likes is associated with persisting 1.07 and 1.11 more modules on the platform, respectively.[5]

Furthermore, while the main specification distinguishes simply $high$ from $low$ initial peer activity, Table B.14 presents an alternative, more fine-grained separation of $high$ (above 66.7 percentile in the number of active users across cohorts), $medium$ (33.4 to 66.6 percentile), and $low$ (bottom 33.3 percentile) number of peers actively engaging in the first period on the platform. We show the results for the model specification containing only our most conservative specification, including a range of demographic and peer controls as well as year-fixed effects.

Compared to being in a cohort with a comparatively low number of actively commenting peers, having more active peers appears to have a positive linear effect. According

[5]To interpret the coefficient of a Poisson regression holding all other factors constant, the following transformation is necessary: $e^{-0.034} \sim 0.97$, $e^{0.07} \sim 1.07$, and $e^{0.108} \sim 1.11$.

to column (1), users in cohorts with a medium-high number of actively commenting peers are about 4pp more likely to give comments in future periods and 7pp more likely to engage in commenting in cohorts with high initial commenting activity. These estimates are significant at the 5 and 1 percent level, respectively. There appears to be no significant effect of high commenting activity on the propensity to distribute likes.

In addition, a high number of peers giving comments early seems to be associated with a small and weakly significant increase in persistence on the platform. In contrast, there is no effect for medium active commenting cohorts. When considering highly active cohorts in terms of giving likes, results in column (2) indicate that only being part of a cohort with a high number of peers distributing likes positively impacts users' likelihood of distributing likes in later modules with a 3.4 pp increase statistically significant at the 1 percent level. The negative results found for cohorts with highly active likers on platform persistence are not significant in this specification. Finally, the positive effect sizes of receiving directed comments and likes, respectively, from peers in the first period on engagement and persistence are almost equivalent to the ones shown in our main specification.

Taken together, the results from these alternative specifications provide support for the robustness of our main findings as presented in Tables 2.1 and 2.2.

**Intensive Margin of Engagement**

For further robustness analysis, we consider the sub-sample of users who received any engagement during the first period on the platform and use the *number* of comments and likes given as the dependent variable. Results from Table B.15 reveal some heterogeneity in the impact of general versus directed peer activity and the engagement type, namely comments and likes. While the models in columns (1), (3), and (5) only include user and peer controls, the models in columns (2), (4), and (6) contain additional year fixed effects.

The estimates suggest that for this sub-sample, there is a negative association between a high number of comment givers in Module 1 and the intensity of later comment engagement (significant at the 5 percent level in column 2). This suggests that exposure to high commenting activity early on may reduce a user's commenting intensity by about 4 comments in later modules. On the contrary, we observe a positive relationship between a high number of actively liking peers in Module 1 and later commenting and liking activity. However, coefficients decrease substantially in size and lack statistical significance in columns (2) and (4) when we include the year FE. This suggests that there is a large variation in the liking activity across cohorts (see also Table B.2). Similar to our main specification, also for this sub-sample, general undirected liking activity is negatively associated with persistence by about half a module (-0.475) in column (6). Hence, while being in a cohort with many actively liking peers could potentially encourage the intensity of this type of interaction, they may also lead to reduced persistence on the platform.

Turning to early directed peer engagements, we find that receiving more comments in the first period positively correlates with future commenting behavior. This relationship is statistically significant at the 5 percent level and shows an increase in later commenting by about one comment (see columns (1) and (2)). Receiving a higher number of likes in Module 1 is positively associated with giving likes in later modules (see columns (3) and (4)) and with program persistence (see columns (5) and (6)). Despite the estimates being significant at the five and one percent level, respectively, the effect sizes are very small in magnitude. This suggests a modest influence of the intensity of early likes on continued engagement and platform persistence. Finally, we observe that the effect of directed commenting in Module 1 on platform persistence seems to be driven by the extensive margin. That is, we do not find any association between the number of comments received in Module 1 and platform persistence.

### 2.4.4 Exploratory Heterogeneity Analysis

The presented main results might vary for different types of users. Hence, we conducted some exploratory heterogeneity analyses based on the following user characteristics: median age, residency in the United States, being on the platform during the COVID-19 pandemic, and user gender. Furthermore, we split future engagement into early versus later periods on the platform. All specifications discussed below include the same user-level controls as in our main specification and year-fixed effects.

#### Age

The first variable for which we consider heterogeneity among users is their age. It could be that different age groups have different levels of familiarity with platform-based communication, i.e., younger users are more familiar with platform interfaces or navigating communication thereon (Venter, 2017) or that career concerns are heterogeneous by age (Xu, Nian, & Cabral, 2020).

Table B.5 depicts the main estimates split by users' *median age* of 32. This split of the sample can serve as a rough proxy of users' career stages. Columns (1) to (3) report estimates for users below the median age, and Columns (4) to (6) depict the equivalent for users above the median age. Looking at general peer activity, we see a clear differential effect by user age. Whereas having a high share of active peers does not seem to have any effect on younger users' platform engagement or persistence, we find that a high share of peers commenting in the first module is associated with a 3.9pp increase in the likelihood of giving comments in later modules for above median age users (significant at the 10% level). Furthermore, our results indicate that a high share of peers giving likes in 1st module significantly negatively affects future commenting behavior (7 pp) and persistence (0.72 modules, i.e., 4.2% of 17 modules) for users above median age. These effects are statistically significant at the 5 and 1 percent level, respectively.

Regarding *directed* peer interactions (see rows 3 and 4), there are positive associations between receiving comments and likes on future engagement and platform persistence for both age groups. The effects appear larger for users above the median age. Table B.4 reports estimates of the interaction effects of the main explanatory variables with an indicator of being below the median age. Being in a cohort with many users giving comments has a 5.8 pp lower likelihood of future distribution of likes for users below the median age. This point estimate is statistically significant at the 10% level. Additionally, many peers distributing likes are associated with a significant 0.5 module higher persistence for users below the median age. Finally, receiving comments from peers in the first module has a lower positive effect for users below the median age of 0.39 modules, which is weakly significant.

**US Residency**

As 57% of users are residents of the US, we split the sample by users in and outside this large group to account for unobserved factors potentially more similar for US users than for others, i.e., communication culture, the popularity of certain social media, education platforms or (partial) time zone proximity.

Table B.6 shows the main estimation results split by US residency status. Columns (1) to (3) report estimates for users residing in the US, and Columns (4) to (6) for users who reside outside the US. Overall, the results for the two groups are somewhat similar. There are only two slight differences. First, the results suggest that an initially high share of commenting peers is associated with a 4 pp higher likelihood of future commenting for US residents. This estimate is statistically significant at the 5 percent level. Conversely, for non-US residents, the point estimate is positive as well and at 2 pp; however, it is not statistically significant. Second, the other weakly significant estimate is the 0.47 module reduction in the average persistence of US residents in a cohort of active like-givers (see Column 3). However, this coefficient is also negative for non-US residents, albeit insignificant.

When turning to the impact of interaction instances directed to users from their peers in the initial period on the platform, the results in Table B.6 indicate economically and statistically significant, positive correlations between the receipt of comments and likes early on future engagement as well as the persistence on the platform. While there are differences in magnitudes for US and non-US residents, the effect sizes are overall comparable.

**COVID-19 Pandemic**

In Table B.7, we present the main estimation results split by cohorts pre- and during the COVID-19 period.[6] The main motivation behind presenting this split is to account for a potentially different selection of users on the platform before or during the global COVID-19 pandemic. Columns (1) to (3) report estimates for users on the platform during the COVID-19 pandemic, and Columns (4) to (6) for users who were on the platform pre-pandemic.

Having a high share of commenting peers in the initial period on the platform significantly increases users' probability of giving comments in the future by 4.2 pp for users in the COVID-19 period. The corresponding point estimate is at 1.9 pp and statistically insignificant for users on the platform pre-Covid. Furthermore, a high share of commenting peers early on is associated with slightly lower persistence on the platform pre-Covid, albeit not statistically significant (see Column 6). While estimation results suggest a slight negative impact of a high initial share of like-givers on prospective engagement and persistence, however, not statistically significant (Columns 4 - 6) pre-pandemic, the corresponding coefficient estimates are negligibly small during the COVID-19 period.

[6] We define user cohorts as in the "COVID-19 period" starting from March 2020 because the WHO declared COVID-19 a global pandemic this month. *Source:* https://www.cdc.gov/museum/timeline/covid19.html, last retrieved on Feb 25, 2024.

Turning from the initial general platform activity to peers' interaction instances received by users, the results in Table B.7 indicate a positive and highly significant relationship between receiving comments and likes early on in users' future platform engagement and persistence both, for users pre- and during COVID-19. The coefficient estimates are slightly larger for users throughout the pandemic. In particular, receiving any comment(s) in the first period increases the likelihood of giving comments and likes in the future by 26.5 pp (Column 2) and 23.2 pp (Columns 3) during the pandemic, respectively, and by 24.2 pp (Column 4) and 19.1 pp (Column 5) pre-pandemic, respectively. Regarding platform persistence, the impact of receiving comments and likes in the initial period is approximately 0.5 and one module greater during COVID-19, respectively (see Columns 3 and 6).

**Gender**

As a next heterogeneity check, we split the sample by gender as a large strand of research in social sciences presents robust evidence for behavioral gender differences. In our context, gender can influence interactions and outcomes in several ways. In online settings like ours, where names and real headshots are used, it is amongst the most salient demographic features. Minority status can make shared demographic traits more noticeable, affecting dynamics within groups (Reagans, 2011; Kleinbaum, Stuart, & Tushman, 2013). Communication styles differ across genders, shaping how contributions are perceived and valued (Kolev, Fuentes-Medel, & Murray, 2019; Exley & Kessler, 2022). Finally, received interactions are shaped by homophily—people's tendency to engage with similar others—and digital misogyny, which can create additional barriers for women (McPherson, Smith-Lovin, & Cook, 2001; Khattab et al., 2020; Strathern & Pfeffer, 2022). Table B.8 shows the main estimation results for women in Columns (1) to (3) and for men in Columns (4) to (6).

Overall, the results suggest several notable gender differences. First, the presence of more comment-giving peers in the first period on the platform increases men's likelihood of actively commenting in future periods by 4.4 pp (Column 4). The coefficient estimates are significant at the 5%-level. Conversely, the point estimate for women is 2.7 pp and insignificant (Column 1). However, while also not statistically significant, only for women, the impact of early active comment-givers on the propensity to distribute likes in the future and on persistence is negative (see Columns 2 and 3).

Second, the main adverse effect of a high share of peers giving likes in the first period on future commenting and persisting appears to be slightly larger and only (weakly) significant for men. Third, while the regression estimates for the association between receiving comments and likes from one's peers in the initial platform period are positive and statistically significant at the 1%-level, the coefficient estimates of receiving comments seem slightly larger for women. The differences are economically significant, particularly for future engagement. In the first period, receiving comment(s) is associated with 30.3 pp and 22.6 pp for women and men, respectively (see Columns 1 and 4). Similarly, the propensity to distribute likes in future periods is 25.4 pp and 19.0 pp higher for women and men, respectively, upon receiving initial comment(s) from peers (see Columns 2 and 5).

Finally, there are differences in the impact of peer users receiving any like(s) in the first period, pointing towards gender differences in reciprocating in online interactions. While the coefficient estimates of receiving likes by peers early suggest a slightly larger, positive impact on women's likelihood to give likes in future periods (see Columns 2 and 5), they are marginally larger for men considering future commenting as an outcome (see Columns 1 and 4).

In Table B.9, we report the interaction effects of our main explanatory variables and an indicator of being female. The estimates suggest no gender differences in early, general peer activity on future engagement and persistence. Looking at peer activity directed to users individually, we find that receiving comments increases the likelihood of giving comments and likes respectively by 6.4 pp and 4.9 pp more for women than men. The impact of receiving early comments and likes on persistence seems larger for women than men but is not statistically significant.

### Early and Later Periods on the Platform

Lastly, we test whether the potential impacts of peers' early platform activity on users' behavior has relatively short or medium-term impacts. For this purpose, we split the future engagement outcomes in *early* versus *late* modules, i.e., periods on the platform. Table B.10 presents results for future commenting and Table B.11 for future liking behaviors.

Overall, the results do not indicate stark differential impacts for splitting the outcome variables by early versus later future periods on the platform, i.e., not pointing towards a relatively short-lived impact of initial general or directed peer interactions on users' prospective engagement.

## 2.5 Discussion and Conclusion

Nowadays, an indispensable share of knowledge sourcing and knowledge exchange in organizations, educational settings, or related to private interests happens on digital platforms. However, the core issue many such platforms share is the decay of user engagement over time. Most of the evidence from the literature on online platforms has focused on interactions surrounding users' *own behavior*, e.g., on determinants of knowledge sharing and sourcing. While there is robust evidence from economics that peers matter in various offline settings, ranging from the traditional classroom to consumption choices, little is known about how peers' activities may impact users' behavior on online knowledge platforms. Recent research has shown that especially interactions that occur early in the platform life cycle may have a lasting impact on future behavior (Claussen, Halbinger, & Hermida Carrillo, 2022). In this paper, we contribute to this emerging body of literature and ask if and how *early* active peers can impact users' future engagement and platform persistence and what type of activities might be most effective in that regard.

We distinguish two types of interactions (comments and likes), as well as two different modes through which platform designers can foster knowledge exchange on their platform (*general* vs *directed* peer activity). The former concerns setting a norm of frequent interaction by encouraging many users to place at least one comment or like in the early stages of the platform life cycle. Our results indicate that this type of norm-setting is ineffective for fostering future engagement in terms of liking and commenting on the platform and that it might even have negative consequences for platform persistence. That is, we do not find any evidence of the effect of *general* undirected peer activity on users' future engagement, and unexpectedly, we find that an initial high share of active peers giving likes significantly decreases platform persistence by about 3%. This crowding-out effect of observing general high 'like activity' seems to be consistent with social comparison theory, according to which users may adapt their behaviors upon observing likes received by others (Gerber, Wheeler, & Suls, 2018).

The other type of social exchange norm we study in this paper concerns interactions in terms of receiving a comment or like on one's own posted content early in the platform life cycle. We find that this type of early *directed* interactions leads to significant increases in the likelihood of future engagement and users' persistence on the platform. For peer activities directed to users individually, we find that receiving early comments and likes correlates with a robust ten-plus percentage point increase in future engagement. We find that this relationship is strongest by interaction type, i.e., receiving comments leads to commenting while receiving likes leads to liking, pointing to reciprocity as a mechanism. Additionally, early comments and likes received are associated with a 7% and 9% higher platform persistence, respectively. Finally, our mechanism analysis shows that receiving early comments that agree and elaborate on a point has the strongest positive impact on all three outcome measures.

This study makes several contributions to the literature. First, our findings have implications for the design and structure of online knowledge exchange platforms. Studies on digital communities have shown that only a small fraction of members typically contribute knowledge therein and that it can be difficult to retain members' engagement over time (Wasko & Faraj, 2005; Ren, Kraut, & Kiesler, 2007; Faraj & Johnson, 2011; Claussen, Kretschmer, & Mayrhofer, 2013; Mickeler et al., 2023). We contribute to this platform literature by studying the timing, type, and content of online peer interactions to deepen the understanding of how early general and directed peer activities influence subsequent user behavior. Based on our findings, designers of online knowledge exchange platforms are advised to harvest better social engagement by promoting directed peer activities early on to enhance their user experience and the value generated by their platforms.

Second, previous evidence from the peer effects literature strongly indicates that social interactions matter for improving learning outcomes. While having been widely studied in economics due to their impact on education and workplace behaviors, most previous work on peer effects has focused on various offline contexts, e.g., in conventional education in a physical classroom (Sacerdote, 2001; Zimmerman, 2003; Ammermueller & Pischke, 2009; Calvó-Armengol, Patacchini, & Zenou, 2009), but also with regards to agricultural technology adoption in developing countries (Foster & Rosenzweig, 1995; Munshi, 2004; Bandiera & Rasul, 2006), entrepreneurship (Lerner & Malmendier, 2013), and consumption behavior (Moretti, 2011). Despite these studies offering ample evidence for peer effects to impact behavioral choices and future (learning) outcomes, the precise mechanisms of how exactly and under which conditions interactions occur remain mostly unobserved in offline settings.

In our study, the unique and detailed data on various online peer interactions and outcomes of interest over a longer period of time allows us to shed light on how and through which channels individuals may benefit from early peers' activity and input. In particular, we show that both *directed* comments and likes early in the platform life cycle are beneficial for future engagement and persistence on online platforms.

Third, our paper has implications for the scholarship on labor force training (Bidwell & Briscoe, 2010; Cappelli, 2015) due to our novel focus on a diverse, global sample of working professionals and executives in this context. Our focus on MBA-type training allows us to assess platform behaviors in a highly labor market-relevant setting. First, the courses' content in and of itself is a domain of knowledge that is highly employable. Second, the sampled courses are feasible alongside full-time occupation. Together, these features bear high relevance for management scholars and practitioners as continuous human capital accumulation is essential to firm growth and innovation (Dragoni et al., 2009).

Moreover, due to the increasing pace at which technical skills become outdated (Deming & Noray, 2018), major multinational corporations such as JP Morgan, Amazon, and Google have launched platform-based training to *upskill* their workforce with the latest business skills and tools (Bidwell et al., 2013; Cappelli, 2015; Deming & Noray, 2018; Illanes et al., 2018; Tamayo et al., 2023). For this purpose, digital training provides a scalable and cost-efficient approach to enable continuous human capital accumulation, essential to firm growth and innovation (Dragoni et al., 2009). Due to the similar structure and shared core interaction features, our insights are likely applicable to, e.g., intra-organizational knowledge platforms as well.

Lastly, we advance management literature's understanding of content-related online communication behaviors by introducing a novel scheme of comment classification in a meaningful way. For this purpose, we apply state-of-the-art Natural Language Processing (NLP) methods on comment texts to classify these into "agreeing" versus "disagreeing" versus "elaborating" comments. This classification intends to bear new insights into communication patterns and their impact on subsequent engagement and persistence. Hereby, we also add to a broader line of works in social science that uses novel methodologies to generate insights from (un-)structured text data (Athey & Imbens, 2019; Gentzkow, Kelly, & Taddy, 2019; Ash & Hansen, 2023).

On a critical note, this paper is subject to several limitations. As an empirical setting to complement our theoretical considerations, we employ a specific knowledge exchange platform in an online education context. As it requires an application and a considerable tuition fee, we cannot rule out a certain degree of selection onto the platform. The sample consists, however, of a diverse and global set of users from over 120 countries, more than 30 different study backgrounds, and working in 52 industries. The share of women is 40%. Hence, the external validity is relatively high within the context of working professionals. Nonetheless, future studies could perhaps investigate if our findings also hold in different platform contexts.

The utilized data source comprises the following shortcomings. First and foremost, data on comprehensive demographic background information is only available for a subset of the sample, limiting the internal validity of these measures. Second, some users drop out during the program. Although the observed attrition does not appear to occur systematically, we cannot rule out unobserved factors contributing to the drop-out of some users. Lastly, we provide purely correlational evidence on the directed peer activity measures, i.e., the *received* comments and likes. Hence, future research may exogenously vary factors that can foster or hinder beneficial, knowledge-sharing interactions in online contexts like the one this project investigates.

Future research in this area may want to investigate heterogeneity in the quality of early user contributions further. As providers and users of knowledge exchange platforms have a particular interest in maximizing "high-quality" content provisions, it would be fruitful to explore tools for detecting high-quality contributions quickly and which behavioral motives play a significant role in motivating or discouraging such "high-quality" contributors early to keep them engaged on the platform over time.

In conclusion, investigating heterogeneities in social interactions, users' motivations, and underlying mechanisms to facilitate beneficial exchange and continued user engagement offers a promising avenue for future work in this field. Our findings underscore the importance of peers' early activity for users' follow-on platform engagement and persistence. Our results highlight the relative benefits of integrating individual members bottom-up versus establishing top-down norms for cohort interactions. Moving forward, designers of online knowledge exchange environments should leverage the potential of active peers by fostering directed interactions *early on*, to foster engagement and enhance the effectiveness and resilience of online communities.

# 3

# The Impact of Competition and Gender Composition on Creative Idea Generation and Selection

## 3.1 Introduction

R&D and innovation are key drivers of economic growth and prosperity (Aghion et al., 1998; Acemoglu et al., 2018). As innovative ideas become increasingly difficult to find (B. F. Jones, 2009; Bloom et al., 2020), societies should enable all individuals with the potential to innovate. However, women remain underrepresented in innovative professions such as inventors, entrepreneurs, and scientists (Bechthold et al., 2021). Even among women who entered innovative professions, their measurable, innovative activity or output is often lower than their male counterparts — a disparity known as the *gender innovation gap*. For instance, women represented 28% of the US science and engineering workforce in 2021.[1] In contrast, the share of women on granted USPTO patents was 10.5% in the same year.[2] Looking at start-ups, female founders represented 23% in the US in 2019.[3] The share of the share of venture capital (VC) funding acquired in the same year by founding teams including at least one woman was 14.5%, and by all-female US startups was only 2.6%.[4] As a result, a significant portion of innovative potential may remain untapped.

Understanding the frictions women in the knowledge industries face is necessary for overcoming them.[5] Previous literature has focused primarily on explanatory factors of why women may not select into such innovative environments (e.g., Niederle & Vesterlund, 2007; Griffith, 2010; Rocha & Van Praag, 2020; Hoisl, Kongsted, & Mariani, 2023). Yet, the innovative processes and behaviors within such environments are not well-understood. Does the often competitive and male-dominated nature of innovative environments discourage women who have already entered them from in-

---

*For valuable comments, I am grateful to Valeria Burdea, Silvia Castro, Lena Greska, Dominik Grothe, Timm Opitz, Nathalie Roemer, Marina Schroeder, seminar participants at LMU Munich, Max Planck Institute for Innovation and Competition, and conference participants at ESA and the WSI Promotionswerkstatt. I used the generative AI tool ChatGPT for coding and editing the experimental instructions for clarity, form, and grammar. I used the AI-based writing assistance tool Grammarly to edit the text for spelling, grammar, and style. I reviewed and edited this content as needed and take full responsibility for the content of this study. This experiment has been pre-registered here: https://osf.io/efgzn/?view_only=4ecb7f4e028b49c6af30fd2ac6c78701.

[1] Source: https://ncses.nsf.gov/pubs/nsf23315/.

[2] https://www.weforum.org/agenda/2024/03/women-inventors-gender-gap-stem-patents/. USPTO: United States Patent Office.

[3] Source: https://www.eib.org/attachments/lucalli/support_for_female_entrepreneurs_en.pdf, last retrieved on Sept 12, 2024.

[4] Source: https://pitchbook.com/news/articles/the-vc-female-founders-dashboard, last retrieved on Sept 12, 2014.

[5] While this study focuses on women, it acknowledges the underrepresentation of other demographic groups by ethnic group, age, or socioeconomic status in innovation. See, e.g., for further reading: Gompers and Wang (2017) and Hofstra et al. (2020).

novating?

In this study, I investigate (1) how the competitive and male-dominated nature of innovative environments impact gender gaps in creative ideation output, (2) which dimensions of creative ideation - originality, flexibility, and validity - are impacted, and (3) shed light on men's and women's idea decisions when selecting ideas for a competition.[6]

Using a field experiment conducted online with white-collar professionals in the US, I randomly vary the gender composition of the environment and the incentives to establish a causal relationship between the separate and combined impacts of these features on creative ideation. The experiment consists of a short demographic survey, a main part, and a subsequent questionnaire. In the main part, participants perform a creative ideation task within a limited time over two rounds and select their self-perceived best idea after each round for evaluation. In particular, I use the "Unusual Uses of Objects"-task to capture key elements of business innovation, requiring participants to think creatively and repurpose existing resources for novel ideas. This task mirrors the flexible problem-solving skills essential in business for developing innovative products and solutions. They are compensated in a linear piece rate scheme plus an additional treatment-variant pay component and receive no performance feedback beyond the raw number of ideas generated per round before selecting ideas.

I find that women outperform men in this creative ideation task. Even under stark competition, women continue to excel by selecting ideas that are not only more original but also achieve higher overall creativity scores compared to men. Yet, these gaps do not result in gender disparities in selecting one's best idea for or winning a competition per se. Importantly, the gender composition of the competitor environment matters. Once gender is revealed, men tend to catch up, selecting better ideas. Looking at underlying channels, women exhibit lower overconfidence at baseline than men, but their confidence increases significantly in male-dominated, competitive environments. Women also perceive gender-balanced environments as more competitive than men. Together, these findings point towards women having some degree of sophistication about their comparative advantage in the creative ideation task and competitor environments of this study.

[6]Note that this work uses a binary differentiation of biological sex rather than a more complex differentiation of gender identities for simplicity and following the convention within the discipline of economics (Brenøe et al., 2022). At the same time, the author acknowledges a generally broader diversity of sexes (Fausto-Sterling, 2012) and genders (e.g., West and Zimmerman (1987)).

These findings offer implications for organizations, particularly as teamwork is ubiquitous in modern organizations and often contains an inherent notion of competition. Given women's strengths in creative ideation, their presence in male-dominated teams should be increased in knowledge-intensive jobs that require creative thinking and problem-solving. In light of gender differences in the reaction to different competitor environments, a one-size-fits-all competition may not produce optimal outcomes. This study did not find gender gaps in creative ideation and selecting one's best idea in gender-blind versus gender-balanced competitor groups, highlighting the importance of equal opportunities. Finally, recognizing how competitive environments affect confidence differently by gender, providing regular performance feedback could help calibrate employees' self-beliefs with their actual capabilities. In sum, organizations should acknowledge gender differences in creative processes and create environments that allow men and women alike to realize their creative potential and ultimately harvest the best innovative ideas.

This study's methodological approach offers two key advantages. It allows control over endogenous selection into specific environments by randomly assigning individuals to competitive and non-competitive settings with varying gender compositions, thus eliminating self-selection biases. The clean and simple experimental design also ensures a clear separation between the effects of competition and competitor composition when comparing creative performance. Additionally, using the same creative ideation task for all participants, the design provides a systematic way to quantify and compare the creativity of innovative ideas — measuring their originality, flexibility, and validity — across individuals and experimental conditions.

A decoupling of the innovation process from its embedded environment is vital for understanding its mechanics. First, the *innovation process* can be conceptualized as a four-stage process consisting of a first idea generation stage, a second idea selection stage, an idea evaluation stage, and a final idea implementation stage (Vahrs & Brem, 2015). It is naturally linked to creativity as innovation requires *creative ideation* at the onset (M. Baer & Frese, 2003; Laske & Schroeder, 2017). Creativity is commonly defined as the generation of ideas that are *novel* and *useful* for achieving a goal at hand (Amabile, 1996; Attanasi et al., 2021). Overall, the evidence on gender gaps in creativity from other disciplines, such as psychology or education research, is mixed. However, evidence from psychology suggests that, if anything, men perform better (Abraham, 2016; Hora et al., 2022). While analyzing incentivized actual decisions, the large experimental economics literature on incentives for creativity has rarely con-

sidered gender gaps explicitly (Attanasi et al., 2021).
Looking more at applied works and outcomes of creative ideation, Hoisl and Mariani (2017) find no gender gaps in the quality of inventions. Recent evidence from the science-of-science field suggests that research topics are different once more women enter a research field or university, respectively (Koning, Samila, & Ferguson, 2021; Truffa & Wong, 2022). Einiö, Feng, and Jaravel (2019) document innovator-consumer homophily, i.e., innovators tend to create products that are more likely to be purchased by consumers similar to them with respect to age, gender, and socioeconomic status. Taken together, this recent evidence hints at potential gender differences in ideation and selection. Yet, systematic evidence on this topic is scant.

*Innovative environments* share several common features. Namely, they are inherently *competitive*, illustrated by winner-takes-it-all priority races in science or competitions like hackathons or start-up pitch contests are actively fostered (Hill & Stein, 2019). Besides, these environments are mostly *male-dominated* (STEMM) fields. Research indicates that women behave differently than men in competitive (Shurchkov, 2012; Iriberri & Rey-Biel, 2017) and male-dominated environments (Ulku-Steiner, Kurtz-Costes, & Kinlaw, 2000; Born, Ranehill, & Sandberg, 2022). Specifically, previous results find that men's presence affects women's decisions to take less risk (Booth, Cardona-Sosa, & Nolen, 2014), opt out of competition more frequently (Booth & Nolen, 2012; Hogarth, Karelaia, & Trujillo, 2012), avoid signaling high career ambitions (Bursztyn, Fujiwara, & Pallais, 2017), and exhibit a lower willingness to lead (Born, Ranehill, & Sandberg, 2022). Yet, the interplay between these two factors remains poorly understood.

This study relates to three strands of literature. First, it adds to the studies investigating behavioral gender differences in mixed- versus single-gender environments (e.g., Bursztyn, Fujiwara, & Pallais, 2017; Born, Ranehill, & Sandberg, 2022). I extend this line of work by providing evidence on how gender composition in competitions may impact the creative performance of men and women differently. My experiment allows me to cleanly separate a competition from a gender composition effect on a highly labor-market-relevant category of work: creative ideation.

Second, it contributes to a large scholarship on incentives for creativity (Attanasi et al. (2021) for a comprehensive review of the experimental literature). Among the studied monetary and non-monetary incentives, competition or tournament incentives have received notable attention (Erat & Gneezy, 2016; Laske & Schroeder, 2017; Bradler, Neckermann, & Warnke, 2019; Charness & Grieco, 2019; Gross, 2020; Kleine, 2021, e.g.). However, this evidence stems mostly from laboratory experiments and rarely discusses gender gaps explicitly. To add to this literature, I conduct my study in a highly relevant yet understudied population: a diverse group of individuals working in "knowledge-intense" jobs. Unlike traditional laboratory experiments that mainly rely on student samples, my study includes professionals from diverse industries, roles, and career stages. This approach provides a more accurate representation of the target population. This sample choice enhances the external validity of this study's findings and strengthens its applicability to other real workplace settings.

A third strand of work studies gender gaps in innovative knowledge work more broadly (e.g. Hoisl & Mariani, 2017; Koning, Samila, & Ferguson, 2021; Iaria, Schwarz, & Waldinger, 2024). So far, this literature has primarily considered the final output of innovative processes, such as scientific publications or patents. This evidence stems from observational data, which does not allow us to draw causal inferences due to endogenous selection and other confounding factors. I complement these works with nuanced micro-level evidence from a large online experiment on men's and women's creative ideation and idea selection - both critical early stages of the innovation process. To my best knowledge, only Irlenbusch et al. (2024) have examined idea selection explicitly when examining path dependencies between knowledge provision and subsequent creative ideation and idea selection in a German student sample.

The remainder of this study is structured as follows. Section 3.2 describes the research design and hypotheses. Section 3.3 presents the experimental results. The final section 3.4 offers a discussion and conclusive remarks.

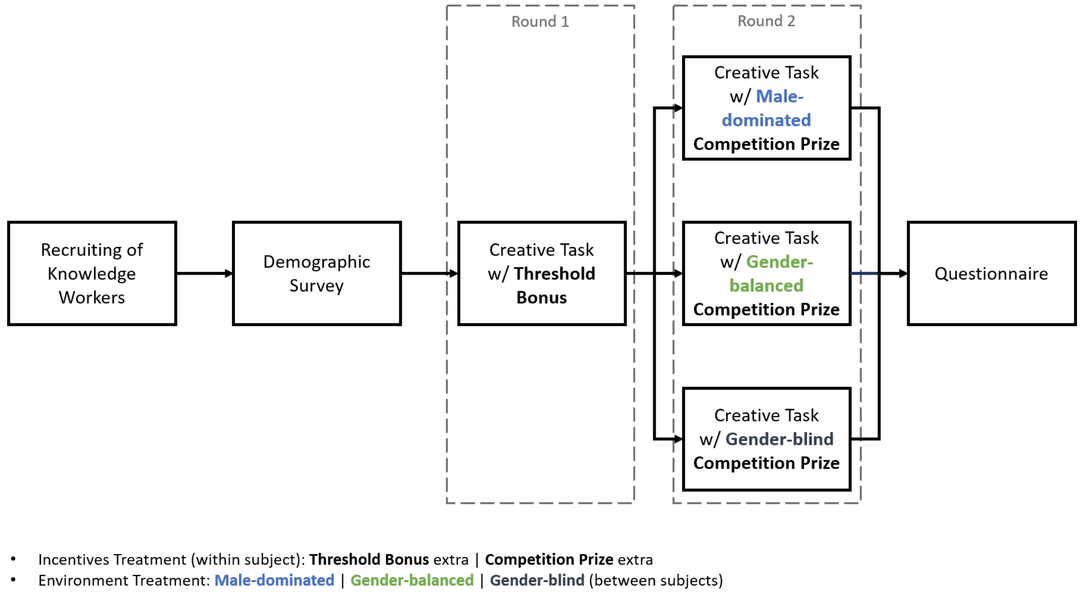
## 3.2 Experimental Design

### 3.2.1 Structure of the Experiment

**Overview** This experiment examines how (i) different incentive structures and (ii) the gender composition of the environment affect gender gaps in creative performance. The interplay of these two factors provides nuanced insights into creative performance in terms of ideation and idea selection.

The experiment consists of three parts. The first part is a short demographic survey used as input data for later. Then, participants encountered the main part, which consisted of two rounds in which participants performed a creative ideation task. They are compensated in a linear piece rate scheme plus an additional treatment-variant bonus component across rounds. The two rounds are presented in random order. Participants do not receive feedback about their task performance during the experiment. In the final part, they complete an extensive questionnaire. Figure 3.1 presents the experimental design.

**Figure 3.1:** Overview




**Demographic Survey** To begin, participants are asked to provide information about demographics: biological sex, age, gender, ethnic group, final high school and university education, field of study, and occupation details. The information about participants' sex is used as input data for displaying the applicable competitor group information later in the experiment.

**Creative Task: Round 1** Participants individually perform the creative ideation task for three minutes and are informed about the incentives beforehand.

Their incentives are to achieve a high score for their creative output that translates linearly into monetary pay (see Sections 3.2.1 and 3.2.3 for details). They are executing the task for three minutes per round. Specifically, I use the *unusual uses* task (Bradler, Neckermann, & Warnke, 2019), where participants are asked to generate as many unique and innovative uses for an ordinary object (e.g., a paper clip) as possible within a limited time (see Section 3.2.2 for details).

After the task execution period, they select one of the previously generated ideas they believe to be the best, i.e., scored the highest without knowing the score. They'll receive a *bonus* payment if the idea score for this selected idea surpasses a predetermined score *threshold*.[7] This bonus component in both rounds reflects the fact that multiple ideas might be generated in an innovative process. Still, only one idea may be executed and evaluated due to time and resource constraints. Thus, Round 1 of the experiment measures task-related performance under a given incentive scheme, thereby allowing for the exploration of potential underlying gender differences in performance.

Once participants have selected their ideas, I elicit their beliefs about the score of this selected idea ($CorrectGuess$).[8] I incentivize these beliefs with a small payment of £ 0.25 for a correct report.[9] Lastly, they provide a short reason for selecting a particular idea in an open text field.

**Creative Task: Round 2** Again, participants perform another instance of the same creative ideation task for another three minutes. Before the task period starts, they are informed about the incentives in place: just as in Round 1, they are incentivized to achieve a high score for their creative output as they are compensated in a linear piece rate scheme. Additionally, they are randomly paired in groups of six competitors. Again, they should select their best idea after the task period. They can only earn a *competition prize* if their selected idea scored highest among their competitor group. After finishing this part, participants report their beliefs about the score of their selected idea. I incentivize these beliefs with a small payment of £ 0.25 for a correct report.[10] Lastly, they provide a short reasoning for selecting a particular idea in an open text field.

**Questionnaire** Lastly, an extensive questionnaire solicits the following information:

[7] Threshold was calibrated to the upper sextile of scores in the pilot study. The objective behind this calibration was to keep the underlying data-generating process of the participants' objective function as similar as possible.

[8] Previous work has found robust evidence for gender gaps in overconfidence, particularly in stereotypically male (e.g., math) tasks. See, for example, Niederle and Vesterlund (2011), Buser et al. (2020), and Bandiera et al. (2022). As it is ex-ante not clear whether these gaps also prevail in creative tasks, subjective confidence is measured in each condition of the experiment.

[9] The incentive structure in Round 1 is: $Payoff_1 = PieceRate * CreativityScore_1 + Bonus * (SelectedIdea_1 > Threshold) + 0.25 * (CorrectGuess_1 = 1)$.

[10] The incentive structure in Round 2 is: $Payoff_2 = PieceRate * CreativityScore_2 + Bonus * (SelectedIdea_2 = WonCompetition) + 0.25 * (CorrectGuess_2 = 1)$.

- Stereotype perceptions of the creativity task (Coffman, 2014).
- Ex-post motivation to perform well and enjoyment of the task.
- Subjective level of task difficulty.
- Experience with this and similar divergent thinking tasks and current and past pursuit of creative hobbies.
- Risk attitudes (Falk et al., 2023).
- Perceived degree of competitiveness (of the environment) and preference for the pay scheme in Round 1 or Round 2 (Buser, Niederle, & Oosterbeek, 2024).
- Participants' reasoning on (i) what their strategy was to obtain a high score and (ii) the reason why they would prefer the competitive or non-competitive pay scheme in two open text fields.

In doing so, the experiment can shed light on several underlying mechanisms through which the task and environment may impact creative performance. More precisely, I explore changes in overconfidence and perceived competitive pressure across competitor groups and genders as potential mechanisms driving gaps in creative idea generation and idea selection.

Aditionally, I ask them about parallel activities during the study and cheating on the creative ideation task. Appendix C.3 contains the full experimental instructions.

### 3.2.2 Task and Performance Measurement

**Unusual Uses Task** This chapter employs the unusual-uses-task to measure creative ideation (Torrance, 1968; Bradler, Neckermann, & Warnke, 2019). In each part, participants are presented with an ordinary object, e.g., a paper clip, for which they are asked to generate as many unique and unusual uses (innovative "ideas") in a limited time as possible. The unusual uses are not restricted to a particular size of the object or the usage of one piece of the item only. For instance, a giant paper clip or 100 paper clips together could be proposed for a new, unusual use. Figure C.3 illustrates several examples of the submitted ideas by object. Each idea can be scored in the creative dimensions $Validity$, $Originality$, and $Flexibility$. These individual scores all generated ideas per experimental sound can be summed up. Hence, a clean, quantitative metric of the aggregate creative output can be calculated. The next paragraph details the measurement of creative idea dimensions and the scoring rule.

**Idea Measurement and Payoff** The creative ideation output, meaning participants' ideas for unusual uses of conventional objects, is evaluated using three standard categories: $Validity$, $Originality$, and $Flexibility$. These reflect important dimensions of creativity that can be aggregated to participants' overall $CreativityScore$ per round. Similar to Bradler, Neckermann, and Warnke (2019), the categories mentioned above are defined as follows:

$Validity$ measures if an idea is feasible, i,e., the *usefulness*, within the framework of defining innovative ideas (Amabile, 1996). It is subject to external evaluators' judgment.[11] Ideas can receive 0 or 1 point. Ideas receive 1 point if they are feasible and not the typical or common usage of an object. For example, "for writing" would be the common purpose of a paper sheet.

$Originality$ can be measured using the statistical infrequency of a stated idea[12] across the entire sample. This dimension refers to the *novelty* aspect of innovative ideas (Amabile, 1996). They receive 0 points in this category if the idea is mentioned more than 10%, one point if the idea is stated less than 10%, two points if it is stated less than 2%, and three points if it is stated less than 0.5% in the participant pool per common object in a given part. This applied scoring rule in the $Originality$ dimension should work as a disincentive to rely on technological means for generating ideas such as large language models, e.g., ChatGPT, which commonly produce responses probabilistically.[13]

$Flexibility$ mirrors participants' variety of ideas by attributing one point for each new "category" a stated idea falls into, e.g., jewelry or tool.[14]

Additionally, I only count ideas that are $valid$, implying that the indicated use is feasible in practice. This is a necessary condition for any idea to be counted favorably into participants' $CreativityScore$.

The subsequent equation describes the construction of the $CreativityScore$ for $i =$

[11] Research assistants blind to this study's research questions rated the ideas as valid or not. For this task, they were given a short description of the experimental task, and all submitted ideas per object and round were displayed, as well as the corresponding (optional) use case explanations.

[12] The submitted ideas will be normalized using state-of-the-art text pre-processing methods such as lemmatization and stemming. The purpose is to "standardize" the terminology.

[13] Source: https://platform.openai.com/docs/advanced-usage/parameter-details, last retrieved on Sept 12, 2024.

[14] Research assistants blind to this study's research questions categorized the ideas. For this task, they were given a short description of the experimental task and displayed all submitted ideas per object and round as well as the corresponding (optional) use case explanations.

$1, 0...., n$ ideas generated in Round 1 or Round 2 of the experiment:

$$\sum_{i=1}^{n} CreativityScore_i = \sum_{i=1}^{n} Validity_i \times (Validitidy_i + Originality_i + Flexibility_i) \tag{3.1}$$

Scoring in *No-competition* works as follows: a creativity score and bonus are awarded if the score is larger than a threshold (pre-calibrated). In *Competition*, a creativity score and prize are awarded if the score of the selected idea is the highest in the competition among 6 participants. The two rewards are equivalent in expected payoffs, keeping the underlying data-generating process of participants' objective function constant between the $No-competition$ and $Competition$ conditions.

### 3.2.3 Treatments and Conditions

In a 2x2x3 mixed design, this study varies the *Incentives* within subjects, while gender and social *Environment* vary between subjects in which the creative ideation task is performed.

**Incentives** The Incentives-treatment randomly varies the compensation for producing creative output between the two main parts of the experiment. They are compensated in a linear piece rate scheme plus an *additional treatment-variant bonus component* varying the incentives. In $No-Competition$ (Round 1), their objective is to achieve a high score for their creative output that translates linearly into monetary pay. This score attributes points for each mentioned dimension of creativity to feasible ideas. After the task period, they selected their most promising idea for the bonus, i.e., the idea they believed to have scored highest. They'll receive a *bonus* if the idea score for this idea surpasses a certain *threshold*.
In $Competition$ (Round 2), participants are paired in groups of six and informed about the competition aspect. Specifically, in addition to paying according to their score for the produced ideas, participants each select their most promising idea to enter the competition pool. They can win a substantial *prize* if their idea has the highest score within their competitor group.
The bonus in Round 1 and the prize in Round 2 are equally high, with £ 0.50.

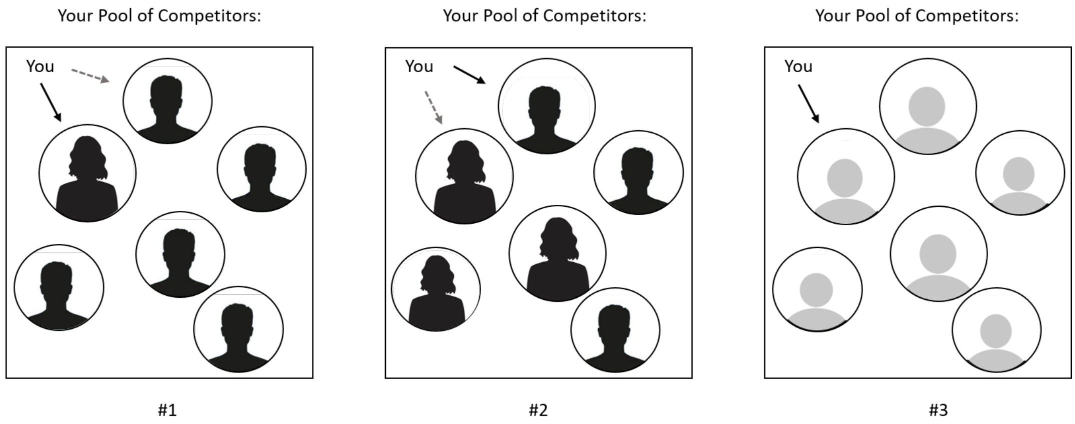
**Environment** The *Environment* treatment randomly assigns participants to competitor groups of different gender compositions: in *MaleDominatedEnv* (five men, one woman), they compete in groups of five men and one woman; *GenderBalanceEnv* (three men, three women), they compete in groups of three men and three women, and in *GenderBlindEnv* (five competitors), they have no information about the gender of their competitors.[15]

The set of competitors in each condition will be shown to participants as one of the three images illustrated in Figure 3.2.[16] The gender of each participant is revealed in *MaleDominatedEnv* and *GenderBalanceEnv*. Hence, as in every treatment condition, groups of six individuals compete, and participants can notice the gender composition of their *Environment* condition. Table C.3 provides an overview of participants' demographic information and relevant questionnaire items by *Environment* treatment condition. Overall, the allocation of participants to conditions is balanced across conditions. With respect to study field background, the only statistically significant differences are that there are significantly more individuals with a 'Business/Law & Public Administration' background in the *GenderBlindEnv* than the *GenderBalanceEnv* condition (p-value=0.034) and fewer individuals with an 'Education Studies'-background in *GenderBlindEnv* compared to *MaleDominatedEnv* (p-value=0.008) and *GenderBalanceEnv* (p-value=0.086). There are significantly fewer participants with a 'Social Science & background in *GenderBlindEnv* than *MaleDominatedEnv* (p-value=0.043). For the ethnic groups, there are signif-

[15]All participants receive the following information together with the displayed image of their competitor pool: *These five other participants have been recruited based on the same filters as you via Prolific: (1) Living in the US, (2) working age population (20 to 65 years), (3) at minimum high school diploma or equivalent. For data protection reasons, we are using anonymized icons. Below is an image of your group of competitors: [image #1 | #2 | #3 of Figure 3.2].*

[16]The competitor groups were constructed from a pretest as follows: The pretest was conducted in May 2024. 120 participants were recruited based on the same filters as the main study sample. They performed the same unusual uses of objects task for four rounds with the everyday objects rubber band, brick, mug, and paper clip. From the data on the objects chosen for the main study, rubber band, and brick, 6 participants were randomly chosen for the *Gender-Blind* competition. Then, three men and three women were randomly chosen for the *Gender-Balanced* competition. Finally, one woman and five men were randomly chosen for the *Male-Dominated* competition. I chose a valid idea for the competition at random from these individuals. Based on matching sex, a main study participant randomly replaced one out of the six individuals in the competitor group of their allocated *Environment* condition. To determine whether the main study participants won the competition, I constructed a binary indicator that equals one if their selected idea was (amongst) the idea(s) with the highest score in their competition group. This approach allowed competitors' performance within a competition group to be constant across participants of the same sex under the condition of the *Environment* treatment. Thereby, noise to a fully random competitor group composition is reduced. The pretest participants were excluded from the possibility of being recruited for the main study.

**Figure 3.2:** Competitor Pool Visualization by Environment Condition




*Notes*. #1 shows the *MaleDominatedEnv* condition, #2 shows the *GenderBalanceEnv* condition, and #3 shows the *GenderBlindEnv* condition. The grey arrows in images #1 and #2 show the counterfactual sex a participant could have in this condition.

icantly less individuals who self-identify as belonging to 'Other & Mixed'[17] ethnic groups in *GenderBlindEnv* than *MaleDominatedEnv* (p-value=0.002) and *GenderBalanceEnv* (p-value=0.074), although the overall share of this group is low. I include demographic controls to account for these differences in the parametric analyses.

### 3.2.4 Procedures

The experiment was conducted online in July of 2024. Through the platform *Prolific Academic*[18], participants are recruited who (1) are residents of the US, UK, or Ireland, (2) have at least a Bachelor's degree, (3) are of working age between 25 to 60, (4) have an approval rate of minimum 95% from previous study participation on the platform. Table C.3 depicts the number of participants per experimental condition and gender. The experiment was programmed using the survey software Qualtrics. The median completion time was 24 minutes, and participants earned a median pay of £ 3.44 ($\sim$ USD 4.52).[19] The study included mandatory comprehension questions to ensure high attention to and comprehension of the experimental instructions.

[17]The other ethnic group options were 'White/Caucasian', 'Black/African American', 'Asian/Pacific Islander', and 'Hispanic/Latino'.

[18]Link to the online labor market platform: https://www.prolific.com/about. Last retrieved on Jan 31, 2024.

[19]The exchange rate was calculated using the following converter provided by Forbes: https://www.forbes.com/advisor/money-transfer/currency-converter/gbp-usd/, last retrieved on Aug 30, 2024.

**Means to combat cheating online** Several means are taken to limit cheating possibilities in the online data collection setting. First, all parts are timed to $n$ minutes. Upon beginning with one part, i.e., a new object that appears, participants have to submit every idea for unusual use of this object to a new page. Each page's time is shown is restricted to 30 previously calibrated seconds. Participants can be faster at submitting more ideas but are auto-forwarded if they are too slow. By capturing the time stamps on each page, unusual patterns might hint at cheating or using generative AI tools (i.e., very slow in the beginning and then "firing" 20 ideas) and potentially exclude such observations.

Second, participants are strongly incentivized via double scoring on the creativity dimension $Originality$ of ideas. This scoring rule rewards rare ideas and penalizes very frequent ideas. Hence, this should disincentivize participants from choosing the publicly available baseline everyone can access. Even though probabilistic, e.g., ChatGPT shows similar output per object in repeated iterations or across participants.

Third, before starting the study, participants are requested to devote their full attention, refrain from using other devices, and remain in the study browser tab as their initial consent to the study terms and conditions.

Finally, at the end of the questionnaire, I ask participants if they had switched browser tabs or used other tools, e.g., Google or ChatGPT, when their response could have no consequence on their payoff.

### 3.2.5 Hypotheses about Behavior in the Experiment

The experiment is designed to investigate whether there are gender gaps in creative ideation and idea selection at baseline (Round 1) and to disentangle additional effects of introducing competition incentives under varying competitor gender compositions.

Against the backdrop of mixed previous evidence regarding gender differences in creativity (J. Baer & Kaufman, 2008; Abraham, 2016; Hora et al., 2022), I test the following first hypothesis:

**Hypothesis 1.** *There are no gender differences in performance in a creative ideation task under piece-rate incentives (Round 1, linear pay scheme plus bonus).*[20]

[20]In the event of finding any average performance difference between men and women in Round 1), I hypothesize that a gender gap is smaller under piece-rate than competition incentives. Put differently, there is no interaction between gender and monetary piece-rate incentives.

Previous research has robustly shown that men outperform women in various competitive settings (Gneezy, Niederle, & Rustichini, 2003; Shurchkov, 2012) and that they shy away from competitions more often when having the choice (Niederle & Vesterlund, 2007). One of the explanations that has been put forward and received much scholarly attention is gender gaps in (over-)confidence. While some studies find that everyone exhibits overconfidence, i.e., no difference across gender (Bandiera et al., 2022), other works indicate stronger male overconfidence, particularly in domains regarded as "traditionally male" (Barber & Odean, 2001; Croson & Gneezy, 2009; Marianne, 2011; Exley & Nielsen, 2024). Another factor contributing to the observed gender gaps in entering and performing in competitive environments may be differing perceptions of the competitive pressure therein (Cai et al., 2019). Thus, I hypothesize:

**Hypothesis 2.** *Men perform better on the creative ideation task under competition incentives than women on average.*

In addition, the gender composition of the competitor group could impact men's and women's idea generation and idea selection differently, thereby exacerbating gender performance gaps. Due to their higher preference for competition, men might be particularly motivated in male-dominated settings and, thus, exert higher effort (Gneezy & Rustichini, 2004). At the same time, women's performance may be hampered in such a setting. They could perform worse under the perception of confirming a stereotype about women being less capable in a task or field, i.e., due to the so-called *stereotype threat theory* (Spencer, Steele, & Quinn, 1999; Inzlicht & Ben-Zeev, 2000; Hoyt & Murphy, 2016). Such a perception may cause a feeling of stress, which in turn can impair performance. In the context of this experiment, this could lead women to generate and select safer, i.e., less original ideas in a male-dominated competition.

**Hypothesis 3a.** *Men perform better on the creative ideation task under competition incentives when being in a male-dominated environment compared to a gender-balanced or gender-blind environment on average.*

**Hypothesis 3b.** *On average, men select better ideas than women under tournament incentives than piece-rate incentives or flat pay.*

No hypotheses are formulated about which idea dimensions drive the gender gaps in outcome measures; there is only a difference on average. It remains an empirical question of which dimension(s) of creativity, namely originality, validity, or flexibility, produce gender gaps under tournament incentives and different social compositions of the competitor environment.

## 3.3 Main Results

### 3.3.1 Creative Idea Generation without Competition

In the following, I shed light on whether there are gender differences in creative idea generation under monetary incentives *without* a competition. For this purpose, I explore differences in the Round 1 performance across genders. Figure C.1 shows the average total $CreativityScore$ and the idea dimensions $Valdity$, $Flexibility$, and $Originality$.

The CI-Barplots in Figure C.1 reveal that women, on average, attain significantly higher scores in total $CreativityScore$ and in all idea dimensions $Validity$, $Flexibility$, and $Originality$. The non-parametric rank-sum tests confirm the statistical significance of these differences for the overall $CreativityScore$ ( p-value=0.0001), the creativity dimensions $Validity$ (p-value<0.0001), $Flexibility$ (p-value=0.0001), and $Originality$ (p-value=0.0002).

Given the substantial variation observed in the 95% Confidence Intervals in Figure C.1, I further investigate the distributions of creative performance measures by gender, as depicted in Figure C.2. The analysis reveals a rightward shift in the distribution of women's performance relative to men's in the overall $CreativityScore$ and across all creative dimensions. I conduct non-parametric Kolmogorov-Smirnov tests to assess the statistical significance of these observed differences. The results indicate significant differences in the distribution of performance between men and women, with p-values of less than 0.001 for the total $CreativityScore$, 0.001 for $Validity$, less than 0.001 for $Flexibility$, and 0.006 for $Originality$.

Hence, I reject Hypothesis 1, which states that there are no gender differences in creative idea generation under monetary incentives without competition in the context of the employed task.

**Result 1:** Under non-competitive, monetary incentives, women outperform men in a creative ideation task.

### 3.3.2 Creative Idea Generation and Selection with Competition

**Estimation Strategy** To estimate the effect of competition incentives and gender composition of the environment on outcome variables in round 2, I use Ordinary Least Squares (OLS) regressions according to versions of the subsequent equation 3.1:

$$\begin{aligned} Y_{j2} = \beta_0 &+ \beta_1 Female_j + \beta_2 MaleDomEnv_2 + \beta_3 GenderBalanceEnv_2 \\ &+ \beta_4 Female \times MaleDomEnv_2 + \beta_5 Female \times GenderBalancedEnv_2 \\ &+ \beta_6 Y_{j1} + \beta_7 Sequence_j + \beta_8 \boldsymbol{X}_j + \varepsilon_j \end{aligned} \tag{3.1}$$

where $Y_{j2}$ is the respective outcome for participant $j$ in round 2. For idea *generation*, the performance in the creative task is measured with the overall $CreativityScore_{j2}$ as the primary outcome and $ValidityScore_{j2}$, $FlexibilityScore_{j2}$, and $OriginalityScore_{j2}$ as secondary outcomes. Conversely, for idea *selection*, the outcomes of interest are the total $CreativityScoreSelected$, $OriginalityScoreSelected$, $BestIdeaSelected$, and $CompetitionWon$ (see Section 3.2.2 for details on the construction of score outcome variables). The two latter outcome variables are binary indicators that equal one if the best idea was selected in terms of the highest score, and the selected idea won the competition, respectively.

Turning to the explanatory variables, $Female$ is an indicator that equals one if a participant identifies as female and zero otherwise. $MaleDomEnv$ is an indicator equal to one if the participant was in a male-majority environment during the competition in Round 2 and zero otherwise. Similarly, the indicator $GenderBalanceEnv$ equals one if the participant was in a competitor group with an equal number of men and women. The $GenderBlindEnv$ is the omitted baseline environment category. $Y_{j1}$ is a participant $j$'s lagged respective performance outcome of round 1. It is a proxy for participants' creative abilities at baseline to isolate the treatment effects of interest on the respective outcome. Besides, I include several additional control variables. $Sequence$ is an indicator that equals one if the sequence of everyday objects is "brick" in round 1 and "rubber band" in round 2 and zero if this order is reversed to control for potential order effects, e.g., learning over time or fatigue. $\boldsymbol{X}$ is a vector of demographic controls including age, age$^2$, and field of study. Finally, $\varepsilon_{it}$ is the residual error term. The model employs heteroscedasticity-robust standard errors.

### 3.3.3 How do Competition Incentives and the Gender Composition of the Competitor Environment Impact Idea Generation?

The presentation of results starts with the first stage of an innovation process, creative idea generation. Figure C.3 depicts the overall mean gender differences in the experiment's Round 2 Competition. Again, women perform slightly better than men in the ideation task. Looking at the corresponding performance distribution overall and by idea dimension in Figure C.4, it becomes evident that the slight performance gap across genders occurs along the entire performance distribution.

When breaking down idea generation performance by competitor environment in Figure C.5, it becomes visible that women perform well across competitor groups on average. Men, on the other hand, appear to excel particularly in $Gender-Balanced$ competitions overall and across idea dimensions. The 95% confidence intervals in the bar plots are relatively wide, suggesting considerable variation in performance for both genders.

Next, I turn to parametric estimates to control for additional factors. Table 3.1 contains the estimated results from OLS regression models of treatment effects on the overall $CreativityScore$ (Column 1) as well as the creative idea dimensions $Validity$ (Column 2), $Flexibility$ (Column 3), and $Originality$ (Column 4). All specifications include demographic controls, Round 1 performance, and task sequence controls.

Overall, the estimates indicate no evidence for a statistically significant impact of either treatment alone or in combination causing a shift in the average creative idea generation, compared to the baseline means in this chapter's context. Looking at performance gaps between men and women, there is no significant gender effect visible in the data as the coefficient for $Female$ remains statistically insignificant across all dimensions. Directionally, women score higher overall (p-value=0.91) and on average in the Validity (p-value=0.95) and Flexibility (p-value=0.34) dimensions (Columns 1-3). The coefficient estimate on the originality score is negative for the female indicator but very small in size (p-value=0.97).

None of the environmental indicators, *Male-Dominated* environment, and *Gender-Balanced* environment, significantly affect idea generation overall or across idea dimensions. For instance, the average performance overall appears to be higher (lower) in the gender-balanced (male-dominated) environment, with a p-value of 0.24 (0.76) compared to the gender-blind baseline environment.

Including the interaction between the environment and being female, the findings indicate that women's performance is higher, on average, in the male-dominated environment compared to the gender-blind one and, vice versa, men's performance is higher in the gender-balanced environment for the overall Score and the idea dimensions. For instance, Column 1 shows that the estimated coefficient for the interaction ($FemalexGenderBalancedEnv$) is -2.661 with a 95%-confidence interval (CI) of [-5.988, 0.66], indicating that women perform about 14% worse on the ideation in a gender-balanced competitor group compared to men in the gender-blind condition. The coefficient estimate is, however, marginally insignificant with p-value=0.114.

In summary, the results in Table 3.1 show no statistically significant gender gap or influence of competitor environment's gender composition on overall creative idea generation or its dimensions — $Validity$, $Originality$, and $Flexibility$. Consequently, Hypothesis 2 and Hypothesis 3 are rejected. These findings are qualitatively and quantitatively robust to using stricter, NLP-adjusted score measures that account for very close synonyms of submitted ideas, i.e., being stricter on the originality dimension (see Table C.6).

**Result 2:** Neither competitive incentives nor the gender composition of competitors contribute to gender differences in performance on a creative ideation task.

### 3.3.4 How do Competition Incentives and the Gender Composition of the Competitor Environment Impact Idea Selection and Winning a Competition?

Moving from the first to the second stage of an innovation process, I next present results from regression 3.1 on participants' idea selection choice and competition outcomes under varying competitor compositions in Table 3.2. All specifications include demographic controls, round 1 outcome equivalents, and task sequence controls.

After each round, participants saw an ordered overview of all the ideas they generated in this round. They were then asked to select their best idea (i.e., idea number and text) for a threshold bonus in round one and a competition prize in round 2. To shed light on details of the idea selection, I measure the selected ideas' overall Creativity Score (Column 1), its Originality Score (Column 2), whether the selected idea was

**Table 3.1:** Treatment Effects on Idea Generation in Round 2 - Total and by Idea Dimension

| | (1) Creativity Score | (2) Validity Score | (3) Flexibility Score | (4) Originality Score |
|---|---|---|---|---|
| Female (1/0) | 0.122 | 0.018 | 0.182 | −0.026 |
| | [-1.993,2.237] | [-0.506,0.541] | [-0.193,0.557] | [-1.328,1.276] |
| | (1.078) | (0.267) | (0.191) | (0.664) |
| MaleDomEnv (1/0) | −0.352 | −0.146 | −0.096 | 0.125 |
| | [-2.630,1.927] | [-0.709,0.417] | [-0.488,0.295] | [-1.526,1.276] |
| | (1.161) | (0.287) | (0.199) | (0.714) |
| GenderBalanceEnv (1/0) | 1.567 | 0.428 | 0.155 | 1.022 |
| | [-1.051,4.186] | [-0.251,1.107] | [-0.275,0.585] | [-0.606,2.650] |
| | (1.334) | (0.346) | (0.219) | (0.830) |
| Female x MaleDomEnv | 1.699 | 0.488 | 0.090 | 1.054 |
| | [-1.542,4.940] | [-0.312,1.288] | [-0.479,0.659] | [-0.925,3.032] |
| | (1.652) | (0.408) | (0.290) | (1.008) |
| Female x GenderBalanceEnv | −2.661 | −0.644 | −0.463 | −1.609 |
| | [-5.988,0.667] | [-1.499,0.211] | [-1.037,0.111] | [-3.645,0.427] |
| | (1.696) | (0.436) | (0.293) | (1.038) |
| Dep. Variable Mean | 18.697 | 4.753 | 3.865 | 10.101 |
| Demographic Controls | Yes | Yes | Yes | Yes |
| Sequence & Round 1 DV | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.299 | 0.321 | 0.408 | 0.229 |
| N | 1076 | 1076 | 1076 | 1076 |

**Notes.** The dependent variable in Column (1) is the total $CreativityScore$; in Column (2), it is the $ValidityScore$; in Column (3), the $FlexibilityScore$ indicates if they selected their highest scoring idea for Round 2 competition; in Column (4), $OriginalityScore$. See Section 3.2.2 for a construction of the performance measures. $Female$ indicates that a participant's sex is female. $MaleDomEnv$ indicates that a participant competes in a male-majority environment in Round 2. $GenderBalanceEnv$ indicates that a participant competes in an environment with equally many men and women. The $GenderBlindEnv$ is the omitted baseline environment category. $Sequence$ indicates that the creative task used "brick" as an object in round 1 and "rubber band" in round 2; it is 0 if the object order is reversed. *Round 1 DV* is the corresponding performance outcome of Round 1. *Demographic Controls* include age, age$^2$, and field of study. 95% confidence intervals in squared brackets. Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

(amongst) their best idea(s) (Column 3)[21], and finally, whether their selected idea was best within their competitor group and won the competition (Column 4).

Starting with whether there are gender gaps at the idea selection stage per se, the estimates in Table 3.2 suggest that women tend to select ideas that score 19.5% higher overall (p-value=0.007) and that are 17.8% more original in the *Gender-Blind* baseline condition (p-value=0.007), as reported in Column (1) and (2), respectively. These economically large effects are statistically significant at the 1% level. Given the scoring rule applied to ideas in this study, these results suggest women are generally more aware of how rare their selected ideas were. However, this gender gap in the selected idea's score does not translate into significant differences in the selection of one's best idea (p-value=0.237) or winning the competition (p-value=0.246) reported in Columns (3) and (4), although both $Female$-coefficients are positive.

**Result 3:** Women select more original and higher-scoring ideas than men.

**Result 4:** There are no gender differences in selecting one's best idea and winning a competition per se.

The gender composition of the competitor environment appears to have a limited impact at the idea selection stage. Only being in a *Male-Dominated* environment decreases the likelihood of winning the competition by 11 percentage points (p-value=0.035), whereas being in a *Gender-Balanced* environment has no significant effect on any measured outcome. As competitor groups were chosen randomly, the mere effect of being in a *Male-Dominated* environment on winning the competition is likely an artifact of this allocation.

When looking at interactions between being $Female$ and in different competitor environments, the estimated coefficients of the idea selection outcomes (Columns 1-3) are negative in both conditions where competitor gender is visible. Compared to women in the *Gender-Blind* baseline condition, women select ideas of about 10.7 percentage points lower scores in the *Male-Dominated* competition on average (p-value=0.093). This effect is statistically significant at the 10% level (Column 1). Despite selecting ideas with a lower score, women are still 54 percentage points more likely to win the competition in this environment than the omitted *Gender-Blind*

[21] The plural refers to the possibility that participants could have several ideas with the same (highest) score.

baseline, statistically significant at the 1-percent level (p-value<0.001) in Column (4). This is likely, however, a mechanical effect because the quality of ideas is still higher if they only select about 10 percentage points lower-quality ideas.

**Result 5:** Women select inferior ideas compared to men in male-dominated competitor environments, whereas this difference is not observed in gender-blind environments.

Conversely, the estimates in the *Gender-Balanced* environment, statistically significant at the 10-percent level (p-value=0.16), indicate that women select 14% less original ideas than in the *Gender-Blind* environment (Column 2). However, this differential choice does not translate to significant consequences for selecting one's best idea (p-value=0.997) or winning the competition (p-value=0.417) noted in Columns (3) and (4), respectively.

Taken together, these results point to nuanced dynamics between gender, competitor environment, and idea selection. Hence, hypothesis 4 is partly rejected. These findings are qualitatively and quantitatively robust to using stricter, NLP-adjusted score measures that account for very close synonyms of submitted ideas, i.e., being stricter on the originality dimension (see Table C.7).

### 3.3.5 Underlying Channels

In the following, I investigate whether participants' (over-)confidence and perceptions about the degree of group competitiveness depend on their gender and the gender composition of the competitor environment.

#### Change in Overconfidence across Rounds

Table 3.3 presents the results from three OLS regression models estimating the effects of gender and the composition of the competitor environment on participants' changes in (over-)confidence of their selected idea's score (standardized outcome). Overconfidence is calculated as the difference between a participant's $belief$ of their selected idea's score minus the $actual$ score of their selected idea. A value $> 0$ indicates overconfidence, and vice versa; a value $< 0$ is underconfidence. Conversely, for accurate beliefs matching their actual scores, this difference is 0. These beliefs were

**Table 3.2:** Treatment Effects on Idea Selection and Competition Outcome

| | (1)<br>Score<br>Selected Idea | (2)<br>Originality<br>Selected Idea | (3)<br>Best Idea<br>Selected | (4)<br>Competition<br>Won |
|---|---|---|---|---|
| Female (1/0) | 0.464*** | 0.321*** | 0.064 | 0.062 |
| | [0.128,0.800] | [0.090,0.553] | [-0.042,0.169] | [-0.043,0.167] |
| | (0.171) | (0.118) | (0.054) | (0.053) |
| MaleDomEnv (1/0) | 0.201 | 0.171 | 0.020 | −0.110** |
| | [-0.143,0.545] | [-0.060,0.403] | [-0.084,0.123] | [-0.212,-0.008] |
| | (0.175) | (0.118) | (0.053) | (0.052) |
| GenderBalanceEnv (1/0) | 0.194 | 0.182 | 0.020 | 0.008 |
| | [-0.160,0.548] | [-0.056,0.420] | [-0.084,0.125] | [-0.095,0.112] |
| | (0.180) | (0.121) | (0.053) | (0.053) |
| Female x MaleDomEnv | −0.409* | −0.259 | −0.037 | 0.540*** |
| | [-0.886,0.068] | [-0.580,0.062] | [-0.184,0.110] | [0.415,0.666] |
| | (0.243) | (0.164) | (0.075) | (0.064) |
| Female x GenderBalanceEnv | −0.335 | −0.298* | 0.000 | −0.061 |
| | [-0.801,0.132] | [-0.617,0.022] | [-0.148,0.147] | [-0.207,0.086] |
| | (0.238) | (0.163) | (0.075) | (0.075) |
| Dep. Variable Mean | 3.348 | 1,803 | .478 | .494 |
| Demographic Controls | Yes | Yes | Yes | Yes |
| Sequence & Round 1 DV | Yes | Yes | Yes | Yes |
| $R^2$ | 0.058 | 0.046 | 0.014 | 0.167 |
| N | 1076 | 1076 | 1076 | 1076 |

**Notes.** The dependent variable in Column 1 $ScoreSelectedIdea$ is participants' overall $CreativityScore$ of the selected idea ranging from 0 to 5; in Column (2), it is their $OriginalityScore$ of their selected idea ranging from 0 to 3; in Column (3) $BestIdeaSelect$ indicates if they selected their highest scoring idea for the Round 2 competition; in Column (4), $CompetitionWon$ indicates if their selected idea. $Female$ indicates that a participant's sex is female. $MaleDomEnv$ indicates that a participant competes in a male-majority environment in Round 2. $GenderBalanceEnv$ indicates that a participant competes in an environment with equally many men and women. The $GenderBlindEnv$ is the omitted baseline environment category. $Sequence$ indicates that the creative task used "brick" as an object in round 1 and "rubber band" in round 2; it is 0 if the object order is reversed. *Round 1 DV* is the DV's equivalent of Round 1. *Demographic Controls* include age, age$^2$, and field of study. 95% confidence intervals in squared brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

incentivized with a small bonus payment in both rounds of the creative ideation task.

The coefficient for being $Female$ is negative throughout and statistically significant across all three models ($p < 0.01$ in Model 1, $p < 0.05$ in Models 2 and 3). Hence, the results indicate that women display significantly lower overconfidence in Round 2 (i.e., with competition) than men in the *Gender-Blind* baseline, even when controlling for prior overconfidence, performance, and the task sequence (Column 3). The magnitude of this effect ranges from $\sigma$ -0.291 (29.1% of a standard deviation) with a 95%-CI of [-0.495 $\sigma$,-0.088 $\sigma$] in Column 1 to -0.242 (24.2 % of a standard deviation) with a 95%-CI of [-0.441 $\sigma$ ,-0.043 $\sigma$] in Column 3 on average.

**Result 6:** In anonymous, gender-blind competitor environments, women exhibit lower levels of overconfidence compared to men.

In both environments, *Male-Dominated* and *Gender-Balanced*, where the gender of competitors is revealed, the coefficients are negative throughout all estimated models. In the most conservative specifications, including all controls, the estimated coefficients of the environment indicators are $\sigma$ -0.124 with a 95%-CI of [-0.324$\sigma$,0.075$\sigma$] for *Male-Dominated* environment (p-value=0.222) and $\sigma$ -0.085 with a 95%-CI of [-0.282 $\sigma$,0.111 $\sigma$] for *Gender-Balanced* environment (p-value=0.392). As these coefficients are statistically insignificant, I cannot reject that the competitor environment of any particular gender composition per se has no direct effect on overconfidence.

The interaction between being $Female$ and being in a *Male-Dominated* environment is positive and statistically significant in all three models (p=0.039 in Column 1, p=047 in Column 2, p=0.055 in Column 3). The coefficient ranges from $\sigma$ 0.309 with a 95%-CI of [0.015 $\sigma$,0.602 $\sigma$] to $\sigma$ 0.277 [-0.005 $\sigma$,0.560 $\sigma$]. This result suggests that the negative impact of being a woman in a competitive setting on overconfidence is reduced, or even slightly reversed, within the context of a *Male-Dominated* environment during this creative ideation task. This finding hints towards women being aware of their comparative advantage in ideation amongst only male competitors.

**Result 7:** Women exhibit significantly higher levels of (over)confidence in male-dominated, competitive environments during a creative ideation task.

While the corresponding coefficients of the interaction between being $Female$ and being in a gender-balanced environment are also positive but insignificant with 16 percent of a standard deviation and a 95%-CI of [-0.116 $\sigma$,0.435 $\sigma$] in the most conservative model of Column 3 (p-value=0.256).

The results provide evidence for gender differences in overconfidence when competing in a creative ideation task. However, this effect is context-dependent on the respective competitor environment: While women appear significantly less overconfident than men in a highly anonymous, gender-blind competition, they adjust their perception when the gender composition of competitors is revealed.

**Perceived Group Competitiveness across Gender and Competition Environment**

The results presented in Table 3.4 examine the treatment effects on *Perceived Group Competitiveness* using OLS regression models. It was measured on a scale of 0 (not at all) to 5 (extremely) and displayed on the same page, announcing the competition incentives and competitor groups.[22] All models include task sequence controls. In Column 2, I also include controls for Round 1 performance and beliefs about the selected idea's score. In Column 3, I additionally control for demographic factors.

Across all specifications, the coefficients for the $Female$ indicator are negative and small, with -8.0% (95%-CI of [-0.287 $\sigma$,0.128 $\sigma$]), -5.7% (95%-CI of [-0.260 $\sigma$,0.146 $\sigma$]), and -3.9% (95%-CI of [-0.260 $\sigma$,0.146 $\sigma$]) of a standard deviation on average in Columns (1), (2), and (3), respectively, and statistically insignificant (p-value=0.713 in the most comprehensive model of Column 3). Hence, I cannot reject that there are no gender differences with respect to perceptions about competitiveness in the *Gender-Blind* baseline competition.

Similarly, the estimated coefficients for the *Male-Dominated* environment indicator are very small on average and not statistically significant with $\sigma$ 0.06 (0.6 percent of a standard deviation) around a 95%-CI of [-0.216 $\sigma$,0.227 $\sigma$] in the most comprehensive model in Column (3) (p-value=0.960). The coefficient estimates of the corresponding indicator for the *Gender-Balanced* indicator range from -12.4% of a standard deviation with a 95%-CI of [-0.342 $\sigma$,0.094 $\sigma$] in Column (1) (p-value=0.263) to -10.1% with a 95%-CI of [-0.316 $\sigma$,0.115 $\sigma$] in Column (3) (p-value=0.360) on average. These findings suggest that the revealed gender composition of competitors does not impact the perceived group competitiveness meaningfully compared to a *Gender-*

[22]The corresponding question was: " How competitive do you perceive your group of participants shown above?". See Appendix C.3 the experiment instructions.

**Table 3.3:** Treatment Effects on Round 2 Overconfidence

| | (1)<br>Overconfidence<br>Round 2 (std) | (2)<br>Overconfidence<br>Round 2 (std) | (3)<br>Overconfidence<br>Round 2 (std) |
|---|---|---|---|
| Female (1/0) | −0.291*** | −0.247** | −0.242** |
| | [-0.495,-0.088] | [-0.441,-0.053] | [-0.441,-0.043] |
| | (0.104) | (0.099) | (0.102) |
| MaleDomEnv (1/0) | −0.126 | −0.128 | −0.124 |
| | [-0.336,0.085] | [-0.327,0.072] | [-0.324,0.075] |
| | (0.107) | (0.101) | (0.102) |
| GenderBalanceEnv (1/0) | −0.090 | −0.088 | −0.085 |
| | [-0.291,0.111] | [-0.281,0.106] | [-0.282,0.111] |
| | (0.102) | (0.099) | (0.100) |
| Female x MaleDomEnv | 0.309** | 0.285** | 0.277* |
| | [0.015,0.602] | [0.004,0.567] | [-0.005,0.560] |
| | (0.150) | (0.144) | (0.144) |
| Female x GenderBalanceEnv | 0.163 | 0.157 | 0.160 |
| | [-0.120,0.446] | [-0.116,0.430] | [-0.116,0.435] |
| | (0.144) | (0.139) | (0.140) |
| Dep. Variable Mean | 0.253 | 0.253 | 0.253 |
| Task Sequence Control | Yes | Yes | Yes |
| Round 1 Performance Control | No | Yes | Yes |
| Round 1 Overconfidence | No | Yes | Yes |
| Demographic Controls | No | No | Yes |
| $R^2$ | 0.014 | 0.086 | 0.094 |
| N | 1076 | 1076 | 1076 |

**Notes.** The dependent variable in all specifications is participants' *Overconfidence* in Round 2. This variable is standardized to have a mean of zero and a standard deviation of 1 in the baseline *Gender−Blind* condition. specified in the column header. *Female* indicates that a participant's sex is female. *MaleDomEnv* indicates that a participant competes in a male-majority environment in Round 2. *GenderBalanceEnv* indicates that a participant competes in an environment with equally many men and women. The *GenderBlindEnv* is the omitted baseline environment category. *Sequence* indicates that the creative task used "brick" as an object in round 1 and "rubber band" in round 2; it is 0 if the object order is reversed. *Round 1 Performance* is the overall *CreativityScore* of Round 1. *Demographic Controls* include age, age$^2$, and field of study. 95% confidence intervals in squared brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

*Blind* baseline. This could also serve as a potential explanation for why the competitor environment has little impact on creative idea generation (Table 3.1).

Regarding gender differences within the varying competitor environment, the estimated coefficients of the interaction terms *Female x MaleDomEnv* and *Female x GenderBalanceEnv* are positive across all models. For the *Male-Dominated* environment, the estimated coefficients range from 13.2% (95%-CI [-0.167 $\sigma$,0.430 $\sigma$]) to 11.4% (95%-CI [-0.183 $\sigma$,0.411 $\sigma$]) of a standard deviation from Column 1 to Column 3 but are not statistically significant at conventional levels(p-value=0.453 in Column 3). Conversely, the interaction term *Female x GenderBalanceEnv* shows a positive and statistically significant effect (p-values<0.1) in all models, with coefficients $\sigma$ 0.271, i.e., 27.1% of a standard deviation (95%-CI of [-0.023$\sigma$,0.566$\sigma$]), $\sigma$ 0.275 (95%-CI of [-0.015$\sigma$,0.565$\sigma$]), and $\sigma$ 0.273 (95%-CI of [-0.019$\sigma$,0.565$\sigma$]) across Columns 1 to 3, respectively. This indicates that women in gender-balanced environments perceive their groups as more competitive compared to the baseline, with an increase in perceived competitiveness of approximately 21.3 percentage points. Conversely, the interaction term *Female x MaleDomEnv* is positive but insignificant compared to a $GenderBlind$ competition at baseline.

**Result 8:** When gender is revealed during a creative ideation task, women perceive gender-balanced environments as more competitive compared to men.

Overall, the results in Table 3.4 suggest that men and women hold different perceptions about a competitor group's degree of competitiveness in environments where gender is revealed when competing in a creative ideation task. In particular, women perceive gender-balanced environments to be more competitive than men in this context.

## 3.4 Discussion and Conclusion

R&D and innovation drive economic growth, yet women continue to participate less in innovative environments than men. This raises concerns about substantial inventive potential remaining untapped. Does the often competitive and male-dominated nature of innovative environments deter women from innovating? While the literature has mostly focused on women's non-selection into innovative environments, behaviors

**Table 3.4:** Treatment Effects on Perceived Group Competitiveness

| | (1) Perceived Group Competitiveness (std) | (2) Perceived Group Competitiveness (std) | (3) Perceived Group Competitiveness (std) |
|---|---|---|---|
| Female (1/0) | −0.080 | −0.057 | −0.039 |
| | [-0.287,0.128] | [-0.260,0.146] | [-0.247,0.169] |
| | (0.106) | (0.104) | (0.106) |
| MaleDomEnv (1/0) | −0.013 | −0.015 | 0.006 |
| | [-0.237,0.211] | [-0.237,0.207] | [-0.216,0.227] |
| | (0.114) | (0.113) | (0.113) |
| GenderBalanceEnv (1/0) | −0.124 | −0.114 | −0.101 |
| | [-0.342,0.094] | [-0.328,0.101] | [-0.316,0.115] |
| | (0.111) | (0.109) | (0.110) |
| Female x MaleDomEnv | 0.132 | 0.126 | 0.114 |
| | [-0.167,0.430] | [-0.171,0.422] | [-0.183,0.411] |
| | (0.152) | (0.151) | (0.151) |
| Female x GenderBalanceEnv | 0.271* | 0.275* | 0.273* |
| | [-0.023,0.566] | [-0.015,0.565] | [-0.019,0.565] |
| | (0.150) | (0.148) | (0.149) |
| Dep. Variable Mean | 3.787 | 3.787 | 3.787 |
| Task Sequence Control | Yes | Yes | Yes |
| Round 1 Performance & Belief | No | Yes | Yes |
| Demographic Controls | No | No | Yes |
| $R^2$ | 0.005 | 0.020 | 0.044 |
| N | 1076 | 1076 | 1076 |

**Notes.** The dependent variable in all specifications is participants' *Perceived Group Competitiveness* in Round 2. The question is ". How competitive do you perceive your group of participants shown above?" and is measured on a scale of 0 (not at all) to 5 (extremely). This variable is standardized to have a mean of zero and a standard deviation of 1 in the baseline $Gender-Blind$ condition. specified in the column header. $Female$ indicates that a participant's sex is female. $MaleDomEnv$ indicates that a participant competes in a male-majority environment in Round 2. $GenderBalanceEnv$ indicates that a participant competes in an environment with equally many men and women. The $GenderBlindEnv$ is the omitted baseline environment category. $Sequence$ indicates that the creative task used "brick" as object in round 1 and "rubber band" in round 2, it is 0 if the object order is reversed. *Round 1 Performance* is the overall $CreativityScore$ of Round 1. $Belief$ is a participant's belief about the score of their selected idea in Round 1. *Demographic Controls* include age, age$^2$, and field of study. 95% confidence intervals in squared brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

within these environments and the dynamics of the innovation process are understudied. Innovation processes typically start with creative acts and commonly occur in competitive and male-dominated innovative (STEM) environments. Yet, there is no systematic evidence of how these aspects interplay, affecting men's and women's creative idea generation and subsequent idea selection.

In this chapter, I investigate (1) how the competitive and male-dominated environments, separately and combined, impact gender gaps in creative ideation, (2) which dimensions of creative ideation, i.e., validity, originality, and flexibility, are impacted, and (3) shed light on men's and women's idea selection decisions. Using an artefactual field experiment online, I randomly vary the gender composition of the environment and the incentive structure to establish a causal link between these features and performance on a creative ideation task.

The findings indicate that women outperform men in a creative ideation task under non-competitive, monetary incentives. When competition is introduced, women continue to excel by selecting more original and higher-scoring ideas than men. Still, these gaps do not result in gender disparities in selecting one's best idea or winning a competition per se. However, the gender composition of the competitor environment plays a role, as men tend to catch up and select better ideas once gender is disclosed, with gaps in idea selection and competition outcomes being the lowest in gender-balanced environments. Looking at underlying channels, women exhibit lower overconfidence than men in anonymous, gender-blind competitions, but their confidence increases significantly in male-dominated, competitive environments. When gender is revealed, women also perceive gender-balanced environments as more competitive than men. These latter findings point towards some degree of sophistication about one's comparative advantage in the utilized creative ideation task.

Using an experimental design provides two significant advantages. First, it allows me to address challenges related to endogenous selection into specific innovative environments and tasks. By utilizing an artefactual field experiment in an online labor market, I can draw from a pool of actual knowledge workers while maintaining control over key features and information structures of interest. Specifically, I can manipulate incentives by exposing individuals to both competitive and non-competitive conditions. Additionally, individuals are randomly assigned to competitive environments with varying gender compositions, whereas, in observational settings, individuals would self-select into a single environment. Moreover, I use the "Unusual Uses of Objects"-task. It captures core business innovation elements, requiring participants

to think "outside the box" by repurposing existing resources to generate novel ideas. Thereby, it mirrors the creative and flexible problem-solving skills necessary in business to develop innovative products and solutions. This approach enables a clear separation of the effects of competition from those of competitor composition, facilitating consistent performance comparisons across different incentive structures and among individuals.
Second, the complexity and specificity of real-world innovations make measuring and comparing innovative ideas challenging in practice. By employing an experimental approach, I can use a creative ideation task to systematically quantify innovative ideas and assess their creative dimensions — validity, flexibility, and originality — ensuring consistent comparison across individuals and experimental conditions.

However, the stylized experimental setting employed in this study comes with several costs. First, one concern might have been the salience of the competitors' gender composition in this digital environment. The fairly anonymous icons (see Figure 3.2) were chosen to introduce as little noise as possible when varying the competitors' gender information.[23] The implemented manipulation checks indicate that participants have been aware of the gender composition (see Table C.4). Besides, on each of Round 2's idea submission pages, participants were reminded about their competitor group with a corresponding image (see experimental instructions in Appendix C.3). The chosen online competition setting is relevant to increasingly important remote teamwork in organizations Dingel (2020) and C. G. Aksoy et al. (2023) and platform-based "gig"-jobs (e.g., Upwork or TaskRabbit).

Second, the utilized *Unusual Uses of Objects Task* can only measure one type of creativity: divergent thinking (Bradler, Neckermann, & Warnke, 2019). Future research may investigate whether and how the gender composition of competitor environments may impact convergent thinking, closed innovation, or other complex, non-routine tasks highly demanded in today's labor market.

Third, with the rapid advancement of large language models like OpenAI's ChatGPT, it remains an open question to what extent these technologies will complement or substitute human efforts in creative tasks. A recent study by Koivisto and Grassini (2023) found that state-of-the-art AI could outperform human participants in divergent thinking on average, yet the top human performers still surpass AI tools. This highlights the unique and complex nature of human idea generation and selection, which cannot

[23]In comparison, the study by Born, Ranehill, and Sandberg (2022) uses physical group interactions of varying gender majority shares in their study on the willingness to lead.

be fully replicated by AI. Moreover, recent evidence from a field experiment suggests that fairly homogenous knowledge workers employ AI tools differently on the job and that the AI's task-solving success differs for seemingly similar tasks (Dell'Acqua et al., 2023). Thus, a promising avenue for future studies is to investigate under which incentives and for which parts or types of creative tasks humans can leverage the power of AI most beneficially.

This study contributes to three key strands of literature. First, it builds on research into behavioral gender differences in mixed- versus single-gender environments, where prior findings suggest that male-dominated settings lead women to exhibit lower risk-taking, competition avoidance, and reduced leadership aspirations (Booth, Cardona-Sosa, & Nolen, 2014; Bursztyn, Fujiwara, & Pallais, 2017). Second, it adds to a large body of works on incentives for creativity, particularly the role of stark competition in driving creative performance, which has produced heterogeneous findings based on context and sample (Bradler, Neckermann, & Warnke, 2019; Attanasi et al., 2021). Third, this study relates to research on gender gaps in knowledge-intensive professions more broadly (Hoisl & Mariani, 2017; Einiö, Feng, & Jaravel, 2019; Koning, Samila, & Ferguson, 2021; Iaria, Schwarz, & Waldinger, 2024).

This research offers three key contributions. First, it provides micro-level evidence on gender differences in creative ideation across varying competitive and gendered environments, disentangling the effects of competition from gender composition. Second, it examines gender differences in idea selection under different incentive structures, an area rarely considered in existing work. Third, by studying experienced knowledge workers across diverse industries rather than relying on student samples, this research enhances external validity and offers insights more directly applicable to real-world workplace settings.

Against the backdrop that even ubiquitous teamwork often has an inherent notion of competition, the results of this study bear several important implications for organizations. First, women outperform men in the creative ideation task at hand. Organizations should recognize women's strengths in creative tasks and ensure they are utilized effectively. This could involve providing environments that foster creativity and working actively towards increasing women's share in male-dominated teams.

Second, the gender composition of the competitive environment has intricate effects on men's and women's creative performance. Women select worse ideas in male-dominated competitions compared to gender-blind settings, while men appear to excel in gender-balanced settings. Thus, organizations should acknowledge that one-size-fits-all competitions might not maximize output for men and women.

Third, when it comes to critical decision-making (e.g., selecting the best innovative ideas) and outcomes (e.g., winning a competition), the study results to not suggest inherent gender gaps in gender-blind and gender-balanced settings. This underscores the importance of equal opportunities and fair performance evaluation criteria in competitive settings. Finally, being aware of how competitive environments can impact confidence differently based on gender, appropriate measurement, and feedback on performance might be a possibility to calibrate employees' beliefs about their abilities better and, potentially, increase work satisfaction.

By applying these insights, organizations can create environments that support men and women alike in realizing their creative potential and ultimately harnessing the best innovative ideas.

# A

# Appendix to Chapter 1

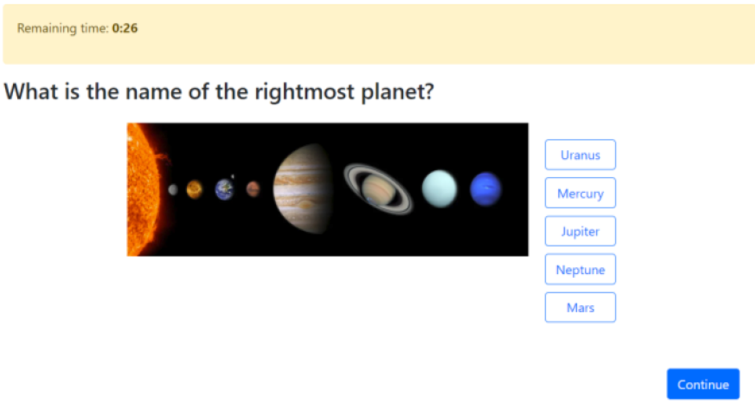
*Reputational Concerns and Advice-Seeking at Work*

# A.1 Main Appendix

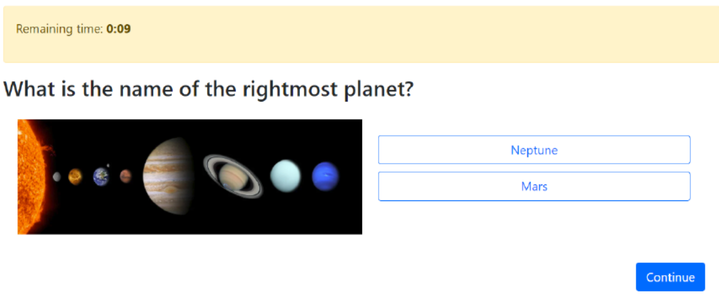
## A.1.1 Additional Figures and Tables

**Figure A.1:** Quiz Example Question Before and After Advice

**(a)** *Example Question. Science & Technology Quiz (Part 1)*



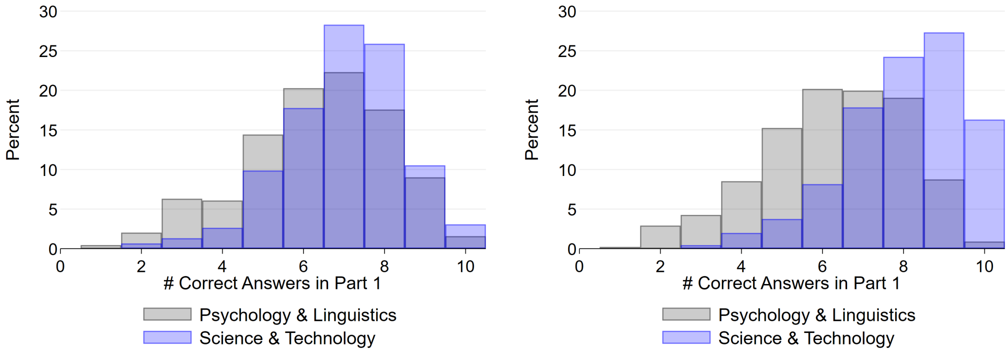

**(b)** *Example Question. Science & Technology Quiz After Advice Is Sought (Part 2)*



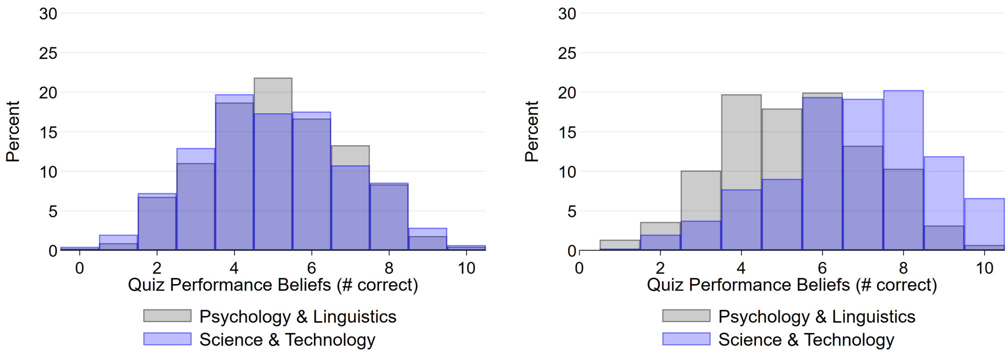

**Notes.** Example of 1/10 image-based quiz questions from the *"Science & Technology"*-quiz in Part 1 (A.1a) and Part 2 (A.1b), i.e., when the Employee had sought advice. Half of the participants were randomly assigned to this topic and the other half to the *"Psychology & Linguistics"*-quiz. There was a time limit of 30 seconds in Part 1 and 15 seconds in Part 2 to answer each question, before participants were auto-advanced.

**Figure A.2:** Actual and Subjective Quiz Score by Topic and Gender

**(a)** *Actual Score of Women (left, n=900) and Men (right, n=900) by Topic*

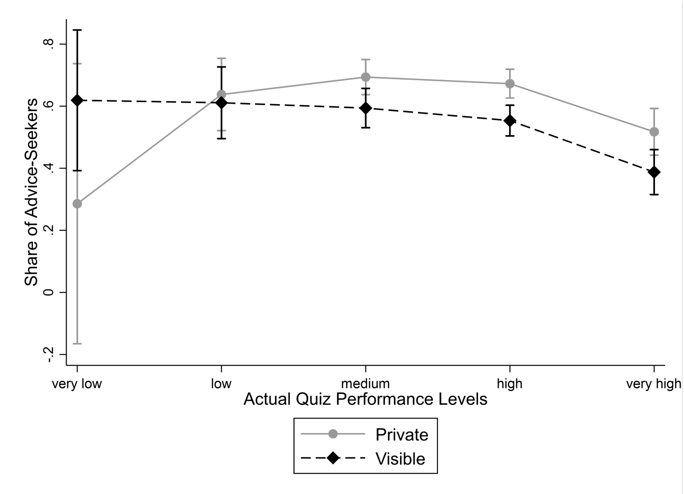



**(b)** *Quiz Score Beliefs of Women (left, n=900) and Men (right, n=900) by Topic*



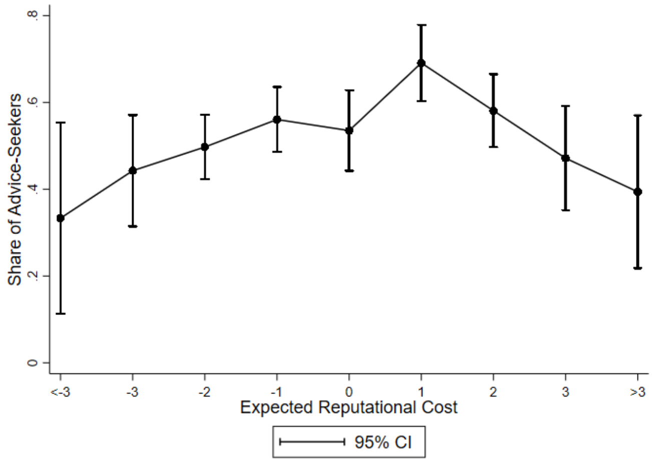

**Notes.** *Actual Score* (A.2a) is the quiz score in Part 1, ranging from 0 to 10. *Quiz Performance Beliefs* (A.2b) is an incentivized belief about the independent quiz score in Part 1. It is the answer to the following question “Guess, how many of your answers are correct?”, ranging from 0 to 10.

**Figure A.3:** Share of Employees who Sought Advice by Actual Performance Level and Treatment




**Notes.** The *Actual Quiz Performance Levels* are based on the Employees' independent quiz score in Part 1, binned in the following way: very low (n=28): 0-2, low (n=141): 3-4, medium (n=492): 5-6, high (n=789): 7-8, very high (n=350): 9-10. The bars represent 95% confidence intervals.

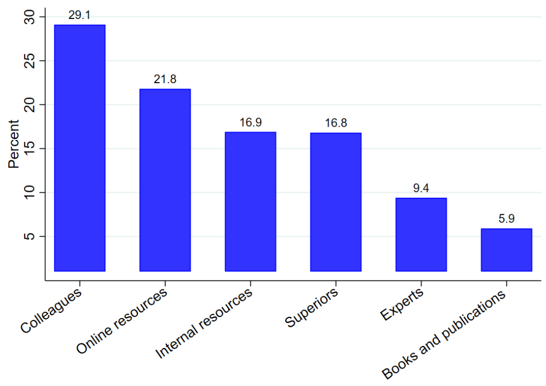
**Figure A.4:** Share of Employees Who Sought Advice by Expected Reputational Cost



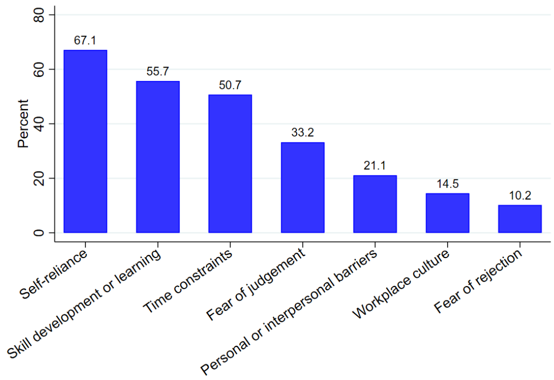

**Notes.** *Expected Reputational Cost* is the belief of Managers' quiz score estimate for a non-seeker 0,1,..,9,10 minus the belief of Managers' quiz score estimate for a seeker 0,1,..,9,10. Positive numbers indicate an expected reputational cost and negative numbers an expected reputational benefit to seeking advice. These incentivized second-order beliefs were elicited in Visible (N=897). The bars represent 95% confidence intervals.

**Figure A.5:** Advice Seeking at Work Survey of Professionals

*Panel A: Typical Sources of Information at Work*




*Panel B: : Reasons for Not Seeking Advice at Work*




**Notes.** These data come from a survey of 500 working professionals. Panel A shows the average percent allocated to an information source, in response to the following question: "Where do you typically source information that can help you solve work-related challenges on tasks or projects? Allocate 100% among these options to show how often you use each. Enter 0 for an option that you never use. Panel B shows how often (in percent) a reason was chosen in response to the following question: "What are typically the reasons for choosing to solve a work-related challenge on your own (e.g., through online resources or search engines), even though you know that asking someone for advice would be quicker? Select all that apply". This question was shown to the 422 participants who had previously reported to have refrained from seeking advice from others at work, even though it would have provided a faster solution to a work-related challenge.

**Table A.1:** Descriptive Statistics for Employees

| | (1) Pooled | (2) Male Employees | (3) Female Employees |
|---|---|---|---|
| *Age (average in years)* | 38.18 (SD=9.3) | 38.23 (SD=9.24) | 38.13 (SD=9.35) |
| *Resident of UK or Ireland* | 100.00% | 100.00% | 100.00% |
| *Employment* | | | |
| Full-time | 65.06% | 75.56% | 54.56% |
| Unemployed | 10.17% | 8.11% | 12.22% |
| Part-time | 13.72% | 6.67% | 20.78% |
| Self-employed | 11.06% | 9.67% | 12.44% |
| *Minimum Bachelor's Degree* | 100.00% | 100.00% | 100.00% |
| *Subject Studied* | | | |
| Humanities | 27.62% | 22.16% | 33.03% |
| Business & economics | 15.69% | 17.87% | 13.53% |
| Other social sciences | 14.24% | 9.86% | 18.58% |
| Engineering & computer science | 15.40% | 24.48% | 6.42% |
| Life science | 11.36% | 9.16% | 13.53% |
| Cognitive science | 2.48% | 1.62% | 3.33% |
| Other natural sciences & math | 9.17% | 11.25% | 7.11% |
| Law | 4.04% | 3.60% | 4.47% |
| # of Employees | 1800 | 900 | 900 |

**Table A.2:** Sample Size per Experimental Condition and Total by Role

**(a)** Employee (N=1,800)

| | | Topic | |
|---|---|---|---|
| | | Science & Technology (n=910) | Psychology & Linguistics (n=890) |
| Advice | Private (n=903) | 224 women<br>233 men | 220 women<br>226 men |
| | Visible (n=897) | 232 women<br>221 men | 224 women<br>220 men |

**(b)** Manager (N=721)

| | | Topic | |
|---|---|---|---|
| | | Science & Technology (n=360) | Psychology & Linguistics (n=361) |
| Advice | Private (n=241) | 62 women<br>60 men | 59 women<br>60 men |
| | Visible (n=480) | 120 women<br>118 men | 121 women<br>121 men |

**Table A.3:** Logit Models Predicting Employees' Propensity to Seek Advice

| DV: *Advice* (1/0) | (1) Pooled | (2) Pooled | (3) Female Employee | (4) Female Employee | (5) Male Employee | (6) Male Employee |
|---|---|---|---|---|---|---|
| *Visible* (1/0) | -.121*** | -.134*** | -.134*** | -.179*** | -.112*** | -.092* |
| | (.024) | (.035) | (.033) | (.048) | (.035) | (.050) |
| *Science & Tech* (1/0) | -.032 | -.046 | -.022 | -.069 | -.034 | -.014 |
| | (.024) | (.035) | (.033) | (.049) | (.037) | (.051) |
| *Visible X Science & Tech* | | .027 | | .087 | | -.040 |
| | | (.048) | | (.066) | | (.070) |
| *Male* (1/0) | -.037 | -.037 | | | | |
| | (.025) | (.025) | | | | |
| Baseline mean *Advice* | .643 | .643 | .700 | .700 | .588 | .588 |
| Subj. performance-level dummies | yes | yes | yes | yes | yes | yes |
| Adjusted $R^2$ | .062 | .062 | .045 | .046 | .075 | .076 |
| # of Employees | 1800 | 1800 | 900 | 900 | 900 | 900 |

**Notes.** Logit marginal effects (dF/dx) in all columns. The dependent variable in all specifications is *Advice* that equals 1 if the Employee sought advice and 0 otherwise. *Visible* is an indicator that equals 1 if the Employee was in the treatment condition where their advice-seeking decision was visible to the Manager and 0 otherwise. *Science & Tech* is an indicator that equals 1 if the Employee encountered the quiz topic "Science & Technology" and 0 when they encountered the "Psychology & Linguistics"-quiz. *Male* indicates Employee's sex equaling 1 if they are male and 0 if they are female. *Baseline mean Advice* is the mean of *Advice* for the (sub-) sample of Employees in each column header. *Subj. performance-level-dummies* bin Employees' incentivized beliefs about their achieved quiz score ranging from 0 to 10 and binned into five levels: 0-2, 3-4, 5-6, 7-8, 9-10. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table A.4:** Linear Probability Models Predicting Employees' Propensity to Seek Advice with Actual Performance-level Controls

| DV: *Advice* (1/0) | (1)<br>Pooled | (2)<br>Female<br>Employee | (3)<br>Male<br>Employee | (4)<br>Female<br>Employee | (5)<br>Male<br>Employee | (6)<br>Pooled |
|---|---|---|---|---|---|---|
| *Visible* (1/0) | -.105***<br>(.023) | -.129***<br>(.032) | -.082**<br>(.033) | -.168***<br>(.045) | -.059<br>(.046) | -.077**<br>(.032) |
| *Science & Tech* (1/0) | -.049**<br>(.024) | -.012<br>(.033) | -.096**<br>(.037) | -.051<br>(.044) | -.073<br>(.049) | |
| *Male (1/0)* | -.071***<br>(.023) | | | | | -.069***<br>(.023) |
| *Visible X Science & Tech* | | | | .077<br>(.064) | -.045<br>(.066) | |
| *Favorable Competence Stereotype* (1/0) | | | | | | -.002<br>(.032) |
| *Visible X Favorable Competence Stereotype* | | | | | | -.055<br>(.046) |
| Private Mean *Advice* | .643 | .700 | .588 | .700 | .588 | .643 |
| *Performance-level-dummies* | yes | yes | yes | yes | yes | yes |
| Adjusted R-sq | .035 | .023 | .032 | .023 | .031 | .034 |
| # of Employees | 1800 | 900 | 900 | 900 | 900 | 1800 |

**Notes.** The dependent variable in all specifications is *Advice* that equals 1 if the Employee sought advice and 0 otherwise. *Visible* indicates that the Employee was in the treatment condition in which the advice-decision was revealed to the Manager. *Science & Tech* indicates that the Employee took the Science & Technology quiz. *Male* indicates that the Employee's sex is male. *Favorable Competence Stereotype* indicates whether a stereotype about competence and an Employee's sex are congruent. For women, it takes the value of 1 in the "Psychology & Linguistics"-quiz and for men in the "Science & Technology"-quiz. *Private mean Advice* is the mean of advice for the (sub-)sample of Employees as described in the column header. *Performance-level-dummies* bin Employees' actual independent quiz score in part 1 (ranging from 0 to 10) into five levels: 0-2, 3-4, 5-6, 7-8, 9-10, with 7-8 as the omitted category. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table A.5:** Correlation Between Advice-Seeking and Additional Variables by Visibility

| Category | Independent Variable (Statement) | (1) Private | (2) Visible |
|---|---|---|---|
| | 1) Usefulness of advice | -.168*** | -.186*** |
| *a. Instrumental value of advice* | | (.013) | (.013) |
| | 2) Importance of luck in quiz (%) | -.071*** | -.035* |
| | | (.017) | (.017) |
| | 3) Male (0/1) | -.061 | -.006 |
| | | (.032) | (.033) |
| *b. Gender and norms* | 4) Continuous gender identity | .012 | -.011 |
| | | (.017) | (.017) |
| | 5) On average, men are less willing to ask others for | .001 | .009 |
| | advice than women. | (.016) | (.016) |
| | 6) In general, it is more socially acceptable for women | -.002 | .018 |
| | to ask for advice than for men. | (.016) | (.016) |
| | 7) Reputation is very important for one's career | .041** | -.001 |
| | advancement and promotions. | (.015) | (.016) |
| *c. Expectations and reputation* | 8) Seeking advice from others can hurt my reputation | -.019 | -.001 |
| | if others have high expectations about my ability. | (.016) | (.016) |
| | 9) Seeking advice from others can hurt my reputation | -.022 | -.023 |
| | if others have low expectations about my ability. | (.015) | (.016) |
| | 10) Overconfidence | -.024 | -.013 |
| | | (.02) | (.019) |
| | 11) Risk | -.004 | -.039* |
| | | (.016) | (.017) |
| *d. Other individual factors* | 12) It is particularly uncomfortable to ask for advice on | -.004 | .008 |
| | a work-related task that others think I am competent at. | (.015) | (.016) |
| | 13) I do not care what others think of me. | -.010 | .011 |
| | | (.015) | (.016) |
| | 14) I feel bad if I cannot accomplish tasks | -.194*** | -.169*** |
| | independently | (.007) | (.006) |
| | 15) Uncomfortable that Manager controls pay | .014 | -.003 |
| | | (.015) | (.016) |
| | 16) Job requires less teamwork | .019 | -.006 |
| *e. Work environment* | | (.016) | (.016) |
| | 17) More/most colleagues are female | .009 | .012 |
| | | (.016) | (.017) |
| | 18) Employment status: | | |
| | i) Unemployed | .037 | .076 |
| | | (.051) | (.053) |
| | ii) Part-Time | .060 | .060 |
| | | (.046) | (.046) |
| | iii) Self-employed | .049 | .008 |
| | | (.047) | (.053) |
| # of Employees | | 903 | 897 |

**Notes.** Individual level regression results from a linear probability model, correlating each independent variable,1-18, with the dependent variable *Advice* (1 if advice was sought, 0 otherwise), conditional on *subjective performance-level-dummies*. All independent variables except for 3) and 18) are standardized (mean zero, standard deviation one). The independent variables are worded in the questionnaire: 1) *"In your view, how useful is it to seek advice in this study?"* with choices: 1: 'Always useful,' 2: 'Only useful if you have an idea about the correct answer,' 3: 'Only useful if you do not know the correct answer,' 4: 'Never useful.' 2) *"When answering a quiz like today's, how important is the role of luck, as opposed to knowledge, in getting the correct answer?"* with choices: 0 % 'no luck' to 100 % 'only luck'. .3) *Male* is an indicator for Employee's sex that equals 1 if the Employee is male and 0 otherwise. 4) *"In general, how do you see yourself? Where would you put yourself on this scale from 0-'very masculine' to 10-'very feminine'?"* 10) *Overconfidence* is the difference between subjective confidence and objective Part 1. 11) *Risk* is a combined, weighted risk measure following Falk et al. 2018 using the weights from an experimental validation procedure of Falk et al. 2016 for staircase risk and willingness to take risks-. 15) *"How comfortable are you with the manager controlling a large part of your earnings for this study? Please indicate your answer on the scale below."* With choices: 1: 'very comfortable' to 7: 'very uncomfortable.' 16) *"Does your job require working in teams?"*, choices: 1: 'Always,' 2: 'Mostly,' 3: 'Balanced shares of team & individual work,' 4: 'Rarely,' 5: 'Never.' 17)*"What would best describe your colleagues at your current workplace?"*, choices: 1: 'By far most of my colleagues are men', 2: 'A somewhat bigger share of my colleagues are men',3: 'Among my colleagues, the share of men and women is about equal, 4: 'A somewhat bigger share of my colleagues are women', 5: 'By far most of my colleagues are women,' 6:'I have no colleagues'. In this specification, 5.4% of Employees with no colleagues are excluded from the regression, leaving N=849 and N=845 Employees in Private and Visible, respectively. 18)*Employment status* is a categorical variable with the choices i) to iii) above and iv) Full-Time, which is the omitted category. All other independent variables 5) - 9) and 12)-14) are measured on a Likert scale, indicating an agreement to a statement on a scale from 1: 'strongly disagree' to 7: 'strongly agree.' Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$,*** $p < 0.005$.

**Table A.6:** Frequencies of Motives for Seeking and Not Seeking Advice by Visibility to Manager

| Motive | Advice | | | No advice | | |
|---|---|---|---|---|---|---|
| | (1) Private | (2) Visible | (3) p-value | (4) Private | (5) Visible | (6) p-value |
| *Information value of advice* | .63 | .60 | .37 | .00 | .01 | .72 |
| *No benefit of advice* | .00 | .00 | - | .21 | .18 | .30 |
| *Manager rewards seeking advice* | .00 | .01 | .01 | .00 | .00 | - |
| *Manager discounts seeking advice* | .00 | .00 | .27 | .03 | .16 | .00 |
| *Preference for independent performance* | .00 | .00 | - | .18 | .17 | .90 |
| *Poor understanding of incentives* | .01 | .00 | .11 | .02 | .03 | .54 |
| *Economic cost-benefit trade-off* | .23 | .19 | .12 | .23 | .27 | .24 |
| *Confidence in own performance* | .00 | .00 | .27 | .44 | .42 | .61 |
| *Little confidence in own performance* | .64 | .68 | .26 | .03 | .03 | .71 |
| *Control own payment* | .00 | .00 | - | .00 | .01 | .45 |
| # of Employees | 581 | 482 | | 322 | 415 | |

**Notes.** Three raters blind to the research question independently classified each response to the question "First, briefly describe why you chose to [not] seek advice?" into the 10 motives. These three (yes/no) classifications were aggregated by taking the median. The median classification was then used to calculate the frequencies of motives for [not] seeking advice among seekers and non-seekers in Private and Visible. P-values of a two-sided test of proportions with H0 that proportions are the same in Private and Visible. Statistic of inter-rater agreement (Krippendorff's alpha) and description of each motive in Tables A.13 and A.14 in the Online Appendix.

**Table A.7:** Frequencies of Reasons Stated in Quiz Score Estimate Descriptions of Managers by Visibility of Advice

| Reason | (1)<br>Pooled | (2)<br>Private | (3)<br>Visible | (4)<br>P-Value |
|---|---|---|---|---|
| *Comparison to self* | .34 | .37 | .33 | .196 |
| *Education* | .41 | .38 | .43 | .168 |
| *Age* | .17 | .17 | .17 | .797 |
| *Knows sex* | .25 | .17 | .29 | .000 |
| *Explanation sex* | .04 | .05 | .04 | .503 |
| *Advice* | .13 | .03 | .19 | .00 |
| *Topic* | .25 | .25 | .25 | .879 |
| *Quiz difficulty* | .13 | .16 | .12 | .108 |
| *Guess* | .18 | .24 | .15 | .005 |
| # of Managers | 721 | 241 | 480 | |

**Notes.** Three raters blind to the research question independently classified each response to the question "We would like to understand how you arrived at your estimate of the quiz-taker's quiz performance without advice. Please briefly describe your thought process:" into the 9 reasons. These three (yes/no)-classifications were aggregated by taking the median. The median classification was then used to calculate the frequencies of reasons in the pooled sample of all Managers and split by visibility. P-values of a two-sided test of proportions with H0 that proportions are the same in Private and Visible. Statistic of inter-rater agreement (Krippendorff's alpha) and description of each reason in Tables A.19 and A.20 in the Online Appendix.

**Table A.8:** Job Characteristics of Professionals Surveyed about Advice-Seeking at Work and of Professionals in a Reference Sample

| | Advice-Seeking at Work Survey | | ACS 2022 |
|---|---|---|---|
| | UK | U.S. | U.S. |
| Age | 37 | 38 | 42 |
| Sector of employment (%) | | | |
| Private | 55.00 | 57.75 | 64.15 |
| Public | 35.00 | 29.00 | 22.68 |
| Not-for-profit | 10.00 | 12.75 | 13.05 |
| Other | | .50 | .12 |
| Self-Employed (%t) | 12.00 | 15.66 | 8.76 |
| Industry of employment (%) | | | |
| Agriculture, Forestry, Fishing & Hunting | | .75 | .68 |
| Mining, Quarrying, & Oil & Gas Extraction | | 1.00 | .25 |
| Utilities | | 1.00 | .85 |
| Construction | | 2.00 | 2.61 |
| Manufacturing | | 6.50 | 8.26 |
| Wholesale Trade | | .25 | 1.83 |
| Retail Trade | | 5.00 | 5.59 |
| Transportation & Warehousing | | 3.00 | 2.23 |
| Information | | 10.00 | 2.99 |
| Finance & Insurance | | 9.50 | 7.37 |
| Real Estate & Rental & Leasing | | 2.00 | 1.82 |
| Professional, Scientific, & Technical Services | | 13.75 | 15.36 |
| Management of Companies & Enterprise | | 1.00 | .20 |
| Administrative & Support/Waste Management | | .75 | 2.22 |
| Educational Services | | 13.75 | 17.56 |
| Health Care & Social Assistance | | 13.25 | 16.17 |
| Arts, Entertainment, & Recreation | | 5.75 | 1.87 |
| Accommodation & Food Services | | 2.75 | 1.82 |
| Other Services (except Public Administration) | | 7.00 | 3.21 |
| Public Administration | | 2.00 | 6.37 |
| Military | | .00 | .71 |
| Length of employment at current employer (in years) | 4.90 | 6.20 | |
| Size of Organization (# of employees) | 677 | 657 | |
| Size of Own Organizational Unit (# of colleagues) | 17 | 14 | |
| Job requires teamwork (%) | | | |
| Always or mostly | 44.00 | 41.00 | |
| Balanced share of team & individual work | 35.00 | 39.50 | |
| Rarely or never | 21.00 | 19.50 | |
| Typical team size at job (# of colleagues) | 7 | 7 | |
| N | 100 | 400 | 509,53 |

**Notes.** This table shows the job characteristics professionals reported in our Advice-Seeking at Work survey, separately for the UK & the U.S. samples. Length of employment, size of the organization, size of own organizational unit, & typical team size at the job are all approximations created from categorical variables, e.g., from the category "10-25 people" for typical team size at the job. For comparison, we also report data from the American Community Survey 2022, accessed through IPUMS. The ACS data were filtered to match the demographic & socioeconomic recruitment filters implemented on Prolific (age 25-60, educational level of at least a Bachelor's degree, currently employed). We do not report the industry of employment for the UK sample, given the small sample size & many answer options to this question.

### A.1.2 Manipulation Check Quiz Topics

Do quiz topics manipulate beliefs about the Manager's stereotypical beliefs about competence? In the final questionnaire, second-order beliefs about Managers' beliefs about competence were elicited in two ways.

First, we elicited beliefs about others' stereotypical views on what women and men know, on average, about different topics with a slightly modified version of the continuous slider measure introduced by (Coffman, 2014).[1] This unincentivized measure ranges from -1 (*most people* think there is a female advantage in knowledge) to 0 (no gender difference) to 1 (*most people* think there is a male advantage in knowledge). Every participant answered these sliders on six different topics.

Second, whenever an Employee was randomly assigned to Private, she was asked to report her belief about the Manager's quiz score estimate for two other participants: a woman and a man. She reported these beliefs for the same quiz topic that she had worked on, such that this measure varies between subjects. This elicitation was incentivized. The outcome *female advantage* is the difference in the reported beliefs and is positive whenever a participant believes that a Manager would estimate that women performed better than a man. According to either measure, we can conclude that the selected quiz topics successfully manipulated participants' beliefs about the Manager's belief about their competence (see averages presented in Table A.9).

**Table A.9:** Manipulation Checks Quiz Topic

| | **Psychology & Linguistics** | | | | **Science & Technology** | | | |
|---|---|---|---|---|---|---|---|---|
| | Slider | *p* | Fem A. | *p* | Slider | *p* | Fem A. | *p* |
| Overall | -.20 | $< .001$ | .126 | .027 | .29 | $< .001$ | -.379 | $< .001$ |
| Women | -.22 | $< .001$ | -.073 | .374 | .28 | $< .001$ | -.638 | $< .001$ |
| Men | -.17 | $< .001$ | .319 | $< .001$ | .30 | $< .001$ | -.129 | .070 |

**Notes.** Averages for the slider measure (slider) and the female advantage measure (Fem A.) P-values for a two-sided t-test against H0 that an average is equal to zero.

[1] Originally, the slider asks participants to report their own views. We, instead, asked about higher-order beliefs: what do you think *most people think?*

# A.2 Online Appendix: Alternative Specifications and Robustness Checks

## A.2.1 Additional Tables

**Table A.10:** LPM Predicting the Willingness to Seek Advice by Quiz Topic

| DV: *Advice* (1/0) | (1)<br>Science & Tech | (2)<br>Science & Tech | (3)<br>Psychology & Linguistics | (4)<br>Psychology & Linguistics |
|---|---|---|---|---|
| *Visible* (1/0) | -.100***<br>(.031) | -.085*<br>(.044) | -.122***<br>(.032) | -.168***<br>(.044) |
| *Male* (1/0) | -.033<br>(.034) | -.017<br>(.046) | -.027<br>(.032) | -.072*<br>(.043) |
| *Visible* X *Male* | | -.031<br>(.063) | | .091<br>(.063) |
| Baseline mean *Advice* | .602 | .602 | .686 | .686 |
| Subj. performance-level-dummies | yes | yes | yes | yes |
| Adjusted $R^2$ | .100 | .099 | .042 | .043 |
| # of Employees | 910 | 910 | 890 | 890 |

**Notes.** Individual level regression coefficients from a linear probability model. The dependent variable in all specifications is *Advice* that equals 1 if the Employee sought advice and 0 otherwise. *Visible* is an indicator that equals 1 if the Employee was in the treatment condition where their advice-seeking decision was visible to the Manager and 1 otherwise. *Male* is an indicator for Employee's sex equaling 1 if they are male and 0 if they are female. *Baseline mean Advice* is the mean of Advice for the (sub-)sample of Employees in each column header. *Subj. performance-level-dummies* bin Employees' incentivized beliefs about their achieved quiz score ranging from 0 to 10 and binned into five levels: 0-2, 3-4, 5-6, 7-8, 9-10. The column headers indicate the randomly assigned quiz topic. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table A.11:** Frequencies of Motives for Seeking and Not Seeking Advice by Employee Sex and Visibility to Manager: Employees Who Did *Seek* Advice

| Motives | Private | | | Visible | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Male | Female | p-value | Male | Female | p-value |
| *Information value of advice* | .69 | .57 | .37 | .66 | .55 | .72 |
| *No benefit of advice* | .00 | .00 | - | .00 | .00 | .30 |
| *Manager rewards seeking advice* | .00 | .00 | .01 | .01 | .02 | - |
| *Manager discounts seeking advice* | .00 | .00 | .27 | .01 | .00 | .00 |
| *Preference independent performance* | .00 | .00 | - | .00 | .00 | .90 |
| *Poor understanding of incentives* | .00 | .01 | .11 | .00 | .00 | .54 |
| *Economic cost-benefit trade-off* | .31 | .16 | .12 | .25 | .14 | .24 |
| *Confidence in own performance* | .00 | .00 | .27 | .00 | .00 | .61 |
| *Little confidence in own performance* | .62 | .67 | .26 | .65 | .70 | .71 |
| *Control own payment* | .00 | .00 | - | .00 | .00 | .45 |
| # Employees | 270 | 311 | | 221 | 261 | |

**Notes.** Median of binary (yes/no)-ratings from three independent raters blind to the research question used to calculate the relative frequencies with which each motive for seeking advice occurred by the gender of the Employee and visibility to the manager in columns (1), (2), (4) and (5). P-values of a proportion test against H0 show that proportions are the same in Visible and Private, as reported in columns (3) and (6). Employees who sought advice only. Description of each motive in Table A.14 in the Online Appendix.

**Table A.12:** Frequencies of Motives for Seeking and Not Seeking Advice by Employee Sex and Visibility to Manager: Employees Who Did *Not Seek* Advice

| Motives | Private | | | Visible | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Male | Female | p-value | Male | Female | p-value |
| *Information value of advice* | .00 | .01 | .37 | .01 | .00 | .72 |
| *No benefit of advice* | .17 | .27 | - | .10 | .28 | .30 |
| *Manager rewards seeking advice* | .00 | .00 | .01 | .00 | .00 | - |
| *Manager discounts seeking advice* | .04 | .01 | .27 | .17 | .14 | .00 |
| *Preference independent performance* | .17 | .19 | - | .15 | .21 | .90 |
| *Poor understanding of incentives* | .02 | .03 | .11 | .01 | .05 | .54 |
| *Economic cost-benefit trade-off* | .24 | .22 | .12 | .32 | .21 | .24 |
| *Confidence in own performance* | .56 | .27 | .27 | .49 | .34 | .61 |
| *Little confidence in own performance* | .02 | .05 | .26 | .01 | .05 | .71 |
| *Control own payment* | .00 | .01 | - | .00 | .02 | .45 |
| # Employees | 189 | 133 | | 220 | 195 | |

**Notes.** Median of binary (yes/no)-ratings from three independent raters blind to the research question used to calculate the relative frequencies with which each motive for not seeking advice occurred by gender of the Employee and visibility to the manager in columns (1), (2), (4) and (5). P-values of a proportion test against H0 show that proportions are the same in Visible and Private, as reported in columns (3) and (6). Employees who did not seek advice only. Description of each motive in Table A.14 in the Online Appendix.

**Table A.13:** Inter-rater Agreement for Classified Motives (Employees)

| Motive | Krippendorff's alpha |
|---|---|
| *Information value of advice* | .302 |
| *No benefit advice* | .719 |
| *Manager rewards seeking advice* | .497 |
| *Manager discounts seeking advice* | .896 |
| *Preference for independent performance* | .814 |
| *Poor understanding of incentives* | .348 |
| *Economic cost-benefit trade-off* | .655 |
| *Confidence in own performance* | .848 |
| *Little confidence own performance* | .713 |
| *Control own payment* | .051 |

**Notes.** Three student assistants blind to the research question independently classified each response to the question "First, briefly describe why you chose to [not] seek advice?" into the 10 motives. These motives were defined by the authors. Krippendorff's alpha is a measure of inter-rater agreement. An alpha of 1 indicates perfect agreement and a value of 0 implies no more agreement than what would be expected by chance. Typically, values between .41–.60 are interpreted to indicate moderate agreement, values between .61-.80 to indicate substantial agreement, and values above .81 to indicate almost perfect agreement

**Table A.14:** Description of Motives for Classification Task (Employees)

| Motive | Description |
|---|---|
| *1. Information value of advice* | Taking the advice option is considered useful: some additional information contained in advice which increases the probability of answering a question correctly. Hence, believed to improve the quiz performance (=score), including the ambition to reach a perfect score. Also, for reassurance in the original answers. *Example: "To increase my chance of making more money"* |
| *2. No benefit of advice* | Taking the advice option is considered useless for improving the quiz performance (=score), e.g., because narrowing down options from 5 to 2 would not help. Dislike of the risk of still getting things wrong. *Example: "Because I was fairly confident on enough of the answers, and I didn't want the Manager to see I'd sought advice."* |
| *3. Manager rewards seeking advice* | Taking the advice option increases the Manager's perception of oneself and thereby the allocated bonus. *Example: "I guessed that the Manager would estimate higher if he/she was shown that I had revised my answers based on advice."* |
| *4. Manager discounts seeking advice* | Taking the advice option decreases the Manager's perception of oneself and thereby the allocated bonus. *Example: "Because I was fairly confident on enough of the answers, and I didn't want the Manager to see I'd sought advice."* |
| *5. Preference for independent performance* | Preference for solving the quiz on one's own without external help of the advice option, i.e., e.g., testing one's knowledge, ownership, "risk it". *Example: "I chose not to seek advice as I like to learn and answer independently. I guessed that the Manager would estimate higher if he/she was shown that I had revised my answers based on advice.* |
| *6. Poor understanding of incentives* | Any statement revealing that one has not understood the incentive structure of the experiment. For instance, thinking the advice comes from the Manager/another candidate. *Example: "I did not seek advice because I don't think another average participant knows more than me and can give good advice on the quiz."* |
| *7. Economic cost-benefit trade-off* | Mentioning a trade-off between any monetary expected benefits (pay) and costs (advice fee) to seeking advice for justifying the advice decision (yes/no). For the non-seekers: considering the fee too high/not worth it. *Example: "For a small one-off fee, there was a good chance of selecting the correct answer in Part 2 and earning more money."* |
| *8. Confidence in own performance* | Stated high confidence in one's knowledge and answers provided independently in Part 1 of the study. *Example: "I am confident in my own abilities."* |
| *9. Little confidence in own performance* | Stated low confidence in one's knowledge and answers provided independently in Part 1 of the study. *Example: "I wasn't confident in the answers I provided in section 1."* |
| *10. Control own payment* | Stated preference for being the person (solely) in control of one's payment with possibly limiting/minimizing anyone else's impact via one's decision. Similar to *5. Preference for independent* performance but with clear monetary component. *Example: "I trust my own judgment. I don't trust others much."* |

**Table A.15:** OLS Regressions Predicting Managers' Quiz Score Estimate Controlling for Manager Sex

| DV: *Manager's (#) Estimate* | (1) Visible Pooled | (2) Visible Female Employee | (3) Visible Male Employee | (4) Private & Visible Female Employee | (5) Private & Visible Male Employee |
|---|---|---|---|---|---|
| *Advice* (1/0) | -.140 | -.214 | -.103 | | |
| | [-.369,.089] | [-.571,.142] | [-.387,.181] | | |
| | (.117) | (.181) | (.144) | | |
| *Visible* (1/0) | | | | -.191** | .033 |
| | | | | (.090) | (.091) |
| *Science & Tech* (1/0) | .239** | .009 | .488*** | .065 | .420*** |
| | (.105) | (.149) | (.151) | (.092) | (.092) |
| *Advice x Science & Tech* | .164 | .215 | .118 | | |
| | (.163) | (.232) | (.234) | | |
| *Female Employee* (1/0) | -.119 | | | | |
| | (.084) | | | | |
| *Male Manager* (1/0) | -.050 | -.014 | -.099 | -.033 | .017 |
| | (.087) | (.123) | (.125) | (.094) | (.096) |
| *Own subj. quiz performance (#)* | .241*** | .264*** | .222*** | .251*** | .198*** |
| | (.023) | (.034) | (.032) | (.027) | (.025) |
| Mean *Estimate* | 5.554 | 5.325 | 5.779 | 5.661 | 5.750 |
| Adjusted $R^2$ | .252 | .259 | .249 | .256 | .226 |
| # of Managers | 477 | 239 | 238 | 360 | 357 |

**Notes.** The dependent variable in all specifications is the Manager's *Estimate* of a matched Employee's quiz score. This variable is standardized to have a mean of zero and standard deviation of 1. *Advice t* indicates that the matched Employee sought advice on the quiz. *Visible* indicates that Managers observed the matched Employee's advice-seeking decision. *Science & Tech* indicates that Manager and the matched Employee took the Science & Technology quiz. *Female Employee* indicates that the matched Employee is a woman. *Male Manager* indicates that the Manager's sex is male. As four Managers have indicated a non-binary sex or refused to answer this question, the sample size is slightly lower than in Table 2. *Own subj. quiz performance* is the Manager's subjective belief of their own quiz performance and ranges from 0 to[2] 10. *Mean Estimate* is the overall mean of the Managers' estimate for the sample specified in the column header. Results presented in Columns (1)-(3) are restricted to Managers who were randomly assigned to Visible, while columns (4) and (5) include all Managers who were matched with female and male Employees, respectively. 95% confidence intervals in squared brackets. Robust standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

**Table A.16:** OLS Regressions Predicting Manager's Estimate in Private

| DV: *Manager's Estimate (#)* | (1)<br>Private<br>Pooled | (2)<br>Private<br>Female Employee | (3)<br>Private<br>Male Employee |
|---|---|---|---|
| *Female Employee* (1/0) | .296 | | |
| | (.237) | | |
| *Science & Tech* (1/0) | .258 | -.065 | .294 |
| | (.229) | (.252) | (.231) |
| *Female Employee x Science & Tech* | .278 | | |
| | (.330) | | |
| *Own subj. quiz performance* | .338*** | .409*** | .279*** |
| | (.044) | (.061) | (.063) |
| *Baseline mean Estimate* | 5.587 | 5.481 | 5.694 |
| # of Managers | 241 | 121 | 120 |
| $R^2$ | .216 | .242 | .198 |

**Notes.** Managers' individual-level regression coefficients from an OLS regression model. Managers in *Private* only. The dependent variable in all specifications is a Manager's *Estimate* (standardized, mean zero, standard deviation one) of a matched Employee's quiz score ranging between 0 and 10. *Female Employee* is an indicator that equals one if the sex of the matched Employee is female and zero otherwise. *Science & Tech* is an indicator that equals one if the matched Employee and Manager encountered the quiz topic Science & Technology and zero when they encountered the Psychology & Linguistics quiz. *Own subj. quiz performance* is the Manager's subjective belief of their own quiz performance and ranges between 0 and 10. This belief elicitation was incentivized. *Baseline mean Estimate* is the Mean of *Estimate* for the subsample of Managers matched with the respective type of Employee in each column header. Robust standard errors in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

**Table A.17:** Linear Probability Models Predicting the Likelihood of Manager's Estimate of Employee Performance being Correct

| DV: *Correct Estimate* (1/0) | (1)<br>Pooled | (2)<br>Pooled | (3)<br>Visible | (4)<br>Visible |
|---|---|---|---|---|
| *Visible* (1/0) | .002 | .003 | | |
| | (.026) | (.026) | | |
| *Science & Tech* (1/0) | -.023 | -.023 | -.013 | -.013 |
| | (.025) | (.025) | (.031) | (.031) |
| *Male Manager* (1/0) | .022 | | .033 | .033 |
| | (.026) | | (.032) | (.032) |
| *Same-sex pairing* (1/0) | | -.002 | | |
| | | (.025) | | |
| *Advice* (1/0) | | | .009 | .008 |
| | | | (.031) | (.033) |
| *Advice mentioned in statement* (1/0) | | | | .002 |
| | | | | (.043) |
| *Correct guess own performance* (1/0) | -.008 | -.009 | -.012 | -.012 |
| | (.032) | (.032) | (.039) | (.039) |
| *Own subj. quiz performance (#)* | .018*** | .019*** | .014* | .014* |
| | (.007) | (.007) | (.008) | (.008) |
| Constant | .031 | .035 | .040 | .040 |
| | (.040) | (.043) | (.045) | (.045) |
| *Correct Estimate* Mean | .129 | .129 | .133 | .133 |
| Adjusted $R^2$ | .007 | .006 | .001 | -.001 |
| # of Managers | 717 | 717 | 477 | 477 |

**Notes.** The dependent variable in all specifications is an indicator of the Manager's *Correct Estimate* of a matched Employee's quiz score. It takes on the value 1 if the estimate is correct and 0 otherwise. *Visible* indicates that Managers observed the matched Employee's advice-seeking decision. *Science & Tech* indicates that Manager and the matched Employee took the Science & Technology quiz. *Male Manager* indicates that the Manager's sex is male. As four Managers have indicated non-binary sex or refused to answer this question, the sample size is slightly smaller than in Table 2. *Same-sex pairing* indicates that the matched Manager and Employee have the same sex. *Advice* indicates that the matched Employee sought advice on the quiz. *Advice mentioned in statement* indicates that Managers have mentioned the term "advice" when reasoning about how they arrived at their estimate for the matched employee's quiz score. The wording of the question was: "We would like to understand how you arrived at your estimate of the quiz-taker's quiz performance without advice. Please, briefly describe your thought process:". *Correct guess own performance* indicates that Managers have correctly guessed their own quiz score. *Own subj. quiz performance* is the Manager's subjective belief of their own quiz performance and ranges from 0 to 10. *Mean Estimate* is the overall mean of the Managers' estimate for the sample specified in the column header. Results presented in Columns (1) and (2) include all Managers, while columns (3) and (4) are restricted to Managers in the *Visible*– Condition only. respectively. 95% confidence intervals in squared brackets. Robust standard errors in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

**Table A.18:** Frequencies of Reasons Stated in Manager's Description of Their Quiz Score Estimate

| | Male Employees | | | Female Employees | | |
|---|---|---|---|---|---|---|
| Reason | Visible | Private | P-Value | Visible | Private | P-Value |
| *Comparison to self* | .310 | .367 | .278 | .340 | .380 | .454 |
| *Education* | .414 | .383 | .574 | .448 | .372 | .166 |
| Age | .146 | .192 | .272 | .187 | .157 | .485 |
| *Knows sex* | .280 | .183 | .045 | .299 | .149 | .002 |
| *Explanation sex* | .038 | .058 | .370 | .033 | .033 | .995 |
| *Advice* | .205 | .017 | .000 | .166 | .033 | .000 |
| *Topic* | .247 | .242 | .914 | .261 | .256 | .915 |
| *Quiz difficulty* | .117 | .175 | .132 | .120 | .149 | .448 |
| *Guess* | .163 | .233 | .108 | .145 | .248 | .016 |
| # Managers | 239 | 120 | | 241 | 121 | |

**Notes.** Median of binary (yes/no)-ratings from three independent raters blind to the research question used to calculate the relative frequencies of reasons mentioned in Managers' open text descriptions of their reasoning behind their quiz score estimate of a matched Employee. The wording of the question was: "We would like to understand how you arrived at your estimate of the quiz-taker's quiz performance without advice. Please, briefly describe your thought process:". P-values of a proportion test against H0 that proportions are the same in Visible and Private. Description of each reason in Table A.20 in the Online Appendix.

**Table A.19:** Inter-rater Agreement for Classified Reasons (Managers)

| Reason | Krippendorf's alpha |
|---|---|
| *Comparison to self* | .868 |
| *Education* | .678 |
| *Age* | .724 |
| *Knows sex* | .885 |
| *Explanation sex* | .772 |
| *Advice* | .929 |
| *Topic* | .705 |
| *Quiz difficulty* | .541 |
| *Guess* | .523 |

**Notes.** Three student assistants blind to the research question independently classified each response to the following prompt "We would like to understand how you arrived at your estimate of the quiz-taker's quiz performance without advice. Please, briefly describe your thought process:" into the 9 reasons. These reasons were defined by the authors. Krippendorff's alpha is a measure of inter-rater agreement. An alpha of 1 indicates perfect agreement, and a value of 0 implies no more agreement than what would be expected by chance. Typically, values between .41–.60 are interpreted to indicate moderate agreement, values between .61–.80 to indicate substantial agreement, and values above .81 to indicate almost perfect agreement.

**Table A.20:** Description and Examples of Reasons for Classification Task (Managers)

| Reason | Description | Examples |
|---|---|---|
| *1.Comparison to self* | Manager compares profile to himself/herself or their knowledge/abilities/how difficult they found the quiz. Manager took their own performance as a reference point. | *"Based on my own performance, similar background and age. That's all I had to go on, so I went with it."*<br>*"I just guessed based on my own performance."* |
| *2. Education* | Manager refers to the information on the employee's education level e.g., "a-levels". | *"I wondered what A levels this quiz-taker had taken, as this might include some areas of knowledge pertinent to the questions."*<br>*"Looked at level of education"*<br>*"She is intelligent."* |
| *3. Age* | Manager explains that s/he considered the employee's age when making an estimate. | *"Took into account age, education and type of questions"*<br>*"She could have been younger than me."* |
| *4. Knows sex* | Manager's explanation shows that s/he knows the Employee's sex, for example, by using pronouns. | *"She could have been younger than me"*<br>*"He is male so probably more interested in the subject, profile suggests he is intelligent."* |
| *5. Explanation sex* | Manager explains that s/he considered the employee's sex when making an estimate. | *"He is male so probably more interested in the subject, profile suggests he is intelligent."*<br>*"As a woman, she may be interested in psychology."* |
| *6. Advice* | Manager explains that s/he based the estimate (also) on the Employee's decision to seek advice. | *"I took in consideration his age, his education and that he sought advice*<br>*"I noted that the quiz-taker took advice and therefore the choices of each were narrowed to 2, he had taken a-levels and I guessed that he might have studied the subject in question."* |
| *7. Topic* | Manager mentions that his/her estimate is (also) based on the quiz topic ("Science & Technology" or "Psychology & Linguistics"). The manager could also mention the topic indirectly (e.g. "the subject was hard") | *"The Science & Tech knowledge he has pretty much is the deciding factor."*<br>*"Knowledge of psychology methods and authors amongst the general population is low so I would imagine most answers were guesswork."*<br>*"I assume that she knows little about the topic."* |
| *8. Quiz difficulty* | Manager comments on their perception of the difficulty of the quiz. | *"It's a fairly difficult quiz so I went with slightly more than half"* |
| *9. Guess* | Manager states no reason other than a (random) guess. | *"It was a guess"*<br>*"Picked an average score"* |

### A.2.2 Calibration of the Knowledge Task

For constructing the knowledge task of this study, we collected over 160 questions from various domains and gathered the performance data from 119 individuals in the same subject pool as the main study.[3] Simultaneously, we elicited beliefs about the average performance of men and women in several domains of knowledge ("Art & Art History", "Geography & Geology", "History", "Information Science & Technology", "Linguistics & Language Use", "Literature", "Philosophy", "Physics & Astrophysics", "Politics", "Popular Culture", "Psychology" and "Sports".

Participants reported their beliefs on a scale from -1 (women know more on average) to 1 (men know more on average), where 0 reflects parity Coffman (2014). We selected the most stereotypical and potentially labor-market-relevant domains of knowledge to combine them into one topic, namely "Science & Technology" as well as "Psychology & Linguistics". In our sample, stereotypes assign women and men, respectively, a knowledge advantage. We curated the final set of questions per quiz to be of comparable difficulty. Based on the knowledge of women and men in our pre-test sample, we selected 10 questions per topic. The questions were selected such that women and men gave 6 correct answers, on average, and had the same modal number of correct answers (7).

All reported p-values are from a two-sided t-test for an equality of means. Looking at the actual quiz performance in Part 1 of the main study, the "Science & Technology" quiz —with an average score of 7.5 correct answers—turned out slightly easier than the "Psychology & Linguistics" quiz, with an average score of 6.3 ($p$<0.001). Overall, the modal quiz scores are 7 and 8, respectively. On the "Psychology & Linguistics" quiz, women and men performed equally well, on average: women with an average score of 6.4 and men with an average score of 6.3 ($p$=0.69). The modal score is 7 for women and 6 for men. On the "Science & Technology" quiz, women had an average score of 7.0, whereas men performed slightly better with an average score of 8.0 ($p$<0.001). The modal score is 9 for men and 7 for women.

[3] Participants of any of our quiz calibration pilots or pre-tests were excluded from our main data collection.

## A.3 Instructions

### A.3.1 Employee-Version

*Screens marked with a and b reflect treatment variations, i.e., each participant saw either version of the screen in line with the random treatment assignment.*

#### *Screen 1*

**Welcome!**

**Overview:** You are participating in a research study on economic decision-making. It consists of two main parts in which you make decisions for an extra bonus payment and a final questionnaire. Your bonus payment depends on your decisions and the decisions of other participants throughout the study. Therefore, it is important that you read all instructions carefully. Your responses are anonymous.

**Payment:** You will receive a participation payment of **1.50 GBP.** In addition, you can earn an extra bonus payment of up to **4.25 GBP.** The payment will be processed within the next 4 working days. Please be advised that only complete submissions will be paid.

**General Rules of Conduct:** This study will take about 10 minutes. We ask for your full attention. Please find a quiet space to complete the study and do not use other devices, talk to other people, use social media, etc. Please remain solely in this browser tab for the entire time of the study.

**Comprehension Checks:** During the study, you will be asked several comprehension questions referring to the instructions on the same screen. You can only proceed once you answer them correctly.

**Consent:** I have read and understood the information above. I agree to comply with these rules of conduct and want to participate in this study.

Please click “Continue” to proceed.

*[Button: Continue]*

#### *Screen 2*

**Initial Survey**
To begin, please answer a short survey about yourself.

> **# 1**
> How old are you?
>
> [          ]
>
> **# 2**
> What is your sex?

◯ Female

◯ Male

**# 3**
Have you completed A levels or an equivalent level of education that qualifies you for university studies?

◯ Yes

◯ No

**# 4**
Were you a university student at some point in time during your life, including current enrollment?

◯ Yes

◯ No

**# 5**
Are you currently employed?

◯ Full-Time

◯ Part-Time

◯ Self-employed

◯ No

*[Button: Continue]*

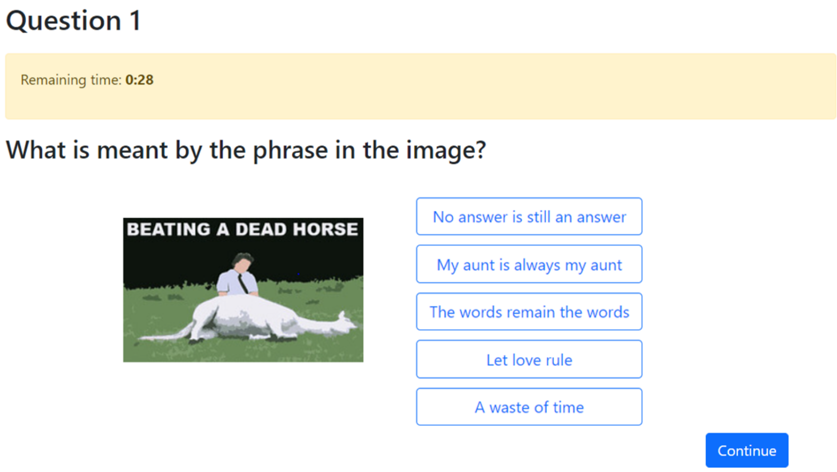
***Screen 3a: Psychology & Linguistics***

**Part 1: The Quiz**

In Part 1, you will take a general knowledge quiz on **Psychology & Linguistics**.

You will answer 10 multiple-choice questions. You will receive **0.10 GBP** for each correct answer and no extra payment for wrong answers.

The questions are all structured in the same way: a question text, a corresponding image, and 5 possible answers, one of which is correct (see an example picture).




To submit an answer, choose one and click "Submit". You will then advance to the next question. You will have **30 seconds** to submit your answer, before you are automatically advanced.

**We encourage you to submit your best guess before the time runs out, even if you do not know the correct answer with certainty**. You have a 20%-chance to simply guess correctly.

### Comprehension Question

Is this statement true or false?

Even if I do not know the answer to a question, submitting any guess will increase the chance to receive 0.10 GBP from 0% to 20%.

- ◯ True
- ◯ False

*[Button: Continue]*

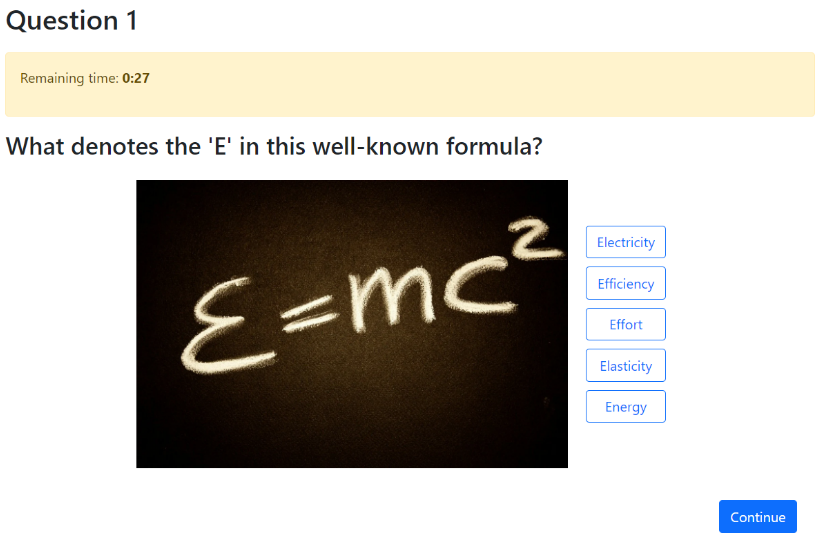
### *Screen 3b: Science & Technology*

### Part 1: The Quiz

In Part 1, you will take a general knowledge quiz on **Science & Technology.**

You will answer 10 multiple-choice questions. You will receive **0.10 GBP** for each correct answer and no extra payment for wrong answers.

The questions are all structured in the same way: a question text, a corresponding image, and 5 possible answers, one of which is correct (see an example picture).




To submit an answer, choose one and click “Submit”. You will then advance to the next question. You will have 30 seconds to submit your answer, before you are automatically advanced.

**We encourage you to submit your best guess before the time runs out, even if you do not know the correct answer with certainty.** You have a 20%-chance to simply guess correctly.

### Comprehension Question

Is this statement true or false?

Even if I do not know the answer to a question, submitting any guess will increase the chance to receive 0.10 GBP from 0% to 20%.

◯ True

◯ False

*[Button: Continue]*

*[Quiz — Part 1: 10 sequential screens with multiple choice questions as in the example screen above, participants have 30 seconds to answer each question, after they submitted an answer or timed out, they see the next question of the quiz or proceed with the rest of the experiment.]*

### *Screen 4*

### Guess, how many of your answers are correct?

You just answered 10 questions on Psychology & Linguistics. How many of your answers are correct? You will receive an extra **correct-guess-pay of 0.25 GBP** if your guess equals your actual number of correct answers.

Please enter a number from 0 (no correct answers) to 10 (all correct answers):

| |
|---|
| |

At the end of the study, you will see how many of your answers are correct. You will also have the option to learn the correct answers if you want.

*[Button: Continue]*

### *Screen 5: Psychology & Linguistics, text otherwise equivalent*

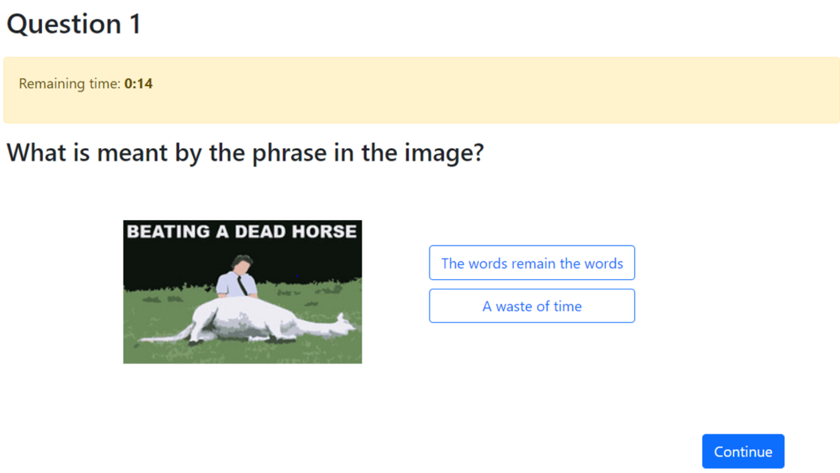
### Part 2: Revision with Advice

In Part 2, you can seek advice on this Psychology & Linguistics quiz and revise your answers. Seeking advice costs **a one-time fee of 0.08 GBP (8 pence)** and it can help you to submit more correct answers for payment.

### Advice

***If you decide to seek advice,*** you will answer each question again but with only 2 instead of 5 answer choices (see an example below). For this simplified quiz, **the computer randomly removed 3 wrong answers.** Your initial answer will be highlighted if it is among the remaining choices. You will have *15 seconds* to review and revise each question. If you run out of time or proceed without a revision, your initial answer will count for payment. Your payment for correct answers will equal **0.10 GBP times your performance in Part 2 after the revision of your answers** in the simplified quiz.

***If you decide not to seek advice,*** you will proceed to the final questionnaire. Your payment for correct answers will equal **0.10 GBP times your quiz performance in Part 1, which was already recorded.**




### Comprehension Questions

With advice, each question has only 2 instead of 5 answer choices. This increases the chance to simply guess correctly from 20% to . . .

◯ ...25%

◯ ...50%

**Consider the following situation:** ***a person has sought advice and revised their answers to two questions. This yielded two additional correct answers.***

How much more will that person earn for the 2 correct answers submitted during the revision?

◯ 0.00 GBP since correct answers after advice do not count.

◯ 0.20 GBP since that person gets 0.10 GBP for two more correct answers.

How much does this person pay for seeking advice?

◯ 0.16 GBP

◯ 0.08 GBP

*[Button: Continue]*

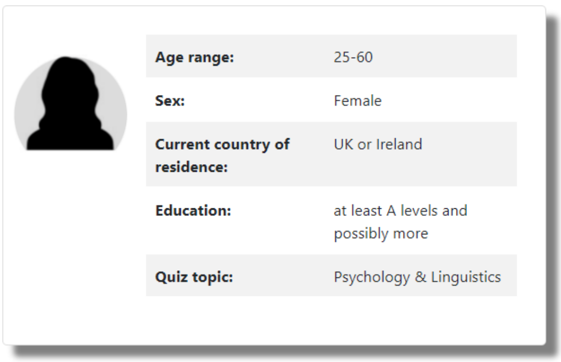
### *Screen 6a: Private*

### Additional Pay

You can earn more money, which will be added to your other payments. A "manager" will estimate your prior knowledge on Science & Technology, as recorded with your quiz performance in Part 1. The manager is another participant who reported to Prolific to have experience in a managerial position and who will complete a different version of this study.

Your manager will estimate how many questions you have answered correctly **without advice in Part 1**, based on your profile as shown below. You will earn **0.30 GBP** times the manager's estimate. This pay is higher, the more questions the manager thinks you have answered correctly without advice in Part 1. It can be as high as **3.00 GBP**–if the manager thinks that you have answered all 10 questions correctly in Part 1,– and as low as **0.00 GBP.**

The manager will *only see your profile* to make their estimate of your quiz performance without advice in Part 1 and have no other information about you. Your profile does **not** show your upcoming decision on whether to seek advice.




The manager will earn **3.00 GBP** if their estimate of your quiz performance without advice in Part 1 equals your actual performance, and **0.00 GBP** otherwise. Therefore, the manager is motivated to estimate your quiz performance without advice in Part 1 correctly.

The manager will take the same quiz and read a complete summary of your version of the study.

### Comprehension Questions

The manager estimates your...

◯ quiz performance without advice in Part 1.

◯ your quiz performance after a revision with advice in Part 2.

**Consider the following situation: a manager estimates that a person has answered 5 questions correctly without advice in Part 1.**

This person will then receive the following pay for the manager's estimate of their quiz performance in Part 1:

◯ 3.00 GBP

◯ 1.00 GBP

◯ 1.50 GBP

*[Button: Continue]*

### *Screen 6b: Visible*

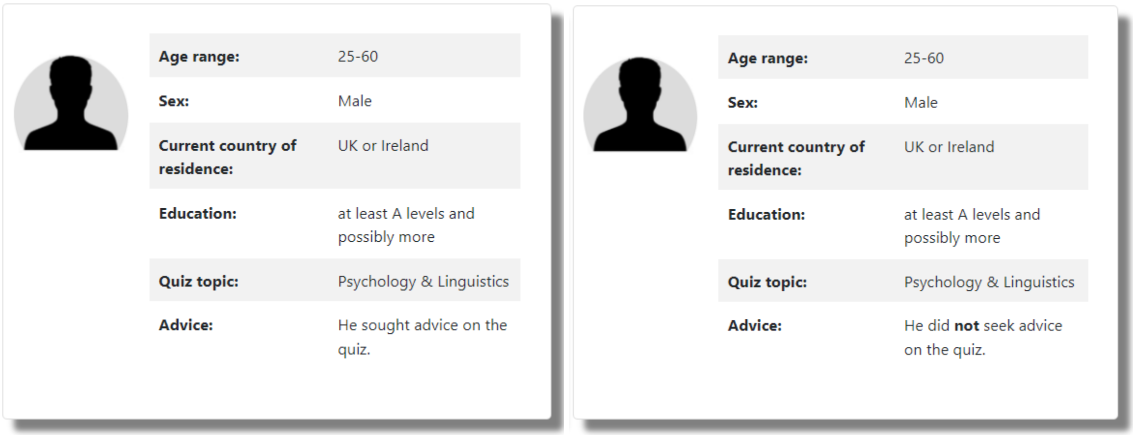
### Additional Pay

You can earn more money, which will be added to your other payments. A "manager" will estimate your prior knowledge on **Psychology & Linguistics**, as recorded with your quiz performance in Part 1. The manager is another participant who reported to Prolific to have experience in a managerial position and who will complete a different version of this study.

Your manager will estimate how many questions you have answered correctly **without advice in Part 1**, based on your profile as shown below. You will earn **0.30 GBP** times the manager's estimate. This pay is higher, the more questions the manager thinks you have answered correctly without advice in Part 1. It can be as high as **3.00 GBP**–if the manager thinks that you have answered all 10 questions correctly in Part 1–and as low as **0.00 GBP**.

The manager will *only see your profile* to make their estimate of your quiz performance without advice in Part 1 and have no other information about you. If you decide to seek advice, the manager sees the left profile, if you decide not to seek advice, the manager sees the right profile.
The manager will earn **3.00 GBP** if their estimate of your quiz performance without advice in Part 1 equals your actual performance, and **0.00 GBP** otherwise. Therefore, the manager is motivated to estimate your quiz performance without advice in Part 1 correctly.

The manager will take the same quiz and read a complete summary of your version of the study.




### Comprehension Questions

The manager estimates your...

◯ quiz performance without advice in Part 1.

◯ your quiz performance after a revision with advice in Part 2.

**Consider the following situation: a manager estimates that a person has answered 5 questions correctly without advice in Part 1.**

This person will then receive the following pay for the manager's estimate of their quiz performance in Part 1:

◯ 3.00 GBP

◯ 1.00 GBP

◯ 1.50 GBP

*[Button: Continue]*

*[After answering the comprehension questions correctly (6a,b), participants were reminded of the quiz questions. They saw the list of questions without corresponding pictures or answer options for 15 seconds.]*

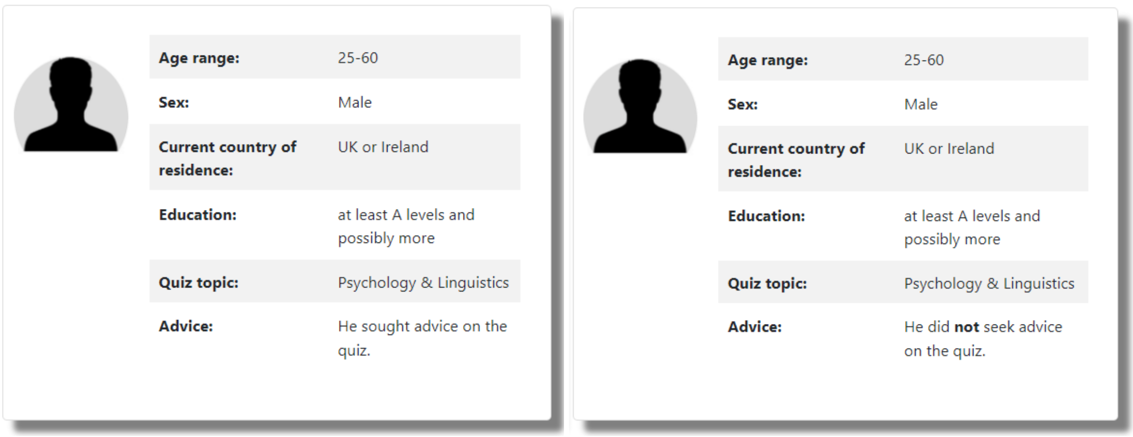
### *Screen 7b: Visible (In Private, profiles do not contain the last row, see Screen 6a)*

### Do You Want to Seek Advice?

You can now decide to seek advice.

- If you seek advice, you pay a **one-time advice fee of 0.08 GBP** and can revise your answers in the simplified quiz.
- **A correct answer** pays **0.10 GBP**, regardless of whether you arrived at it without advice in Part 1 or after seeking advice in Part 2.

- You will earn **0.30 GBP** times your manager's estimate of your quiz performance without advice in Part 1. This estimate ranges from 0 (no correct answers) to 10 (all correct answers).
- To make this estimate, the manager will *only* see the information in your profile, either the left or the right one.



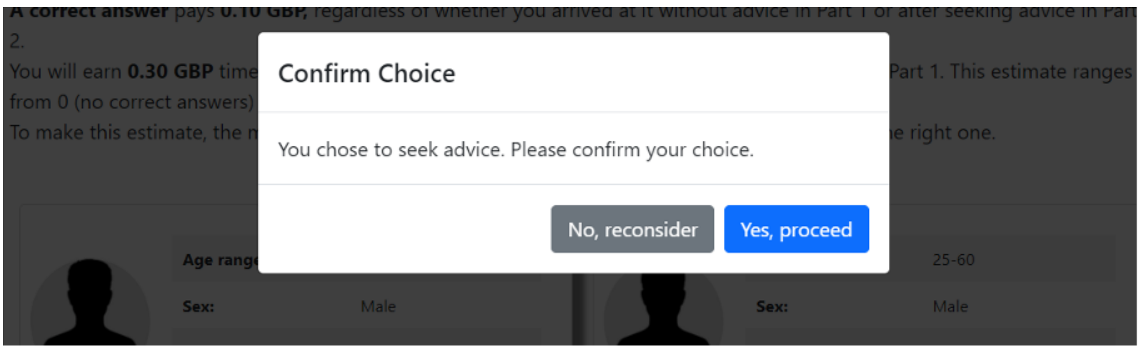

At the end of the study, you will see how many questions you answered correctly in Part 1 and, if applicable, in Part 2 after the revision. **Regardless of your decision now, at the end of the study you can see the correct answers to all the questions.**

**Do you want to seek advice and revise your answers in the simplified quiz?**

◯ Proceed without advice

◯ Seek advice

After making the choice, participants saw the following popup. They had to confirm the choice the make of could reconsider.




*[Quiz Part 2: If participants chose to seek advice, they saw the same 10 general knowledge questions again with 2 instead of 5 answer options. They had 15 seconds per question. If they chose not to seek advice, they proceeded with the questionnaire.]*

### *Screen 8*

### Questionnaire (1/8)

This short questionnaire is the final part. Afterwards, you will see an overview of your earnings and can learn the correct answers.

First, please briefly describe why you chose to not seek advice:

In your view, how useful is it to seek advice in this study?

- ◯ Always useful
- ◯ Only useful if you have an idea about the correct answer
- ◯ Only useful if you do not have an idea about the correct answer
- ◯ Never useful

*[Button: Continue]*

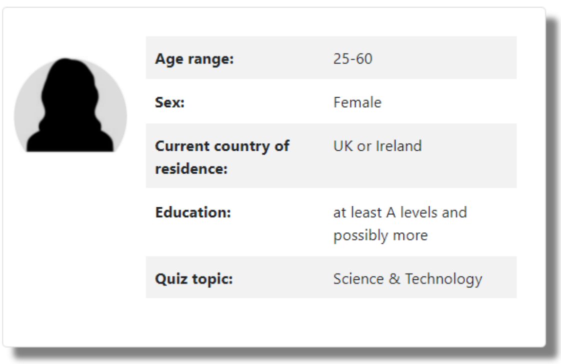
### *Screen 9a: Private*

### Questionnaire (2/10)

In this part of the questionnaire, you can **earn extra payment by guessing the manager's estimate correctly.**

We ask you to provide two guesses. One of them is randomly selected by the computer to count for payment. If you guess the manager's estimate correctly, you will receive an **extra correct-guess-pay of 0.25 GBP.**

**Guess 1**
Consider a participant with the same profile as you. Below is the profile that the manager sees when estimating their quiz performance without advice in Part 1. This is the only information that the manager sees about them:



| | |
|---|---|
| **Age range:** | 25-60 |
| **Sex:** | Female |
| **Current country of residence:** | UK or Ireland |
| **Education:** | at least A levels and possibly more |
| **Quiz topic:** | Science & Technology |

In your view, what will the manager estimate this participant's quiz performance without advice in Part 1? Please enter a number from 0 (no correct answers) to 10 (all correct answers):

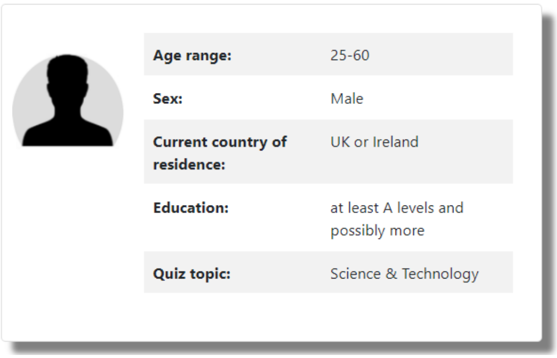
*[Button: Continue]*

### *Screen 10a: Private*

## Questionnaire (3/10)

You will receive **0.25 GBP** if this guess is correct and randomly selected to count for the correct-guess-pay.

**Guess 2**
Consider another participant. Below is the profile that the manager sees when estimating their quiz performance without advice in Part 1. This is the only information that the manager sees about them.




In your view, what will the manager estimate this participant's quiz performance without advice in Part 1? Please enter a number from 0 (no correct answers) to 10 (all correct answers):

*[Button: Continue]*

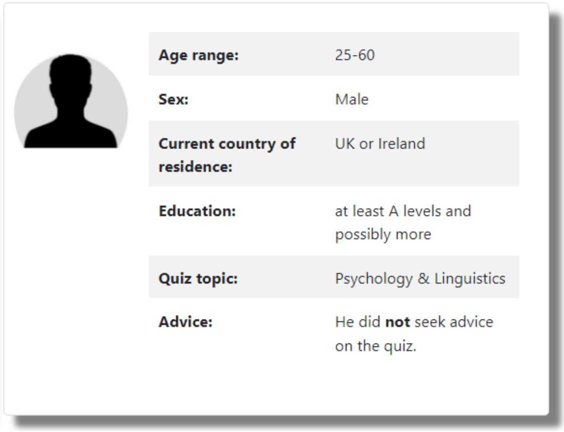
### *Screen 9b: Visible*

## Questionnaire (2/10)

In this part of the questionnaire, you can **earn extra payment by guessing the manager's estimate correctly**.

We ask you to provide two guesses. *One of them* is randomly selected by the computer to count for payment. If you guess the manager's estimate correctly, you will receive an **extra correct-guess-pay of 0.25 GBP**.

**Guess 1**
Consider a participant who chose, like you, not to seek advice on the quiz. Below is the profile that the manager sees when estimating their quiz performance without advice in Part 1. This is the only information that the manager sees about them.



| | |
|---|---|
| **Age range:** | 25-60 |
| **Sex:** | Male |
| **Current country of residence:** | UK or Ireland |
| **Education:** | at least A levels and possibly more |
| **Quiz topic:** | Psychology & Linguistics |
| **Advice:** | He did **not** seek advice on the quiz. |

In your view, what will be the manager's estimate of this participant's quiz performance without advice in Part 1? Please enter a number from 0 (no correct answers) to 10 (all correct answers):

*[Button: Continue]*

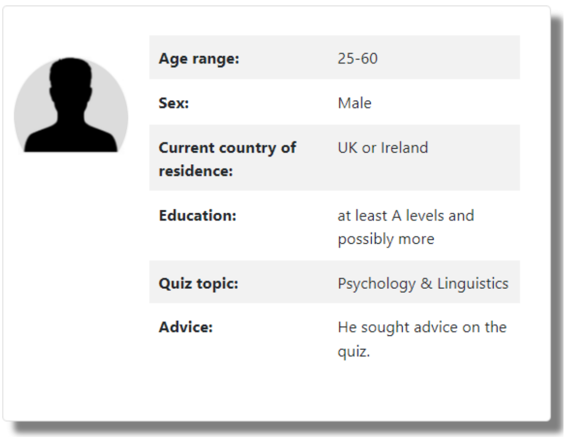
### *Screen 10b: Visible*

## Questionnaire (3/10)

You will receive **0.25 GBP** if this guess is correct and randomly selected to count for the correct-guess-pay.

**Guess 2**
Consider another participant who chose, **unlike you**, to seek advice on the quiz in Part 2. This participant's profile is, otherwise, the same as the one you just looked at.

Below is the profile that the manager sees when estimating their quiz performance without advice in Part 1. This is the only information that the manager sees about them:



| | |
|---|---|
| **Age range:** | 25-60 |
| **Sex:** | Male |
| **Current country of residence:** | UK or Ireland |
| **Education:** | at least A levels and possibly more |
| **Quiz topic:** | Psychology & Linguistics |
| **Advice:** | He sought advice on the quiz. |

In your view, what will be the manager's estimate of this participant's quiz performance without advice in Part 1? Please enter a number from 0 (no correct answers) to 10 (all correct answers):

*[Button: Continue]*

### *Screen 11*

## Questionnaire (4/10)

**Your Relative Performance**
You guessed that you submitted 6 correct answers without advice.

Think about all other study participants with the same profile (25-60 years old, male, UK or Ireland resident, at least A levels or equivalent and possibly more) who took the same Psychology & Linguistics quiz as you.

How does your quiz performance without advice in Part 1 compare to theirs?

My performance is in the...

- ◯ Top 25%
- ◯ Somewhere between the top 25% and the bottom 25% but closer to the top 25%
- ◯ Somewhere between the top 25% and the bottom 25% but closer to the bottom 25%
- ◯ Bottom 25%

*[Button: Continue]*

### *Screen 12*

## Questionnaire (5/10)

What gender are you currently?

- ◯ Man (including Trans Male/Trans Man)
- ◯ Women (including Trans Female/ Trans Women)
- ◯ Non-binary
- ◯ Rather not say

In general, how do you see yourself? Where would you put yourself on this scale from 0-"very masculine" to 10-"very feminine"? Please click the blue bars to reveal the sliders and indicate your answer.

very masculine — very feminine
0 — 10

Your value: **8**

In general, how willing or unwilling you are to take risks? Where would you put yourself on this scale from 0-"very unwilling to take risks" to 10-"very willing to take risks"? Please click the blue bar to reveal the slider.

very unwilling to take risks | very willing to take risks
0 — 10

Your value: **5**

*[Button: Continue]*

### *Screen 13*

### Questionnaire (6/10)

Please imagine the following situation: You can choose between a sure payment of a particular amount of money, **or** a lottery, where you would have an equal chance of getting 300 GBP or getting nothing. We will present to you five different situations.

1) What would you prefer: **a 50 percent chance of winning 300 GBP** when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of **160 GBP as a sure payment**?

- ◯ lottery
- ◯ sure payment

*[Button: Continue]*

### *Screen 14*

### Questionnaire (6/10)

Please imagine the following situation: You can choose between a sure payment of a particular amount of money, **or** a lottery, where you would have an equal chance of getting 300 GBP or getting nothing. We will present to you five different situations.

3) What would you prefer: **a 50 percent chance of winning 300 GBP** when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of **80 GBP as a sure payment**?

- ◯ lottery
- ◯ sure payment

*[Button: Continue]*

### *Screen 15*

### Questionnaire (6/10)

Please imagine the following situation: You can choose between a sure payment of a particular amount of money, **or** a lottery, where you would have an equal chance of getting 300 GBP or getting nothing. We will present to you five different situations.

4) What would you prefer: **a 50 percent chance of winning 300 GBP** when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of **60 GBP as a sure payment**?

- ◯ lottery
- ◯ sure payment

*[Button: Continue]*

### *Screen 16*

### Questionnaire (6/10)

Please imagine the following situation: You can choose between a sure payment of a particular amount of money, **or** a lottery, where you would have an equal chance of getting 300 GBP or getting nothing. We will present to you five different situations.

5) What would you prefer: **a 50 percent chance of winning 300 GBP** when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of **50 GBP as a sure payment**?

- ◯ lottery
- ◯ sure payment

*[Button: Continue]*

### *Screen 17*

### Questionnaire (7/10)

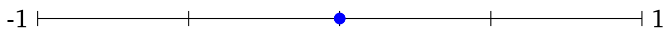
For each of the topics listed below, tell us whether **most people** think that men or women, on average, know more about it.

Indicate your answer on the scale below, where 0 means no gender difference.

The bigger the gender difference, the more you should move the slider in that direction.

### Arts & Literature

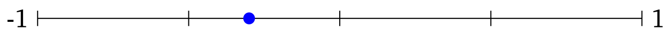
(*-1: Women know more, on average; 0: no difference; 1: Men know more, on average*)




Your value: **0.0**

### Science & Technology

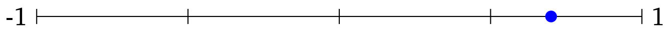
(*-1: Women know more, on average; 0: no difference; 1: Men know more, on average*)




Your value: **-0.3**

### Psychology & Linguistics

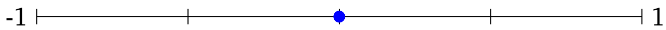
(*-1: Women know more, on average; 0: no difference; 1: Men know more, on average*)




Your value: **0.7**

### History & Politics

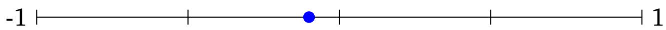
(*-1: Women know more, on average; 0: no difference; 1: Men know more, on average*)




Your value: **0.0**

### Pop Culture

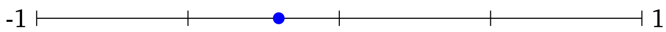
(*-1: Women know more, on average; 0: no difference; 1: Men know more, on average*)




Your value: **-0.1**

### Sports

(*-1: Women know more, on average; 0: no difference; 1: Men know more, on average*)




Your value: **-0.2**

*[Button: Continue]*

### *Screen 18*

### Questionnaire (8/10)

For each statement, please indicate your agreement with it on a scale from **1 - strongly disagree** to **7 - strongly agree**. There are **no right or wrong answers**.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Reputation is very important for one's career advancement and promotions. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Seeking advice from others can hurt my reputation if others have low expectations about my ability. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| I do not care what others think of me. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| It is particularly uncomfortable to ask for advice on a work-related task that others think I am competent at. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| In general, it is more socially acceptable for women to ask for advice than for men. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| I feel bad if I cannot accomplish tasks independently. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Seeking advice from others can hurt my reputation if others have high expectations about my ability. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Please select '6' in this row. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| On average, men are less willing to ask others for advice than women. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |

*[Button: Continue]*

***Screen 19***

**Questionnaire (9/10)**

What would best describe your colleagues at your current workplace? If you are currently unemployed, please answer with respect to your most recent workplace.

◯ By far most of my colleagues are men

◯ A somewhat bigger share of my colleagues are men

◯ Among my colleagues, the share of men and women is about equal

◯ A somewhat bigger share of my colleagues are women

◯ By far most of my colleagues are women

◯ I have no colleagues

Does your job require working in teams? If you are unemployed, please answer with respect to your most recent job.

◯ Always

◯ Mostly

◯ Balanced shares of team & individual work

◯ Rarely

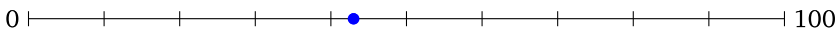
◯ Never

*[Button: Continue]*

***Screen 20***

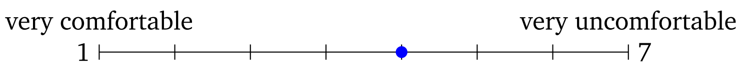
**Questionnaire (10/10)**

When answering a quiz like today's, how important is the *role of luck, as opposed to knowledge*, in getting the correct answer? Please indicate the role of luck on this scale from 0% (no luck) to 100% (only luck).




Your value: **43**

How comfortable were you with the manager controlling a large part of your earnings for this study? Where would you put yourself on this scale from 1-"very comfortable" to 7-"very uncomfortable"?




Your value: 4

When answering the quiz, did you look up answers in any way or ask someone for help? Please respond truthfully to this question. It has NO impact on your payment or your future invitations to participate in studies on Prolific.

◯ Yes

◯ No

Did you have enough time to complete the tasks?

◯ Yes, more than enough time

◯ Yes, just enough time

◯ No, just not enough time

◯ No, by far not enough time

Were all 10 images in the quiz and the screenshots of the profiles displayed correctly? If not, please briefly describe what was wrong in the comments field below.

◯ Yes

◯ No

Do you have any comments about the content of this study you would like to share with us?

[ ]

*[Button: Continue]*

***Screen 21***

**Overview of Your Earnings**

Your participation pay is **1.50 GBP**.

Here is an overview of your extra bonus payments:

- ***Quiz Performance & Advice Seeking***
  - Your performance in Part 1 is 1 correct answer. You did not seek advice on the quiz.
  - Hence, your bonus payment from this part of the study amounts to **0.10 GBP**.
- ***Pay for Manager's Estimate***
  - The manager will review your profile shortly and, depending on the manager's estimate of your quiz performance without advice in Part 1, you will receive a pay ranging from **0.00 GBP to 3.00 GBP**.

- ***Correct-guess-pay***
  - You estimated that you answered 6 questions correctly in Part 1. Since you answered 1 question correctly, you do not receive the correct-guess-pay of **0.25 GBP**.
  - You provided two guesses about the manager's estimate. **Guess 2** has been randomly selected by the computer as the one-that-counts for payment. You will receive an additional **0.25 GBP** if your guess was correct.

**Summary**

Your **minimum payment** is **1.60 GBP**, and it can increase **up to 4.85 GBP**, depending on your manager's estimate and your guess of the other employee's behavior. The payment will be administered within the next 4 working days.

If you want, you can find the correct answers to the 10 general knowledge questions on Psychology and Linguistics **here**.

*[Button: Continue]*

### A.3.2 Manager-Version

***Screen 1***

**Welcome!**

**Overview:** You are participating in a research study on economic decision-making. It consists of two main parts in which you make decisions for an extra bonus payment and a final questionnaire. Your bonus payment depends on your decisions and the decisions of other participants throughout the study. Therefore, it is important that you read all instructions carefully. Your responses are anonymous.

**Payment:** You will receive a participation payment of **1.00 GBP.** In addition, you can earn an extra bonus payment of up to **4.50 GBP.** The payment will be processed within the next 4 working days. Please be advised that only complete submissions will be paid.

**General Rules of Conduct:** This study will take about 10 minutes. We ask for your full attention. Please find a quiet space to complete the study and do not use other devices, talk to other people, use social media, etc. Please remain solely in this browser tab for the entire time of the study.

**Comprehension Checks:** During the study, you will be asked several comprehension questions referring to the instructions on the same screen. You can only proceed once you answer them correctly.

**Consent:** I have read and understood the information above. I agree to comply with these rules of conduct and want to participate in this study.

Please click "next" to proceed.

*[Button: Next]*

### *Screen 2*

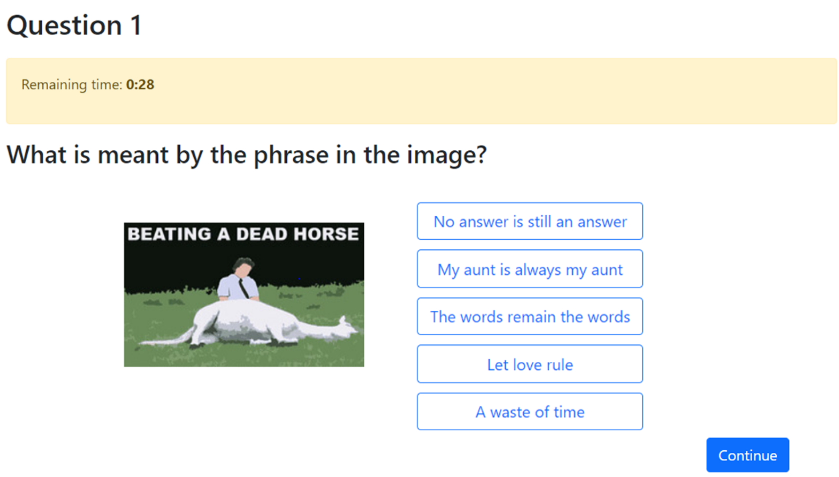
What is your Prolific ID? *Please note that this response should auto-fill with the correct ID.*

### *Screen 3a: Psychology & Linguistics*

### Part 1

In part 1, you will take a general knowledge quiz on **Psychology & Linguistics**. You will answer 10 multiple-choice questions. You will receive **0.10 GBP** for each correct answer and no extra payment for wrong answers.

The questions are all structured in the same way: a question text, a corresponding image, and 5 answer choices, one of which is correct (see an example picture).




To submit an answer, choose one answer choice and click "submit". You will then advance to the next question. You will have **30 seconds** to submit your answer, before you are automatically advanced.

**We encourage you to submit your best guess before the time runs out, even if you do not know the correct answer with certainty.** You have a 20%-chance to simply guess correctly.

### Comprehension Question

Even if I do not know the answer to a question, submitting any guess will increase the chance to receive 0.10 GBP from 0% to 20%.

- ◯ True
- ◯ False

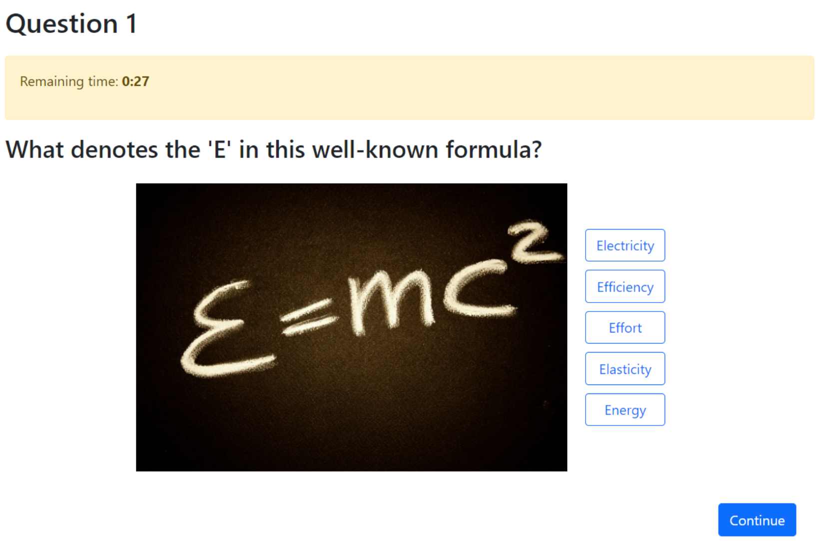
*[Button: Next]*

### *Screen 3b: Science & Technology*

### Part 1: The Quiz

In Part 1, you will take a general knowledge quiz on **Science & Technology.**You will answer 10 multiple-choice questions. You will receive **0.10 GBP** for each correct answer and no extra payment for wrong answers.

The questions are all structured in the same way: a question text, a corresponding image, and 5 possible answers, one of which is correct (see an example picture).




To submit an answer, choose one and click "submit". You will then advance to the next question. You will have **30 seconds** to submit your answer, before you are automatically advanced.

**We encourage you to submit your best guess before the time runs out, even if you do not know the correct answer with certainty.** You have a 20%-chance to simply guess correctly.

### Comprehension Question

Even if I do not know the answer to a question, submitting any guess will increase the chance to receive 0.10 GBP from 0% to 20%.

◯ True

◯ False

*[Button: Next]*

### *Screen 4a: Psychology & Linguistics*

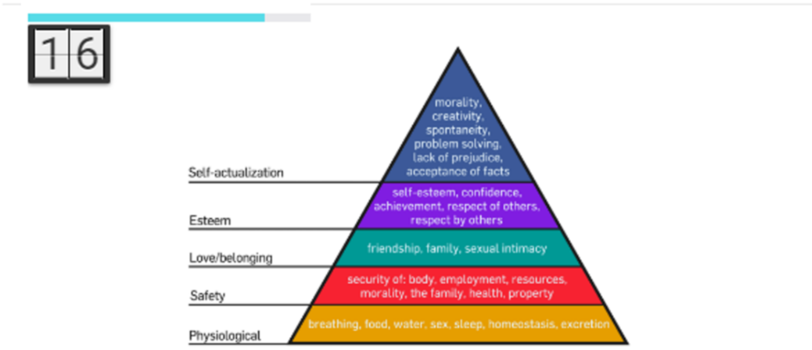
*[Screen 4b looks identical for quiz type "Science & Technology"]*

### Start the Quiz

When you are ready, click "next" to begin the general knowledge quiz on Psychology & Linguistics.
*[Button: Next]*

### *Screen 5a*

*[Question 1 out of 10 Questions of the Psychology & Linguistics quiz. For each question, its text referred to a picture. The picture was instrumental to providing a correct answer (see Figure A1 for an example). The pictures made it essentially impossible to search online for the correct answer given the time limit of 30 seconds.]*




What is the basic idea behind this well-known psychological theory?

- ◯ To preserve this psychologist's theory forever
- ◯ It depicts that everyone has desires
- ◯ The road to enlightenment consists of many small steps
- ◯ It shows the three factors underlying self-actualization, one on each corner
- ◯ As one takes care of basic needs, the "higher needs" become more relevant

### *Screen 5b*

*[Question 1 out of 10 Questions of the Psychology & Linguistics quiz. For each question, its text referred to a picture. The picture was instrumental to providing a correct answer (see Figure A1 for an example). The pictures made it essentially impossible to search online for the correct answer, given the time limit of 30 seconds.]*

16

This physicist is most well-known for his research on . . . ?

- ◯ Higgs boson
- ◯ Condensed matter
- ◯ Black moons
- ◯ Supermassive compact objects
- ◯ Black holes

*[Button: Next]*

### *Screen 6a*

**Guess, how many of your answers are correct?**

You just answered 10 questions on Psychology & Linguistics. How many of your answers are correct? You will receive **an extra correct-guess-pay of 0.25 GBP** if your guess equals your actual number of correct answers.

Please enter a number from 0 (no answers correct) to 10 (all correct answers):

At the end of the study, you will see how many of your answers are correct. You will also have the option to learn the correct answers if you want.
*[Button: Next]*

### *Screen 6b*

**Guess, how many of your answers are correct?**

You just answered 10 questions on Science & Technology. How many of your answers are correct? You will receive **an extra correct-guess-pay of 0.25 GBP** if your guess equals your actual number of correct answers.

Please enter a number from 0 (no answers correct) to 10 (all correct answers):

At the end of the study, you will see how many of your answers are correct. You will also have the option to learn the correct answers if you want.
*[Button: Next]*

### *Screen 7a*

**Part 2**

In part 2, you are asked to **estimate how many questions another participant answered correctly in the same quiz** you just took. We will refer to this participant as the *quiz-taker*. The quiz-taker is another Prolific participant who completed a different version of this study.

For each correctly answered question, the quiz-taker also got **0.10 GBP**. They did not receive any feedback on their quiz performance.

Afterwards, quiz-takers could decide whether to seek advice on the **Psychology & Linguistics quiz** and **revise their answers**.

- If a quiz-taker **decided to seek advice**, they revisited every question with only 2 instead of 5 answer choices. For this simplified quiz, the computer had randomly removed 3 wrong answers to each question. A quiz-taker had 15 seconds to review and revise each question.
- If a quiz-taker **did not seek advice**, they proceeded to the final questionnaire.

Seeking advice cost a **one-time fee of 0.08 GBP**. Advice could help a quiz-taker to submit more correct answers and increase their payment. They received **0.10 GBP**, regardless of whether they submitted correct answers before or after advice.

When taking the quiz for the first time, quiz-takers did *not know* that they could later *revise* their answers in the simplified quiz with advice.

### *Screen 7b*

**Part 2**

In part 2, you are asked to **estimate how many questions another participant answered correctly in the same quiz** you just took. We will refer to this participant as the *quiz-taker*. The quiz-taker is another Prolific participant who completed a different version of this study.

For each correctly answered question, the quiz-taker also got **0.10 GBP**. They did not receive any feedback on their quiz performance.

Afterwards, quiz-takers could decide whether to seek advice on the **Science & Technology quiz** and **revise their answers**.

- If a quiz-taker **decided to seek advice**, they revisited every question with only 2 instead of 5 answer choices. For this simplified quiz, the computer had randomly removed 3 wrong answers to each question. A quiz-taker had 15 seconds to review and revise each question.

- If a quiz-taker **did not seek advice**, they proceeded to the final questionnaire.

Seeking advice cost a **one-time fee of 0.08 GBP**. Advice could help a quiz-taker to submit more correct answers and increase their payment. They received **0.10 GBP**, regardless of whether they submitted correct answers before or after advice.

When taking the quiz for the first time, quiz-takers did *not know* that they could later *revise* their answers in the simplified quiz with advice.

### *Screen 8a*

### Your Task: Estimate Quiz-Taker's Quiz Performance without Advice

Your task is to estimate the prior knowledge of a quiz-taker on Psychology & Linguistics, as recorded with their quiz performance without advice.

You were *randomly* matched with the quiz-taker by a computer. To make your estimate you will see the quiz-taker's profile.

### Your Estimate and Payment

Your estimate affects **your own payment and the payment of the quiz-taker** as follows:

If your estimate equals the number of correct answers that the quiz-taker submitted without advice, you will receive an estimation **bonus of 3 GBP**. This estimation bonus will be added to your final payment.

**The quiz-taker** will receive **0.30 GBP** times your estimate of their quiz performance without advice, regardless of whether your estimate is correct. This was known to the quiz-taker. The quiz-taker also knew what profile you will see of them, before deciding whether to seek advice.

Your estimate may affect the payment of more than one quiz-taker with identical profiles. Yet, the quiz performance of only one of these randomly matched quiz-takers determines if you receive the estimation bonus.

### Comprehension Questions

Your task is to estimate how many correct answers (0-10) the quiz-taker submitted…

◯ ...without advice

◯ ....after the revision with advice

Is this statement true or false?

To get the estimation bonus of 3 GBP, you need to provide a correct estimate of the quiz-taker's quiz performance without advice.

◯ True

◯ False

***Screen 8b***

### Your Task: Estimate Quiz-Taker's Quiz Performance without Advice

Your task is to estimate the prior knowledge of a quiz-taker on Science & Technology, as recorded with their quiz performance without advice.

You were *randomly* matched with the quiz-taker by a computer. To make your estimate you will see the quiz-taker's profile.

### Your Estimate and Payment

Your estimate affects **your own payment and the payment of the quiz-taker** as follows:

If your estimate equals the number of correct answers that the quiz-taker submitted without advice, you will receive an estimation **bonus of 3 GBP**. This estimation bonus will be added to your final payment.

**The quiz-taker** will receive **0.30 GBP** times your estimate of their quiz performance without advice, regardless of whether your estimate is correct. This was known to the quiz-taker. The quiz-taker also knew what profile you will see of them, before deciding whether to seek advice.

Your estimate may affect the payment of more than one quiz-taker with identical profiles. Yet, the quiz performance of only one of these randomly matched quiz-takers determines if you receive the estimation bonus.

### Comprehension Questions

Your task is to estimate how many correct answers (0-10) the quiz-taker submitted…

◯ ...without advice

◯ ....after the revision with advice

Is this statement true or false?

To get the estimation bonus of 3 GBP, you need to provide a correct estimate of the quiz-taker's quiz performance without advice.

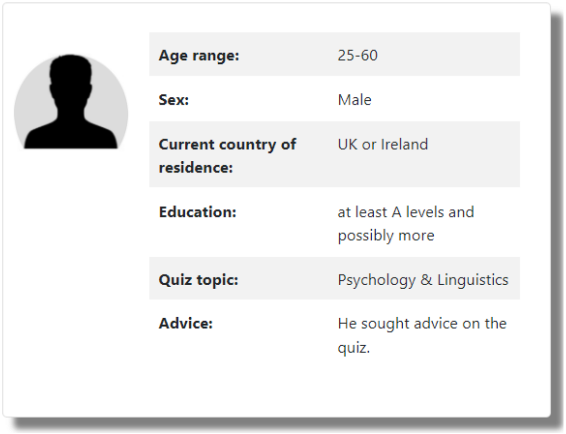
◯ True

◯ False

### *Screen 9a*

**Your performance estimate**

You have been randomly matched with a quiz-taker with the following profile:



| | |
|---|---|
| **Age range:** | 25-60 |
| **Sex:** | Male |
| **Current country of residence:** | UK or Ireland |
| **Education:** | at least A levels and possibly more |
| **Quiz topic:** | Psychology & Linguistics |
| **Advice:** | He sought advice on the quiz. |

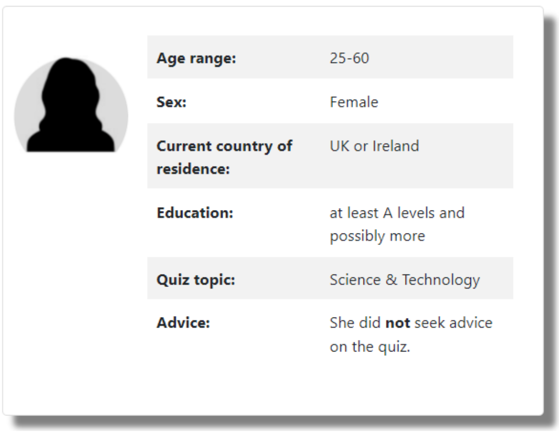
How many correct answers did the quiz-taker submit without advice? If your estimate is correct, you will get the estimation bonus of **3 GBP**.

Please enter a number from 0 (no correct answers) to 10 (all answers correct):

### *Screen 9b*

**Your performance estimate**

You have been randomly matched with a quiz-taker with the following profile:



| | |
|---|---|
| **Age range:** | 25-60 |
| **Sex:** | Female |
| **Current country of residence:** | UK or Ireland |
| **Education:** | at least A levels and possibly more |
| **Quiz topic:** | Science & Technology |
| **Advice:** | She did **not** seek advice on the quiz. |

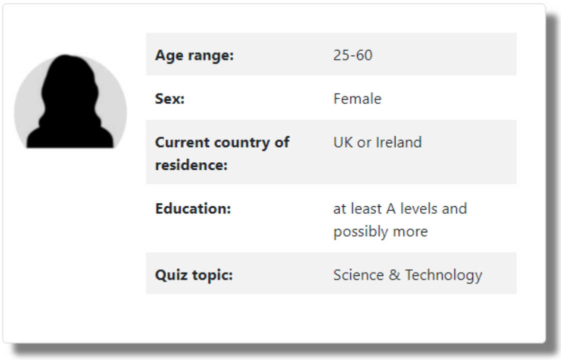
How many correct answers did the quiz-taker submit without advice? If your estimate is correct, you will get the estimation bonus of **3 GBP**.

Please enter a number from 0 (no correct answers) to 10 (all answers correct):

### *Screen 9c*

**Your performance estimate**

You have been randomly matched with a quiz-taker with the following profile:




How many correct answers did the quiz-taker submit without advice? If your estimate is correct, you will get the estimation bonus of **3 GBP**.

Please enter a number from 0 (no correct answers) to 10 (all answers correct):

### *Screen 9d*

**Your performance estimate**

You have been randomly matched with a quiz-taker with the following profile:


How many correct answers did the quiz-taker submit without advice? If your estimate is correct, you will get the estimation bonus of **3 GBP**.

Please enter a number from 0 (no correct answers) to 10 (all answers correct):

*[Note that screen 9a – 9d exist all with opposite Employee sex and quiz category. Screen 91 & 9b additionally exist with opposite advice-seeking decisions amounting to a total of 12 different screens (i.e., treatment conditions).]*

### *Screen 10*

**Questionnaire**

This short questionnaire is the final part. Afterwards, you will see an overview of your earnings and can learn the correct answers.

*[Button: Next]*

### *Screen 11*

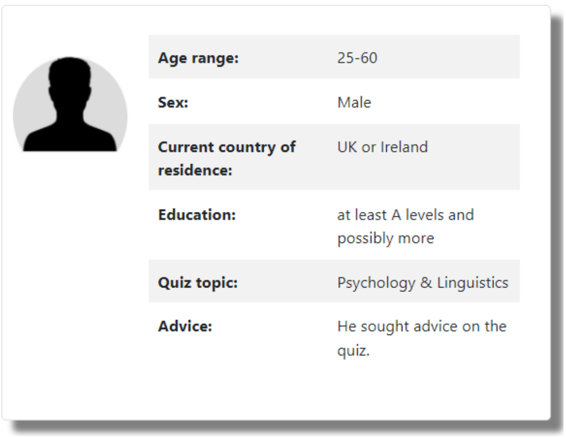
We would like to understand how you arrived at your estimate of the quiz-taker's quiz performance without advice. Please, briefly describe your thought process:

*[Button: Next]*

### *Screen 12a*

*[exists for all different 12 profile versions, see screen 9).]*

The next two questions refer to a quiz-taker with this profile:




For quiz-takers, seeking advice cost a one-time fee of 0.08 GBP and could increase paid performance in the quiz by 0.10 GBP per correct answer. After advice, each question had only 2 answer choices instead of 5.

In your opinion, how many answers must a quiz-taker with this profile *not* know to decide to seek advice on the Psychology & Linguistics quiz?

Please enter a number from 0 (no answers) to 10 (all answers):

What do you think is the likelihood that a quiz-taker with this profile cheated on this quiz (e.g., by googling answers or asking others for help)?

*Please indicate the likelihood on this scale from 0 % to 100 %.*

0% 50% 100%

### *Screen 13*

In your opinion, how big is the role of luck, as opposed to knowledge, in answering a multiple-choice general knowledge quiz like the one you took today?

*Please indicate the role of luck on this scale from 0 % (no luck) to 100 % (only luck).*

0%- no luck 50% 100%- only luck

### *Screen 14*

For each of the topics listed below, tell us whether **you** believe that **men or women**, on average, **know more** about it. There are no right or wrong answers.

Indicate your answer on the scale below, where 0 means no gender difference. The bigger the gender difference, the more you should move the slider in that direction.

Please note that you must move the slider in any case to validate your response. The slider has to be moved even if you want to place your response at the original position.

### Arts & Literature

(*-1: Women know more, on average; 0: no difference; 1: Men know more, on average*)

-1 0 1

### Science & Technology

(*-1: Women know more, on average; 0: no difference; 1: Men know more, on average*)

-1 0 1

**Psychology & Linguistics**

(*-1: Women know more, on average; 0: no difference; 1: Men know more, on average*)

-1 0 1

**History & Politics**

(*-1: Women know more, on average; 0: no difference; 1: Men know more, on average*)

-1 0 1

**Pop Culture**

(*-1: Women know more, on average; 0: no difference; 1: Men know more, on average*)

-1 0 1

**Sports**

(*-1: Women know more, on average; 0: no difference; 1: Men know more, on average*)

-1 0 1

*[Button: Next]*

### *Screen 15*

Please evaluate several statements in terms of how well they apply to you or others in general. For each statement, please indicate whether you agree with it or not on the provided scale (ranging from **1 = strongly disagree** to **7 = strongly agree**). Keep in mind that there are no right or wrong answers.

| | strongly disagree | disagree | somewhat disagree | neutral | somewhat agree | agree | strongly agree |
|---|---|---|---|---|---|---|---|
| Reputation is very important for one's career advancement and promotions. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Seeking advice from others can hurt my reputation if others have low expectations about my ability. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| I do not care what others think of me. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| It is uncomfortable to ask for advice on a work-related task that others think I am competent at. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| In general, it is more socially acceptable for women to ask for advice more than men. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| I feel bad if I cannot accomplish tasks independently. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Seeking advice from others can hurt my reputation if others have high expectations about my ability. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Please select 'agree'. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| On average, men are less willing to ask others for advice than women. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |

*[Button: Next]*

### *Screen 16*

Finally, please answer a few questions about yourself. Afterwards, you will see an overview of your earnings.

How old are you?

[ ]

What gender are you currently?

- ◯ Man (including Trans Male/Trans Man)
- ◯ Woman (including Trans Female/Trans Woman)
- ◯ Non-binary
- ◯ Would rather not say

Were you a university student at some point in time during your life, including current enrollment?

- ◯ Yes
- ◯ No

Are you currently employed?

◯ Yes, full time

◯ Yes, part time

◯ Self-employed

◯ No

*[Button: Next]*

### *Screen 17*

What is your current occupation?

[ ]

Do you have **managerial responsibilities** at your current employment that would involve, for instance, directly supervising others, the ability to hire or terminate other employees, etc.?

◯ Yes

◯ No

*[Button: Next]*

### *Screen 18*

Do you have any comments for us?

[ ]

*[Button: Next]*

### *Screen 19*

Thank you for participating in our study.

### Payment

Your participation pay is **1 GBP**.

You answered 10 questions correctly and receive **1 GBP** for answering the quiz.

You guessed that you answered 8 questions correctly, therefore you **do not** receive 0.25 GBP.

The computer will compare your estimate of the quiz-taker's quiz performance with the quiz-taker's actual data.

Therefore, *should your estimate be correct*, your total earnings from this study will be **5 GBP**.

Your payment will be administered within the next 4 working days. Please click the button to finish the study.

If you want, you can find the correct answers to the 10 general knowledge questions on Science & Technology here.

*[Button: Complete]*

### A.3.3 Advice-Seeking-at-Work Survey

#### *Screen 1*

**Welcome!**

In this survey you will encounter questions about yourself and your professional behaviors. Your responses are anonymous. The survey will take about 5 minutes and you will receive a participation payment of £${e://Field/participation_pay}. Please be advised that only complete submissions will be paid.

**General Rules:** Please give your full attention and find a quiet space. Avoid using other devices social media or having conversations during the study.

**Verification:** You will encounter two attention checks. Passing at least one of them is required for payment.

**Consent:** I have read and understood the information above. I agree to comply with these rules of conduct and want to participate in this study. Please indicate your choice:

◯ I agree

◯ I do not agree

What is your Prolific ID? Please note that this response should auto-fill with the correct ID.

[ ]

*[Button: Next]*

#### *Screen 2*

To begin please answer the following questions about yourself.

What is your current age in years?

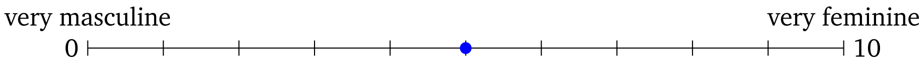
[ ]

What is your sex?

◯ Male

◯ Female

Generally speaking how do you see yourself where would you put yourself on this scale from 0=very masculine to 10=very feminine?




Do you have a job outside of taking surveys?

◯ Yes: full-time (35+ hours a week)

◯ Yes: part-time (less than 35 hours a week)

◯ No: homemaker

◯ No: currently seeking employment

◯ No: student

◯ No: retired

◯ No: other

Are you self-employed? *[Display if Do you have a job outside of taking surveys? is Yes]*

◯ Yes

◯ No

Do you have any experience being in a management position?

◯ Yes

◯ No

*[Button: Next]*

### *Screen 3*

In which sector do you work? *[Display if Do you have a job outside of taking surveys? is Yes]*

◯ Public sector

◯ Private sector

◯ Not-for-profit sector

◯ Other

In which sector did you most recently work? *[Display if Do you have a job outside of taking surveys? is No]*

◯ Public sector

◯ Private sector

◯ Not-for-profit sector

◯ Other

Which of the following categories best describes the industry you primarily work in? *[Display if Do you have a job outside of taking surveys? is Yes]*

◯ Agriculture Forestry Fishing and Hunting

◯ Mining Quarrying and Oil and Gas Extraction

◯ Utilities

◯ Construction

◯ Manufacturing

◯ Wholesale Trade

◯ Retail Trade

◯ Transportation and Warehousing

◯ Information

◯ Finance and Insurance

◯ Real Estate and Rental and Leasing

◯ Professional Scientific and Technical Services

◯ Management of Companies and Enterprise

◯ Administrative and Support and Waste Management and Remediation Services

◯ Educational Services

◯ Health Care and Social Assistance

◯ Arts Entertainment and Recreation

◯ Accommodation and Food Services

◯ Other Services (except Public Administration)

◯ Public Administration

Which of the following categories best describes the industry your most recent employment was primarily in? *[Display if Do you have a job outside of taking surveys? is No]*

◯ Agriculture Forestry Fishing and Hunting

◯ Mining Quarrying and Oil and Gas Extraction

◯ Utilities

◯ Construction

◯ Manufacturing

◯ Wholesale Trade

◯ Retail Trade

◯ Transportation and Warehousing

◯ Information

◯ Finance and Insurance

◯ Real Estate and Rental and Leasing

◯ Professional Scientific and Technical Services

◯ Management of Companies and Enterprise

◯ Administrative and Support and Waste Management and Remediation Services

◯ Educational Services

◯ Health Care and Social Assistance

◯ Arts Entertainment and Recreation

◯ Accommodation and Food Services

◯ Other Services (except Public Administration)

◯ Public Administration

For how long have you been working for your current employer? *[Display if Do you have a job outside of taking surveys? is Yes]*

◯ less than a year

◯ 1-2 years

◯ 2-5 years

◯ 5-10 years

◯ 10+ years

For how long did you work for your most recent employer? *[Display if Do you have a job outside of taking surveys? is No]*

- ◯ less than a year
- ◯ 1-2 years
- ◯ 2-5 years
- ◯ 5-10 years
- ◯ 10+ years

How many people work at your organization? *[Display if Do you have a job outside of taking surveys? is Yes]*

- ◯ 1-9
- ◯ 10-49
- ◯ 50-249
- ◯ 250-999
- ◯ More than 1000
- ◯ I don't know

How many people worked at your most recent employer? *[Display if Do you have a job outside of taking surveys? is No]*

- ◯ 1-9
- ◯ 10-49
- ◯ 50-249
- ◯ 250-999
- ◯ More than 1000
- ◯ I don't know

How many colleagues work in the same organizational unit as you? *[Display if Do you have a job outside of taking surveys? is Yes]*

- ◯ Up to 5 people
- ◯ Between 5 and 10 people
- ◯ Between 10 and 25 people
- ◯ More than 25 people

◯ Not applicable

How many colleagues worked in the same organizational unit as you at your most recent employer? *[Display if Do you have a job outside of taking surveys? is No]*

◯ Up to 5 people

◯ Between 5 and 10 people

◯ Between 10 and 25 people

◯ More than 25 people

◯ Not applicable

Does your job require working in teams? *[Display if Do you have a job outside of taking surveys? is Yes]*

◯ Always

◯ Mostly

◯ Balanced share of team & individual work

◯ Rarely

◯ Never

Did your previous job require working in teams? *[Display if Do you have a job outside of taking surveys? is No]*

◯ Always

◯ Mostly

◯ Balanced share of team & individual work

◯ Rarely

◯ Never

What is the typical team size at your job? *[Display if Do you have a job outside of taking surveys? is Yes]*

◯ 2-3 people

◯ 3-5 people

◯ 5-10 people

◯ 10-25 people

◯ More than 25 people

◯ Not applicable

What was the typical team size at your most recent job? *[Display if Do you have a job outside of taking surveys? is No]*

◯ 2-3 people

◯ 3-5 people

◯ 5-10 people

◯ 10-25 people

◯ More than 25 people

◯ Not applicable

*[Button: Next]*

### *Screen 4*

Where do you typically source information that can help you solve work-related challenges on tasks or projects?

*Allocate 100% among these options to show how often you use each. Enter 0 for an option that you never use.*

Colleagues : [ ]

Superiors : [ ]

Online Resources and Search Engines : [ ]

Experts (external or internal) : [ ]

Internal Resources (e.g. knowledge repositories forums internal documentation) : [ ]

Books and Publications : [ ]

Total : [ ]

Have you ever chosen to solve a work-related challenge on your own (e.g. through online resources or search engines), *even though you knew that asking someone for advice would have been quicker?*

◯ Yes

◯ No

What are typically the reasons for choosing to solve a work-related challenge on your own (e.g. through online resources or search engines), *even though you know that asking someone for advice would be quicker?* Select all that apply. [*Display if Have you ever chosen to solve a work-related challenge on your own... is Yes*]

◯ Fear of judgement

◯ Skill development or learning

◯ Self-reliance

◯ Workplace culture

◯ Fear of rejection

◯ Time constraints (either for you or the person you could ask for help)

◯ Personal or interpersonal barriers

If an important reason is missing from the list please enter it here. [*Display if Have you ever chosen to solve a work-related challenge on your own... is Yes*]

[ ]

Generally speaking why do you ask someone for advice on a work-related challenge?

*Rank the following reasons from the most important (rank 1) to the least important (rank 6).*

◯ ______ Accessing knowledge or experience

◯ ______ Learning and development

◯ ______ Seeking feedback or reassurance

◯ ______ Relationship building

◯ ______ Showing engagement

◯ ______ Time constraints

Does this list miss an important reason? If yes please write it here.

[ ]

Think about how much advice you seek at work. Are you in general satisfied with it?

◯ Yes

◯ No I could seek more

◯ No I could seek less

*[Button: Next]*

### *Screen 5*

Please evaluate several statements in terms of how well they apply to you or others in general on a scale from 1 = strongly disagree to 7 = strongly agree.

There are no right or wrong answers.

| | strongly disagree | disagree | somewhat disagree | neither agree/ nor disagree | somewhat agree | agree | strongly agree |
|---|---|---|---|---|---|---|---|
| Reputation is very important for one's career advancement and promotions. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Seeking advice from others can hurt my reputation if others have low expectations about my ability. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| I do not care what others think of me. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| It is uncomfortable to ask for advice on a work-related task that others think I am competent at. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| In general, it is more socially acceptable for women to ask for advice than for men. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| I feel bad if I cannot accomplish tasks independently. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Seeking advice from others can hurt my reputation if others have high expectations about my ability. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| I currently do not pay attention to the questions I am being asked in the survey. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| On average, men are less willing to ask others for advice than women. | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

*[Button: Next]*

***Screen 6***

Think about the organization you are currently primarily working for. Please indicate the extent to which each of the following statements applies to this organization on a scale from **1 = very inaccurate to 7 = very accurate.** *[Display if Do you have a job outside of taking surveys? is Yes]*

| | very inaccurate | inaccurate | somewhat inaccurate | neither inaccurate nor accurate | somewhat accurate | accurate | very accurate | NA |
|---|---|---|---|---|---|---|---|---|
| If you make a mistake it is often held against you. | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Members of this organizationare able to bring up problems and tough issues. | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| People in this organization sometimes reject others for being different. | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| It is safe to take risks in this organization. | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Please select accurate in this row. | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| It is difficult to ask other members of this organization for help. | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| No one in this organization would deliberately act in a way that undermines my efforts. | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| At my organization, my unique skills and talents are valued and utilized. | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Think about the organization you are currently primarily working for. Please indicate the extent to which each of the following statements applies to this organization on a scale from **1 = very inaccurate to 7 = very accurate.** *[Display if Do you have a job outside of taking surveys? is No]*

| | very inaccurate | inaccurate | somewhat inaccurate | neither inaccurate nor accurate | somewhat accurate | accurate | very accurate | NA |
|---|---|---|---|---|---|---|---|---|
| If you make a mistake it is often held against you. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Members of this organization are able to bring up problems and tough issues. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| People in this organization sometimes reject others for being different. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| It is safe to take risks in this organization. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Please select accurate in this row. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| It is difficult to ask other members of this organization for help. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| No one in this organization would deliberately act in a way that undermines my efforts. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| At my organization, my unique skills and talents are valued and utilized. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |

*[Button: Next]*

### *Screen 7*

Do you have any comments for us?

Thank you for completing the survey. We will validate your submission and process the payment through Prolific within 4 working days.

*[Button: Complete]*

# B

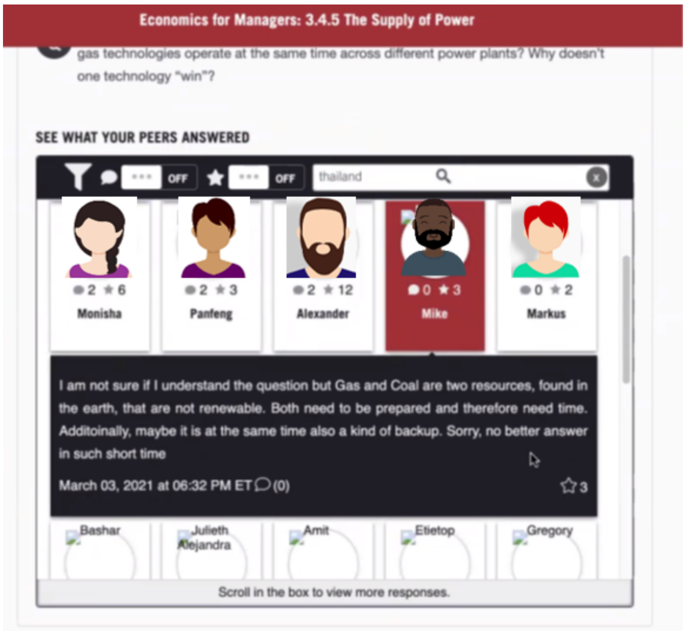
# Appendix to Chapter 2

*Breaking the Ice: Can Initial Peer Activity Enhance Platform Engagement and Persistence?*

## B.1 Additional Figures and Tables

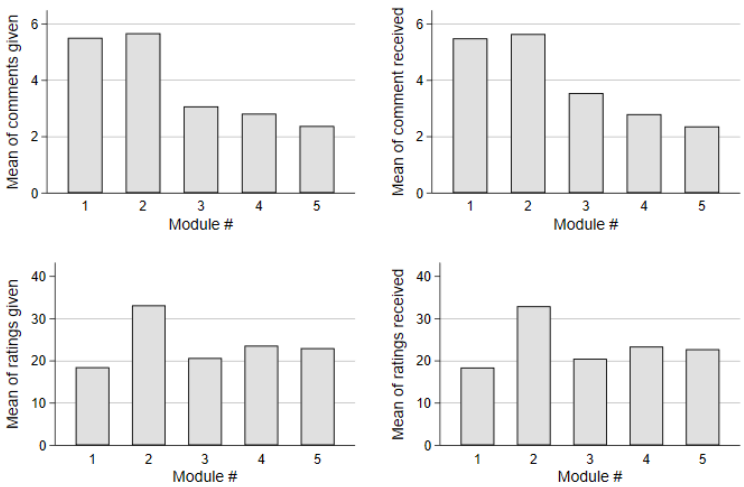
## B.2 Figures

**Figure B.1:** Interaction frequencies by type and module




**Notes.** On the platform, users upload actual headshots of themselves visible to their peers. We replaced them manually with icons for data protection purposes.

**Figure B.2:** Interaction frequencies by type and module




**Notes.** All 3 courses combined and modules 1-5 aggregated.

## B.3 Tables

**Table B.1:** Summary Statistics

| | N | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|---|
| *Dependent Variables* | | | | | | |
| Comment(s) given in m2-7 (1/0) | 12,687 | 0.66 | 1 | 0.47 | 0 | 1 |
| Like(s) given in m2-7 (1/0) | 12,687 | 0.85 | 1 | 0.36 | 0 | 1 |
| Persistence (# modules) | 12,687 | 15.64 | 17 | 3.64 | 1 | 17 |
| *Independent Variables* | | | | | | |
| *A: Peer Interaction Variables* | | | | | | |
| High # comment givers (1/0) | 12,687 | 0.53 | 1 | 0.50 | 0 | 1 |
| High # like givers (1/0) | 12,687 | 0.62 | 1 | 0.48 | 0 | 1 |
| Comment(s) received m1 (1/0) | 12,687 | 0.83 | 1 | 0.38 | 0 | 1 |
| Like(s) received m1 (1/0) | 12,687 | 0.94 | 1 | 0.23 | 0 | 1 |
| *B: Control Variables* | | | | | | |
| Female (%) | 12,687 | 0.40 | 0 | 0.49 | 0 | 1 |
| Age (yrs) | 12,687 | 33.98 | 32 | 8.49 | 20 | 76 |
| US resident (1/0) | 12,687 | 0.57 | 1 | 0.50 | 0 | 1 |
| Official language English (1/0) | 12,687 | 0.62 | 1 | 0.49 | 0 | 1 |
| # same age peers (+/-2 yrs) | 12,687 | 87.53 | 80 | 54.79 | 1 | 290 |
| # peers from same country | 12,687 | 134.44 | 170 | 119.54 | 1 | 392 |
| # peers w/ same gender | 12,687 | 197.19 | 195 | 53.13 | 87 | 293 |
| # peers w/same citizenship | 12,687 | 90 | 60 | 87.15 | 1 | 303 |
| Cohort size (N) | 12,687 | 384.11 | 376 | 73.60 | 271 | 566 |
| Submission before deadline (h) | 12,687 | 43.77 | 18.63 | 74.15 | 0 | 3022.47 |
| Quiz score in fin. accounting m1 | 12,687 | 8.45 | 85 | 16.62 | 0 | 100 |

-

**Table B.2:** Summary Statistics of Peer Interaction Variables

| | N | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|---|
| *A: Cohort-level Variables* | | | | | | |
| High # comment givers (1/0) | 12,687 | 0.53 | 1 | 0.50 | 0 | 1 |
| Module 1 %-share comment givers | 12,687 | 0.64 | 0.65 | 0.07 | 0.44 | 0.75 |
| High # like givers (1/0) | 12,687 | 0.62 | 1 | 0.48 | 0 | 1 |
| Module 1 %-share like givers | 12,687 | 0.81 | 0.83 | 0.06 | 0.63 | 0.90 |
| *B: User-Level Commenting* | | | | | | |
| Total # of comments received | 12,687 | 22.15 | 12 | 34.65 | 0 | 762 |
| Total # of comments given | 12,687 | 21.80 | 6 | 51.24 | 0 | 1008 |
| Comment(s) given in m2-7 (1/0) | 12,687 | 0.66 | 1 | 0.47 | 0 | 1 |
| Comment(s) received m1 (1/0) | 12,687 | 0.83 | 1 | 0.38 | 0 | 1 |
| *C: User-level Liking* | | | | | | |
| Total # of likes received | 12,687 | 13.47 | 66 | 202.71 | 0 | 3031 |
| Total # of likes given | 12,687 | 131.79 | 56 | 263.48 | 0 | 13 531 |
| # Likes given m2-7 | 12,687 | 112.81 | 43 | 225.98 | 0 | 9542 |
| Like(s) received m1 (1/0) | 12,687 | 0.94 | 1 | 0.23 | 0 | 1 |

**Table B.3:** Main Results controlling for Users' Own Module 1 Interaction Behaviors

| | (1) *Comments given (m2-7)* | (2) *Likes given (m2-7)* | (3) *Persistence (# modules)* |
|---|---|---|---|
| *High # comment givers (1/0)* | 0.002 | −0.009 | 0.069 |
| | (.013) | (.010) | (.113) |
| *High # like givers (1/0)* | −0.041** | 0.001 | −0.563*** |
| | (.018) | (.014) | (.174) |
| *Comment(s) received m1 (1/0)* | 0.056*** | 0.034*** | 0.875*** |
| | (.011) | (.009) | (.113) |
| *Likes(s) received m1 (1/0)* | 0.047*** | 0.094*** | 1.354*** |
| | (.018) | (.017) | (.212) |
| *Comment(s) given m1 (1/0)* | 0.438*** | 0.115*** | 0.260*** |
| | (.010) | (.008) | (.077) |
| *Like(s) given m1 (1/0)* | 0.206*** | 0.413*** | 0.827*** |
| | (.011) | (.012) | (.108) |
| Means Dep. Variable | 0.658 | 0.847 | 15.643 |
| Controls | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.348 | 0.337 | 0.147 |
| N | 12,687 | 12,687 | 12,687 |

**Notes.** The dependent variable in column (1) *Comments given m2-7* equals 1 if any comments were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in column (2) *Likes given m2-m7* equals 1 if any likes were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable column (3) *Persistence* is the number of completed modules of the entire program (0-17). *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. *Comment(s) given m1* equals 1 if users gave any comment(s) in module one and 0 otherwise. *Like(s) given m1* equals 1 if users distributed any like(s) in the first module and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table B.4:** Heterogeneity Analysis: Interaction effect with Below Median Age

| | (1) *Comments given m2-7 (1/0)* | (2) *Likes given m2-7 (1/0)* | (3) *Persistence (# modules)* |
|---|---|---|---|
| *Age below median (1/0)* | −0.112*** | −0.052 | 0.250 |
| | (.038) | (.032) | (.417) |
| *High # comment givers (1/0)* | 0.047** | 0.009 | 0.301 |
| | (.024) | (.025) | (.193) |
| *High # comment givers x Age below median* | −0.045 | −0.058* | −0.361 |
| | (.030) | (.032) | (.226) |
| *High # like givers (1/0)* | −0.053** | −0.034 | −0.755*** |
| | (.027) | (.028) | (.221) |
| *High # like givers x Age below median* | 0.044 | 0.046 | 0.497** |
| | (.031) | (.033) | (.238) |
| *Comment(s) received m1 (1/0)* | 0.206*** | 0.172*** | 1.238*** |
| | (.017) | (.016) | (.157) |
| *Comment(s) received m1 x Age below median* | −0.006 | −0.017 | −0.388* |
| | (.024) | (.022) | (.216) |
| *Like(s) received m1 (1/0)* | 0.080*** | 0.086*** | 1.663*** |
| | (.029) | (.025) | (.325) |
| *Like(s) received m1 x Age below median* | 0.043 | −0.010 | −0.366 |
| | (.039) | (.034) | (.424) |
| Means Dep. Variable | 0.658 | 0.847 | 15.643 |
| Controls | Yes | Yes | Yes |
| Year-FE | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.103 | 0.086 | 0.138 |
| N | 12,687 | 12,687 | 12,687 |

**Notes.** The dependent variable in column (1) *Comments given m2-7* equals 1 if any comments were given in later modules of the program and 0 otherwise. The dependent variable in column (2) *Likes given m2-7* equals 1 if any likes were given in later modules and 0 otherwise. *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. The dependent variable *Persistence* is the number of completed modules of the entire program (0-17). *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

**Table B.5:** Heterogeneity Analysis: Split by Median Age

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Age below Median | | | Age above Median | | |
| | *Comments m2-7* (1/0) | *Likes m2-7* (1/0) | *Persistence* (# modules) | *Comments m2-7* (1/0) | *Likes m2-7* (1/0) | *Persistence* (# modules) |
| *High # comment givers m1 (1/0)* | .019 | -.030 | -.094 | .039* | -.000 | .111 |
| | (.019) | (.020) | (.147) | (.023) | (.024) | (.214) |
| *High # like givers m1* (1/0) | .032 | .014 | .014 | -.070** | -.008 | -.721*** |
| | (.034) | (.035) | (.281) | (.027) | (.029) | (.254) |
| *Comment(s) received m1* (1/0) | .238*** | .191*** | 1.428*** | .265*** | .224*** | 2.109*** |
| | (.018) | (.016) | (.171) | (.016) | (.015) | (.176) |
| *Like(s) received m1* (1/0) | .213*** | .150*** | 3.247*** | .188*** | .167*** | 4.113*** |
| | (.025) | (.020) | (.326) | (.026) | (.020) | (.358) |
| Dep. Variable Mean | .658 | .847 | 15.643 | .658 | .847 | 15.643 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | .074 | .043 | .094 | .080 | .052 | .120 |
| N | 6,061 | 6,061 | 6,061 | 7,266 | 7,266 | 7,266 |

**Notes.** The dependent variable in columns (1) and (4) *Comments given m2-7* equals 1 if any comments were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in columns (2) and (5) *Likes given m2-m7* equals 1 if any likes were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in columns (3) and (6) *Persistence* is the number of completed modules of the entire program (0-17). *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table B.6:** Heterogeneity Analysis: Split by US Residency

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | US residents | | | Non-US residents | | |
| | *Comments m2-7 (1/0)* | *Likes m2-7 (1/0)* | *Persistence (# modules)* | *Comments m2-7 (1/0)* | *Likes m2-7 (1/0)* | *Persistence (# modules)* |
| *High # comment givers (1/0)* | .040** | -.010 | -.156 | .020 | -.027 | -.000 |
| | (.018) | (.019) | (.159) | (.024) | (.026) | (.185) |
| *High # like givers (1/0)* | -.023 | .021 | -.474* | -.045 | -.014 | -.206 |
| | (.027) | (.029) | (.260) | (.032) | (.033) | (.256) |
| *Comment(s) received m1 (1/0)* | .247*** | .195*** | 1.938*** | .259*** | .228*** | 1.614*** |
| | (.016) | (.014) | (.171) | (.018) | (.016) | (.177) |
| *Like(s) received m1 (1/0)* | .199*** | .147*** | 3.844*** | .207*** | .173*** | 3.322*** |
| | (.023) | (.019) | (.325) | (.027) | (.022) | (.365) |
| Dep. Variable Mean | .658 | .847 | 15.643 | .658 | .847 | 15.643 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | .069 | .039 | .109 | .091 | .064 | .105 |
| N | 7,598 | 7,598 | 7,598 | 5,729 | 5,729 | 5,729 |

**Notes.** The dependent variable in columns (1) and (4) *Comments given m2-7* equals 1 if any comments were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in columns (2) and (5) *Likes given m2-m7* equals 1 if any likes were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in columns (3) and (6) *Persistence* is the number of completed modules of the entire program (0-17). *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, English as an official language (yes/no), # of the same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table B.7:** Heterogeneity Analysis: Split by COVID-19 Impact

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | COVID-19 Period | | | Pre-COVID-19 Period | | |
| | *Comments m2-7 (1/0)* | *Likes m2-7 (1/0)* | *Persistence (# modules)* | *Comments m2-7 (1/0)* | *Likes m2-7 (1/0)* | *Persistence (# modules)* |
| *High # comment givers m1 (1/0)* | .042** | -.014 | .031 | .019 | .010 | -.472 |
| | (.016) | (.017) | (.132) | (.032) | (.033) | (.296) |
| *High # like givers m1* (1/0) | .000 | .000 | .000 | -.025 | -.013 | -.136 |
| | (.) | (.) | (.) | (.028) | (.029) | (.250) |
| *Comment(s) received m1* (1/0) | .265*** | .232*** | 2.103*** | .244*** | .191*** | 1.541*** |
| | (.018) | (.016) | (.195) | (.016) | (.014) | (.160) |
| *Like(s) received m1* (1/0) | .204*** | .166*** | 4.105*** | .204*** | .157*** | 3.316*** |
| | (.028) | (.023) | (.403) | (.023) | (.018) | (.303) |
| Dep. Variable Mean | .658 | .847 | 15.643 | .658 | .847 | 15.643 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | .063 | .040 | .117 | .081 | .050 | .095 |
| N | 7,271 | 7,271 | 7,271 | 6,056 | 6,056 | 6,056 |

**Notes.** The dependent variable in columns (1) and (4) *Comments given m2-7* equals 1 if any comments were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in columns (2) and (5) *Likes given m2-m7* equals 1 if any likes were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in columns (3) and (6) *Persistence* is the number of completed modules of the entire program (0-17). *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, English as an official language (yes/no), # of the same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table B.8:** Heterogeneity Analysis: Split by Gender

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Women | | | Men | | |
| | *Comments m2-7* | *Likes m2-7* | *Persistence* | *Comments m2-7* | *Likes m2-7* | *Persistence* |
| *High # comment givers m1* (1/0) | .027 | -.018 | -.171 | .044** | .001 | .014 |
| | (.022) | (.023) | (.180) | (.020) | (.021) | (.169) |
| *High # like givers m1* (1/0) | -.022 | .032 | -.331 | -.047* | -.035 | -.644** |
| | (.033) | (.035) | (.301) | (.028) | (.029) | (.254) |
| *Comment(s) received m1* (1/0) | .303*** | .253*** | 2.162*** | .226*** | .190*** | 1.610*** |
| | (.020) | (.017) | (.224) | (.015) | (.014) | (.151) |
| *Like(s) received m1* (1/0) | .197*** | .177*** | 3.541*** | .211*** | .149*** | 3.666*** |
| | (.029) | (.023) | (.414) | (.023) | (.019) | (.306) |
| Dep. Variable Mean | .658 | .847 | 15.643 | .658 | .847 | 15.643 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | .087 | .056 | .109 | .071 | .042 | .106 |
| N | 5,188 | 5,188 | 5,188 | 7,841 | 7,841 | 7,841 |

**Notes.** The dependent variable in columns (1) and (4) *Comments given m2-7* equals 1 if any comments were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in columns (2) and (5) *Likes given m2-m7* equals 1 if any likes were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in columns (3) and (6) *Persistence* is the number of completed modules of the entire program (0-17). *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include (male/female), age, English as an official language (yes/no), residency in the US (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table B.9:** Heterogeneity Analysis: Interaction with Gender

| | (1) *Comments given m2-7 (1/0)* | (2) *Likes given m2-7 (1/0)* | (3) *Persistence (# modules)* |
|---|---|---|---|
| *High # comment givers (1/0)* | 0.027 | −0.008 | 0.060 |
| | (.021) | (.022) | (.153) |
| *High # comment givers x Female* | −0.017 | −0.042 | 0.036 |
| | (.031) | (.032) | (.227) |
| *High # like givers (1/0)* | −0.031 | −0.022 | −0.377* |
| | (.027) | (.028) | (.212) |
| *High # like givers x Female* | 0.001 | 0.025 | −0.313 |
| | (.034) | (.035) | (.252) |
| *Comment(s) received m1 (1/0)* | 0.181*** | 0.147*** | 0.984*** |
| | (.015) | (.014) | (.132) |
| *Comment(s) received m1 x Female* | 0.064** | 0.049** | 0.202 |
| | (.025) | (.023) | (.236) |
| *Like(s) received m1 (1/0)* | 0.105*** | 0.066*** | 1.317*** |
| | (.025) | (.022) | (.258) |
| *Like(s) received m1 x Female* | −0.004 | 0.040 | 0.386 |
| | (.040) | (.034) | (.445) |
| *Female (1/0)* | −0.021 | −0.057 | −0.093 |
| | (.043) | (.038) | (.459) |
| Means Dep. Variable | 0.658 | 0.847 | 15.643 |
| Controls | Yes | Yes | Yes |
| Year-FE | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.101 | 0.084 | 0.137 |
| N | 12,687 | 12,687 | 12,687 |

**Notes.** The dependent variable in column (1) *Comments given m2-7* equals 1 if any comments were given in later modules of the program and 0 otherwise. The dependent variable in column (2) *Likes given m2-7* equals 1 if any likes were given in later modules and 0 otherwise. *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. The dependent variable *Persistence* is the number of completed modules of the entire program (0-17). *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table B.10:** Heterogeneity Analysis: Future Commenting Engagement - split by Early vs. Late Modules

| | (1) *Comments given early: m2-3 (1/0)* | (2) *Comments given early: m2-3 (1/0)* | (3) *Comments given early: m4-7 (1/0)* | (4) *Comments given early: m4-7 (1/0)* |
|---|---|---|---|---|
| *High # comment givers (1/0)* | 0.016 | 0.020 | −0.009 | −0.004 |
| | (.016) | (.015) | (.017) | (.016) |
| *High # like givers (1/0)* | −0.025 | 0.003 | −0.013 | 0.010 |
| | (.022) | (.017) | (.023) | (.017) |
| *Comment(s) received m1 (1/0)* | 0.209*** | 0.212*** | 0.191*** | 0.193*** |
| | (.013) | (.013) | (.012) | (.012) |
| *Like(s) received m1 (1/0)* | 0.083*** | 0.084*** | 0.083*** | 0.083*** |
| | (.020) | (.020) | (.017) | (.017) |
| Means Dep. Variable | 0.616 | 0.616 | 0.472 | 0.472 |
| Controls | Yes | Yes | Yes | Yes |
| Year FE | No | Yes | No | Yes |
| Adjusted $R^2$ | 0.095 | 0.094 | 0.096 | 0.095 |
| N | 12,687 | 12,687 | 12,687 | 12,687 |

**Notes.** The dependent variable in columns (1) and (2) *Comments given m2-3* equals 1 if any comments were given in later early modules 2 to 3 in the program and 0 otherwise. The dependent variable in columns (3) and (4) *Comments given m4-7* equals 1 if any comments were given in later modules 4 to 5/7 in the program and 0 otherwise. *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table B.11:** Heterogeneity Analysis: Future Liking Engagement - split by Early vs. Late Modules

| | (1) *Likes given early: m2-3 (1/0)* | (2) *Likes given early: m2-3 (1/0)* | (3) *Likes given late: m4-7 (1/0)* | (4) *Likes given late: m4-7 (1/0)* |
|---|---|---|---|---|
| *High # comment givers (1/0)* | 0.004 | 0.004 | −0.001 | 0.001 |
| | (.012) | (.012) | (.014) | (.013) |
| *High # like givers (1/0)* | 0.019 | 0.034*** | −0.005 | 0.019 |
| | (.017) | (.013) | (.020) | (.015) |
| *Comment(s) received m1 (1/0)* | 0.139*** | 0.140*** | 0.156*** | 0.158*** |
| | (.011) | (.011) | (.012) | (.012) |
| *Like(s) received m1 (1/0)* | 0.153*** | 0.154*** | 0.149*** | 0.151*** |
| | (.020) | (.020) | (.020) | (.020) |
| Means Dep. Variable | 0.822 | 0.822 | 0.741 | 0.741 |
| Controls | Yes | Yes | Yes | Yes |
| Year FE | No | Yes | No | Yes |
| Adjusted $R^2$ | 0.095 | 0.094 | 0.104 | 0.103 |
| N | 12,687 | 12,687 | 12,687 | 12,687 |

**Notes.** The dependent variable in columns (1) and (2) *Likes given m2-3* equals 1 if any likes were given in later early modules 2 to 3 in the program and 0 otherwise. The dependent variable in columns (3) and (4) *Likes given m4-7* equals 1 if any likes were given in later modules 4 to 5/7 in the program and 0 otherwise. *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts *Comment(s) received m1* equals 1 if any comments from peers were received and 0 otherwise. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table B.12:** Mechanism: Regressions of Outcomes on Different Comment Types

| | (1) *Comments given m2-7 (1/0)* | (2) *Comments given m2-7 (1/0)* | (3) *Likes given m2-7 (1/0)* | (4) *Likes given m2-7 (1/0)* | (5) *Persistence (# modules)* | (6) *Persistence (# modules)* |
|---|---|---|---|---|---|---|
| *High # comment givers m1 (1/0)* | .031** | .015 | -.007 | -.004 | .096 | .063 |
| | (.015) | (.016) | (.011) | (.012) | (.108) | (.114) |
| *High # like givers m1 (1/0)* | -.010 | -.024 | .032*** | .027* | -.429*** | -.493*** |
| | (.016) | (.022) | (.012) | (.016) | (.125) | (.175) |
| *Pure agree comment(s) received m1 (1/0)* | .086*** | .083*** | .049*** | .047*** | .340*** | .349*** |
| | (.009) | (.009) | (.006) | (.006) | (.060) | (.061) |
| *Agree & elaborate comment(s) received m1 (1/0)* | .105*** | .104*** | .068*** | .067*** | .720*** | .726*** |
| | (.011) | (.011) | (.008) | (.008) | (.083) | (.083) |
| *Disagree & elaborate comment(s) received m1 (1/0)* | .058*** | .058*** | .026*** | .026*** | .287*** | .290*** |
| | (.009) | (.009) | (.007) | (.007) | (.070) | (.070) |
| *Like(s) received m1 (1/0)* | .114*** | .111*** | .152*** | .151*** | 1.423*** | 1.421*** |
| | (.020) | (.020) | (.019) | (.019) | (.210) | (.211) |
| Dep. Variable Mean | .658 | .658 | .847 | .847 | 15.643 | 15.643 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | No | Yes | No | Yes | No | Yes |
| Adjusted $R^2$ | .100 | .105 | .090 | .091 | .140 | .141 |
| N | 12,687 | 12,687 | 12,687 | 12,687 | 12,687 | 12,687 |

**Notes.** The dependent variable in columns (1) and (2) *Comments given m2-7* equals 1 if any comments were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in columns (4) and (4) *Likes given m2-m7* equals 1 if any likes were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in columns (5) and (6) *Persistence* is the number of completed modules of the entire program (0-17). *High # comment givers m1* equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. *High # like givers m1* equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts. *Pure agree comment(s) received m1* equals 1 if any purely agreeing comments from peers were received and 0 otherwise. *Agree & elaborate comment(s) received m1* equals 1 if any agreeing and elaborating comments from peers were received and 0 otherwise. *Disagree & elaborate comment(s) received m1* equals 1 if any disagreeing comments from peers were received and 0 otherwise. The omitted baseline comment category is "other comments" that fall in neither of the three aforementioned categories. *Like(s) received m1* equals 1 if any likes from peers were received and 0 otherwise. Control variables include (male/female), age, English as an official language (yes/no), residency in the US (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). 95% confidence intervals in brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Table B.13:** Robustness 1 - Count Model for Platform Persistence

| | (1) *Comments given m2-7 (1/0)* | (2) *Likes given m2-7 (1/0)* | (3) *Persistence* (# modules) |
|---|---|---|---|
| *High # comment givers m1 (1/0)* | 0.072*** | 0.023 | 0.369* |
| | (.025) | (.019) | (.198) |
| *Medium # comment givers m1 (1/0)* | 0.044** | 0.011 | 0.170 |
| | (.022) | (.017) | (.176) |
| *High # like givers m1 (1/0)* | 0.007 | 0.034*** | −0.138 |
| | (.017) | (.013) | (.131) |
| *Medium # of like givers m1 (1/0)* | −0.002 | 0.018 | −0.150 |
| | (.014) | (.011) | (.110) |
| *Comments received m1 (1/0)* | 0.203*** | 0.121*** | 1.052*** |
| | (.013) | (.011) | (.111) |
| *Likes received m1 (1/0)* | 0.104*** | 0.145*** | 1.453*** |
| | (.020) | (.019) | (.212) |
| Means Dep. Variable | 0.658 | 0.847 | 15.643 |
| Controls | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.101 | 0.090 | 0.136 |
| N | 12,687 | 12,687 | 12,687 |

**Notes.** Estimates are derived from a Poisson count model. The dependent variable in all columns, Persistence, is the number of completed modules of the entire program (0-17). High # comment givers m1 equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. High # like givers m1 equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts Comment(s) received m1 equals 1 if any comments from peers were received and 0 otherwise. Like(s) received m1 equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). Robust standard errors in parentheses.* $p < .10$, ** $p < .05$, *** $p < .01$.

**Table B.14:** Robustness 2 - Alternative Construction of Cohort Activity

| | (1 | (2) |
|---|---|---|
| | *Persistence* | *Persistence* |
| | *(# modules)* | *(# modules)* |
| *High # comment givers m1 (1/0)* | .006 | .006 |
| | 0.007) | 0.007) |
| *High # like givers m1 (1/0)* | −0.037*** | −0.034*** |
| | 0.011) | 0.012) |
| *Comments received m1 (1/0)* | 0.070*** | .070*** |
| | 0.008) | 0.008) |
| *Likes received m1 (1/0)* | 0.107*** | 0.108*** |
| | 0.016) | 0.016) |
| Means Dep. Variable | 15.641 | 15.643 |
| Pseudo-$R^2$ | .021 | .021 |
| N | 12,687 | 12,687 |

**Notes.** The dependent variable in column (1) Comments given m2-7 equals 1 if any comments were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable in column (2) Likes given m2-7 equals 1 if any likes were given in later modules 2 to 5/7 in the program and 0 otherwise. The dependent variable column (3) Persistence is the number of completed modules of the entire program (0-17). High # comment givers m1 equals 1 if the number of peers giving comments in module 1 is above the 66.6 centile across all cohorts and 0 otherwise. Medium # of comment givers m1 equals 1 if the number of peers giving comments is within the 33.3 to 66.6 centile across cohorts and 0 otherwise. High # like givers m1 equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts. Medium # of like givers m1 equals 1 if the number of peers giving likes is within the 33.3 to 66.6 centile across cohorts and 0 otherwise. Low # of comment givers and low # of like givers are the omitted baseline categories, respectively. Comment(s) received m1 equals 1 if any comments from peers were received and 0 otherwise. Like(s) received m1 equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score is first submitted module (m1 in financial accounting course). Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

**Table B.15:** Robustness 3 - Main Estimates at the Intensive Margin for Subset of Users who received at least 1 Comment & 1 Like in the First Module

| | (1) *Comments given m2-7 (#)* | (2) *Comments given m2-7 (#)* | (3) *Likes given m2-7 (#)* | (4) *Likes given m2-7 (#)* | (5) *Persistence (#)* | (6) *Persistence (#)* |
|---|---|---|---|---|---|---|
| *High # comment givers m1 (1/0)* | −2.507 | −4.420** | 2.776 | 0.732 | 0.046 | 0.013 |
| | (1.698) | (1.760) | (7.299) | (7.326) | (.104) | (.111) |
| *High # like givers m1 (1/0)* | 5.195*** | 0.419 | 23.316*** | 8.089 | −0.342*** | −0.475*** |
| | (1.846) | (2.291) | (7.793) | (10.352) | (.121) | (.173) |
| *Comments received m1 (#)* | 1.149*** | 1.115*** | 0.810 | 0.790 | −0.004 | −0.003 |
| | (.167) | (.167) | (.508) | (.514) | (.006) | (.006) |
| *Likes received m1 (#)* | −0.101** | −0.101** | 0.360** | 0.355** | 0.013*** | 0.013*** |
| | (.041) | (.041) | (.160) | (.160) | (.002) | (.002) |
| Means Dep. Variable | 0.658 | 0.658 | 0.847 | 0.847 | 15.643 | 15.643 |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | No | Yes | No | Yes | No | Yes |
| Adjusted $R^2$ | 0.048 | 0.052 | 0.039 | 0.039 | 0.087 | 0.089 |
| N | 10,275 | 10,275 | 10,275 | 10,275 | 10,275 | 10,275 |

**Notes.** The dependent variable in columns (1) and (2) Comments given m2-7 is the number of comments given in these later modules The dependent variable in columns (3) and (4) Likes given m2-7 is the number of likes given in these later modules. The dependent variable in columns (5) and (6), Persistence, is the number of completed modules of the entire program (0-17). High # comment givers m1 equals 1 if the number of peers giving comments in module 1 is above the median of all cohorts and 0 otherwise. High # like givers m1 equals 1 if the number of peers giving likes in module 1 is above the median of all cohorts Comment(s) received m1 equals 1 if any comments from peers were received and 0 otherwise. Like(s) received m1 equals 1 if any likes from peers were received and 0 otherwise. Control variables include gender (male/female), age, living in the US (yes/no), English as an official language (yes/no), # of same gender, similar age (+/- 2 years), same country & same citizenship peers, procrastination (submission hours to quiz deadline), cohort size, quiz score in first submitted module (m1 in financial accounting course). Robust standard errors in parentheses.

## B.4 Comment Type Examples

**I. Agreeing and Elaborating Comments:**

- "[Name], I definitely agree with you that I would go with self-finance over external financing (...)."
- ”Hi [Name], very true. Vacation Rentals can be a nightmare. So when the prices go up, the alternative would be to cancel the vacation. It is not a necessity. ”

**II. Purely Agreeing Comments:**

- ”Great answer”
- ”100% agree with you”

**III. Disagreeing and Elaborating Comments:**

- ”Hi [Name], I would disagree. The original question indicates that we specifically want to find out how many people purchase warranties from their dealership. (...) ”
- ”There can still be a relationship between the two variables if the value of the correlation coefficient is 0, just not a linear one.”

**IV. Other Comments:**

- ”Good attempt [Name]. Although this was a tough cold call.”
- ”Good examples, although you lost me on cosmetics there...”
- ”Cars have so many alternatives, so I don’t agree.”

# C

# Appendix to Chapter 3

*The Impact of Competition and Gender Composition on Creative Idea Generation and Selection*

# C.1 Additional Figures and Tables

## C.1.1 Figures

**Figure C.1:** CI-Barplots of Creative Performance in Round 1 by Gender

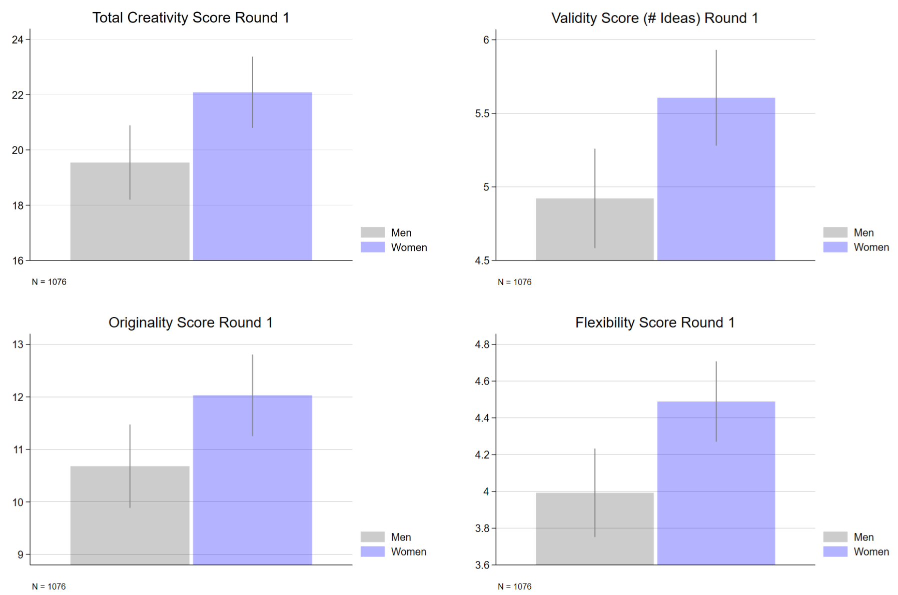



**Figure C.2:** Distributions of Creative Performance in Round 1 by Gender

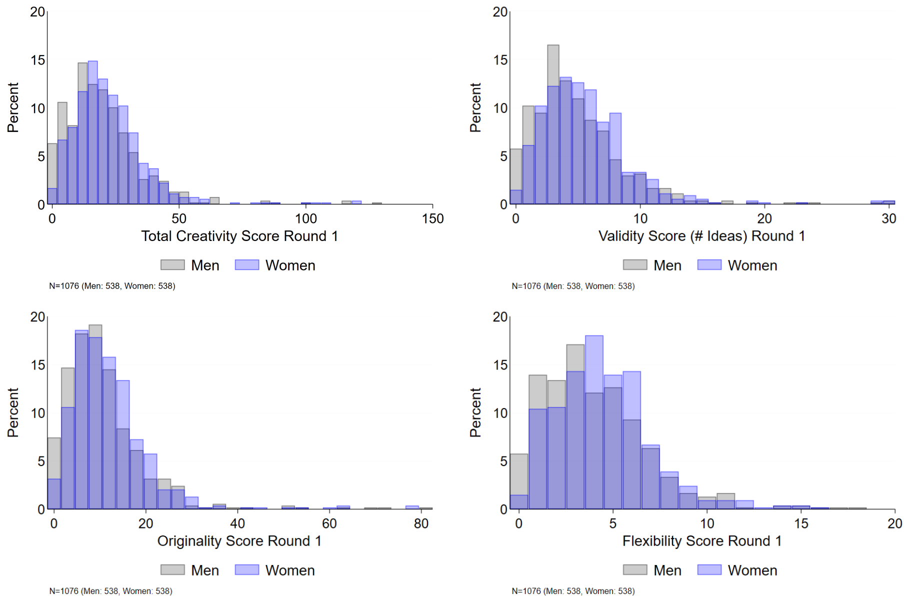


**Figure C.3:** CI-Barplots of Creative Performance in Round 2 by Gender

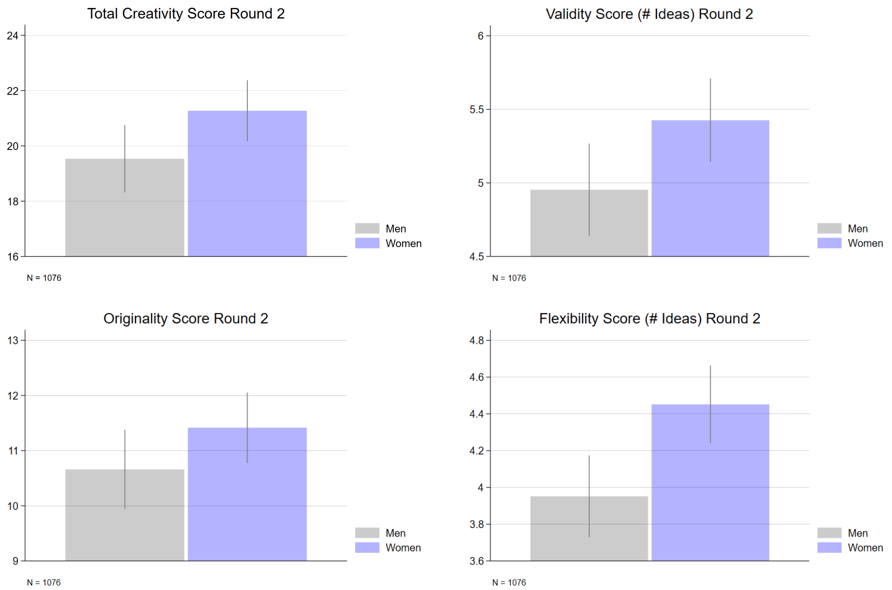



**Figure C.4:** Distributions of Creative Performance in Round2 by Gender

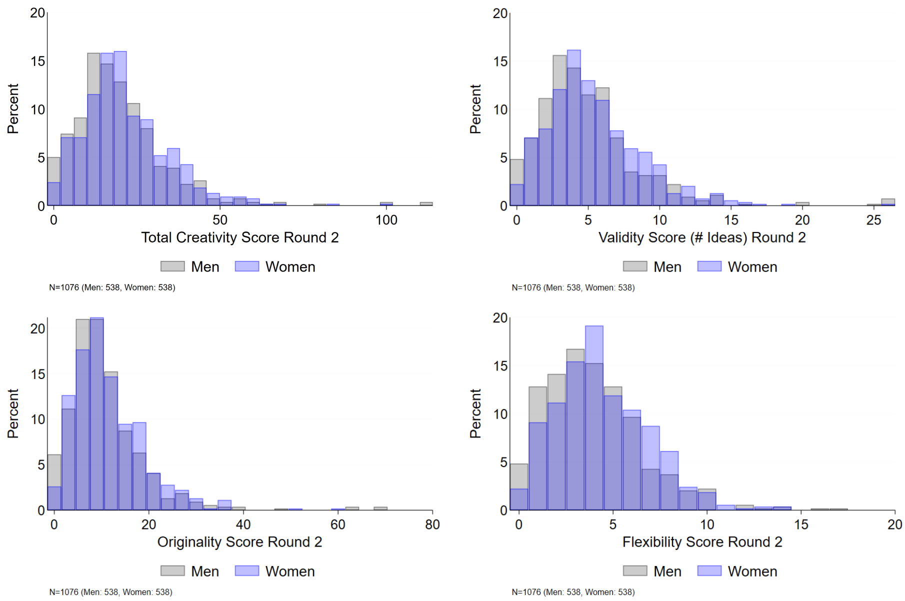


**Figure C.5:** Round 2 Creativity Score Overall and Dimensions by Competitor Environment and Gender

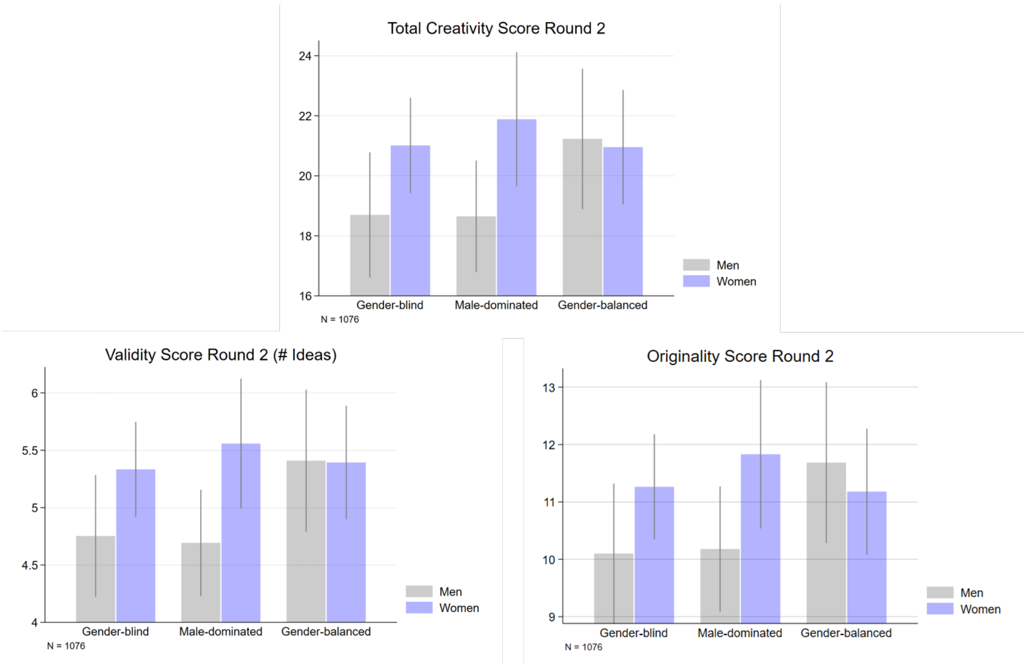



**Figure C.6:** Word Cloud of Task Strategy Replies

### C.1.2 Tables

**Table C.1:** Sample Size per Experimental Condition, Gender, and Total

| | Environment | | | |
|---|---|---|---|---|
| | GenderBlind | MaleDominated | GenderBalanced | Total |
| Women | 183 | 172 | 183 | 538 |
| Men | 178 | 179 | 181 | 538 |
| Total | 361 | 351 | 364 | 1076 |

**Table C.2:** Examples of Ideas and Categories for the Unusual Uses Task by Object

| | Everyday Object | |
|---|---|---|
| | Brick | Rubber Band |
| Frequent Answers | Doorstop | Slingshot |
| | Paperweight | Hair tie |
| | Weapon | Ball |
| Frequent Categories | Barrier | Defense |
| | Weight | Accessories |
| | Defense | Toy |
| Original Answers | Hammer | Bookmark |
| | Fire pit | Cable organizer |
| | Stepping stool | Cat toy |
| Very Original Answers | Exfoliation | Zip lock |
| | Barbecue base | Tourniquet |
| | Figure stand | Curtain holder |
| Invalid Answers | Bread* | Chair* |
| | Dating tool* | Office* |
| | Smart Brick* | Robotics* |

**Notes.** *Answers without any (plausible) explanation or infeasible uses with the given object could be deemed invalid by the evaluating research assistants. This procedure was communicated to participants in the experimental instructions (see Appendix C.3).

**Table C.3:** Balance Table

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Treatment Conditions | | | p-value | p-value | p-value |
| | BL | MD | GB | (BL-MD) | (BL-GB) | (MD-GB) |
| *Basic Demographic Information* | | | | | | |
| Age (years) | 39 | 40 | 39 | .690 | .388 | .228 |
| Female | .507 | .490 | .503 | .652 | .910 | .734 |
| Unemployed | .119 | .131 | .104 | .630 | .529 | .268 |
| Min. Bachelor's Degree | .997 | .997 | .995 | .984 | .568 | .584 |
| *Subject Studied* | | | | | | |
| Arts and Design | .050 | .037 | .041 | .402 | .576 | .774 |
| Business, Law, & Publ. Admin. | .285 | .245 | .217 | .223 | .034 | .375 |
| Education Studies | .042 | .057 | .091 | .341 | .008 | .086 |
| Environmental Studies | .019 | .017 | .016 | .819 | .768 | .949 |
| Health & Medicine | .089 | .080 | .104 | .670 | .473 | .255 |
| Logistics & Construction | .014 | .011 | .005 | .769 | .250 | .387 |
| STEM | .255 | .245 | .250 | .762 | .881 | .877 |
| Social Sciences & Humanities | .183 | .245 | .225 | .043 | .156 | .534 |
| Unknown | .064 | .063 | .049 | .955 | .406 | .442 |
| *Ethnic Groups* | | | | | | |
| White/Caucasian | .518 | .462 | .508 | .132 | .793 | .212 |
| Black/African American | .271 | .296 | .275 | .463 | .922 | .523 |
| Asian/Pacific Islander | .133 | .105 | .115 | .257 | .473 | .671 |
| Hispanic/Latino | .039 | .040 | .033 | .940 | .674 | .621 |
| Other & Mixed | .039 | .097 | .069 | .002 | .074 | .171 |
| *Further Measures* | | | | | | |
| Own Competitiveness (0-10) | 6.410 | 6.510 | 6.451 | .686 | .699 | .995 |
| Own Risk Assessment (0-10) | 5.842 | 5.698 | 5.942 | .327 | .539 | .119 |
| Creative Hobby | .468 | .504 | .492 | .335 | .525 | .738 |

**Notes.** All variables presented in columns (1)–(3) represent means within each treatment condition. If no unit of measurement is specified in parentheses, variables are measured as percentage shares (%) of the corresponding treatment condition indicated in the column header. The treatment condition abbreviations are as follows: BL = $GenderBlind$, MD = $MaleDominated$, GB = $GenderBalanced$. Columns (4)–(6) report p-values from two-sample Mann-Whitney-U-tests for continuous variables and Chi$^2$-tests for binary variables, comparing differences between the specified groups. See experimental instructions in Appendix C.3).

**Table C.4:** Perceived Average Gender Advantage in Task and by Object

| | Task in general | | Object A: Brick | | Object B: Rubber Band | |
|---|---|---|---|---|---|---|
| | Slider | p-value | Slider | p-value | Slider | p-value |
| Overall | -.051 | .000 | .223 | .000 | -.179 | .000 |
| Women | -.157 | .000 | .179 | .000 | -.217 | .000 |
| Men | .054 | .005 | .268 | .000 | -.141 | .000 |

**Notes.** Averages for the slider measure (slider). P-values for a two-sided t-test against H0 that an average is equal to zero. See Appendix C.3 for the experimental instructions.

**Table C.5:** Most Frequent Trigrams in Open Text Field on Task Strategy

| # | Overall | Women | Men |
|---|---|---|---|
| 1 | think outside box | think outside box | think outside box |
| 2 | strategy generate idea | way use item | try think outside |
| 3 | try think outside | look around room | strategy generate idea |
| 4 | way use item | way could use | object could use |
| 5 | look around room | strategy generate idea | whatever come mind |
| 6 | object could use | think different way | could use object |
| 7 | way could use | anything come mind | use item past |
| 8 | whatever come mind | strategy try think | way use item |
| 9 | could use object | think way use | many idea possible |
| 10 | think different way | tried think way | rubber band brick |

**Table C.6:** Treatment Effects on Idea Generation in Round 2 (AI-adjusted Measures)

| | (1) Total Score | (2) Validity | (3) Similarity | (4) Originality |
|---|---|---|---|---|
| Female (1/0) | .413 | .018 | .182 | .269 |
| | -1.509,2.334] | -.506,.541] | -.193,.557] | -.836,1.374] |
| | (.979) | (.267) | (.191) | (.563) |
| MaleDomEnv (1/0) | -.171 | -.146 | -.096 | .057 |
| | -2.252,1.910] | -.709,.417] | -.488,.295] | -1.139,1.252] |
| | -1,061 | (.287) | (.199) | (.609) |
| GenderBalanceEnv (1/0) | 1,306 | .428 | .155 | .747 |
| | -1.077,3.689] | -.251,1.107] | -.275,.585] | -.631,2.125] |
| | (1,215) | (.346) | (.219) | (.702) |
| Female x MaleDomEnv | 1.274 | .488 | .090 | .638 |
| | -1.664,4.212] | -.312,1.288] | -.479,.659] | -1.033,2.310] |
| | (1.497) | (.408) | (.290) | (.852) |
| Female x GenderBalanceEnv | -2,132 | -.644 | -.463 | -1,054 |
| | -5.183,.919] | -1.499,.211] | -1.037,.111] | -2.797,.690] |
| | (1.555) | (.436) | (.293) | (.889) |
| Dep. Variable Mean | 18.263 | 5.047 | 4.150 | 9.083 |
| Demographic Controls | Yes | Yes | Yes | Yes |
| Sequence & Round 1 DV | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | .328 | .321 | .408 | .269 |
| N | 1076 | 1076 | 1076 | 1076 |

**Notes.** The dependent variable in Column (1) is the total $CreativityScore^x$; in Column (2), it is the $ValidityScore$; in Column (3), the $FlexibilityScore$ indicates if they selected their highest scoring idea for Round 2 competition; in Column (4), $OriginalityScore^x*$. See Section 2.2. for a construction of the performance measures. $Female$ indicates that a participant's sex is female. $MaleDomEnv$ indicates that a participant competes in a male-majority environment in Round 2. $GenderBalancedEnv$ indicates that a participant competes in an environment with equally many men and women. The $GenderBlindEnv$ is the omitted baseline environment category. $Sequence$ indicates that the creative task used "brick" as an object in round 1 and "rubber band" in round 2; it is 0 if the object order is reversed. *Round 1 Performance* is the corresponding performance outcome of Round 1. *Demographic Controls* include age, age$^2$, and field of study. 95% confidence intervals in squared brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $^x$These two scores are based on an alternative construction of the originality category accounting for very close synonyms with NLP. See Appendix C.2 for a description.

**Table C.7:** Treatment Effects on Idea Selection and Competition Outcome (AI-adjusted Measures)

| | (1)<br>Score<br>Selected Idea | (2)<br>Originality<br>Selected Idea | (3)<br>Best Idea<br>Selected | (4)<br>Competition<br>Won |
|---|---|---|---|---|
| Female (1/0) | .391** | .252** | -.037 | .092* |
| | .088,.695] | .052,.453] | -.141,.067] | -.013,.198] |
| | (.155) | (.102) | (.053) | (.054) |
| MaleDomEnv (1/0) | .182 | .157 | .031 | -.133*** |
| | -.132,.496] | -.047,.360] | -.073,.134] | -.233,-.033] |
| | (.160) | (.104) | (.053) | (.051) |
| GenderBalanceEnv (1/0) | .150 | .141 | -.013 | .053 |
| | -.171,.470] | -.068,.350] | -.116,.090] | -.051,.157] |
| | (.163) | (.106) | (.053) | (.053) |
| Female x MaleDomEnv | -.341 | -.194 | .032 | .615*** |
| | -.778,.096] | -.481,.093] | -.114,.178] | .492,.738] |
| | (.223) | (.146) | (.075) | (.063) |
| Female x GenderBalanceEnv | -.127 | -.091 | .106 | -.135* |
| | -.554,.300] | -.376,.194] | -.040,.251] | -.282,.011] |
| | (.218) | (.145) | (.074) | (.075) |
| Dep. Variable Mean | 3.230 | 1.601 | .413 | .474 |
| Demographic Controls | Yes | Yes | Yes | Yes |
| Task Sequence & Round 1 DV | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | .061 | .046 | -.003 | .191 |
| N | 1076 | 1076 | 1076 | 1076 |

**Notes.** The dependent variable in Column 1 $ScoreSelectedIdea^x$ is participants' overall $CreativityScore$ of the selected idea ranging from 0 to 5; in Column (2), it is their $OriginalityScore^x$ of their selected idea ranging from 0 to 3; in Column (3) $BestIdeaSelect^x$ indicates if they selected their highest scoring idea for the Round 2 competition; in Column (4), $CompetitionWon^x$ indicates if their selected idea. $Female$ indicates that a participant's sex is female. $MaleDomEnv$ indicates that a participant competes in a male-majority environment in Round 2. $GenderBalancedEnv$ indicates that a participant competes in an environment with equally many men and women. The $GenderBlindEnv$ is the omitted baseline environment category. $Sequence$ indicates that the creative task used "brick" as an object in round 1 and "rubber band" in round 2; it is 0 if the object order is reversed. *Round 1 Performance* is the DV's equivalent of Round 1. *Demographic Controls* include age, age$^2$, and field of study. 95% confidence intervals in squared brackets. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $^x$These two scores are based on an alternative construction of the originality category accounting for very close synonyms with NLP. See Appendix C.2 for a description.

## C.2 Description of NLP Tool SpaCy for Determining Idea Synonyms

To account for the fact that the submitted ideas per object, i.e., the brick(s) or rubber band(s), include many synonyms for the same use, for instance, "hair binder" and "hair tie" as new uses for the rubber band, I employ NLP to find very close synonyms. With these preprocessed ideas, I re-calculate the ideas' originality and overall creativity scores and re-run the main analyses as a robustness exercise. The subsequent section describes the procedures to construct the synonyms in detail.

I use SpaCy's “en_core_web_lg” model to semantically group words together into synonym buckets. SpaCy is a NLP library, and the "en_core_web_lg" model is one of the pretrained models provided for processing English text[1]. Each "idea" is put through a "processing pipeline", for which the steps are detailed below:

1. **Tok2Vec:** this step tokenizes the text into words, punctuation marks, and individual units. For us, this step only breaks our code into words, since preprocessing has already been applied to remove punctuation.

2. **Lemmatizer:** reduces words to their base or dictionary form. While this is done within our preprocessing, it is also embedded in the pipeline itself.

3. **Ner:** named entity recognition. This portion evaluates the context around each token and classifies entities based on that.

For this task, the large model was chosen, since it performed better on a direct comparison, and because it includes a greater specificity for word vectors, which allow the model to compare the semantic similarity between words to a greater degree of accuracy. These word vectors are then used to calculate similarity scores between words, and within our code, those similarity scores are used to determine which synonym bucket a certain word should belong in.

The code employs SpaCy's *.similarity method*, which takes the tokens generated by the pipeline, computes the 300-dimensional similarity vector for each token, and then uses cosine similarity to calculate the distance between the resulting vectors. These similarity scores are then combined into a single number.

Each submitted idea is put through the NLP processing pipeline. Then, the code loops through the keys in a dictionary, putting each separate key through the pipeline. A variable "max_similiarity_score” is kept, which keeps the maximum of the similarity score calculation between each word and a certain key. Once the code has iterated through all keys if the max_similiarity_score is greater than the threshold set. After careful inspection of several thresholds, 0.78 is chosen as the best fit. Lastly, the word is added to that dictionary key. If not, it creates its own dictionary key.

[1] https://spacy.io/models, last retrieved on Sept 2, 2024.

## C.3 Instructions

Appendix C.3 includes the instructions of the experiment. Treatment-specific parts are shown in *italics*, and the corresponding treatment is clearly indicated.

***[Screen 1]***

### Consent Form

**Overview**: You are participating in a research study on economic decision-making. It consists of a brief survey, a main part, and a questionnaire. Your decisions and other participants' decisions determine a bonus you can earn. Your responses are anonymous, and the data gathered for this study will be stored securely. Participation in the study is voluntary; you can exit it anytime without giving a reason.

**Payment:** You will receive a guaranteed participation pay of GBP 2.25. In addition, you can earn a bonus of up to GBP 7.50. The payment will be processed within the next 10 working days. We compensate for complete submissions only.

**General Rules of Conduct:** This study will take about 22 minutes. Please give your full attention and find a quiet space. Avoid using other devices, social media, or having conversations during the study. Please remain in this browser tab throughout the study. We have implemented code to check for browser tab switches.

**Comprehension Checks:** There will be several comprehension questions throughout the study. They refer to the instructions on the same screen. You can only proceed once you answer the comprehension checks correctly.

**Consent:** I have read and understood the information above. I agree to comply with these rules of conduct and want to participate in this study.

◯ Yes, I want to participate.

◯ No, I want to exit.

*[Button: Next]*

***[Screen 2]***

### Prolific ID

**What is your Prolific ID (PID)?**
Please note that this field should auto-fill with the correct PID.

________________________________________

*[Screen 3]*

## Demographics

**Survey on Demographic Information**
Please provide some information about yourself.

**How old are you?**

________________________________________

**Have you completed A-levels or an equivalent level of education qualifying you for university studies?**

- ◯ Yes
- ◯ No

**If Yes: Were you a university student at some point in time during your life, including current enrollment?**

- ◯ Yes
- ◯ No

**If Yes: Which subject are you studying/did you study?**

________________________________________

**Do you have a job outside of taking surveys?**

- ◯ Yes, Full-time
- ◯ Yes, Part-time
- ◯ No, Unemployed

**If Yes: Are you self-employed?**

- ◯ Yes, I'm an entrepreneur/founder
- ◯ Yes, other types of self-employment
- ◯ No

**If Yes: What is your current occupation?**

________________________________________

**In which sector do you work?**

◯ Public sector

◯ Private sector

◯ Not-for-profit sector

◯ Other

**What is your sex?**

◯ Female

◯ Male

**In general, how do you see yourself? Where would you put yourself on this scale from "0-Very masculine" to "10-Very feminine"? Please indicate your response below.**

◯ 0 (very masculine) ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 ◯ 6 ◯ 7 ◯ 8 ◯ 9 ◯ 10 (very feminine)

**Please select your ethnic group(s) from the options below. You may choose more than one if applicable. If your ethnicity is not listed, please select 'Other'.**

◯ White/Caucasian

◯ Black/African-American

◯ Hispanic/Latino

◯ Asian/ Pacific Islander

◯ Native American/Alaska Native

◯ Middle Eastern/North African

◯ Other

*[Button: Next]*

***[Screen 4]***

## Task Description

**Main Part: Description of the Brainstorming Task**
This part of the study consists of two rounds. You will have 3 minutes per round. There will be a short break after every round.

Your task is a brainstorming task. The goal is to list as many new and innovative uses (= "ideas") as possible for an everyday object such as a rubber tire. Do not restrict yourself to familiar uses or a specific size. You can also list uses that require several objects of its kind (e.g., several small and large rubber tires used jointly).

Each round, you are presented with another everyday object. You will receive the same task in each round with varying everyday objects.

There will be two open-text fields for each idea per page:

- **Your idea:** Enter your new and innovative idea here. [2 words max.]
- **Short description:** Enter your elaboration if necessary here.

Every idea is submitted on a new page. Please aim to spell the ideas correctly. The two-word limit is applied strictly for scoring, i.e., ideas expressed in more words will not count for payment. No hyphens "-" are considered. Please remain in the same browser tab throughout.

The next page describes the scoring rule for the bonus you can win in part 1.

**Comprehension Checks:** Please indicate whether the following two statements about the instructions above are true.

1. You will have 3 minutes per round.

   ◯ True
   ◯ False

2. You are asked to list as many conventional uses for a given everyday object as you can think of.

   ◯ True
   ◯ False

*[Button: Next]*

*[Screen 5]*

## Task Evaluation in Round 1

Your bonus for this round includes two elements: the (I) Innovative Idea Score and (II) Idea Selection Bonus.

**I) Innovative Idea Score**
Your submitted ideas for everyday objects are scored based on three criteria: Validity, Originality, and Similarity[2]

- **#1** Validity: measures if an idea, i.e., a stated use, is feasible.
  - Valid ideas earn 1 point each; invalid ideas earn 0 points.
  - Only valid ideas are eligible for scoring and payment.
  - Validity is determined anonymously by external raters.
  - Describe your idea in a few words in the 2nd text field below the submitted idea if the use is not straightforward.
- **#2** Originality: measures how unique an idea is among all participants.
  - 0 points if more than 10% of the participants submitted the same idea.
  - 1 point if 2% to 10% of the participants submitted the same idea.
  - 2 points if less than 2% of the participants submitted the same idea.
  - 3 points if less than 0.5% of participants submitted the same idea.
- **#3** Similarity: measures the number of distinct categories ideas in each round fall into, e.g., "tools" or "clothing".
  - 1 Point for each new category your next idea introduces (see example).
  - Each idea is assigned only one category determined simultaneously by a NLP tool for consistency.

**Total Score Per Round**

- Your total score per idea is calculated as
  Score = Validity × (Validity + Originality + Similarity) Points.
- Maximum score per idea: 5 points.
- The total score per round is the sum of the scores of all valid ideas.
- Exchange rate: 1 point = GBP 0.02.

**Tip:** Use the entire 3 minutes per round to generate ideas for maximum payoff.

[2]This creativity dimension is referred to as $Flexibility$ in the main study.

**II) Idea Selection Bonus**
After the round, you will see an overview of your submitted ideas. Select your best idea to win an additional bonus of GBP 0.50 if it scores a full 5 points.

(Note that the similarity point is given to all selected ideas for fairness and comparability.)

## Comprehension Checks

1. Do you have to provide an explanation for all the ideas you submit?

   ◯ Yes
   ◯ No

2. How many points will you score in the "originality" dimension if 8% of participants submitted the same idea as you?

   ◯ 0 Points
   ◯ 1 Point
   ◯ 2 Points
   ◯ 3 Points

3. It is my best strategy for maximizing my payoff to generate as many ideas as possible in the next 3 minutes.

   ◯ True
   ◯ False

4. The exchange rate of score points to GBP is: 1 Point = GBP 0.01.

   ◯ True
   ◯ False

5. You can win a bonus of 0.5 GBP if you select your BEST idea after the round but only if its score is 5?

   ◯ True
   ◯ False

*[Button: Next]*

*[Screen 6]*

## Example of the Scoring Rule

Let's assume you submit the following three ideas for new and innovative uses for rubber tire(s): (1) sled, (2) flower box, and (3) target.

- **Validity:** The ideas "sled" or "flower box" are clear answers, whereas "target" would require further explanation; you provided: "ball game with the tire as a target."
  - As all three ideas are feasible, you would earn 3 points for "Validity".
- **Originality:** Let's assume that 15% of participants also submitted "flower box", 8% also submitted "(ball sports) target", and no one submitted "sled".
  - Hence, based on our scoring rule, you would get 0+1+3 = 4 points for originality.
- **Similarity:** The ideas of "sled" and "target" fall into the category of "sports devices" and the idea of "flower box" falls into the category of "decoration."
  - Hence, you would earn 1+1 = 2 points for similarity because the three ideas fall into two distinct categories.

**Overall Score:** Your total score per round is computed by multiplying the validity score by the sum of all ideas' validity, originality, and similarity scores.

**Example Calculation:**

- Score = Validity × (Validity + Originality + Similarity) = 3 × (3 + 4 + 2) = 3 × 9 = 27 Points in total.
- Given the Exchange Rate: 1 Point = GBP 0.02, you would earn GBP 0.54 here.

## Comprehension Checks

1. If your idea is invalid, what would be your score for that idea?

   ◯ 0
   ◯ 1
   ◯ It depends on the points of other criteria.

2. What is the correct payoff calculation for each idea?

   ◯ Validity × (Validity + Originality + Similarity)
   ◯ Originality × (Validity + Similarity)
   ◯ Validity + Originality + Similarity
   ◯ Similarity × 2 × (Originality + Validity)

*[Button: Next]*

*[Screen 7]*

## Next Round

The next round of the brainstorming task will start soon. Please remain in this browser tab throughout the study. Please use the entire 3 minutes to generate ideas for maximizing your payoff. Do not skip ahead.

Please click "Next" to proceed.

*[Button: Next]*

*[This is an example of the sequence where the everyday object **Rubber Band** was used in Round 1 and **Brick** in Round 2. This object sequence was also randomized across participants.]*

***[Screen 8: Example of an idea submission page during Round 1]***

## Your Idea

**THE OBJECT: rubber band**

Use #1

Your idea: enter your new and innovative idea here. [2 words max.]

- Your Idea: ______________________________________________

Short description: enter your elaboration, if necessary, here.

- Short Description: ______________________________________________

*[Button: Next]*

*[Screen 9]*

## Select Your Best Idea

The table below shows all of the ideas and elaboration you entered for **rubber band**.

1. *Item 1*
2. *Item 2*
3. *Item 3*
4. ...

Please select your best idea from the above to win a bonus of GBP 0.50.
Remember, you only receive the bonus if your selected idea receives a score of 5.

Idea Number (#): ____________________________________________

Your Idea (two words max): ____________________________________________

*[Screen 8]*

## Your Idea

**THE OBJECT: brick**

Use #1

Your idea: enter your new and innovative idea here. [2 words max.]

- Your Idea: ____________________________________________

Short description: enter your elaboration, if necessary, here.

- Short Description: ____________________________________________

*[Button: Next]*

*[Screen 9]*

## Select Your Best Idea

The table below shows all of the ideas and elaboration you entered for **brick**.

1. *Item 1*
2. *Item 2*
3. *Item 3*
4. ...

Please select your best idea from the above to win a bonus of GBP 0.50.
Remember, you only receive the bonus if your selected idea receives a score of 5.

Idea Number (#): ____________________________________________

Your Idea (two words max): ____________________________________________

## Your Score Guess

**What do you think is your score for the selected idea? Please insert an integer between 0 and 5.**
If you guess correctly, you will earn another bonus of 0.25 GBP.

____________________________________________

**Reminder of the Scoring Rule:** The total score per round is calculated by multiplying the validity score with the sum of all generated ideas' validity, originality, and similarity scores.

Score = Validity × (Validity + Originality + Similarity).

- Validity: Measures if an idea is feasible. 1 Point for feasible ideas.
- Originality: Measures how unique an idea is among all participants. More original, i.e., rarer ideas, get more Points. (0/1/2/3 Points for if $\leq$ 10% $\leq$ 5% $\leq$ 2% $\leq$ 0.5% of participants submit the same idea).
- Similarity: Measures the number of distinct categories ideas in each round fall into, e.g., "tools" or "clothing". 1 Point for each new category the next idea introduces. The similarity point is given to all selected ideas for fairness and comparability.

*[Button: Next]*

*[In the following, the Gender-Balanced condition of the Environment treatment is displayed for Screen 10 and Screen 12. Depending on the participants' indicated sex and treatment allocation, the competitor groups would have looked different. See Figure 3.2 for the different competitor*

*group images. This is an example of the sequence where the everyday object* ***Brick*** *was used in Round 1 and* ***Rubber Band*** *in Round 2. This object sequence was also randomized across participants. ]*

***[Screen 10]***

## Task Evaluation in Round 2

Your bonus for this round consists of two elements: the I) Innovative Idea Score and II) Competition Prize for Selected Idea detailed below.

**I) Innovative Idea Score**
In each round, your submitted ideas for everyday objects are scored based on three criteria: Validity, Originality, and Similarity.

- Validity: Measures if an idea is feasible. 1 Point for feasible ideas. The validity of all ideas will be determined anonymously by external raters. Please describe the possible use in a few words, if necessary, in the text field below.
- Originality: Measures how unique an idea is among all participants. More original, i.e., rarer ideas, get more Points. (0/1/2/3 Points for if $\leq$ 10% $\leq$ 5% $\leq$ 2% $\leq$ 0.5% of participants submit the same idea).
- Similarity: Measures the number of distinct categories ideas in each round fall into, e.g., "tools" or "clothing". 1 Point for each new category the next idea introduces. The similarity point is given to all selected ideas for fairness and comparability.

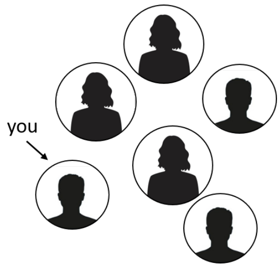
**Reminder of the Scoring Rule:** The total score per round is calculated by multiplying the validity score with the sum of all generated ideas' validity, originality, and similarity scores. Score = Validity × (Validity + Originality + Similarity).

**II) Competition Prize for Selected Idea**
At the end of this round, you will see an overview of all your submitted ideas. Please select your best idea. You and five other study participants have been randomly chosen to compete for a prize in round 2 of the same brainstorming task. We randomly select one of your competitors' ideas. If your selected idea receives the highest score among these six ideas, you will win an additional bonus of GBP 0.50. (The 1 Point in the similarity criterion is added to all selected ideas entering the competition for consistency.)

These five other participants have been recruited based on the same filters as you via Prolific:

- Living in the US
- Working age population (20 to 65 years)
- At minimum high school diploma or equivalent

For data protection reasons, we are using anonymized icons. Below is an image of your group of competitors:




**Comprehension Questions**

1. How many competitors are you going to face?

- ◯ 4
- ◯ 5
- ◯ 6

2. How competitive do you perceive your group of participants shown above?

- ◯ 1 (not at all)
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5 (extremely)

3. How confident are you that you will win this competition with your selected idea (in %)? A higher % indicates higher confidence in winning.

- ◯ 0% ... 100%

4. What is the ratio of women to men among your competitor group above? For example, a ratio of "2:5" indicates that a group of 7 people consists of 2 women and 5 men.

- ◯ 2:1
- ◯ 1:1
- ◯ 1:5
- ◯ 2:6
- ◯ unknown

*[Button: Next]*

*[Screen 11]*

**The next round of the brainstorming task will start soon.**

Please remain in this browser tab throughout the study. Please use the entire 3 minutes to generate ideas for maximizing your payoff. Do not skip ahead.

Please click "Next" to proceed.

*[Button: Next]*

*[Screen 12]*

## Your Idea

**THE OBJECT: rubber band**

Use #1

Your idea: enter your new and innovative idea here. [2 words max.]

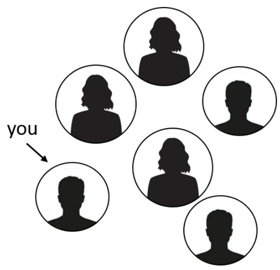
- Your Idea: ______________________________

Short description: enter your elaboration, if necessary, here.

- Short Description: ______________________________

## Reminder

**This is a reminder that you compete for a prize with these other participants as your competitors:**




*[Button: Next]*

*[Screen 13]*

## Select Your Best Idea

The table below shows all of the ideas and elaboration you entered for **brick**.

1. *Item 1*
2. *Item 2*
3. *Item 3*
4. ...

Please select your best idea from the above to win a bonus of GBP 0.50.
Remember, you only receive the bonus if your selected idea receives a score of 5.

Idea Number (#): ____________________________________________

Your Idea (two words max): ____________________________________________

## Your Score Guess

**What do you think is your score for the selected idea? Please insert an integer between 0 and 5.**
If you guess correctly, you will earn another bonus of 0.25 GBP.

____________________________________________

**Reminder of the Scoring Rule:** The total score per round is calculated by multiplying the validity score with the sum of all generated ideas' validity, originality, and similarity scores.

Score = Validity × (Validity + Originality + Similarity).

- Validity: Measures if an idea is feasible. 1 Point for feasible ideas.
- Originality: Measures how unique an idea is among all participants. More original, i.e., rarer ideas, get more Points. (0/1/2/3 Points for if $\leq$ 10% $\leq$ 5% $\leq$ 2% $\leq$ 0.5% of participants submit the same idea).
- Similarity: Measures the number of distinct categories ideas in each round fall into, e.g., "tools" or "clothing". 1 Point for each new category the next idea introduces. The similarity point is given to all selected ideas for fairness and comparability.

*[Button: Next]*

*[Screen 14]*

## Survey - Part 1a

You have just performed a brainstorming task where you had to find new and innovative uses for everyday objects.

What was your strategy for generating ideas in the task you just completed? Please briefly describe your thought process:

_______________________________________________

Which statement below best describes how you think you performed compared to other participants in this study task overall:

◯ I think I performed better than most (I'm among the top 25% of all participants).

◯ I think I performed better than about half (I'm among the top 50% of all participants).

◯ I think I performed worse than about half (I'm among the bottom 50% of all participants).

◯ I think I performed worse than most (I'm among the bottom 25% of all participants).

Have you encountered this or a similar task before (e.g., in a previous online study or an application process)?

◯ Yes

◯ No

How much did you enjoy performing this task (i.e., thinking of new and innovative uses for everyday objects)?

◯ Not at all.

◯ Slightly.

◯ Moderately.

◯ Quite a bit.

◯ Very much.

How difficult did you find this task (i.e., thinking of new and innovative uses for everyday objects)?

◯ 0 (not at all) ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 ◯ 6 ◯ 7 (very much)

Did you find the brainstorming task more difficult for any of the two everyday objects you encountered? Please indicate below.

◯ Brick

◯ Rubber band

◯ No difference

In the second round, you competed in the brainstorming task for a prize with a randomly selected group of other participants. How competitive is the shown participant group? A higher number on the scale below indicates a higher perceived competitiveness.

◯ 1 (not at all) ◯ 2 ◯ 3 ◯ 4 ◯ 5 (extremely)

How creative would you rate yourself?

◯ Not at all

◯ Slightly

◯ Moderately

◯ Considerably

◯ Exceptionally

Are you currently pursuing or have you pursued any creative hobbies?

◯ Yes

◯ No

If Yes, which creative hobby/hobbies have you pursued? (If it is more than one, just list the most important three separated with a comma.)

__________________________________________

If Yes, for how many years have you pursued your creative hobby or hobbies? (If it is more than one, just indicate the years of the hobby most important to you.)

__________________________________________

*[Button: Next]*

*[Screen 15]*

## Survey - Part 1b

For the brainstorming task you just performed in general, tell us whether most people believe that men or women, on average, are better at it:

Indicate your answer on the scale below, where 0 means no gender difference. The larger the gender gap, the more you should move the slider in that direction. You must also move the slider once to respond with "0".

(-1: Women are better, on average; 0: no gender difference; 1: Men are better, on average)

For a given object, do you think men or women, on average, are better at coming up with “new and innovative" uses?

-1 1

Your value: **0**

For a given object, do you think men or women, on average, are better at coming up with “new and innovative" uses?

-1 1

Your value: **0**

For a given object, do you think men or women, on average, are better at coming up with “new and innovative" uses?

-1 1

Your value: **0**

For a given object, do you think men or women, on average, are better at coming up with “new and innovative" uses?

-1 1

Your value: **0**

For a given object, do you think men or women, on average, are better at coming up with “new and innovative" uses?

-1 1

Your value: **0**

For a given object, do you think men or women, on average, are better at coming up with “new and innovative" uses?

-1 1

Your value: **0**

For a given object, do you think men or women, on average, are better at coming up with "new and innovative" uses?

-1 1

Your value: **0**

*[Button: Next]*

***[Screen 16]***

## Survey - Part 2

How competitive do you consider yourself to be?
Please choose a value on the scale below, where the value 0 means 'not competitive at all' and the value 10 means 'very competitive'.

◯ 0 (not at all) ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 ◯ 6 ◯ 7 ◯ 8 ◯ 9 ◯ 10 (very competitive)

Imagine the following hypothetical scenario:

You participate in an experiment where you are paid for your performance in a simple task. This task consists of adding up sets of five two-digit numbers without the help of a calculator (so, for example, 43+82+11+94+68=?). You have five minutes to solve as many problems of this type as possible.

You have to choose how you want to be paid for your performance in this task:

◯ Option 1: Piece rate: you receive GBP 1.00 for each correctly solved problem;

◯ Option 2: Competition: you compete against three other people. If you perform better than all three, you receive GBP 4.00 per correctly solved problem; otherwise, you receive nothing.

Your three opponents will be randomly selected among the other participants of this study. Which option would you choose?

◯ Option 1- Piece Rate

◯ Option 2- Competition

In the described task (adding up sets of five two-digit numbers), how well would you perform compared to other participants? In particular, compared to 10 randomly selected participants of this study, do you think you would come 1st, 2nd, 3rd, . . . , 10th?

◯ 1st ... 10th

How do you see yourself: Are you generally willing to take risks, or do you try to avoid taking risks? Please use the scale below, where a higher number indicates a higher willingness to take risks.

◯ 0 (completely unwilling) ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 ◯ 6 ◯ 7 ◯ 8 ◯ 9 ◯ 10 (very willing)

*[Button: Next]*

***[Screen 17]***

## Survey - Part 3

Did you use other devices or switch browser tabs while participating in this study, e.g., during the brainstorming tasks?

Please note that answering this question truthfully will have no negative consequences for your payment.

◯ No

◯ Yes

If your previous answer was "Yes", please specify the outside help you have used during participation in this study below:

________________________________________

What do you think was the objective of this study?

________________________________________

Do you have any comments for us?

________________________________________

*[Button: Next]*

***[Screen 18]***

## End

Thank you for participating in this research study.

Your submitted ideas will be evaluated now, and you will receive your payment within 10 business days. You are guaranteed a participation pay of GBP 2.25.

# List of Figures

# List of Tables

# Bibliography


Abraham, A. (2016). Gender and creativity: An overview of psychological and neuroscientific literature. *Brain Imaging and Behavior*, *10*(2), 609–618.

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., & Kerr, W. (2018). Innovation, reallocation, and growth. *American Economic Review*, *108*(11), 3450–3491.

Agarwal, R., Audretsch, D., & Sarkar, M. (2007). The process of creative construction: Knowledge spillovers, entrepreneurship, and economic growth. *Strategic Entrepreneurship Journal*, *1*(3–4), 263–286.

Aghion, P., & Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, *60*(2), 323–351.

Aghion, P., Howitt, P., Brant-Collett, M., & García-Peñalosa, C. (1998). *Endogenous growth theory*. MIT press.

Akcigit, U., Grigsby, J., & Nicholas, T. (2017). The rise of american ingenuity: Innovation and inventors of the golden age. *National Bureau of Economic Research Working Paper*, (w23047).

Aksoy, B., Chadd, I., & Koh, B. H. (2023). Sexual identity, gender, and anticipated discrimination in prosocial behavior. *European Economic Review*, *154*, 104427.

Aksoy, C. G., Barrero, J. M., Bloom, N., Davis, S. J., Dolls, M., & Zarate, P. (2023). Working from home around the globe: 2023 report. *EconPol Policy Brief*.

Alavi, M., & Leidner, D. E. (2001). Research commentary: Technology-mediated learning—a call for greater depth and breadth of research. *Information Systems Research*, *12*(1), 1–10.

Albanesi, S., Olivetti, C., & Petrongolo, B. (2023). Families, labor markets, and policy. In *Handbook of the economics of the family* (pp. 255–326, Vol. 1). North-Holland.

Al-Shatti, E., & Ohana, M. (2021). Impression management and career related outcomes: A systematic literature review. *Frontiers in Psychology*, *12*, 694–701.

Alston, M. (2019). The (perceived) cost of being female: An experimental investigation of strategic responses to discrimination. *Working Paper*.

Amabile, T. M. (1996). Creativity and innovation in organizations. *Harvard Business School Background Note*, (396-239).

Ammermueller, A., & Pischke, J.-S. (2009). Peer effects in european primary schools: Evidence from the progress in international reading literacy study. *Journal of Labor Economics*, *27*(3), 315–348.

Anger, S. (2008). Overtime work as a signaling device. *Scottish Journal of Political Economy*, *55*(2), 167–189.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Ardichvili, A., Page, V., & Wentling, T. (2003). Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *Journal of Knowledge Management*, *7*(1), 64–77.

Ash, E., & Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, *15*(1), 659–688.

Ashford, S. J., Blatt, R., & VandeWalle, D. (2003). Reflections on the looking glass: A review of research on feedback-seeking behavior in organizations. *Journal of Management*, *29*(6), 773–799.

Athey, S., & Imbens, G. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*, 685–725.

Attanasi, G., Chessa, M., Gil-Gallen, S., & Llerena, P. (2021). A survey on experimental elicitation of creativity in economics. *Revue D'Économie Industrielle*, 273–324.

Auriol, E., Friebel, G., & Pechlivanos, L. (2002). Career concerns in teams. *Journal of Labor Economics*, *20*(2), 289–307.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, *118*(4), 1279–1333. https://doi.org/10.1162/003355303322552801

Baek, J., & Shore, J. (2020). Forum size and content contribution per person: A field experiment. *Management Science*, *66*(12), 5906–5924.

Baek, K., Holton, A., Harp, D., & Yaschur, C. (2011). The links that bind: Uncovering novel motivations for linking on facebook. *Computers in Human Behavior*, *27*(6), 2243–2248.

Baer, J., & Kaufman, J. C. (2008). Gender differences in creativity. *The Journal of Creative Behavior*, *42*(2), 75–105.

Baer, M., & Frese, M. (2003). Innovation is not enough: Climates for initiative and psychological safety, process innovations, and firm performance. *Journal of Organizational Behavior*, *24*(1), 45–68.

Bandiera, O., Parekh, N., Petrongolo, B., & Rao, M. (2022). Men are from mars, and women too: A bayesian meta-analysis of overconfidence experiments. *Economica*, *89*, S38–S70.

Bandiera, O., & Rasul, I. (2006). Social networks and technology adoption in northern mozambique. *The Economic Journal*, *116*(514), 869–902.

Barber, B. M., & Odean, T. (2001). The internet and the investor. *Journal of Economic Perspectives*, *15*(1), 41–54.

Barron, K., Ditlmann, R., Gehrig, S., & Schweighofer-Kodritsch, S. (2024). Explicit and implicit belief-based gender discrimination: A hiring experiment. *Management Science*.

Bechthold, L. A., Chugunova, M., Friess, S., Hoisl, K., & Rose, M. E. (2021). Women in creative labor: Inventors, entrepreneurs and academics. In *Fundamental Questions: Gender Dimensions in Max Planck Research Projects* (pp. 135–154). Nomos Verlagsgesellschaft.

Bénabou, R., Jaroszewicz, A., & Loewenstein, G. (2022). It hurts to ask. *NBER Working Paper, No. w30486*.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., et al. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*(1), 6–10.

Benson, A., Li, D., & Shue, K. (2024). Potential and the gender promotions gap. *SSRN Working Paper, No. 4747175*.

Bettinger, E., Liu, J., & Loeb, S. (2016). Connections matter: How interactive peers affect students in online college courses. *Journal of Policy Analysis and Management*, *35*(4), 932–954.

Bidwell, M., & Briscoe, F. (2010). The dynamics of interorganizational careers. *Organization Science*, *21*(5), 1034–1053.

Bidwell, M., Briscoe, F., Fernandez-Mateo, I., & Sterling, A. (2013). The employment relationship and inequality: How and why changes in employment practices are reshaping rewards in organizations. *Academy of Management Annals*, *7*(1), 61–121.

Bloom, N., Jones, C. I., Van Reenen, J., & Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, *110*(4), 1104–44. https://doi.org/10.1257/aer.20180338

Blunden, H., Logg, J. M., Brooks, A. W., John, L. K., & Gino, F. (2019). Seeker beware: The interpersonal costs of ignoring advice. *Organizational Behavior and Human Decision Processes*, *150*, 83–100.

Bohren, J. A., Imas, A., & Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American Economic Review*, *109*(10), 3395–3436.

Booth, A., Cardona-Sosa, L., & Nolen, P. (2014). Gender differences in risk aversion: Do single-sex environments affect their development? *Journal of Economic Behavior & Organization*, *99*, 126–154.

Booth, A., & Nolen, P. (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, *81*(2), 542–555.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, *109*(3), 739–73.

Borgatti, S. P., & Cross, R. (2003). A relational view of information seeking and learning in social networks. *Management Science*, *49*(4), 432–445.

Born, A., Ranehill, E., & Sandberg, A. (2022). Gender and willingness to lead: Does the gender composition of teams matter? *Review of Economics and Statistics*, *104*(2), 259–275.

Bostwick, V. K., & Weinberg, B. A. (2022). Nevertheless she persisted? Gender peer effects in doctoral stem programs. *Journal of Labor Economics*, *40*(2), 397–436.

Bradler, C., Neckermann, S., & Warnke, A. J. (2019). Incentivizing creativity: A large-scale experiment with performance bonuses and gifts. *Journal of Labor Economics*, *37*(3), 793–851.

Brenøe, A. A., Heursen, L., Ranehill, E., & Weber, R. A. (2022). Continuous gender identity and economics. *AEA Papers and Proceedings*, *112*, 573–577.

Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart people ask for (my) advice: Seeking advice boosts perceptions of competence. *Management Science*, *61*(6), 1421–1435.

Bursztyn, L., Fujiwara, T., & Pallais, A. (2017). Acting wife: Marriage market incentives and labor market investments. *American Economic Review*, *107*(11), 3288–3319.

Buser, T., Grimalda, G., Putterman, L., & van der Weele, J. (2020). Overconfidence and gender gaps in redistributive preferences: Cross-country experimental evidence. *Journal of Economic Behavior & Organization*, *178*, 267–286.

Buser, T., Niederle, M., & Oosterbeek, H. (2024). Can competitiveness predict education and labor market outcomes? Evidence from incentivized choice and survey measures. *Review of Economics and Statistics*, 1–45.

Cai, X., Lu, Y., Pan, J., & Zhong, S. (2019). Gender gap under pressure: Evidence from china's national college entrance examination. *Review of Economics and Statistics*, *101*(2), 249–263.

Calvó-Armengol, A., Patacchini, E., & Zenou, Y. (2009). Peer effects and social networks in education. *The Review of Economic Studies*, *76*(4), 1239–1267.

Camerer, C., & Malmendier, U. (2007). Behavioral organizational economics. In *Behavioral economics and its applications* (pp. 235–290). Princeton University Press.

Campbell, E. L., & Hahl, O. (2022). He's overqualified, she's highly committed: Qualification signals and gendered assumptions about job candidate commitment. *Organization Science*, *33*(6), 2451–2476.

Campos-Mercade, P., & Mengel, F. (2024). Non-bayesian statistical discrimination. *Management Science*, *70*(4), 2549–2567.

Cappelli, P. H. (2015). Skill gaps, skill shortages, and skill mismatches: Evidence and arguments for the united states. *ILR Review*, *68*(2), 251–290.

Castro, S., Englmaier, F., & Guadalupe, M. (2022). Fostering psychological safety in teams: Evidence from an rct. *SSRN Working Paper, No. 4141538*.

Chandrasekhar, A. G., Golub, B., & Yang, H. (2018). Signaling, shame, and silence in social learning. *NBER Working Paper, No. w25169*.

Charness, G., Cobo-Reyes, R., Meraglia, S., & Sánchez, Á. (2020). Anticipated discrimination, choices, and performance: Experimental evidence. *European Economic Review*, *127*, 103473.

Charness, G., & Grieco, D. (2019). Creativity and incentives. *Journal of the European Economic Association*, *17*(2), 454–496.

Chen, D. L., Schonger, M., & Wickens, C. (2016). Otree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

Chen, L., Baird, A., & Straub, D. (2019). Why do participants continue to contribute? Evaluation of usefulness voting and commenting motivational affordances within an online knowledge community. *Decision Support Systems*, *118*, 21–32.

Claussen, J., Halbinger, M., & Hermida Carrillo, A. (2022). Thanks for fixing my 3d printer: How and when users reciprocate tool-related assistance in user innovation communities. *SSRN Working Paper, No. 4286464*.

Claussen, J., & Halbinger, M. A. (2021). The role of pre-innovation platform activity for diffusion success: Evidence from consumer innovations on a 3d printing platform. *Research Policy*, *50*(8), 103943.

Claussen, J., Kretschmer, T., & Mayrhofer, P. (2013). The effects of rewarding user engagement: The case of facebook apps. *Information Systems Research*, *24*(1), 186–200.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, *129*(4), 1625–1660.

Coffman, K. B., Exley, C. L., & Niederle, M. (2021). The role of beliefs in driving gender discrimination. *Management Science*, *67*(6), 3551–3569.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, *47*(2), 448–474.

de Janvry, A., He, G., Sadoulet, E., Wang, S., & Zhang, Q. (2023). Subjective performance evaluation, influence activities, and bureaucratic work behavior: Evidence from china. *American Economic Review*, *113*(3), 766–799.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. *HBS TOM Unit Working Paper*, (24-013).

Deming, D. J., & Noray, K. L. (2018). Stem careers and the changing skill requirements of work. *NBER Working Paper, No. 25065*.

Dingel, J. (2020). How many jobs can be done at home?

Dragoni, L., Tesluk, P. E., Russell, J. E., & Oh, I.-S. (2009). Understanding managerial development: Integrating developmental assignments, learning orientation, and access to developmental opportunities in predicting managerial competencies. *Academy of Management Journal*, *52*(4), 731–743.

Drucker, P. F. (1999). *Management challenges for the 21st century*. HarperBusiness.

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. In *Handbook of development economics* (pp. 3895–3962). Elsevier.

Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annu. Rev. Organ. Psychol. Organ. Behav.*, *1*(1), 23–43.

Einiö, E., Feng, J., & Jaravel, X. (2019). Social push and the direction of innovation. *SSRN Working Paper, No. 3383703*.

Erat, S., & Gneezy, U. (2016). Incentives for creativity. *Experimental Economics*, *19*, 269–280.

Espinosa, M., & Stanton, C. T. (2022). Training, communications patterns, and spillovers inside organizations. *NBER Working Paper, No. w30224*.

European Investment Bank. (2022). *Support for female entrepreneurs: Evidence from europe* (tech. rep.) (Accessed: 2024-09-17). European Investment Bank. https:

//www.eib.org/attachments/lucalli/support_for_female_entrepreneurs_en.pdf

Exley, C. L., & Kessler, J. B. (2022). The gender gap in self-promotion. *The Quarterly Journal of Economics*, *137*(3), 1345–1381.

Exley, C. L., & Nielsen, K. (2024). The gender gap in confidence: Expected but not accounted for. *American Economic Review*, *114*(3), 851–885.

Falk, A., Becker, A., Dohmen, T., Huffman, D., & Sunde, U. (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science*, *69*(4), 1935–1950.

Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization science*, *22*(5), 1224–1239.

Faraj, S., & Johnson, S. L. (2011). Network exchange patterns in online communities. *Organization Science*, *22*(6), 1464–1480.

Faraj, S., von Krogh, G., Monteiro, E., & Lakhani, K. R. (2016). Special section introduction—online community as space for knowledge flows. *Information Systems Research*, *27*(4), 668–684.

Fausto-Sterling, A. (2012). *Sex/gender: Biology in a social world*. Routledge.

Feld, J., & Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, *35*(2), 387–428.

Feld, J., & Zölitz, U. (2022). The effect of higher-achieving peers on major choices and labor market outcomes. *Journal of Economic Behavior & Organization*, *196*, 200–219.

Foster, A. D., & Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, *103*(6), 1176–1209.

Gallen, Y., & Wasserman, M. (2021). Informed choices: Gender gaps in career advice.

Gallus, J., & Frey, B. S. (2016). Awards: A strategic management perspective. *Strategic Management Journal*, *37*(8), 1699–1714.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, *108*(5), 874–904.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574.

Gerber, J. P., Wheeler, L., & Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin*, *144*(2), 177.

Gibbons, R., & Murphy, K. J. (1992). Optimal incentive contracts in the presence of career concerns: Theory and evidence. *Journal of Political Economy*, *100*(3), 468–505.

Gibbs, M., Merchant, K. A., Stede, W. A. V. d., & Vargus, M. E. (2004). Determinants and effects of subjectivity in incentives. *The Accounting Review*, *79*(2), 409–436.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, *118*(3), 1049–1074.

Gneezy, U., & Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review*, *94*(2), 377–381.

Gompers, P. A., & Wang, S. Q. (2017). *Diversity in innovation* (tech. rep.). National Bureau of Economic Research.

Griffith, A. L. (2010). Persistence of women and minorities in stem field majors: Is it the school that matters? *Economics of Education Review*, *29*(6), 911–922.

Gross, D. P. (2020). Creativity under fire: The effects of competition on creative production. *Review of Economics and Statistics*, *102*(3), 583–599.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*(4), 1009–1055.

Heikensten, E., & Isaksson, S. (2019). Simon says: Examining gender differences in advice seeking and influence in the lab. *SSRN Working Paper, No. 3273186*.

Hill, R., & Stein, C. (2019). Scooped! estimating rewards for priority in science. *Job Market Paper*.

Hofstra, B., Kulkarni, V. V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity–innovation paradox in science. *Proceedings of the National Academy of Sciences*, *117*(17), 9284–9291.

Hogarth, R. M., Karelaia, N., & Trujillo, C. A. (2012). When should i quit? Gender differences in exiting competitions. *Journal of Economic Behavior & Organization*, *83*(1), 136–150.

Hoisl, K., Kongsted, H. C., & Mariani, M. (2023). Lost marie curies: Parental impact on the probability of becoming an inventor. *Management Science*, *69*(3), 1714–1738.

Hoisl, K., & Mariani, M. (2017). It's a man's job: Income and the gender gap in industrial research. *Management Science*, *63*(3), 766–790.

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies*, *66*(1), 169–182.

Hora, S., Badura, K. L., Lemoine, G. J., & Grijalva, E. (2022). A meta-analytic examination of the gender difference in creative performance. *Journal of Applied Psychology*, *107*(11), 1926.

Hoyt, C. L., & Murphy, S. E. (2016). Managing to clear the air: Stereotype threat, women, and leadership. *The leadership quarterly*, *27*(3), 387–399.

Hwang, E. H., Singh, P. V., & Argote, L. (2015). Knowledge sharing in online communities: Learning to cross geographic and hierarchical boundaries. *Organization Science*, *26*(6), 1593–1611.

Iaria, A., Schwarz, C., & Waldinger, F. (2024). Gender gaps in academia: Global evidence over the twentieth century. *SSRN Working Paper, No. 4150221*.

Illanes, P., Lund, S., Mourshed, M., Rutherford, S., & Tyreman, M. (2018). Retraining and reskilling workers in the age of automation. *McKinsey Global Institute, 8*.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological science*, *11*(5), 365–371.

Iriberri, N., & Rey-Biel, P. (2017). Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision. *Journal of Economic Behavior & Organization*, *135*, 99–111.

Irlenbusch, B., Laske, K., Schroeder, M., & Stoczek, R. (2024). Path dependencies and innovation - An experimental study of the effect of knowledge on idea generation and selection. *Unpublished Manuscript*.

Irlenbusch, B., & Sliwka, D. (2006). Career concerns in a simple experimental labour market. *European Economic Review*, *50*(1), 147–170.

Jin, G. Z., Jones, B., Lu, S. F., & Uzzi, B. (2019). The reverse matthew effect: Consequences of retraction in scientific teams. *Review of Economics and Statistics*, *101*(3), 492–506.

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.

Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research*, *15*(2), 194–210.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Katok, E., & Siemsen, E. (2011). Why genius leads to adversity: Experimental evidence on the reputational effects of task difficulty choices. *Management Science*, *57*(6), 1042–1054.

Kessel, D., Mollerstrom, J., & Van Veldhuizen, R. (2021). Can simple advice eliminate the gender gap in willingness to compete? *European Economic Review*, *138*, 103777.

Khattab, J., Van Knippenberg, D., Pieterse, A. N., & Hernandez, M. (2020). A network utilization perspective on the leadership advancement of minorities. *Academy of Management Review*, *45*(1), 109–129.

Khobzi, H., Lau, R. Y., & Cheung, T. C. (2019). The outcome of online social interactions on facebook pages: A study of user engagement behavior. *Internet Research*, *29*(1), 2–23.

Kleinbaum, A. M., Stuart, T. E., & Tushman, M. L. (2013). Discretion within constraint: Homophily and structure in a formal organization. *Organization Science*, *24*(5), 1316–1336.

Kleine, M. (2021). No eureka! incentives hurt creative breakthrough irrespective of the incentives' frame. *Incentives Hurt Creative Breakthrough Irrespective of the Incentives' Frame (May 18, 2021). Max Planck Institute for Innovation & Competition Research Paper*, (21-15).

Koch, A. K., Morgenstern, A., & Raab, P. (2009). Career concerns incentives: An experimental test. *Journal of Economic Behavior & Organization*, *72*(1), 571–588.

Koivisto, M., & Grassini, S. (2023). Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific reports*, *13*(1), 13601.

Kolev, J., Fuentes-Medel, Y., & Murray, F. (2019). Is blinded review enough? How gendered outcomes arise even under anonymous evaluation. *Working Paper No. w25759*.

Koning, R., Samila, S., & Ferguson, J.-P. (2021). Who do we invent for? Patents by women focus more on women's health, but few women get to invent. *Science*, *372*(6548), 1345–1348.

Kretschmer, T., Leiponen, A., Schilling, M., & Vasudeva, G. (2022). Platform ecosystems as meta-organizations: Implications for platform strategies. *Strategic Management Journal*, *43*(3), 405–424.

Kreyer, A.-C., & Wang, L. X. (2022). Collaborating neuroscience online: The case of the human brain project forum. *PLOS ONE*, *17*(12), e0278402.

Kudashvili, N., & Lergetporer, P. (2022). Minorities' strategic response to discrimination: Experimental evidence. *Journal of Public Economics*, *208*, 104630.

Kwon, K., Park, S. J., Shin, S., & Chang, C. Y. (2019). Effects of different types of instructor comments in online discussions. *Distance Education*, *40*(2), 226–242.

Laske, K., & Schroeder, M. (2017). Quantity, quality and originality: The effects of incentives on creativity.

Lee, F. (1997). When the going gets tough, do the tough ask for help? Help seeking and power motivation in organizations. *Organizational Behavior and Human Decision Processes*, *72*(3), 336–363.

Lee, F. (2002). The social costs of seeking help. *The Journal of Applied Behavioral Science*, *38*(1), 17–35.

Lerner, J., & Malmendier, U. (2013). With a little help from my (random) friends: Success and failure in post-business school entrepreneurship. *The Review of Financial Studies*, *26*(10), 2411–2452.

Lev, B. (2000). *Intangibles: Management, measurement, and reporting*. Rowman & Littlefield.

Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, *53*(1), 1–18.

Liljenquist, K. A. (2010). Resolving the impression management dilemma: The strategic benefits of soliciting others for advice. *Doctoral Dissertation, Northwestern University*.

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in moocs: Motivations and self-regulated learning in moocs. *The Internet and Higher Education*, *29*, 40–48.

Loh, J. (2022). Selection, consumption, and new music exploration in an online social network: A dyadic approach.

Loh, J., & Kretschmer, T. (2023). Online communities on competing platforms: Evidence from game wikis. *Strategic Management Journal*, *44*(2), 441–476.

Makos, A., Oztok, M., Zingaro, D., & Hewitt, J. (2013). Use of a "like" button in a collaborative online learning environment.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542.

Marianne, B. (2011). New perspectives on gender. In *Handbook of labor economics* (pp. 1543–1590, Vol. 4). Elsevier.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.

Mickeler, M., Khashabi, P., Kleine, M., & Kretschmer, T. (2023). Knowledge seeking and anonymity in digital work settings. *Strategic Management Journal*.

Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2022). Managing self-confidence: Theory and experimental evidence. *Management Science*, *68*(11), 7793–7817.

Moretti, E. (2011). Social learning and peer effects in consumption: Evidence from movie sales. *The Review of Economic Studies*, *78*(1), 356–393.

Morris, S. (2001). Political correctness. *Journal of Political Economy*, *109*(2), 231–265.

Munshi, K. (2004). Social learning in a heterogeneous population: Technology diffusion in the indian green revolution. *Journal of Development Economics*, *73*(1), 185–213.

National Science Foundation. (2023). *Diversity and stem: Women, minorities, and persons with disabilities in science and engineering* (tech. rep. No. NSF 23-315) (Accessed: 2024-09-17). National Science Foundation, National Center for Science and Engineering Statistics. https://ncses.nsf.gov/pubs/nsf23315/

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, *122*(3), 1067–1101.

Niederle, M., & Vesterlund, L. (2011). Gender and competition. *Annu. Rev. Econ.*, *3*(1), 601–630.

Ottaviani, M., & Sørensen, P. N. (2006). The strategy of professional forecasting. *Journal of Financial Economics*, *81*(2), 441–466.

PitchBook. (2024). *The vc female founders dashboard* [Accessed: 2024-09-17]. https://pitchbook.com/news/articles/the-vc-female-founders-dashboard

Prendergast, C., & Stole, L. (1996). Impetuous youngsters and jaded old-timers: Acquiring a reputation for learning. *Journal of Political Economy*, *104*(6), 1105–1134.

Price, B., & Price, B. (2013). The changing task composition of the us labor market: An update of autor, levy, and murnane (2003). *Unpublished Manuscript*.

Rabin, M. (2002). A perspective on psychology and economics. *European Economic Review*, *46*(4–5), 657–685.

Reagans, R. (2011). Close encounters: Analyzing how social similarity and propinquity contribute to strong network connections. *Organization Science*, *22*(4), 835–849.

Ren, Y., Harper, F. M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., & Kraut, R. E. (2012). Building member attachment in online communities: Applying theories of group identity and interpersonal bonds. *MIS Quarterly*, 841–864.

Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies*, *28*(3), 377–408.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, *111*(12), 4403–4408.

Rietveld, J., & Schilling, M. A. (2021). Platform competition: A systematic and interdisciplinary review of the literature. *Journal of Management*, *47*(6), 1528–1563.

Rocha, V., & Van Praag, M. (2020). Mind the gap: The role of gender in entrepreneurial career choice and social influence by founders. *Strategic Management Journal*, *41*(5), 841–866.

Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, *1*(4), 990–1029.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, *98*(5, Part 2), S71–S102.

Rosendahl Huber, L., N Lane, J., & Lakhani, K. R. (2020). Learning with people like me: The role of age-similar peers on online business course engagement.

Rosette, A. S., Mueller, J. S., & Lebel, R. D. (2015). Are male leaders penalized for seeking help? The influence of gender and asking behaviors on competence perceptions. *The Leadership Quarterly*, *26*(5), 749–762.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, *116*(2), 681–704.

Sandvik, J. J., Saouma, R. E., Seegert, N. T., & Stanton, C. T. (2020). Workplace knowledge flows. *The Quarterly Journal of Economics*, *135*(3), 1635–1680.

Sarsons, H. (2019). Interpreting signals in the labor market: Evidence from medical referrals. *Working Paper*.

Sarsons, H., Gërxhani, K., Reuben, E., & Schram, A. (2021). Gender differences in recognition for group work. *Journal of Political Economy*, *129*(1), 101–147.

Schumpeter, J. A. (1942). *Capitalism, socialism and democracy*. Harper & Brothers.

Shurchkov, O. (2012). Under pressure: Gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association*, *10*(5), 1189–1213.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology*, *35*(1), 4–28.

Strathern, W., & Pfeffer, J. (2022). Identifying different layers of online misogyny. *Working Paper. arXiv:2212.00480*.

Surma, J. (2016). Social exchange in online social networks. The reciprocity phenomenon on facebook. *Computer Communications*, *73*, 342–346.

Tamayo, J., Doumi, L., Goel, S., Kovács-Ondrejkovic, O., & Sadun, R. (2023). Reskilling in the age of ai. *Harvard Business Review*, *21*.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, *1*(1), 39–60.

Torrance, E. P. (1968). A longitudinal examination of the fourth grade slump in creativity. *Gifted Child Quarterly*, *12*(4), 195–199.

Truffa, F., & Wong, A. (2022). Undergraduate gender diversity and direction of scientific research. *Working Paper*.

Ulku-Steiner, B., Kurtz-Costes, B., & Kinlaw, C. R. (2000). Doctoral student experiences in gender-balanced and male-dominated graduate programs. *Journal of Educational Psychology*, *92*(2), 296–307.

Venter, E. (2017). Bridging the communication gap between generation y and the baby boomer generation. *International Journal of Adolescence and Youth*, *22*(4), 497–507.

Wang, C., Zhang, X., & Hann, I.-H. (2018). Socially nudged: A quasi-experimental study of friends' social influence in online product ratings. *Information Systems Research*, *29*(3), 641–655.

Wasko, M. M., & Faraj, S. (2005). Why should i share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, 35–57.

West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender & Society*, *1*(2), 125–151.

World Economic Forum. (2024, March). *Why closing the gender gap among inventors is critical for innovation* [Accessed: 2024-09-17]. https://www.weforum.org/agenda/2024/03/women-inventors-gender-gap-stem-patents/

Xu, L., Nian, T., & Cabral, L. (2020). What makes geeks tick? A study of stack overflow careers. *Management Science*, *66*(2), 587–604.

Zhang, C., Hahn, J., & De, P. (2013). Research note—continued participation in online innovation communities: Does community response matter equally for everyone? *Information Systems Research*, *24*(4), 1112–1130.

Zhang, D. J., Allon, G., & Van Mieghem, J. A. (2017). Does social interaction improve learning outcomes? Evidence from field experiments on massive open online courses. *Manufacturing & Service Operations Management*, *19*(3), 347–367.

Ziegele, M., Weber, M., Quiring, O., & Breiner, T. (2018). The dynamics of online news discussions: Effects of news articles and reader comments on users' involvement, willingness to participate, and the civility of their contributions. *Information, Communication & Society*, *21*(10), 1419–1435.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics*, *85*(1), 9–23.