The Making-Of: Innovation

Understanding and Designing the Environment for Non-R&D Innovation

A thesis submitted in fulfillment
of the requirements for the degree of Dr. rer. pol.
from the Faculty of Economic Sciences
University of Göttingen

by

Elaine Horstmann

Göttingen, 2021

First Supervisor: Prof. Dr. Kilian Bizer
Second Supervisor: Prof. Dr. Margarete Boos
Third Supervisor: Jun.-Prof. Dr. Holger Rau

Acknowledgements

Just as innovation is not the result of an isolated individual effort, neither was this thesis written in isolation. I am very thankful for all the support during this intense phase of my life, even though I cannot name everyone individually.

At first, I would like to thank my supervisors Prof. Dr. Kilian Bizer, Prof. Dr. Margarete Boos and Jun.-Prof. Dr. Holger Rau, for making this PhD project possible and for leaving me free to find my own way and create experiences on my own.

Prof. Dr. Kilian Bizer, I thank you for giving me the space to run free and pursue the questions I was most interested in. Thank you for always providing me with enough resources to conduct my experiments, cooperating with whom I wanted to and bandying research ideas when we should have been discussing topics for the next lecture. Your bi-annual doctoral seminars mostly drove me nuts in the days before, but have been the gentle push I needed to get things done. Finally, thanks for supporting my research stay in Aarhus and several conference attendance that enabled me to experience the fun part of academia even though it was tough.

Especially two colleagues accompanied me during my thesis project. Dr. Ann-Kathrin Blankenberg, thank you for supporting me in the beginning and advocating for me. It was a pleasure to attend conferences in Montreal and Tucson with you and if it was not for you, I probably would have missed the world of research conferences all together. Alexander Erlei, thank you so much for the last two years of coffee, rice cakes and being my sparring partner. Shooting research ideas and vividly discussing economics, politics, psychology, and all the other stuff made my ideal of scientific discussions come true. Thanks for motivating me to always go an extra round and creating new, comforting routines when the workplace fell apart.

My time at the BSS in Aarhus has been so special and pivotal for me. Jacob Orquin (PhD) thank you so much for offering your Neuroeconomics course back in 2015 which completely changed research directions for me and wanted me to come back to BSS. Your way of teaching and organizing your research group – closely related in terms of content and yet different at the same time – is a role model for me. You always pushed me to think big and have been critical with my ideas which make me work on stating my research ideas more precisely and sharpening my arguments. Dorthe Døjbak Håkonsson (PhD) thank you for introducing me to ICOA, the invigorating discussions we had and making my stay at BSS unforgettable.

Finally, the support of my family and partner was essential during this project. Malte, thank you so much for being my partner and on my side, even in the most desperate moments. You are always there for me and put things into perspective when I lost it. You perfectly ground me when I go crazy and keep me from overdoing it. Thanks for granting me the freedom I needed and creating a home and safe space to always fall back on. My parents and my sister, thanks for being the best and unconventional family I could wish for. You provided me all the tools I needed to pursue my goals and never give up. You are my role models in performing against all odds and going beyond your own limits, while also freeing up space to relax and not take things too serious. It all started with ancient Egypt and ended up with behavioral economics, Mama and Papa, I dedicate this thesis to you.

Contents

Chapter 1

Introduction

When was the last time you used *PayPal*? In the third quarter of 2021, PayPal registers 416 million active accounts worldwide, about 4.9 billions cashless payments, and a total payment volume of USD 310 billion (PayPal 2021). Back in 1998, when PayPal was founded under its original name Confinity, this development was not foreseeable. Back then, Confinity was a software company that developed security software and e-mail payment systems, struggling for success. In 1999, the company decided to shift gears and launched a platform for digital payment (PayPal), starting its success story only one year later after merging with Elon Musk's online bank X.com in 2000 (Rawat 2021). What began as a completely different business venture evolved over a very short time period, on the back of a few individuals, into one of the most successful companies worldwide. In other words, Confinity innovated, and they did so without any traditional R&D structures.

Maybe, you have also heard of *Wise* (formerly TransferWise). Two Estonians working in London, one, a former Deloitte manager helping banks and insurers to transform their processes and systems named Kristo Käärmann, and the other, Taavet Hinrikus, a former director of strategy and Skype's first-ever employee, joined forces after being annoyed by high bank fees when transferring money abroad and created an innovative financial service (*Wise* 2021). While working, living, or studying abroad is common for more and more people, transferring money internationally or across currency-borders is still extremely costly. Recognizing that problem, Wise started as a company dedicated to provide international and inter-currency money transfers at low costs. Today, more than 10 million people worldwide used Wise to send money internationally with an average monthly transaction volume of 6 billion USD (*Wise* 2021). Again, it only took two Estonians who identified a problem, and subsequently applied their cumulative experiential domain-knowledge to create an innovative solution that has grown into a large, global financial service provider.

Finally, we turn our eyes to calligraphy cut, a hairdressing company in Germany. One day, while cutting flowers, the coiffeur Frank Brormann realized that the cross-section could be enlarged by an oblique cut, making the fibers more elastic and flexible, which is often a problem in hair-cutting (*Calligraphy Cut* 2021). Intrigued by his discovery, he invented a new, innovative hairdressing tool – the *calligraph* – that allows the coiffeur to cut the hair at a constant 21 degree angle, extending the cross-section and thus significantly improving the final product, the haircut. Today, the tool has earned him several innovation awards.

All three innovation examples have been developed outside an R&D lab or other forms of formalized innovation structure, pointing to the variety and impact of non-R&D innovation rooted in diverse knowledge sources: experienced knowledge from a mediocre business model of a software company re-combined with an online bank resulting in a new, improved enterprise, cumulative knowledge gained from day-to-day work re-used to solve an initially personal problem and leading to start-up in a different field, or being inspired in everyday life to make an innovative contribution to a standard and routine work step in one's craftsmanship. This thesis aims to contribute to the literature on non-R&D innovation by broadening and deepening the appreciation of learning and knowledge creation processes crucial for employee and individual-driven innovation attainment. It empirically examines organizational design elements and innovation management tools to encourage innovative behavior within a firm.

1 Conceptions of learning behind non-R&D innovations

Non-R&D innovations emerge without formalized innovation structures or earmarked resources, mostly as a by-product of day-to-day work outside of R&D labs. Both, the lack of formalized structures and the vast heterogeneity of the products, processes and organizations comprised within 'by-products of day-to-day work' severely exacerbates issues of detection and measurement. This has lead to the widespread use of insufficiently differentiated proxies. As most work uses firm-level data that classifies any innovation created by a firm performing R&D as an R&D-innovation, researchers often neglect the true origin of knowledge creation, i.e. R&D or non-R&D work (Lee and Walsh 2016). This leads to a subsequent juxtaposition of large, high-tech firms with R&D vs. small, low-tech firms without R&D, eroding any within-firm heterogeneity in the source of innovation and disregarding non-R&D innovation as a phenomenon also found in large, high-tech firms (Arundel et al. 2008; Barge-Gil et al. 2011; Lee 2015; Lee and Walsh 2016; Rammer et al. 2009). One way to remedy this deficiency is to focus on differences in the nature of knowledge and learning as environmental factors for R&D and non-R&D innovations, rather than the size and capital of organizations.

In today's knowledge economy, learning and knowledge creation are decisive for a firm's market performance, which inextricably connects them to innovation performance (Amara et al. 2008; Argote and Fahrenkopf 2016; Argote and Ingram 2000; Argote and Miron-Spektor 2011; Lundvall and Nielsen 2007; McIver et al. 2013). Therefore, several types and distinctions of knowledge or learning, their loci within an organization, and their impact on innovation have been discussed, ranging from regional innovation systems over organizational levels to its very micro-level characteristics, like tacit vs. explicit knowledge. While these concepts have slightly different distinctions and classifications, they share similar underlying characteristics. The concept of knowledge bases is widely discussed in the context of regional innovation systems. Here, knowledge can be of 'analytical' nature where knowledge creation is primarily based on formal models and scientific protocols accompanied by codified and explicit knowledge (Asheim 2007; Asheim and Coenen 2006). Analytical knowledge thus comprises more formally organized knowledge processes and is often the foundation for patents or licensing activities (Asheim 2007). In contrast, 'synthetic' knowledge is more context-specific and sticky, as it describes the application and novel combination of existing knowledge to solve specific problems, accompanied by knowledge creation through testing, experimentation and practical work (Asheim and Coenen 2006). Moreover, the properties of synthetic knowledge bases are of a more tacit nature, although tacitness is not exclusive for synthetic knowledge as both kinds of knowledge are involved in the innovation creation process (Johnson et al. 2002; Nonaka et al. 2000).

On an organizational level, Lee and Walsh (2016) distinguish between general and visible knowledge environments, two concepts that are closely linked to analytical and synthetic knowledge bases. General knowledge – similar to analytical knowledge bases – is defined as abstract, codified and articulated in universal terms, thereby applicable in diverse contexts (Lee and Walsh 2016; Pavitt 1984). In R&D contexts, general knowledge is more relevant and fits the formalized structure and target. In contrast, context-specific, sticky knowledge is predominant in routinized regimes (Winter 1984), making skilled non-R&D workers effective in less general knowledge environments (Lee and Walsh 2016). Thus, environments characterized by lower relevance of general knowledge are equivalent to synthetic knowledge bases, whereas the second dimension, visible knowledge, is closely related to the engineer-based description of synthetic knowledge (Asheim 2007), since visibility delineates tighter links between actions and outcomes, highlighting how actions directly affect outcome (Lee and Walsh 2016). While high visibility of problems will also affect learning in formal R&D structures, its relevance is generally more important in non-R&D contexts (ibid.).

Finally, another approach to understand and distinguish learning and knowledge creation processes in R&D and non-R&D contexts is the differentiation between learning by *science, technology, and innovation* (STI) and *doing, using, and interacting* (DUI) as proposed by Jensen et al. (2007). These two ideal-typical modes share similar characteristics as those described above, where STI relates to analytical, codified and general knowledge and DUI to synthetic, sticky and context-specific knowledge. As the stylized STI-DUI-conception proposes the most tractable framework of learning processes behind non-R&D innovation, the DUI-mode is used as the theoretical foundation for this thesis, especially in the first three essays. Learning by doing describes knowledge accumulation by daily learning, learning from working experience or carrying out a task (Alhusen et al. 2021; Arrow 1962; Jensen et al. 2007; Thompson 2010). When consuming or applying a respective good or service, knowledge is gained by learning by using (Lundvall 2016; Lundvall and Johnson 1994). As this is not restricted to final products but also comprises working with prototypes, learning by using can occur inside, e.g. using prototypes or machines/equipment (Amara et al. 2008; Rosenberg 1982), or outside the firm, e.g. by co-creation with users (Von Hippel 1986). Finally, learning by interacting illustrates learning and innovation opportunities by interacting with colleagues (internal), or external actors like suppliers or public institutions (Aslesen et al. 2012; Fitjar and Rodrıguez-Pose 2013; Haus-Reve et al. 2019). Based on these three core elements, this thesis develops and tests a new real-effort task to experimentally examine what drives or inhibits non-R&D innovation and ultimately come to a deeper understanding of the kinds of innovations that surround us in our daily work-routines.

2 Innovation management tools

As non-R&D innovation emerge in less innovation-focused environments, the design of work processes and structures often fails to account for innovation attainment, highlighting the importance of management tools to create innovation-fostering spheres (Barge-Gil et al. 2011; Hidalgo and Albors 2008; Rammer et al. 2009; Volberda et al. 2013). There is a substantial number of techniques and tools for innovation management (for a review, see e.g. Hidalgo and Albors (2008)), ranging from knowledge and human resource management to information and communication technologies. This thesis discusses and applies three specific interventions: goal-setting, monetary incentive schemes to direct effort and explore new approaches, and the delegation of compensation decision processes to workers.

Goal setting theory is a widely adopted motivational technique that enhances performance effectiveness and stimulates problem-solving. Goals enhance the visibility of problems and thus, point out what needs to be attained and where to allocate resources (c.f., Locke and Latham 2013c). The best known goals are performance and learning goals. While the former direct attention and effort to outcome performance, the latter motivate to explore new or innovative approaches to a task and foster learning (Latham 2004; Seijts and Latham 2005; Seijts et al. 2013). Moreover, goals differ in terms of difficulty and reference points. The most prominent types are do-your-best goals, where no reference point is set, but workers are encouraged to give their best. Contrary, specific, challenging goals set a point of reference which is usually not easily attainable (Locke and Latham 2002; Seijts and Latham 2005). Therefore, setting goals in non-R&D innovation contexts can help individuals by making targets or problems more visible, shift focus, and motivate perseverance in the search for new approaches that solve problems.

Another widely accepted and adopted tool to incentivize motivation and effort is the application of monetary incentives (Camerer and Hogarth 1999; Jenkins et al. 1998). Monetary incentives can also direct effort and attention to what should be obtained through increases in motivation, which drives effort and consecutively provokes performance gains (Bonner and Sprinkle 2002). From the various options to design and implement monetary incentives (c.f., Bonner and Sprinkle 2002), we choose very two common but opposing schemes: fixed and piece-rate payment (e.g., Cadsby et al. 2007; Lazear 1986). Similar to learning goals, a fixed payment emphasizes the input. Piece-rate payment accentuates the output, analogous to performance goals. Thus, reduced outcome focus and lower performance pressure associated with a fixed payment scheme is hypothesized to release capacities for the creative multistage process of innovation (Amabile et al. 1996; Grabner 2014; Holmstrom and Milgrom 1991; Manso 2011; Shalley et al. 2004). On the other hand. the output focus of piece-rate payments can distract individuals from searching for new approaches, diverting attention from innovative solutions (Ederer and Manso 2013). However, piece-rate payments are not necessarily detrimental for innovative solutions, as they fulfill the monotonicity condition and thus indirectly incentivize efficiency improvements in order to increase output and monetary payoff (Smith 1976). Improving performance effectiveness also comprise opportunities for innovative solution paths to match such effectiveness.

Third, participatory mechanisms that involve participants as stakeholders in determining organisation frameworks by giving them more opportunities for self-management and self-determination have been shown to exhibit positive effects on long-run effort and performance (Franke et al. 2016; Harrison and Freeman 2004; Semler 1989, 2007; Sliwka 2001). One mode of participation is the involvement in or delegation of compensation decision processes (Jeworrek and Mertins 2019; Shaw 2021). Prior research finds positive effects of such interventions on motivation and creativity (Brück et al. 2021; Charness et al. 2012; Harrison and Freeman 2004; Mellizo et al. 2014). Moreover, increasing self-determination by involving participants has been discussed to enhance intrinsic motivation (e.g., Gagné and Deci 2005), which, in turn, has often been proposed to be crucial for creativity and thus innovation (Amabile and Gitomer 1984; Bailyn 1985; Paolillo and Brown 1978).

Besides being documented in the literature, qualitative interviews with practitioners discussed in chapter two show that the positive impact of these tools and mechanisms is acknowledged by practitioners and applied in firms. To deepen our understanding of how these tools influence non-R&D innovation, this thesis experimentally applies them to a new real-effort task which was explicitly developed on the basis of learning by doing, using, and interacting.

3 Innovation units: the individual and the team

The smallest unit and origin of innovation, either R&D or non-R&D, are individuals or teams. Thus, to complete the spectrum of analysis and better understand the differentiated effects of innovation management tools on individual or team behavior, abstract frameworks and general mechanisms provide insights into single, critical aspects of innovation. Following the experimental application of the three management tools on non-R&D DUI-innovation, this thesis uses two abstract frameworks that examine exploration-exploitation – a trade-off crucial for every innovation process – and cooperation – which is represented in the 'interacting' part of DUI and essential for any innovation – as fundamental prerequisites of impacting innovation attainment. First, we elicit general resource and risk trade-offs within an urn framework that is embedded into different learning modes, and second, we use a public good game to illustrate a classical social dilemma of cooperation. This reduction allows – similar to reducing non-R&D innovation processes to learning by doing, using and, interacting, and then further reducing these learning form into an experimental real-effort task – to investigate pivotal mechanisms by excluding confounding co-variables and exploiting the highly controllable setting of lab experiments with opportunities to create counterfactual designs.

Chapter five acknowledges that investments into innovation exploration involve risk and uncertainty, as bad outcomes are not unlikely and the exact probability of success is unknown. Thus, many organizations tend to over-exploit (March 1991). For them, innovation activities are an ambidextrous endeavour, as they have to balance both extremes, where exploring and searching for new knowledge, technologies or creations comes at the opportunity costs of securing relatively sure gains through exploitation, i.e. the use and refinement of existing knowledge (Andriopoulos and Lewis 2009; Greve 2007; Guisado-González et al. 2017; He and Wong 2004; Kim et al. 2012; March 1991). This trade-off is particularly crucial for non-R&D innovations, as they are a by-product of learning from working outside of R&D and thus, by definition, have an immanent focus on exploitation. Following this intuition, we build on the insight of chapter two demonstrating that people intuitively use experiential knowledge to develop efficiency-enhancing routines and can be guided towards innovative behavior through goals, i.e. descriptive information. These results are, however, confined to a very specific organizational decision environment. One, subjects involved in DUI-like routine tasks explicitly accumulate cumulative knowledge through experience. Two, there is only very limited autonomous resource allocation (e.g. through trade-offs), and three, so far, the experiments abstracted from uncertainty and risk. Therefore, this thesis embeds the learning structure of non-R&D DUI-innovation into multi-armed-bandit tasks that represent risky choice environments under incomplete information. This allows for a quantification of inherent (dis-)advantages DUI-mode tasks might have in motivating individual risk-taking, which is pivotal for employee-driven innovation attainment.

Theoretically, chapter five builds on the well-established description-experience gap in risky choices (e.g., Hertwig and Erev 2009; Wulff et al. 2018), which differentiates between two learning modes: learning from experience or learning from description. The first mode mostly causes under-weighting of rare events (Erev and Barron 2005; Yechiam and Busemeyer 2006), whereas the second elicits over-weighting of rare events (Tversky and Kahneman 1992). Courses of action that disproportionately result in disappointing returns, with rare instances of large payoffs, tend to be neglected under decisions from experience (Denrell 2007). However, DUI-like innovations depend on precisely these courses of action. Thus, we combine the two learning modes with choice bracketing as another low-intervention management tool aimed at steering behavior through changes in choice architecture (e.g., Moher and Koehler 2010; Read et al. 1999). When presenting choices simultaneously, their consequences are assessed together, which yields increased self-control and decreased cognitive load and thus, leads to higher utility, more expected value maximization and less unfavourable gambling e.g. in the context of lotteries (Haisley et al. 2008; Koch and Nafziger 2019; Read et al. 1999). This simultaneous presentation is an example of broad bracketing. Making several decisions one-by-one and evaluating each action in isolation refers to narrow bracketing. By interacting two learning modes with two presentation modes, we empirically examine whether organizations can foster individual risk-taking in DUI-like contexts through adjustments in choice architecture.

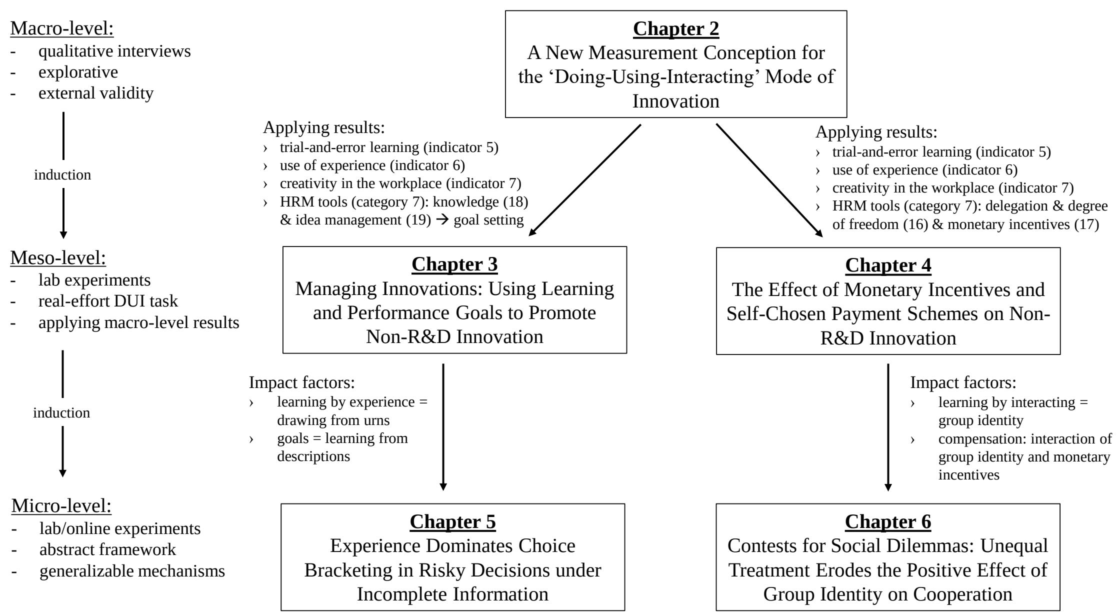
Chapter six considers cooperation as an important facilitator for innovation in R&D and non-R&D contexts, since knowledge and information sharing among individuals, teams, firms or public institutions holds great potential for innovation (Alves et al. 2007; Jorde and Teece 1990; Parrilli and Radicic 2021; Teece 1992). However, cooperation, i.e. sharing knowledge or information in the context of innovation, entails the risk of being exploited or out-competed by partners. Therefore, teamwork introduces the social dilemma of being better off when free-riding but full cooperation by every partner would be socially optimal in facilitating innovation (e.g., Chaudhuri 2011; Ledyard 1995; Zelmer 2003). A classic experimental design that illustrates such social dilemma are public good games, which have a long tradition in economics (Chaudhuri 2011). Relevant to non-R&D innovation through learning by doing, using, and interacting and the context of innovation management tools discussed above, three mechanisms are combined and tested in a repeated standard public good game: group identity, between-group contests and prize sharing rules. All three have previously shown free-riding reducing effects.

Group identity is a crucial factor for the output performance of teamwork, as in-group favoritism reduces competition between group members (Tajfel 1974) and has a positive effect on work motivation, performance, cooperation and knowledge sharing (Ashforth et al. 2011; Chen and Li 2009; Eckel and Grossman 2005; Goette et al. 2012; Van Knippenberg 2000). Between-group contests have a similar strengthening effect on within-group cooperation (Ahn et al. 2011; Burton-Chellew et al. 2010; Kugler et al. 2010; Tan and Bolle 2007), as they force the rational self-interested individual to cooperate with one's group members to win the conflict (Bornstein and Erev 1994; Sherif et al. 1961). Finally, the effect of between-group competition is affected by the prize sharing rule that determines how team members benefit from winning (Gunnthorsdottir and Rapoport 2006; Kugler et al. 2010). The two most prominent rules divide the prize either equally or proportionally among members of the winning team (Abbink et al. 2010; Ke et al. 2013). Thereby, results consistently prove higher levels of cooperation for proportional than equal prize sharing, but cooperation with equal prize sharing is steadily higher than the theoretically predicted equilibrium (Hoffmann and Thommes 2020; Sheremeta 2018). In bringing these three mechanisms together and investigating their combined impact on cooperation within teams, our design allows for a more precise understanding of monetary incentives as an innovation management tool when applied to teams with presumably high group identity, as it is the case in work groups or teams.

4 Structure of this thesis

This thesis aims to consolidate all aspects discussed above and thereby contribute to a broader and deeper appreciation of learning and knowledge creation processes behind non-R&D innovation as well as the organizational design and management tools useful in encouraging non-R&D innovations. Figure (1) illustrates the structure of this thesis and provides a guide through the different essays by contently tying them together.

Figure 1 Structure and nexus of this thesis



Starting on a 'macro-level' by using qualitative interviews, the first essay explores how firms without formal R&D-structures innovate. This serves the purpose to combine the theoretical concept of learning by doing, using, and interacting with qualitative information from 'the real world', and thus provides external validity for the concept itself and the following experiments. Besides, this essay contributes to the issue of empirically capturing non-R&D innovation by providing a new measurement conception.

By distilling and applying the results from the qualitative interviews and the literature, I develop a new real-effort task to bring DUI-like, non-R&D innovation into the lab and thus, test proposed innovation management tools on a 'meso-level' of abstraction. The task itself is closely related to three indicators derived from the interviews in chapter 2: *trial-and-error learning*, *use of experience* and *creativity in the workplace*. In chapter 3, the effect of goal-setting, as a specific tool for indicators *knowledge and idea management*, is experimentally tested. In a second experiment (chapter 4), the indicators *monetary incentives* and *delegation and degree of freedom* are examined, where the latter is implemented by delegating the compensation decision (self-chosen compensation schemes).

Finally, on a 'micro-level', results from DUI-task experiments are condensed onto an abstract and generalizable level. Accordingly, the effect of choice bracketing, a choice-architectural element to asses courses of action simultaneously or sequentially, on decisions from descriptions and/or experience is analyzed in chapter 5. The last chapter (6) embeds the results from chapter 4 in a broader context, as the effect of implemented monetary incentives is shown to be sensitive to social (identity) structures, i.e. groups vs. individual.

5 Summary of essays

A New Measurement Conception for the 'Doing-Using-Interacting' Mode of Innovation

Motivation. A significant share of innovation does not originate from R&D (c.f., Lee and Walsh 2016). For example, in Germany about 22% of innovations stem from firms not relying on traditional R&D (Peters et al. 2017). However, as most established measures are designed to capture innovation activity only in form of formal R&D, e.g., patents, R&D expenditures, or academic workforce (c.f., Hall and Jaffe 2018; Mairesse and Mohnen 2010), they neglect this significant share of non-R&D innovation. The aim of this essay is to develop a new measurement conception to capture non-R&D activities that result in innovation and making them tangible. For this purpose, we use the DUI-STI-approach by Jensen et al. (2007) to make the underlying knowledge creation processes tractable.

This two-dimensional, but not mutually, excluding concept distinguishes between learning form *science, technology, and innovation*, as the primary learning mode in R&D contexts, and learning from *doing, using, and interacting*, especially retrieved in non-R&D environments. As the former innovation-mode is mostly depicted by established measures, the latter lacks a comprehensive measurement approach. DUI comprises learning and knowledge creation processes based on 'doing', i.e. learning form cumulative work experience, developing production-specific skills, and interchanging with colleagues with the firm (Arrow 1962; Thompson 2010). Additionally, the 'using' part is primarily ascribed to learning from feedback by intermediate or end-users and their engagement in ameliorating or co-creating products and processes (Rosenberg 1982; Von Hippel 1976, 1986). Finally, the third core element 'interacting' describes learning by interacting with other firms, such as suppliers, or external actors, e.g. public institutions (Fitjar and Rodrıguez-Pose 2013; Haus-Reve et al. 2019; Lundvall and Johnson 1994). Knowledge created by these three broader learning processes is primarily harnessed to create incremental innovation, like cost reductions or quality improvement, but it can also cause new products or services, especially when highly consumer-specific (Apanasovich et al. 2016; Jensen et al. 2007; Nunes and Lopes 2015; Parrilli and Heras 2016).

To address the lack of a theoretical measurement framework and comprehensive measuring tools for DUI-mode innovation, we develop and contribute a novel measurement conception. First, we create a framework from the DUI-literature that serves as an overarching foundation from which we derive starting points to measure innovation by distinguishing between knowledge flows and facilitators. In a second step, this framework is applied and guides semi-structured interviews with firms and regional innovation consultants. Finally, we combine the interviews with the DUI-literature to recommend indicators and measurement approaches for DUI-mode innovation.

Method. We approached 49 small and medium sized enterprises (SMEs) with one to 250 employees from the 'German Mittelstand', which is emphasized to be exemplary for its strong innovative performance mostly without formal R&D or relevant resources. Complementarily, we conduct semi-structured interviews with 32 innovation consultants to consider overarching patterns in regional particularities. Firms and innovation consultants are chosen from three different regions in Germany, particularly Göttingen, Hanover and East-Thuringia.

Results. From these 81 in-depth interviews, and in combination with the DUI-literature, we derive a comprehensive set of 47 indicators clustered into 15 categories, encompassing both established and new DUI indicators. This ready-to-use set of indicators represents both underlying knowledge flows and the intensity of relevant facilitators, relevant to reach economic application in the firm. Therefore, this set of indicators serves two purposes: its quantitative assessment refines the (theoretical) understanding of DUI-mode activities and provides policy-makers with information, where and how to potentially support such innovation endeavors in non-R&D contexts.

Managing Innovations: Using Learning and Performance Goals to Promote Non-R&D Innovation

Motivation. In the third chapter I partially apply results from the qualitative interviews to design an experimental real-effort DUI-task. I then exploit this new framework to test how goal setting, as a low-threshold management tool, can facilitate non-R&D innovations by increasing visibility of problems and re-focus attention to knowledge accumulation. The new laboratory task is based on the three core elements of learning by doing, using, and interacting, used as a formalization approach to make knowledge creation and learning in non-R&D contexts tractable (Jensen et al. 2007). Moreover, this chapter discusses the increasing recognition of non-R&D innovation in economic innovation research and identifies differences between knowledge environments and knowledge creation paths for R&D and non-R&D innovation. In order to cope with the lack of formalized R&D-structures and resources, innovation management tools and organizational designs – as derived from the interviews in category 7 of the DUI-mode innovation indicator – that exploit local knowledge structures are all the more important for these traditionally underrepresented firms and sectors (Brown and Duguid 1991; Hervas-Oliver et al. 2015; Lee and Walsh 2016; Rammer et al. 2009). Following this, I experimentally analyze goal setting and its impact on non-R&D innovation from learning by cumulative experience.

Goal setting is generally used to increase the *visibility* of knowledge acquisition opportunities and problems, which translates into tighter links between worker action and outcomes by allowing for more effective learning from immediate feedback and perceiving one's own actions as affecting the outcome of interest (Lee and Walsh 2016). Goal setting theory primarily distinguishes goals based on targets and levels of difficulty or external reference points (c.f., Locke and Latham 2013c). While targets direct focus and attention to especially performance outcome or learning, do-your-best goals – due to their lack of an external reference point – are discussed as easier to attain than specific, challenging goals, which not only set an expected threshold, but one that is difficult to match or even exceed (Locke and Latham 2013a).

Method. In a 2 (learning vs. performance) x 2 (specific, challenging vs. do-your-best) design, I experimentally test the effect of four different goals on innovative behavior in the lab. Groups of three (80 groups in total) are asked to decipher four-character letter-strings based on numbers from a given interval within 45 minutes. The cipher, i.e. the letter-number-translation-pattern, represents the innovation. This design illustrates learning by doing, as players learn form an initial trail-and-error process and working on the task.

It describes learning by using, as players can constantly improve their search pattern by applying, testing instantly, and reviewing any idea about the letter-number-translation-pattern. Finally, learning by interacting is represented by a chat box, which allows for knowledge sharing and discussions on task strategies within groups.

Groups with learning goals are instructed to reduce their number of translation trials either as low as they can (do-your-best) or to only one (specific, challenging). Groups with performance goals are either advised to do their best and translate as many letter-strings as possible, or translate 13 letter-strings (specific, challenging)[1].

Results. Results provide clear evidence that specific, challenging learning goals by far surpass their alternatives in facilitating innovation and learning. Moreover, learning goals with a do-your-best frame appear somewhat beneficial in adjusting the knowledge gaining process by making opportunities for process optimization more visible and subsequently shifting attention to problem-solving. Increased visibility for learning opportunities induces players to allocate more time to communication and solution testing at the cost of short-term output. These initial investments pay off in the long-rung, as the likelihood of a group being innovative increases, which positively impacts output. In contrast, performance goals seem to suppress initial learning as participants speed up the trial-and-error process, resulting in growing short-term output at the cost of innovation and its positively correlated long-term output. Therefore, results provide strong evidence that learning goals, and specific, challenging ones in particular, support knowledge acquisition and thus innovative behavior within non-R&D-like, routine tasks. Tracing back to the second chapter, fitting innovation management tools like setting specific, challenging learning goals can help to compensate for the absence of formal R&D-structure.

The Effect of Monetary Incentives and Self-Chosen Payment Schemes on Non-R&D Innovation

Motivation. The fourth chapter uses a slightly modified version of the experimental real-effort DUI-task to test further innovation management tools derived from the interviews (chapter 2). Specifically, interviewees emphasized the value of monetary incentives and employee agency – like worker-driven wage setting or degrees of freedom – for innovation. In the literature, monet-

[1] In laboratory experiments, difficult goals are usually set at the 90$^{\text{th}}$ percentile (Locke and Latham 2013b). A pilot study with 19 groups assigned with a do-your-best performance goal, revealed the 90$^{\text{th}}$ percentile at 13 letter-strings.

ary incentives are broadly accepted in promoting motivation, performance and effort, although their impact on creative behavior is ambiguous (Bonner and Sprinkle 2002; Camerer and Hogarth 1999; Jenkins et al. 1998). We experimentally compare a fixed with a piece-rate payment scheme. The former provides participants with performance-independent compensations, which is supposed to alleviate performance pressure and thereby release capacities for the multistage process of creativity and innovation (Grabner 2014; Holmstrom and Milgrom 1991; Manso 2011). The piece-rate scheme provides individuals with a clear incentive to direct attention towards output. As a result, individuals may feel compelled to search for new solutions and thus to be creative or innovative. However, the output focus can also distract from exploring new approaches (c.f., Amabile et al. 1996; Shalley et al. 2004), resulting in detrimental effects for creativity and innovation (Ederer and Manso 2013).

Besides specific incentives schemes, new participatory mechanisms that involve employees in determining organisation frameworks have been argued to have a positive effect on motivation and performance (e.g., Charness et al. 2012; Faillo and Piovanelli 2017; Mellizo et al. 2014). Delegating the salary determination process positively impacts motivation and creativity, as it satisfies the basic psychological needs – autonomy, competence, and relatedness – of intrinsic motivation (Brück et al. 2021; Charness et al. 2012; Harrison and Freeman 2004; Mellizo et al. 2014; Ryan and Deci 2000). In combining different payment schemes with the opportunity to democratically vote for either one of them, we examine the impact of assigned and self-chosen payment schemes in non-R&D innovation.

Method. We use a lab experiment and apply a 2 (piece-rate vs. fixed payment) x 2 (voting before vs. voting after) between-subject design to the real-effort DUI-task already used in the prior chapter, to investigate its impact on non-R&D innovation. Groups of three (80 groups in sum) have to translate three-character letter-strings into numbers from a given interval within 30 minutes. Once more, the cipher illustrates the innovation. In the first two treatments, groups are assigned to either a pay for performance, i.e., piece-rate, or a fixed payment scheme. In the other two, groups either vote before or after the translation task for one of the payment schemes. The scheme with the majority of votes within a group is implemented.

Results. Compared to Ederer and Manso (2013), we find no innovation-reducing effect of assigning pay-for-performance compared to a fixed payment. Thus, monetary incentives appear less decisive for evoking non-R&D innovation. Moreover, our results give no reason to confirm the crowding-out effect of monetary incentives on intrinsic motivation. As we find no within-group effect of democratic voting, we conclude that democratizing the compensation decision process is neither facilitating, nor inhibiting non-R&D innovation. Finally, a homogeneous majority of players vote for a fixed payment prior the task, indicating a preference for performance independent compensation schemes for ambiguous tasks.

In sum, we find no experimental evidence for monetary incentives or delegation and degree of freedom affecting non-R&D innovation. Therefore, we cannot confirm a generalizable effect of monetary incentives or compensation decision delegation. However, due to abstraction and implementation in a laboratory setting, some conditional co-variables could have been cut out, which could make these two mechanisms still suitable in promoting non-R&D innovation in firm, but not a generally transferable intervention.

Experience Dominates Choice Bracketing in Risky Decisions under Incomplete Information

Motivation. After relying on a very specific organizational decision environment that abstracts from autonomous resource allocation, uncertainty, and risk, the fifth chapter analyzes how an individual's learning mode affects their preferences to engage in risky actions without sure gains. The proper allocation of scarce resources under incomplete information about the future is fundamental to economic progress and innovation attainment. Individuals broadly draw on three different modes to acquire and process probabilistic information: personal experience, descriptive information, or a combination. A long-standing literature shows that individual risky choices differ systematically conditional on their respective learning mode (Erev et al. 2010; Hertwig et al. 2004; Hertwig and Erev 2009; Ungemach et al. 2009; Wulff et al. 2018). In particular, rare events tend to be over-weighted when individuals learn from description (Tversky and Kahneman 1992), but under-weighted when they learn from experience (Erev and Barron 2005; Weber et al. 2004; Yechiam and Busemeyer 2006). Moreover, events that often appear disappointing, but sometimes return disproportionately high payoffs, are more likely to be discarded under experience (Denrell 2007) – a feature that is especially crucial for DUI-like innovation, ans they primarily stem from learning through cumulative experience. However, much of the literature relies on two-option two-outcome gambles where the probabilities of at least one potential action are public knowledge. For more complex environments, little is known about the qualitative difference between the learning modes. Furthermore, recent evidence even suggests that experience might decrease probability sensitivity in the absence of sampling biases, leading to over-weighting of rare events and thus potential disregard of known, safe paths in favor of risk-seeking behavior (Aydogan 2021; Glöckner et al. 2016).

To expand on the current literature and explore low-threshold interventions that can facilitate risky choices, we apply choice bracketing to risky environments where subjects receive incomplete information over three options with multiple outcomes. Choice bracketing refers to the way different options are presented to a decision-maker (Haisley et al. 2008; Moher and Koehler 2010; Read et al. 1999). In broad bracketing, choices are presented simultaneously, inducing individuals to asses the consequences of all options together. Prior literature suggests that this kind of bracketing leads to higher utility, more expected value maximization, increased self-control and decreased cognitive load (Koch and Nafziger 2019; Read et al. 1999). However, it also motivates diversification (Read and Loewenstein 1995), which can have a detrimental effect in cases of one option dominating the alternatives. Contrary, under narrow bracketing, subjects make several decisions one-by-one, assessing each action in isolation.

In combining the description-experience paradigm with choice bracketing, this paper serves two purposes: first, it is the first study to examine how different learning modes influence the effect of choice bracketing on risk-behavior. Second, it expands on the description-experience gap by constructing a more complex decision-environment that induces more subjectivity and probability representations while simultaneously offering a classic trade-off known from the innovation literature: safe exploitation gains vs. low-risk incremental changes vs. high-risk disruptive changes.

Method. We conduct five online experiments on Amazon Mechanical Turk (MTurk), where participants are asked to allocate three points between three options – low-risk, high-risk, and safe –, which never change between experiments. Options can be imagined as urns or money bags, containing several coins that follow different distributions. The safe option always returns a value of six, i.e. contains only coins with a value of six. The low-risk option resembles a normal distribution from a relatively narrow interval ($E(x_1) = 6.5, \sigma^2 = 3.25$) and the high-risk option simulates a heavy-tailed distribution from a comparably broad interval ($E(x_2) = 7, \sigma^2 = 145.7$). By allocating points to the respective option, coins are drawn with one point equal to one draw. Moreover, all five studies implement broad choice bracketing by providing a *simultaneous*, and narrow choice bracketing by designing a *sequential* decision frame. In *simultaneous*, participants make only one decision on how to allocate their three points, whereas in *sequential*, subjects make three consecutive decisions without intermediary feedback, i.e.

investing one point into an option each time. Between experiments, we vary the provision of information through either statistical descriptions (learning from description), sampling through options (learning from experience), or both. Overall, the data set includes 1,313 participants.

Results. Our results show that the effect of choice bracketing depends on the learning mode. When descriptions are the only or primary source for learning and decision making, broad choice bracketing – simultaneous choice – reduces investments in high-risk high-reward options in favor of less-volatile or safe returns. However, once sampled experience becomes the dominant learning mode, choice bracketing effects on high-risk investments vanish. Instead, learning from experience triggers risk-aversion as participants disproportionately invest into the low-risk at the expense of the high-risk option after experiencing a relatively bad mean from the latter. We also find evidence that learning from experience induces satisficing preferences, with people being more attentive towards securing a relatively safe mean instead of chasing large returns.

Contests for Social Dilemmas: Unequal Treatment Erodes the Positive Effect of Group Identity on Cooperation

Motivation. As we find no effect of different assigned or self-chosen payment schemes on eliciting non-R&D innovation in chapter 4, we choose a more abstract and basic experimental design to investigate cooperative behavior. Cooperation is an underlying mechanism for (non-)R&D innovation in general, as innovations are not developed in isolation, and for learning by doing, using, and interacting in particular, cooperation is a necessary precondition for the interaction-part. Besides, cooperation often reduces the hazard of exploitation and free-riding, similar to the innovation context where knowledge sharing is potentially risky due to its competitive environment. The social dilemma applies as well, since firms are better off when employees cooperate, but due to free-riding and the peril of out-competing, they often have an incentive not to. We argue that the social dilemma is moderated by group identity, which naturally occurs in firms, and is discussed as an essential diver for human decision making (Akerlof and Kranton 2005, 2008). Following up chapter 2 and 4, we combine the effect of monetary incentives and different levels of group identity to examine their effects on cooperation by using a standard public good game. This setting is complemented by a competitive component.

Competition between groups is discussed as diminishing free-riding and thus, benefit cooperation (Ahn et al. 2011; Burton-Chellew et al. 2010; Kugler et al. 2010; Tan and Bolle 2007). However, the level of cooperation is sensitive to the prize sharing rule, as individuals show higher levels of cooperation if the prize for between-group competition is shared proportionally compared to equally (Gunnthorsdottir and Rapoport 2006; Ke et al. 2013; Kugler et al. 2010). Most experiments do not account for group identity, although there are strong arguments that group identity should behaviorally interact with different group-level incentive schemes. For example, a team with a long shared history or a strong sense of cohesion will respond differently to monetary incentives than groups with members who barely know each other. In a group with strong social cohesion, suffering from feelings of guilt due to receiving a larger share from a proportionally shared prize will be more pronounced as in groups with low or almost no group identity (Everett et al. 2015; Güth et al. 2009; Morell 2019; Ockenfels and Werner 2014). Therefore, monetary incentives should not only match the production target, but also the respective group identity levels.

Method. We extend the design of Gunnthorsdottir and Rapoport (2006) by introducing group identity resulting in a 2 (low vs. high group identity) x 2 (equal vs. proportional profit sharing) design. Groups of four (80 groups in sum) – with partners matching – participate in a within-group public good game over a period of ten rounds. In every round, the same two groups compete for an exogenously given prize, which is shared either equally or proportional to an individual's contribution to the within-group public good. A probabilistic function determines the winning group, where groups can increase their probability of winning by enlarging their public good. Group identity is manipulated by the team-building puzzle task of Eckel and Grossman (2005). In treatments with low group identity, participants receive five pieces of a puzzle they individually and separately have to compose prior the public good game. Therefore, partners matching is the only group identity inducing aspect. In contrast, groups in high identity treatments jointly compose the puzzle. For this purpose, group members are seated together, with each group member receiving five puzzle pieces. We made sure that no member could assemble the puzzle without exchanging pieces with others to ensure this a group task. Participants are instructed that by the end of the task, each member shall have a completed puzzle.

Results. Overall, we find that monetary incentives, i.e. the prize sharing rule are sensitive to the level of group identity. As theoretically predicted, group identity positively impacts cooperation in contests with equal prize sharing. In contrast, proportional profit sharing attenuates the effect of group identity. We discuss and theoretically present a diminishing effect of proportional prize sharing on inequity aversion in groups with high identity, resulting in inverting the positive effect of group identity. We conclude that effective incentive schemes should consider social incentives and multi-level conflicts emerging from group structures.

References

Abbink, Klaus, Jordi Brandts, Benedikt Herrmann and Henrik Orzen (2010). 'Intergroup Conflict and Intra-Group Punishment in an Experimental Contest Game'. In: *American Economic Review* 100 (1), 420–447.

Ahn, Tho-Keyong, R. Mark Isaac and Timothy C. Salmon (2011). 'Rent seeking in groups'. In: *International Journal of Industrial Organization* 29 (1), 116–125.

Akerlof, George A. and Rachel E. Kranton (2005). 'Identity and the Economics of Organizations'. In: *The Journal of Economic Perspectives* 19 (1), 9–32.

— (2008). 'Identity, Supervision, and Work Groups'. In: *The American Economic Review* 98 (2), 212–217.

Alhusen, Harm, Tatjana Bennat, Kilian Bizer, Uwe Cantner, Elaine Horstmann, Martin Kalthaus, Till Proeger, Rolf Sternberg and Stefan Töpfer (2021). 'A new measurement conception for the 'doing-using-interacting'mode of innovation'. In: *Research Policy* 50 (4), 104214.

Alves, Jorge, Maria José Marques, Irina Saur and Pedro Marques (2007). 'Creativity and innovation through multidisciplinary and multisectoral cooperation'. In: *Creativity and Innovation Management* 16 (1), 27–34.

Amabile, Teresa M., Regina Conti, Heather Coon, Jeffrey Lazenby and Michael Herron (1996). 'Assessing the work environment for creativity'. In: *Academy of Management Journal* 39 (5), 1154–1184.

Amabile, Teresa M. and Judith Gitomer (1984). 'Children's artistic creativity: Effects of choice in task materials'. In: *Personality and Social Psychology Bulletin* 10 (2), 209–215.

Amara, Nabil, Réjean Landry, Nizar Becheikh and Mathieu Ouimet (2008). 'Learning and novelty of innovation in established manufacturing SMEs'. In: *Technovation* 28 (7), 450–463.

Andriopoulos, Constantine and Marianne W. Lewis (2009). 'Exploitation-exploration tensions and organizational ambidexterity: Managing paradoxes of innovation'. In: *Organization Science* 20 (4), 696–717.

Apanasovich, Natalja, Henar Alcalde Heras and Mario Davide Parrilli (2016). 'The impact of business innovation modes on SME innovation performance in post-Soviet transition economies: The case of Belarus'. In: *Technovation* 57-58, 30–40.

Argote, Linda and Erin Fahrenkopf (2016). 'Knowledge transfer in organizations: The roles of members, tasks, tools, and networks'. In: *Organizational Behavior and Human Decision Processes* 136, 146–159.

Argote, Linda and Paul Ingram (2000). 'Knowledge transfer: A basis for competitive advantage in firms'. In: *Organizational Behavior and Human Decision Processes* 82 (1), 150–169.

Argote, Linda and Ella Miron-Spektor (2011). 'Organizational Learning: From Experience to Knowledge'. In: *Organization Science* 22 (5), 1123–1137.

Arrow, Kenneth J. (1962). 'The Economic Implications of Learning by Doing'. In: *The Review of Economic Studies* 29 (3), 155–173.

Arundel, Anthony, Catalina Bordoy and Minna Kanerva (2008). 'Neglected innovators: How do innovative firms that do not perform R&D innovate'. In: *INNO-Metrics Thematic Paper*.

Asheim, Bjørn (2007). 'Differentiated knowledge bases and varieties of regional innovation systems'. In: *Innovation* 20 (3), 223–241.

Asheim, Bjørn T. and Lars Coenen (2006). 'Contextualising regional innovation systems in a globalising learning economy: On knowledge bases and institutional frameworks'. In: *The Journal of Technology Transfer* 31 (1), 163–173.

Ashforth, Blake E., Kristie M. Rogers and Kevin G. Corley (2011). 'Identity in organizations: Exploring cross-level dynamics'. In: *Organization Science* 22 (5), 1144–1156.

Aslesen, Heidi Wiig, Arne Isaksen and James Karlsen (2012). 'Modes of Innovation and Differentiated Responses to Globalisation-A Case Study of Innovation Modes in the Agder Region, Norway'. In: *Journal of the Knowledge Economy* 3 (4), 389–405.

Aydogan, Ilke (2021). 'Prior Beliefs and Ambiguity Attitudes in Decision from Experience'. In: *Management Science*.

Bailyn, Lotte (1985). 'Autonomy in the industrial R&D lab'. In: *Human Resource Management* 24 (2), 129–146.

Barge-Gil, Andres, Maria Jesus Nieto and Lluis Santamaria (2011). 'Hidden innovators: The role of non-R&D activities'. In: *Technology Analysis & Strategic Management* 23 (4), 415–432.

Bonner, Sarah E. and Geoffrey B. Sprinkle (2002). 'The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research'. In: *Accounting, Organizations and Society* 27 (4-5), 303–345.

Bornstein, Gary and Ido Erev (1994). 'The enhancing effect of intergroup competition on group performance'. In: *International Journal of Conflict Management* 5 (3), 271–283.

Brown, John Seely and Paul Duguid (1991). 'Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation'. In: *Organization Science* 2 (1), 40–57.

Brück, Christian, Thorsten Knauer, Harald Meier and Anja Schwering (2021). 'Self-set salaries and creativity'. In: *Journal of Business Economics* 91 (1), 91–121.

Burton-Chellew, Maxwell N., Adin Ross-Gillespie and Stuart A. West (2010). 'Cooperation in humans: competition between groups and proximate emotions'. In: *Evolution and Human Behavior* 31 (2), 104–108.

Cadsby, C. Bram, Song Fei and Francis Tapon (2007). 'Sorting and incentive effects of pay for performance: An experimental investigation'. In: *Academy of Management Journal* 50 (2), 387–405.

Calligraphy Cut (2021). https://de.calligraphy-cut.com/ueber-uns-2/ and https://de.calligraphy-cut.com/unternehmen/. Retrieved December 2, 2021.

Camerer, Colin F. and Robin M. Hogarth (1999). 'The effects of financial incentives in experiments: A review and capital-labor-production framework'. In: *Journal of Risk and Uncertainty* 19 (1), 7–42.

Charness, Gary, Ramón Cobo-Reyes, Natalia Jiménez, Juan A. Lacomba and Francisco Lagos (2012). 'The hidden advantage of delegation: Pareto improvements in a gift exchange game'. In: *American Economic Review* 102 (5), 2358–79.

Chaudhuri, Ananish (2011). 'Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature'. In: *Experimental Economics* 14 (1), 47–83.

Chen, Yan and Sherry Xin Li (2009). 'Group identity and social preferences'. In: *American Economic Review* 99 (1), 431–57.

Denrell, Jerker (2007). 'Adaptive learning and risk taking'. In: *Psychological Review* 114 (1), 177.

Eckel, Catherine C. and Philip J. Grossman (2005). 'Managing diversity by creating team identity'. In: *Journal of Economic Behavior & Organization* 58 (3), 371–392.

Ederer, Florian and Gustavo Manso (2013). ‘Is pay for performance detrimental to innovation?’ In: *Management Science* 59 (7), 1496–1513.
Erev, Ido and Greg Barron (2005). ‘On adaptation, maximization, and reinforcement learning among cognitive strategies’. In: *Psychological Review* 112 (4), 912.
Erev, Ido et al. (2010). ‘A choice prediction competition: Choices from experience and from description’. In: *Journal of Behavioral Decision Making* 23 (1), 15–47.
Everett, Jim A. C., Nadira S. Faber and Molly Crockett (2015). ‘Preferences and beliefs in ingroup favoritism’. In: *Frontiers in Behavioral Neuroscience* 9, 15.
Faillo, Marco and Costanza Piovanelli (2017). ‘Wage delegation and intrinsic motivation: an experimental study’. In: *CEEL working papers* No. 1701.
Fitjar, Rune Dahl and Andrés Rodrıguez-Pose (2013). ‘Firm collaboration and modes of innovation in Norway’. In: *Research Policy* 42 (1), 128–138.
Franke, Jörg, Ruslan Gurtoviy and Vanessa Mertins (2016). ‘Workers’ participation in wage setting: A gift-exchange experiment’. In: *Journal of Economic Psychology* 56, 151–162.
Gagné, Marylène and Edward L. Deci (2005). ‘Self-determination theory and work motivation’. In: *Journal of Organizational Behavior* 26 (4), 331–362.
Glöckner, Andreas, Benjamin E Hilbig, Felix Henninger and Susann Fiedler (2016). ‘The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice.’ In: *Journal of Experimental Psychology: General* 145 (4), 486.
Goette, Lorenz, David Huffman, Stephan Meier and Matthias Sutter (2012). ‘Competition Between Organizational Groups: Its Impact on Altruistic and Antisocial Motivations’. In: *Management Science* 58 (5), 948–960.
Grabner, Isabella (2014). ‘Incentive system design in creativity-dependent firms’. In: *The Accounting Review* 89 (5), 1729–1750.
Greve, Henrich R. (2007). ‘Exploration and exploitation in product innovation’. In: *Industrial and Corporate Change* 16 (5), 945–975.
Guisado-González, Manuel, Jennifer González-Blanco and José Luis Coca-Pérez (2017). ‘Analyzing the relationship between exploration, exploitation and organizational innovation’. In: *Journal of Knowledge Management* 21 (5), 1142–1162.
Gunnthorsdottir, Anna and Amnon Rapoport (2006). ‘Embedding social dilemmas in intergroup competition reduces free-riding’. In: *Organizational Behavior and Human Decision Processes* 101 (2), 184–199.
Güth, Werner, Matteo Ploner and Tobias Regner (2009). ‘Determinants of in-group bias: Is group affiliation mediated by guilt-aversion?’ In: *Journal of Economic Psychology* 30 (5), 814–827.
Haisley, Emily, Romel Mostafa and George Loewenstein (2008). ‘Myopic risk-seeking: The impact of narrow decision bracketing on lottery play’. In: *Journal of Risk and Uncertainty* 37 (1), 57–75.
Hall, Bronwyn H. and Adam B. Jaffe (2018). ‘Measuring Science, Technology, and Innovation: A Review’. In: *Annals of Science and Technology Policy* 2 (1), 1–74.
Harrison, Jeffrey S. and R. Edward Freeman (2004). ‘Is organizational democracy worth the effort?’ In: *The Academy of Management Executive* 18 (3), 49–53.
Haus-Reve, Silje, Rune Dahl Fitjar and Andrés Rodrıguez-Pose (2019). ‘Does combining different types of collaboration always benefit firms? Collaboration, complementarity and product innovation in Norway’. In: *Research Policy* 48 (6), 1476–1486.
He, Zi-Lin and Poh-Kam Wong (2004). ‘Exploration vs. exploitation: An empirical test of the ambidexterity hypothesis’. In: *Organization Science* 15 (4), 481–494.

Hertwig, Ralph, Greg Barron, Elke U. Weber and Ido Erev (2004). 'Decisions from experience and the effect of rare events in risky choice'. In: *Psychological Science* 15 (8), 534–539.

Hertwig, Ralph and Ido Erev (2009). 'The description–experience gap in risky choice'. In: *Trends in Cognitive Sciences* 13 (12), 517–523.

Hervas-Oliver, Jose-Luis, Francisca Sempere-Ripoll, Carles Boronat-Moll and Ronald Rojas (2015). 'Technological innovation without R&D: unfolding the extra gains of management innovations on technological performance'. In: *Technology Analysis & Strategic Management* 27 (1), 19–38.

Hidalgo, Antonio and Jose Albors (2008). 'Innovation management techniques and tools: a review from theory and practice'. In: *R&D Management* 38 (2), 113–127.

Hoffmann, Christin and Kirsten Thommes (2020). 'Combining Egalitarian and Proportional Sharing Rules in Team Tournaments to Incentivize Energy-Efficient Behavior in a Principal-Agent Context'. In: *Organization & Environment*, 1–25.

Holmstrom, Bengt and Paul Milgrom (1991). 'Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design'. In: *Journal of Law, Economics, & Organization* 7, 24–52.

Jenkins, G. Douglas Jr., Atul Mitra, Nina Gupta and Jason D. Shaw (1998). 'Are financial incentives related to performance? A meta-analytic review of empirical research'. In: *Journal of Applied Psychology* 83 (5), 777.

Jensen, Morten Berg, Björn Johnson, Edward Lorenz and Bengt Åke Lundvall (2007). 'Forms of knowledge and modes of innovation'. In: *Research Policy* 36 (5), 680–693.

Jeworrek, Sabrina and Vanessa Mertins (2019). 'Wage delegation in the field'. In: *Journal of Economics & Management Strategy* 28 (4), 656–669.

Johnson, Björn, Edward Lorenz and Bengt-Åke Lundvall (2002). 'Why all this fuss about codified and tacit knowledge?' In: *Industrial and Corporate Change* 11 (2), 245–262.

Jorde, Thomas M. and David J. Teece (1990). 'Innovation and cooperation: implications for competition and antitrust'. In: *Journal of Economic Perspectives* 4 (3), 75–96.

Ke, Changxia, Kai A. Konrad and Florian Morath (2013). 'Brothers in arms – An experiment on the alliance puzzle'. In: *Games and Economic Behavior* 77 (1), 61–76.

Kim, Changsu, Jaeyong Song and Atul Nerkar (2012). 'Learning and innovation: Exploitation and exploration trade-offs'. In: *Journal of Business Research* 65 (8), 1189–1194.

Koch, Alexander K. and Julia Nafziger (2019). 'Correlates of narrow bracketing'. In: *The Scandinavian Journal of Economics* 121 (4), 1441–1472.

Kugler, Tamar, Amnon Rapoport and Asya Pazy (2010). 'Public good provision in inter-team conflicts: Effects of asymmetry and profit-sharing rule'. In: *Journal of Behavioral Decision Making* 23 (4), 421–438.

Latham, Gary P. (2004). 'The motivational benefits of goal-setting'. In: *Academy of Management Executive* 18 (4), 126–129.

Lazear, Edward P (1986). 'Salaries and Piece Rates'. In: *The Journal of Business* 59 (3), 405–431.

Ledyard, John O. (1995). 'Public Goods: Some Experimental Results'. In: *Handbook of Experimental Economics*. Ed. by John H. Kagel and Alvin E. Roth. Princeton University Press, 111–168.

Lee, You-Na (2015). 'Evaluating and extending innovation indicators for innovation policy'. In: *Research Evaluation* 24 (4), 471–488.

Lee, You-Na and John P. Walsh (2016). 'Inventing while you work: Knowledge, non-R&D learning and innovation'. In: *Research Policy* 45 (1), 345–359.

Locke, Edwin A. and Gary P. Latham (2002). 'Building a practically useful theory of goal setting and task motivation: A 35-year odyssey'. In: *American Psychologist* 57 (9), 705–717.

— (2013a). 'Goal Setting Theory, 1990'. In: *New Developments in Goal Setting and Task Performance*. Ed. by Edwin A. Locke and Gary P. Latham. Paperback. New York: Taylor & Francis. Chap. 1, 3–15.

— (2013b). 'Goal Setting Theory: The Current State'. In: *New Developments in Goal Setting and Task Performance*. Ed. by Edwin A. Locke and Gary P. Latham. Paperback. New York: Taylor & Francis. Chap. 37, 623–630.

— (2013c). *New Developments in Goal Setting and Task Performance*. Paperback. New York: Routledge.

Lundvall, Bengt-Åke (2016). *The Learning Economy and the Economics of Hope*. Electronic. New York: Anthem Press.

Lundvall, Bengt-Åke and Björn Johnson (1994). 'The learning economy'. In: *Journal of Industry Studies* 1 (2), 23–42.

Lundvall, Bengt-Åke and Peter Nielsen (2007). 'Knowledge management and innovation performance'. In: *International Journal of Manpower* 28 (3-4), 207–223.

Mairesse, Jacques and Pierre Mohnen (2010). 'Using innovation surveys for econometric analysis'. In: *Handbook of the Economics of Innovation*. Ed. by Bronwyn H. Hall and Nathan Rosenberg. Vol. 2. Elsevier, 1129–1155.

Manso, Gustavo (2011). 'Motivating innovation'. In: *The Journal of Finance* 66 (5), 1823–1860.

March, James G. (1991). 'Exploration and exploitation in organizational learning.' In: *Organization Science* 2 (1), 71–87.

McIver, Derrick, Cynthia A. Lengnick-Hall, Mark L. Lengnick-Hall and Indu Ramachandran (2013). 'Understanding work and knowledge management from a knowledge-in-practice perspective'. In: *Academy of Management Review* 38 (4), 597–620.

Mellizo, Philip, Jeffrey Carpenter and Peter Hans Matthews (2014). 'Workplace democracy in the lab'. In: *Industrial Relations Journal* 45 (4), 313–328.

Moher, Ester and Derek J. Koehler (2010). 'Bracketing effects on risk tolerance: Generalizability and underlying mechanisms'. In: *Judgment and Decision Making* 5 (5), 339.

Morell, Alexander (2019). 'The short arm of guilt – An experiment on group identity and guilt aversion'. In: *Journal of Economic Behavior & Organization* 166, 332–345.

Nonaka, Ikujiro, Ryoko Toyama and Noboru Konno (2000). 'SECI, Ba and leadership: a unified model of dynamic knowledge creation'. In: *Long Range Planning* 33 (1), 5–34.

Nunes, Sérgio and Raul Lopes (2015). 'Firm Performance, Innovation Modes and Territorial Embeddedness'. In: *European Planning Studies* 23 (9), 1796–1826.

Ockenfels, Axel and Peter Werner (2014). 'Beliefs and ingroup favoritism'. In: *Journal of Economic Behavior & Organization* 108, 453–462.

Paolillo, Joseph G. and Warren B. Brown (1978). 'How organizational factors affect R&D innovation'. In: *Research management* 21 (2), 12–15.

Parrilli, Mario Davide and Henar Alcalde Heras (2016). 'STI and DUI innovation modes: Scientific-technological and context-specific nuances'. In: *Research Policy* 45 (4), 747–756.

Parrilli, Mario Davide and Dragana Radicic (2021). 'Cooperation for innovation in liberal market economies: STI and DUI innovation modes in SMEs in the United Kingdom'. In: *European Planning Studies*, 1–24.

Pavitt, Keith (1984). 'Sectoral patterns of technical change: towards a taxonomy and a theory'. In: *Research Policy* 13 (6), 343–373.

PayPal (2021). 'PayPal Holdings Inc., Third Quarter 2021 Results'. In: *PayPal, November 08, 2021.* Retrieved December 2, 2021 from https://s1.q4cdn.com/633035571/files/doc_financials/2021/q3/Q3-21-PayPal-Earnings-Release.pdf.

Peters, Bettina, Mark J. Roberts, Van Anh Vuong and Helmut Fryges (2017). 'Estimating dynamic R&D choice: an analysis of costs and long-run benefits'. In: *The RAND Journal of Economics* 48 (2), 409–437.

Rammer, Christian, Dirk Czarnitzki and Alfred Spielkamp (2009). 'Innovation success of non-R&D-performers: substituting technology by management in SMEs'. In: *Small Business Economics* 33 (1), 35–58.

Rawat, Ayush Singh (2021). 'The Success Story of PayPal'. In: *Analytissteps, February 16, 2021.* Retrieved December 2, 2021 from https://www.analyticssteps.com/blogs/success-story-paypal.

Read, Daniel and George Loewenstein (1995). 'Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices.' In: *Journal of Experimental Psychology: Applied* 1 (1), 34.

Read, Daniel, George Loewenstein and Matthew Rabin (1999). 'Choice Bracketing'. In: *Journal of Risk and Uncertainty* 19 (1-3), 171–197.

Rosenberg, Nathan (1982). *Inside the Black Box: Technology and Economics*. New York: Cambridge University Press.

Ryan, Richard M. and Edward L. Deci (2000). 'Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being'. In: *American Psychologist* 55 (1), 68.

Seijts, Gerard and Gary P. Latham (2005). 'Learning versus performance goals: When should each be used?' In: *Academy of Management Executive* 19 (1), 124–131.

Seijts, Gerard H., Gary P. Latham and Meredith Woodwark (2013). 'Learning Goals: A Qualitative and Quantitative Review'. In: *New Developments in Goal Setting and Task Performance*. Ed. by Edwin A. Locke and Gary P. Latham. Paperback. New York: Taylor & Francis. Chap. 13, 195–212.

Semler, Ricardo (1989). 'Managing without managers'. In: *Harvard Business Review* 67 (5), 76–84.

— (2007). 'Out of this world: Doing things the Semco way'. In: *Global Business and Organizational Excellence* 26 (5), 13–21.

Shalley, Christina E., Jing Zhou and Greg R. Oldham (2004). 'The effects of personal and contextual characteristics on creativity: Where should we go from here?' In: *Journal of Management* 30 (6), 933–958.

Shaw, Dougal (2021). 'The company where colleagues decide your salary'. In: *BBC News, May 5 2021.* Retrieved November 6, 2021 from https://www.bbc.com/news/business-56915767.

Sheremeta, Roman M. (2018). 'Behavior in group contests: A review of experimental research'. In: *Journal of Economic Surveys* 32 (3), 683–704.

Sherif, M. Muzafer, O. J. Harvey, B. Jack White, William R. Hood and Carolyn W. Sherif (1961). *Intergroup conflict and cooperation: The Robbers Cave experiment*. Vol. 10. University Book Exchange Norman, OK.

Sliwka, Dirk (2001). 'On the costs and benefits of delegation in organizations'. In: *Journal of Institutional and Theoretical Economics* 157 (4), 568–590.

Smith, Vernon L. (1976). 'Experimental economics: Induced value theory'. In: *The American Economic Review* 66 (2), 274–279.

Tajfel, Henri (1974). 'Social identity and intergroup behaviour.' In: *Social Science Information/Sur les Sciences Sociales* 13 (2), 65–93.

Tan, Jonathan H. W. and Friedel Bolle (2007). 'Team competition and the public goods game'. In: *Economics Letters* 96 (1), 133–139.

Teece, David J. (1992). 'Competition, cooperation, and innovation: Organizational arrangements for regimes of rapid technological progress'. In: *Journal of Economic Behavior & Organization* 18 (1), 1–25.

Thompson, Peter (2010). 'Learning by doing'. In: *Handbook of the Economics of Innovation*. Ed. by Bronwyn H. Hall and Nathan Rosenberg. Vol. 1. Elsevier, 429–476.

Tversky, Amos and Daniel Kahneman (1992). 'Advances in prospect theory: Cumulative representation of uncertainty'. In: *Journal of Risk and Uncertainty* 5 (4), 297–323.

Ungemach, Christoph, Nick Chater and Neil Stewart (2009). 'Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)?' In: *Psychological Science* 20 (4), 473–479.

Van Knippenberg, Daan (2000). 'Work motivation and performance: A social identity perspective'. In: *Applied Psychology* 49 (3), 357–371.

Volberda, Henk W., Frans A.J. Van Den Bosch and Cornelis V. Heij (2013). 'Management Innovation: Management as Fertile Ground for Innovation'. In: *European Management Review* 10, 1–15.

Von Hippel, Eric (1976). 'The dominant role of users in the scientific instrument innovation process'. In: *Research Policy* 5 (3), 212–239.

— (1986). 'Lead Users : A Source of Novel'. In: *Management Science* 32 (7), 791–805.

Weber, Elke U., Sharoni Shafir and Ann-Renee Blais (2004). 'Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation.' In: *Psychological Review* 111 (2), 430.

Winter, Sidney G. (1984). 'Schumpeterian competition in alternative technological regimes'. In: *Journal of Economic Behavior & Organization* 5 (3-4), 287–320.

Wise (2021). https://wise.com/de/about/our-story and https://wise.com/de/about/wise-reviews. Retrieved December 2, 2021.

Wulff, Dirk U., Max Mergenthaler-Canseco and Ralph Hertwig (2018). 'A meta-analytic review of two modes of learning and the description-experience gap.' In: *Psychological Bulletin* 144 (2), 140.

Yechiam, Eldad and Jerome R. Busemeyer (2006). 'The effect of foregone payoffs on underweighting small probability events'. In: *Journal of Behavioral Decision Making* 19 (1), 1–16.

Zelmer, Jennifer (2003). 'Linear public goods experiments: A meta-analysis'. In: *Experimental Economics* 6 (3), 299–310.

Chapter 2

A New Measurement Conception for the ‘Doing-Using-Interacting’ Mode of Innovation

A New Measurement Conception for the 'Doing-Using-Interacting' Mode of Innovation

Harm Alhusen, Tatjana Bennat, Kilian Bizer, Uwe Cantner, Elaine Horstmann, Martin Kalthaus, Till Proeger, Rolf Sternberg, Stefan Töpfer

Abstract

The 'doing-using-interacting' (DUI) mode of innovation describes informal innovative activities and it can be juxtaposed with the 'science-technology-innovation' (STI) mode based on deliberate research and development. While both modes contribute substantially but differently to technological progress, our empirical understanding of DUI mode innovative activity suffers from the lack of a comprehensive measurement approach. While empirical measurement of the STI mode is well established, empirical indicators for DUI activities are scarce and no consensus has emerged concerning its constituting learning processes. We propose a new measurement conception for innovative activity and based on 81 in-depth interviews with German firms and regional innovation consultants. We derive fifteen categories of DUI mode learning processes and a comprehensive set of 47 indicators comprising both established and new DUI indicators for empirical measurement. This new measurement conception and the respective indicators provide a holistic perspective and their application can be used to increase our understanding of the importance of DUI mode innovative activity, as well as guiding policy-makers.

Published: Harm Alhusen, Tatjana Bennat, Kilian Bizer, Uwe Cantner, Elaine Horstmann, Martin Kalthaus, Till Proeger, Rolf Sternberg and Stefan Töpfer (2021). A New Measurement Conception for the 'Doing-Using-Interacting' Mode of Innovation. In: *Research Policy*, 50(4), 104214.

Chapter 3

Managing Innovations: Using Learning and Performance Goals to Promote Non-R&D Innovation

Managing Innovations: Using Learning and Performance Goals to Promote Non-R&D Innovation

Elaine Horstmann

Abstract

The increasing recognition of non-R&D innovations in economic innovation research calls for more specific insights into the functioning of innovation managements tools. Non-R&D-innovations emerge from learning and knowledge creation as a by-product of employees working in different functions across a firm, extending the source of innovation to the whole organisation. This article develops and experimentally applies a new real-effort task designed to represent non-R&D-innovations rooted in learning by doing, using, and interacting. We examine how goal setting as a low-threshold management tool can encourage non-R&D-innovations by increasing the visibility of a problem and shifting attention to knowledge accumulation. In a 2 (learning vs. performance) x 2 (specific, challenging vs. do-your-best) design, we find clear evidence that specific, challenging learning goals by far outperform their three alternatives in fostering innovative behavior. Indeed, for the three other goals, innovation attainment is very low. Overall, learning goals increase initial investments into learning and group communication at the expense of short-term output, thereby enabling participants to realize their innovative potential and increase output in the long run. Insights into group communication reveal that groups with learning goals allocate more time to spread strategic and solution-orientated information among group members.

This work was financially supported by the German Federal Ministry of Education and Research, Grant Number 16IFI005.

1 Introduction

Economic innovation research increasingly recognizes the importance of non-R&D work in fostering the creation of new processes and products (Alhusen et al. 2021; Jensen et al. 2007; Parrilli and Heras 2016; Thomä 2017). While traditional R&D projects are still the main driver of innovative activities (Cappelen et al. 2012; Cohen 2010; Dimos and Pugh 2016), current evidence suggests not only that around 11 to 12% of patented innovations emerge from non-R&D contexts, but that these innovations do not significantly differ from R&D inventions in scope and impact (Arora et al. 2016; Lee 2015; Lee and Walsh 2016). However, they do exhibit distinct prerequisites for their successful elicitation. Innovation outcomes from non-R&D work are less formally organized and stem more from direct experience and learning by cumulative experience, like learning by doing, using and interacting (Jensen et al. 2007; Levitt and March 1988). Often, learning happens as a by-product of working (Brown and Duguid 1991; Ellström 2010; Pruett and Thomas 2008). In contrast, R&D research relies on codified and general knowledge, usually divorced from a particular product or process (Lee and Walsh 2016). Since many firms, in particular those of small or medium size, do not have or cannot afford formally organized R&D structures (Kesting and Ulhøi 2010), management tools that leverage local knowledge structures to promote on-the-job innovations are all the more important for these traditionally underrepresented sectors or innovators (Hervas-Oliver et al. 2015; Rammer et al. 2009). In this paper, we introduce and experimentally explore how goal setting as a low-threshold intervention can facilitate non-R&D innovations from learning by cumulative experience.

According to Lee and Walsh (2016), one primary distinction between R&D and non-R&D learning lies in their disparate knowledge environment. For R&D, abstract knowledge is higher valued than practical knowledge, favoring science-based innovation emerging from codified scientific and technological knowledge, high skilled academic workforce and patents (Jensen et al. 2007; Parrilli and Heras 2016). In contrast, non-R&D learning draws from routinized regimes, sticky knowledge, experimentation as well as trial-and-error processes that focus on 'know-how' and 'know-who' (Brown and Duguid 1991; Jensen et al. 2007; Lundvall 2016; Pavitt 1984; Winter 1984). This categorization closely relates to the seminal work of Jensen et al. (2007), who name these opposite, but not mutually exclusive, knowledge creation modes learning by doing, using, and interacting (DUI) and learning by science, technology, and innovation (STI). Because non-R&D learning predominantly builds on the DUI mode, meaning the cumulative experience of individual workers or teams within an organization, as gathered through their everyday, routine work, non-R&D innovations are much more dependent on the *visibility* of knowledge as well as the visibility of problems. Here, visibility directly translates into tight links between worker action and outcomes, allowing individuals to effectively learn from immediate feedback and perceive their own actions as affecting a specific outcome of interest (Lee and Walsh 2016).

One innovation management tool to increase the visibility of problems and therefore enhance learning by doing, using and interacting, is to set goals. Goal-setting is a widely adopted motivational technique that aims to increase an employee's performance effectiveness and motivate problem-solving. The most prominent goals are performance and learning goals. Performance goals draw attention towards performance outcomes and foster motivation or effort, while learning goals emphasize knowledge and skill acquisition as well as understanding task mechanisms (Latham 2004; Seijts et al. 2013; Seijts and Latham 2005).

Goals can further be distinguished between do-your-best and specific goals. A specific goal implements a clear marker of progress, while a do-your-best goal does not follow an external standard and allows for more flexible reference points that determine acceptable performance levels (Locke and Latham 2002; Mento et al. 1987; Seijts and Latham 2005). Specific, challenging goals are more effective if individuals possess the necessary knowledge and ability to attain a goal. Otherwise, a do-your-best goal can be more effective (Latham 2004). That is because focusing on performance outcomes while lacking the required knowledge or abilities can be threatening to individuals, inducing anxiety and thus leading to systematic struggles in exploration, search and discovery. This, in turn, inhibits the acquisition of efficient task strategies, worsening performance (Latham 2004; Locke and Latham 2002; Seijts et al. 2013). Similarly, learning goals can overcome the limitations of performance goals specifically when they are rooted in a lack of knowledge or skills. In such scenarios, irrespective of the problem, individuals should first focus on discovering and mastering the performance process (Kanfer and Ackerman 1989; Latham and Saari 1982), and only afterwards think about optimizing their performance routines or increase their performance effort (Kanfer and Ackerman 1989; Miron-Spektor and Beenen 2015; Seijts and Latham 2005). Learning goals support knowledge and skill acquisition, as individuals pay more attention to task mechanisms, aiming to gain an understanding of their problem, their environment, and their tool-kit. Thus, learning goals can foster a necessary condition of innovative activities: learning and acquiring abilities.

Following the reasoning above, we argue that goal-setting can positively influence an individual's innovation attainment by steering attention to different aspects of the innovation process, and subsequently inducing conscious goals, which shift an individual's focus to solve a specific, formulated task and thereby induce particular mindsets (c.f., Locke and Latham 2013c, 1990b). Depending on the available information and skill endowment, either learning or performance goals might be more suited to increase innovation output. Therefore, this paper experimentally tests both goal-types, while further differentiating between do-your-best and specific goals. We aim to unify the motivational technique of goal setting with non-R&D innovations to explore whether performance and learning goals affect innovative activities that rely on learning, using and interacting. For that purpose, we develop a new experimental task that models the DUI-mode's core elements in a laboratory setting[1]. This allows us to isolate core properties of DUI-mode innovations and thereby add to the existing literature on non-R&D innovations, exploiting the controlled setting of a laboratory experiment.

Our results show that learning goals motivate individuals to concentrate on acquiring task-specific skills and, as a consequence, increase innovation attainment. In particular, combining a learning frame with a clearly defined reference point (specific, challenging goal) leads to a substantial fraction of groups achieving their innovative potential. However, in the short-term, this comes at the cost of output, as participants with learning goals allocate more time to single steps in the working process. In contrast, performance goals increase short-term output, but almost no group exhibits innovative behavior. Finally, out chat data reveal groups with learning goals, and specific, challenging ones in particular, to adopt a very different communication strategy and allocate more time to spread strategic information about solutions for complex, non-obvious problems in a non-R&D-like context.

[1] All supplementary material can be obtained from our online repository: https://researchbox.org/496&PEER_REVIEW_passcode=ODEJDL. It contains the list of letter-strings, its related translation, a screen-shot of the task-screen, instructions, post-experimental questionnaire and examples for chat data coding.

2 Theoretical Background

2.1 Knowledge Creation Background for Innovations

Employees invent while they work and innovative ideas can emerge from learning opportunities across the organization, not just R&D (Lee and Walsh 2016). However, the type of knowledge and learning, and oftentimes the type of innovation as well, differs between R&D and non-R&D. One way to make these different learning modes tractable and implementable for an experimental setting is to use the classification of Jensen et al. (2007). In their seminal paper, they distinguish between two learning modes: learning by science, technology, and innovation (STI) and learning by doing, using, and interacting (DUI). Notably, this distinction is very similar to other knowledge categorizations as found e.g. in Asheim and Gertler (2005), who conceptually divide knowledge bases into analytical (science-based) and synthetic (experience-based) knowledge.

The STI-mode is described as explicit, analytical, codified, scientific and technological knowledge, which is fostered by formal processes of R&D and in innovations units. This usually involves large shares of high skilled academic workers or a legal department focusing on patent law. Furthermore, it refers to "know-why" and "know-what" and global knowledge (Aslesen et al. 2012; Guo et al. 2010; Jensen et al. 2007).

The DUI-mode, on the other hand, is described as an informal process of learning and experience-based "know-how" and "know-who" (Aslesen et al. 2012; Guo et al. 2010; Jensen et al. 2007; Thomä 2017). Knowledge is created by skilled workers through practical skills and on the job training (Aslesen et al. 2012; Isaksen and Karlsen 2010; Jensen et al. 2007; Lundvall 2016; Thomä 2017). (Tacit) Knowledge transfer is enabled trough informal communication and communities of practice, which involve building functioning structures and relationships (Guo et al. 2010; Jensen et al. 2007; Thomä 2017). Furthermore, learning and knowledge creation mainly derives from applied research and development, day-to-day activities (routines) as well as non-scientific drivers, namely learning by doing, using, and interacting (Apanasovich et al. 2016; Isaksen and Karlsen 2010; Jensen et al. 2007; Parrilli and Heras 2016). Individuals or groups operating within DUI-modes of learning thus attain innovations through their cumulative experience.

Following earlier work on non-R&D innovations and the DUI-mode, neither formal R&D departments, innovation units, large shares of high skilled academic workers nor a legal department focusing on patent law are necessary prerequisites for innovations. Instead, providing safe spaces for experimentation or trial-and-error learning, or promoting information transfer and interaction within and across organizations, can provide a sustainable source of innovation potential (Aslesen et al. 2012; Guo et al. 2010; Herstad et al. 2015; Jensen et al. 2007; Lundvall 2016) that is largely independent of a company's formal-structural organization. Even knowledge that is acquired as a byproduct of regular routine activities and learned experience can be transformed into innovations (Argote and Guo 2016; Cyert and March 1963; Feldman and Pentland 2003; Lundvall 2016; March and Simon 1958; Nelson and Winter 1982, 2002). Since the characteristics ascribed to the DUI-mode are (pre)existing in every company and are thus relevant to a wide range of firms and domains, identifying ways to exploit their full potential could lead to significant improvements in innovative activities.

The DUI-mode comprises three key elements (doing, using, interacting), which warrant further clarification. Learning-by-doing, as introduced by Arrow (1962), means repeating the same manufacturing operation, which through repetition steadily evolves over time. This experiential learning and development of an (manufacturing) operation generates productivity growth (ibid.). In line with Arrow, Malerba (1992) refers to learning-by-doing as a company-internal process, which is related to production activity, or as Lundvall and Johnson (1994) name it "production sphere". Learning-by-doing occurs while producing a good or service. Accumulating the gained-while-producing experience is essential for optimizing the production process and thus reducing costs or increasing output. In other words, learning-by-doing means, getting better at what you are doing by actually doing it.

Learning-by-using is associated with the consumption of a produced good (or service) (Lundvall 2016; Lundvall and Johnson 1994). Using technologies, machines, equipment or the (final) product (service) itself creates knowledge (Amara et al. 2008; Rosenberg 1982), which can either improve certain features, or add new features to an existing good (Rosenberg 1982). Learning-by-using can take place both inside and outside the firm (Hippel 1986).

Finally, learning-by-interacting is the most investigated component of the DUI-mode (Apanasovich et al. 2016) and "almost all learning is interactive" (Lundvall 2016, p.177), as individuals are in a constant communicative exchange process with others and not working in isolation. Knowledge can be created by interacting with people inside the firm, e.g. colleagues from the same or different departments, or with external actors such as customers, suppliers or other agents along the firm's supply chain (Aslesen et al. 2012).

In sum, these two modes contrast two opposing (ideal) types of knowledge sources, which still lack clear definitions and segregation and might not be found in pure form in practice. Nevertheless, this differentiation adds a considerable value to innovation research. First, it points out that even companies without formal R&D or formalized innovation processes can be innovative, i.e. these are no necessary preconditions. Second, focusing on R&D generated or patented innovations only leaves out an enormous amount of innovations, especially incremental ones. Still, firms without R&D departments have to be more careful with the trade-off between innovations and efficiency in the production process, as they are usually smaller in size and fewer resources are budgeted towards innovation activity.

2.2 Goal Setting

Goal-setting theory has been developed inductively to explain what prompts individuals to perform better on some work-related tasks than others. In practice, managers often set goals to motivate employees, improve performance or aspire learning. Goal-setting theory is based on a large foundation of research work (Locke and Latham 1990a, 2002, 2006, 2013c). The key idea is that setting a conscious goal affects its achievement[2]. There are mainly four causal mechanisms that explain why goal-setting is effective (Latham 2004; Locke and Latham 1990a, 2013a; Seijts and Latham 2005). First, a goal gives individuals evidence what to focus on. Thus, individuals will pay more attention to goal-relevant at the expense of goal-irrelevant activities. Second, individuals will adjust their level of effort to goal difficulty, with higher levels of difficulty resulting in higher effort. Third, as a consequence thereof, effort persists until the goal is reached: higher goals perpetuate effort. While these three are of motivational nature, the fourth is cognitive. Individuals either use their existing knowledge, or generate new strategies and knowledge to attain it.

The benefits of goal-setting depend on several moderators (c.f., Locke and Latham 1990a, 2002, 2013a). First, goal commitment is necessary, notably if the goal is difficult. Individuals need to be attached and determined to be motivated to increase effort and

[2]However, subconscious goals are effective as well (for an overview see Friedman 2013).

attain the goal. Second, ability affects goal performance. A lack of required knowledge and skills will have a negative effect on goal performance, and high ability predicts goal attainment. Third, feedback helps individuals to adjust their level of effort and/or direction of strategy. This feedback has to be translated into action to enhance performance. Fourth, goal-setting works better for tasks with low complexity, because strategy and routines are easier automatized. Fifth, goals and situational constraints should match. Leaders should provide sufficient resources to attain the goal and remove obstacles. Sixth, valuation and appraisal during goal attainment can affect an individual's performance. A positive relationship between performance success and satisfaction, and otherwise a negative relationship between failure and dissatisfaction can be explained by the fact that goals set standards by which performance is measured and perceived.[3]

Goal-setting takes place in a wide range of fields (c.f., Locke and Latham 2013c) and is also used and discussed to promote creativity (for an overview see Shalley and Koseoglu 2013) and innovations (Hoegl and Parboteeah 2003). Innovations, as well as creativity, are valuable if they are novel and useful (Amabile 1983; Grant and Berry 2011; Miron-Spektor and Beenen 2015). The novelty aspect is associated with learning and can be promoted by a learning (achievement) goal (Hirst et al. 2009; Miron-Spektor and Beenen 2015) or moderated or mediated by learning goal orientation (Hirst et al. 2009, 2011; Simmons and Ren 2009; Yaping et al. 2009). Contrary, a performance (achievement) goal focuses an individuals attention to more practicable solutions, which concurrently demonstrates others their competence and ability (Janssen and Yperen 2004; Miron-Spektor and Beenen 2015; VandeWalle 1997). These are the two kinds of goals we analyze in this paper.

Performance versus Learning Goals

It is well established that specific, challenging performance goals promote effort and therefore increase performance outcomes more than easy or do-your-best goals (c.f., Locke and Latham 2002, 2006, 2013a; Mento et al. 1987; Seijts et al. 2004). When setting specific, high-level performance goals, individuals focus attention on this goal and increase effort to attain it. But at a certain point, an increase in effort alone will not enable individuals to reach the performance goal, as high performance is a function of both motivation and ability (Seijts and Latham 2005). The pressure of a specific, challenging performance goal creates incentives to come up with solutions [4]. It can force people to think "outside the box", do things in different way, use their imagination and discover new strategies(c.f., Kerr and LePelley 2013). This re-thinking of the current state not only helps goal attainment, but might also foster innovation.

However, setting a specific, challenging performance goal on a new and rather complex task can lead to tunnel vision in a way that individuals obsessively focus on outcome and miss paying attention on acquiring required skills to achieve the goal (Locke and Latham 2006; Seijts and Latham 2005). Moreover, according to resource allocation theory, cognitive

[3] According to Locke and Latham (2013a), personality is the seventh moderator variable. But specific, high goals create such strong cues, which directs behavior, so individual differences in personality are covered (Locke and Latham 2013a).

[4] Goals with an extremely high and specific performance target, which seem unattainable from the current point of view are called stretch goals (for an overview see, Kerr and LePelley 2013). Setting these kinds of stretch goals requires a certain environment, empowerment and structural modifications (Kerr and Landauer 2004). Furthermore, sensitive and supportive leadership is needed to deal with the negative effects that accompany seemingly unattainable goals (Kerr and LePelley 2013). Since stretch goals need such special environment and leadership through the process of reaching the goals, we abstain from using stretch goals.

resources are limited. Complex tasks have a steep demand for cognitive resources, because attention is paid to self-regulatory processes, such as self-evaluation or attaining results (e.g., Ackerman et al. 1995; Kanfer and Ackerman 1989; Kanfer et al. 1994). Specific, challenging performance goals "interfere with learning appropriate strategies or procedures that will enable individuals to accelerate their effectiveness" (Seijts et al. 2013, p.195). When setting a specific, challenging performance goal, no or too little attentional resources are left for learning or mastering a complex task. This explains why performance goals often perform worse than learning goals when task complexity is high. Thus, when tasks are complex or new, attention should first be paid to acquire appropriate knowledge, and skills. Then, actors can target to focus on optimizing and routinizing task strategies and processes (Kanfer and Ackerman 1989). For complex tasks, it can therefore be advantageous to set easy or do-your-best performance-goals (Earley et al. 1989; Kanfer and Ackerman 1989; Latham and Locke 2007; Seijts et al. 2004; Winters and Latham 1996), since they do not bind as many cognitive resources, and thus allow for more experimentation and less anxiety. Seijts et al. (2004) find, in accordance with goal-orientation theory by Dweck and Leggett (1988), a positive correlation between trying to acquire new knowledge and skills and performance, when individuals asked to do their best. They also find individuals in a do-your-best condition to perform better than individuals with a specific, challenging performance goal[5].

Another way to support employees to develop new and effective task strategies, i.e. be innovative, is to focus their attention on learning goals. Learning goals are an effective tool, especially with new and complex tasks, where appropriate strategies have to be discovered[6] (c.f., Latham 2004; Masuda et al. 2015a; Seijts et al. 2013; Seijts and Latham 2005). The frame of instructions for learning goals draws attention away from efficiently using extant knowledge, skills, ability and previously learned performance routines. Instead, individuals are framed to search for and develop required task knowledge and skills (Latham and Locke 2007; Seijts et al. 2013; Seijts and Latham 2005). Moreover, this instructed learning goal orientation helps individuals to deal with errors, since errors and setbacks are viewed as essential for the learning process.

Several studies focus on the effects of learning goals, especially in comparison to performance goals. Mostly, specific, challenging learning goals are compared to specific, challenging performance goals and do-your-best performance goals. When task difficulty is high and individuals lack appropriate task strategies, they perform better with a learning goal than with a specific, challenging or do-your-best performance goal (c.f., Drach-Zahavy and Erez 2002; Seijts and Latham 2001; Seijts et al. 2004; Winters and Latham 1996). Whereas, when individuals have been provided with effective task strategies, a learning goal is not only useless, it also falsely frames individuals to pay attention to acquire the knowledge they already possess, which result in inefficiencies (Brown and Latham 2002). Moreover, Kozlowski and Bell (2006) find that congruence between a learning frame and learning goal content to be more favorable for self-regulation processes (like less negate affect, more exploratory behavior and higher self-efficacy) than consensus between performance frame and performance goal content. Furthermore, self-reports for exploratory behavior were higher for individuals with a learning goal and results reveal higher basic performance, i.e. number of correct or incorrect decision in a computer radar-tracking simulation, for individuals with a learning than those with a performance goal.

[5]Note, the do-your-best goal refers to do one's best according to outcome performance.
[6]For a review of learning goal literature see, Seijts et al. (2013).

Finally, Seijts et al. (2013) conduct a meta-analysis to examine the effect size for learning goals in relation to specific, high-level performance and do-your-best performance goals. First, when comparing the effect sizes of these three goals on task performance, they find a significant, but small advantage for learning versus performance goals and specifically a medium, also significant, incremental advantage for learning over do-your-best goals. For task complexity, the results described above are confirmed: learning goals are better than performance goals when task complexity is high and performance goals achieve better performance results when task complexity is low. Third, concerning repetition, i.e. the number of trials, the performance benefit of learning goals compared to performance goals is highest for the first trial and than declines moderately, but still holds over multiple trials. Finally, the benefits of learning over performance goals are stronger with increasing task duration.

So far, all effects of goal setting describe results from goal-setting for individuals. However, organizations increasingly adopt their structures to focus on teams. Setting goals in teams allows for setting goals on an individual as well as a team level (Kramer et al. 2013). When team and individual goals are conflicting, individuals face a social dilemma. Goals where individual and group goals are compatible perform better than those that imply incompatibility between the two dimensions (Seijts et al. 2000). Setting cooperative instead of competitive or independent goals among team members improves team performance, as team members apply their abilities for the mutual benefit of all team members (Tjosvold and Yu 2004). Furthermore, cooperative group goals help group members develop confidence in group efficacy and initiative to persist in task accomplishment and innovation (Wong et al. 2009). In line with these findings, we focus on cooperative group goals to foster knowledge acquisition and non-R&D innovations.

3 Experimental Design

Experiments constitute a core part of empirical economics. They allow researchers to isolate specific mechanisms of actions and causally quantify the impact of different interventions on a number of pre-determined outcomes. In economic innovation research, however, experiments are still rare (Brüggemann et al. 2016), and virtually non-existent in the context of non-R&D analogous learning from cumulative knowledge. We therefore develop a new real-effort task, which captures the key elements of innovations as derived from learning by doing, using, and interacting within repetitive, routine-like tasks. To that end, we introduce an encryption task that asks subject to decipher a letter-string based on numbers. Before the experiment, letters are transformed into numbers using an encryption mechanism (cipher), which represents the innovation. Such a setup allows for a task-based objective measure of innovation and matches the attributes of a creative idea – novelty and usefulness –, which is a core element of innovations (Berg 2014; Ford 1996; Miron-Spektor and Beenen 2015). To mimic innovation as a by-product of normal jobs outside of R&D, participants are neither instructed to find the cipher or an encryption pattern, nor is it mentioned in the instructions.

Subjects are divided into groups of three and recurrently translate a letter-string into numbers. At the beginning, all participants have to experiment and rely on a simple trial-and-error process to find the correct translation. Through immediate feedback, participants learn by doing, i.e. by performing the task. This feedback along with an additional history table of the group's cumulative correct translations enables subjects to acquire task experience and knowledge, which is important to enhance performance and foster learning. Groups can create and constantly improve their search pattern. In the best case, they will explore the cipher, i.e. the innovation. Similar to the development process of an innovation, ideas of every step towards the final innovation (changes in a group's search mechanism) have to be applied, tested instantly, and reviewed. This illustrates learning-by-using: creating a solution proposal form information gained through learning-by-doing, and then learn from testing this possible solution by using it. Finally, our experiment allows participants to communicate with their group members via a chat box, representing learning-by-interacting. The encryption tasked is designed to allow for beneficial group-member-interactions, as subjects can learn from each other, discuss ideas, share knowledge, and suggest solutions.

Thus, our overarching design captures all three key characteristics of non-R&D group innovations as proposed by the DUI-literature (Jensen et al. 2007). Importantly, since we neither frame participants to be innovative, nor to find the innovation, it allows us to better understand innovative behavior outside a firm's formal innovation structure like R&D departments and innovation units. Finally, this task generates quantifiable measures (e.g., being innovative, number of trials (input) and translated letter-strings (output)) while allowing for high controllability of external factors such as communication (chat box).

3.1 Experimental Task

We will now describe the main parameters used in this experiment. However, depending on the research objective, they can be altered and augmented to e.g. explore variability in innovation complexity, restrict communication, or change group compositions.

For our purposes, subjects are randomly matched into groups of three and have 45 minutes to translate four-character letter-strings into numbers. Each letter has to be encoded into a positive integer from the interval [1;193]. Before the experiment, we programmed 40 letter-strings, capping the possible number of strings that could be decoded.[7] Letter-strings were created by randomly stringing together four letters of the German alphabet (26 letters). The sequence of letter-strings was fixed across conditions to avoid endogeneity issues. To support knowledge acquisition, we drew letters such that four letters (*S, W, M and J*) appeared more often in the first ten rounds. Thus, it was easier for groups to directly compare letters over time and decipher the pattern. To finish a round and continue with the next letter-string, all three subjects within a group need to indicate the correct integer for each letter. The total number of translated letter-strings is displayed on the decision screen.

We opted for a flat, time-based payment of €15 for about an hour of work[8]. This avoids any confounding biases due to additional monetary incentives (Jenkins et al. 1998; Lee et al. 1997; Locke 2004; Smith 1976) and generalizes to most real-world contexts where employees are primarily compensated with performance-independent wages. This holds in particular for innovative activities that are beyond the scope of an individual's normal job routine. Thus, we isolate the effect of goal-setting on learning and performance, while replicating the basic motivational effect of paying a baseline wage.

[7]The maximum of translated letter-strings is 22.

[8]The encryption task itself lasts for 45 minutes plus approximately 15 minutes for reading instructions, comprehension questions and a post-experimental questionnaire.

Encryption mechanism. The cipher has two dimensions: a between-letter and a within-letter variation (see Table 1). The between-letter variation describes the relationship between subsequent letters in alphabetical order, whereas the within-letter variation describes the relationship within the same letter across rounds, i.e. conditional on how often a specific letter has appeared over the course of the experiment. First, each letter is assigned a starting number β. For example, "A" starts with $\beta = 16$. The next letter in the alphabet, "B", starts with $\beta = 16 + 7 = 23$. This is the between-letter variation. For each letter of the alphabet, the starting number β equals the starting number of the previous letter plus seven. The within-letter variation follows the following pattern: $\beta \longrightarrow \beta - 15 \longrightarrow \beta_1 + 17$. Taking "A" for example, the within-letter encryption pattern is $16 \longrightarrow 1 \longrightarrow 18$, and re-starts with 16 upon its fourth appearance. When a letter has to be translated for the fourth time, the pattern always repeats itself, starting with β.

Table 1 Encryption of the alphabet

	β	β-15	β_1+17		β	β-15	β_1+17		β	β-15	β_1+17		β	β-15	β_1+17
A	16	1	18	H	65	50	67	O	114	99	116	V	163	148	165
B	23	8	25	I	72	57	74	P	121	106	123	W	170	155	172
C	30	15	32	J	79	64	81	Q	128	113	130	X	177	162	179
D	37	22	39	K	86	71	88	R	135	120	137	Y	184	169	186
E	44	29	46	L	78	95	93	S	142	127	144	Z	191	176	193
F	51	36	53	M	100	85	102	T	149	134	151				
G	58	43	60	N	107	92	109	U	156	141	158				

To make the encryption mechanism more decipherable, the encryption mechanism repeats itself for letters *S, W, M and J* in rounds seven, eight, nine, and ten respectively (Table 2). For example, "S" has to be translated in the first, third and fifth letter-string with numerical values of 142, 127 and 144 respectively. In letter-string seven, "S" has to be translated again and the pattern repeats.

Table 2 First ten letter-strings and their translation

No.	Letter-string	Translation	No.	Letter-string	Translation
1	WESM	W=170, E=44, S=142, M=100	6	OJWE	O=114, J=81, W=172, E=29
2	JUFQ	J=79, U=156, F=51, Q=128	7	SIMB	S=142, I=72, M=102, B=23
3	ZSAH	Z=191, S=127, A=16, H=65	8	NWYA	N=107, W=170, Y=184, A=18
4	JWMA	J=64, W=155, M=85, A=1	9	QMFU	Q=113, M=100, F=36, U=141
5	GPCS	G=58, P=121, C=30, S=144	10	SWJH	S=127, W=155, J=79, H=50

No. refers to the chronological order. The full set of all 40 letter-strings is uploaded in the online repository.

3.2 Treatments

In order to test whether and how goal-setting can encourage innovative behavior, we implement a 2x2-design that distinguishes between learning and performance goals, as well as do-your-best (DYB) and specific, challenging (SC) goals. Performance goals focus on the output side of the production process, while learning goals focus on the input side. A higher number of translated letter-strings represents higher output, allegorically goods or services produced by a firm. A lower number of trials used for correct translations exemplifies lower input, i.e. fewer material needed or wasted to create the innovation, resulting in cost reduction. In distinguishing between do-your-best and specific-challenging goals, we examine whether a broader or more narrow focus supports innovation attainment.

Following the fact that that do-your-best performance goals can improve performance when task complexity is high, groups in the first, the do-your-best performance goal treatment (DYB P) are asked to *do their best in translating as many letter-strings as possible within the 45 minutes.* Specific challenging goals have been shown to increase effort and therefore performance. Thus, the second treatment implements a specific, challenging performance goal (SCP). According to Locke and Latham (2013b), difficult goals are usually set at the 90th percentile in laboratory experiments. Using a pilot study with 69 subjects (57 played DYB P), we identified the 90th percentile of letter-strings translated by groups as 13. Thus, participants are asked to *translate 13 letter-strings or more within 45 minutes.*

Acquiring new skills and knowledge are necessary for innovations (c.f., Miron-Spektor and Beenen 2015). To detect the innovation, participants need to learn about the task and the cipher or encryption pattern. Thus, in the third treatment, participants are instructed with a do-your-best learning goal (DYB L), i.e. *the goal is to reduce the number of trials per letter-string of all group members over time.* Again, to find out whether a do-your-best or a specific-challenging goal fits better to our innovation context, a specific, challenging learning goal (SCL) was assigned in the fourth treatment. Subjects were told that *the goal is to reduce the number of trials per letter-string of all group members to one trial per letter-string.*

3.3 Experimental Procedure

The experiment is programmed in oTree (Chen et al. 2016) and took place in the Laboratory for Behavioral Economics at the University of Goettingen. We used ORSEE (Greiner 2015) for participant recruitment. The pilot study was conducted from the end of November to the beginning of December 2018, while the main experiment was conducted from the end of January to the beginning of Mai 2019. In sum, 339 subjects participated, with 69 participating in the pilot study. Due to technical problems with oTree and the server, we excluded data from 30 main experiment participants. This leaves us with 240 participants in 80 groups for the main experiment, which are equally distributed over treatments. The average age is 24 years, 45.11% are female and 44.17% study at the faculty of business and economics. Besides, 25.83% attended a lecture, where economic experiments were part of the syllabus.[9]

3.4 Measures

Innovation. In our experiment, detecting the cipher is equivalent to creating an innovation. On an individual level, we consider participants as innovators if they used only one trial to translate a letter-string. Due to the group structure and communication option in our experiment, the main tests will be on group level. When all group members simultaneously manage to reduce their trials per letter-string to only one for at least one letter-string, we consider the group as innovative. In total, 55 individuals from 28 groups reduced their number of trials to only one. In 12 groups all group members *simultaneously* needed – at least once – only one trial to translate a letter-string.
Learning. Learning performance is measured by the number of trials per group and letter-string. The number of trials is the sum of trials required by all three group members.

[9]We share complementary information about the chat data, additional tables and further explanations in the online repository.

However, if all three group members inserted the correct numbers by their first trial, the number of trials is indicated as one instead of three. We did that, to make instructions in the fourth treatment more clear. For sub-analyses, we also use the individual number of trials per participant and cluster errors on group level.

Performance. Outcome performance is measured by the accumulated number of letter-strings translated within the 45 minutes. We only count letter-strings that have been correctly translated by all three group members in the given time. This means for example, if two group members have inserted all four correct numbers of a letter-string, but the third member only has inserted three correct numbers by the moment the time expires, the word does not count.

Goal commitment. To check for goal commitment, we adapted the five-item standard questionnaire with a five-point Likert scale (1 = strongly disagree; 5 = strongly agree) from Klein et al. (2001) for team goals (moderate Cronbach's alpha of $\alpha = 0.6922$). Furthermore, we constructed an additive index with all questions equally weighted. Mean group goal acceptance is 3.89, indicating that individuals broadly accepted the assigned goal and were somewhat attached to it. There are no significant treatment differences (Wald-test: $F(3, 235) = 1.25, p = 0.291$).

Sense of cohesiveness and applying abilities. Having a sense of cohesiveness and a mindset that all group members' abilities are beneficial for team goal achievement are important for cooperation and goal attainment. Participants rated on a five-point Likert scale ($\alpha = 0.8234$) their sense of cohesiveness and attitude towards applying abilities (items were adopted from Tjosvold and Yu 2004). The mean of an equally-weighted additive index is 4.04, indicating that groups developed a sense of cohesiveness and value others' abilities. There are no treatment differences, revealing all goals affected a group's cohesiveness equally (Wald-test: $F(3, 235) = 1.07, p = 0.364$).

Goal difficulty. Participants rated goal difficulty on average with 4.6 on a seven-point Likert scale (1 = not difficult at all, 7 = extremely difficult), meaning goals are perceived as moderately difficulty. Participants, who were assigned a specific, challenging goal, perceive their goal as significantly more difficult ($\overline{x}_{\text{SC}} = 5.04$) than those assigned a do-your-best goal ($\overline{x}_{\text{DYB}} = 4.17; t(238) = -4.445, p < 0.000$). Furthermore, learning goals are perceived as significantly more difficult ($\overline{x}_{\text{L}} = 4.96$) than performance goals ($\overline{x}_{\text{P}} = 4.25; t(238) = -3.548, p < 0.000$).

Chat data. Lastly, we use chat data that is coded by modifying the decision function coding system of Poole and Roth (1989) and thereby develop 14 initial subcategories. Applying these 14 categories to our chat data and then reducing them to those relevant for complementing the experimental data leaves us with four final categories: knowledge sharing, task strategy, progress orientation and group-directed strategy. To increase the reliability of our coding, we use the independent coder method and compare the results from two coders for consistency. If needed, adjustments were made by the author of this paper. The modified coding scheme, including all subcategories that we used for the categories, coding examples, and descriptive statistics, can be obtained from the online repository.

4 Results

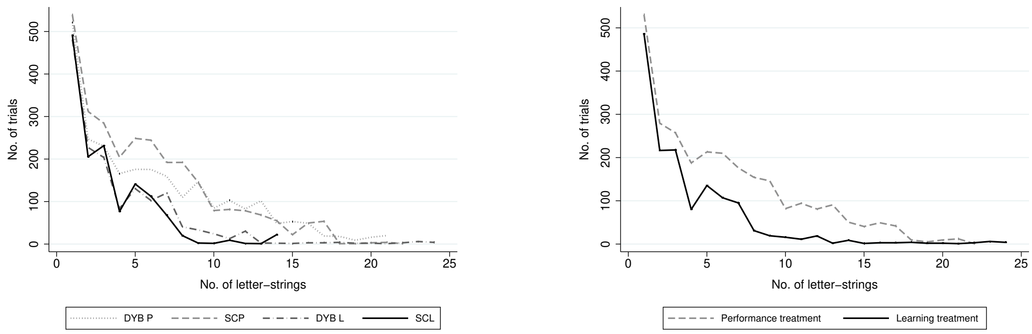
We first report the results on subjects' innovation and learning performance. These are the two most important aspects in determining the role of goal-setting for non-R&D

innovations. We then focus on overall outcome performance, discuss trade-offs, and complement the analysis with chat-level data.

4.1 Innovation and learning performance

Innovation and learning performance is measured by the number of trials per translated letter-string. Figure 1 (left) shows the average number of trials groups needed to translate each respective letter-string over time. As expected, trials decrease with cumulative experience. While the graphs are very similar at the margins, groups who receive a learning goal appear to exhibit steeper learning at the beginning (right figure). At the beginning, all groups possess next to no knowledge and have to rely on trial-and-error processes. However, after a few successful translations, learning goals increase a group's focus on gaining task relevant knowledge.

Figure 1 Average number of trials by treatment and learning over time.



Over time, groups with performance goals become more experienced and effective as well, leading to a convergence towards the end of the experiment.

To test for treatment differences, we first run a linear regression of the average number of trials on a categorical treatment variable and test for significance using a Wald-test. Results reject the hypothesis that all four treatments are equal ($F(3,686)=6.31$, $p<0.000$). Comparing the average number of trials groups needed over time (see Table 3), we document a significant difference between learning and performance goals ($t(688)=3.53$, $p<0.000$). Furthermore, groups with DYB-goals needed on average significantly less trials than groups with SC-goals ($t(688)=-2.46$, $p=0.014$). Excluding the first letter-string to control for the early trial-and-error period only increases the difference (learning vs. performance: $t(608)=5.42$, $p<0.000$; DYB- vs. SC-goals: $t(608)=-2.88$, $p=0.004$).

Table 3 Average trials, number of letter-strings and effort per group by treatments

	Treatment			
Variable	(1)	(2)	(3)	(4)
Average trials	193.45	236.17	157.31	176.35
	(171.38)	(185.47)	(164.04)	(184.80)
Number of letter-strings	9.75	9.5	8.3	7
	(4.53)	(4.21)	(4.65)	(3.25)
Effort	7.86	8.59	18.65	33.00
	(11.99)	(15.44)	(29.92)	(58.94)

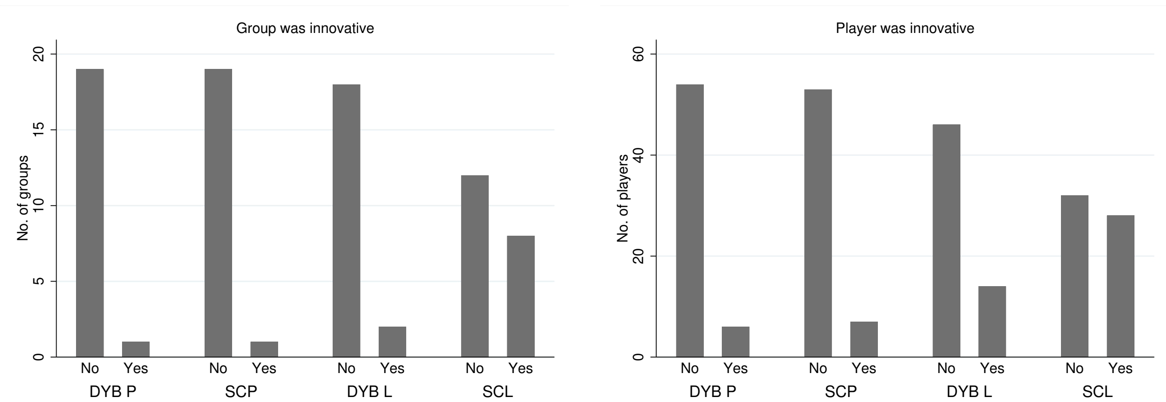
Standard deviation in parentheses.

After establishing that learning goals appear to foster incremental learning, we now analyze groups' and individual innovativeness concerning the cipher. We define groups as innovative if – at least once – all three group members simultaneously needed only one trial per letter-string. This conservative measure minimizes the risk of false-positives.

Overall, only 15% of all groups (12 out of 80) were innovative, confirming the difficulty of our innovation task (see figure 2). A χ^2-test confirms significant differences over all treatments ($\chi^2(3) = 13.33; p = 0.004$). As depicted in figure 2, the specific-challenging learning goal seems to clearly outperform all other alternatives with an innovation share of 40%. The second best performer are do-you-best learning goals (10%), whereas there are almost no innovative groups in the performance goal treatments. Overall, out of 40 groups with a learning goal, 25% were innovative. That share drops to 5% for performance goals ($\chi^2(1) = 6.27; p = 0.012$). Similarly, 22.5% (7.5%) of groups with a SC-goal (DYB) were innovative ($\chi^2(1) = 3.53; p = 0.060$). These differences are solely driven by groups in the SCL treatment.

This conservative definition of innovation on group level might overlook innovation on an individual level, since one non-innovative player can curtain the other two. The right side of figure 2 shows the number of innovative players by treatment. We categorize players as innovative, if they – at least once – used only one trial per letter-string. Results are similar to group level analysis ($\chi^2(3) = 29.13; p < 0.000$), with SCL dominating all conditions. DYB-L has twice as many innovative players than T1 and T2. Thus, the overall results do not change. Learning goals, but in particular challenging, specific learning goals, foster knowledge acquisition and thus innovative behavior within routine, non-R&D-like tasks.

Figure 2 Number of innovative groups (left) and players (right) by treatment



Finally, we run a random effects GLS regression on the number of trials per letter-string with standard errors clustered on group level (see table 4). The analysis confirms the positive effect of learning goals (*learning* is 1 for learning goals and 0 otherwise) on the number of trials needed by an individual. Participants with a learning goal on average need about 20 trials less than those with a performance goal (model 3). The specific-challenging learning goal reduces the average number of trials by another 20 (see interaction term, where *goal type* is 1 for SC and 0 for DYB goals). Lastly, subjects' accumulated knowledge over time reduces the average number of trials by another six trials per additionally translated letter-string.

For the logit regression, we use a dummy variable which is one if a player used only one trial to translate the respective letter-string and zero otherwise. Results are in line

Table 4 Random effects GLS and logit regression on the number of trials

	No. of trials			Innovative letter-string		
	RE			*Logit*		
Variable	(1)	(2)	(3)	(1)	(2)	(3)
Learning	-29.05***	-19.00**	-19.58**	2.70***	2.15**	2.70***
Goal type	6.13	16.07*	14.83*	1.03**	0.37	1.13*
Letter-strings	-5.77***	-5.85***	-5.80***	0.27***	0.28***	0.28***
Learning*Goal type		-20.52*	-19.57*		1.01	
Goal commitment			3.79			-0.70*
Cohesiveness & applying abilities			-1.81			-0.36
Constant	136.36***	132.15***	128.58***	-8.46***	-8.12***	-5.19*
Obs.	2070	2070	2064	2073	2073	2067

$^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.10$. Standard errors are clustered on group level.

with the previous analysis, showing positive effects of learning and SC-goals on individual innovative behavior. Also, an increase in the number of letter-strings raises the log-odds of individual innovative performance, which is consistent with the right figure 2, where the number of trials decrease with increasing number of translated letter-strings.

Overall, our results suggest that setting a learning goal motivates participants to focus on acquiring task-specific skills and enables them to be innovative. Furthermore, if a learning goal is challenging and has a clearly defined reference point, it is a lot more effective in realizing a group's or individual's innovative potential. Performance goals, on the other hand, are less successful – if at all – at eliciting innovative behavior.

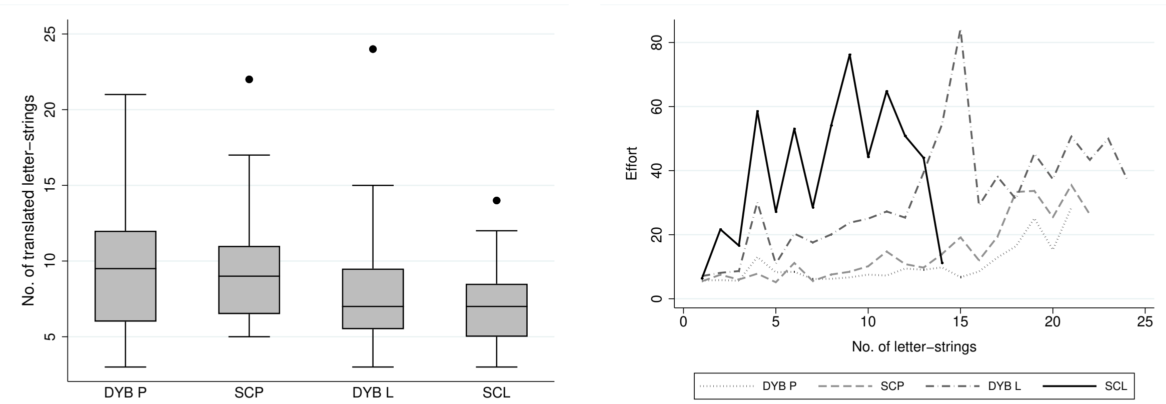
4.2 Outcome performance and effort

We now turn to groups' overall performance within the 45 minutes. While learning goals, in particular specific ones, do appear to increase learning as well as the likelihood that a group or an individual discovers the innovation, these processes might well come with some cost such as short-term output losses.

We measure outcome performance by the number of letter-strings a group translated within 45 minutes. Table 3 depicts averages per treatment and indeed shows an effect of setting a performance goal. When instructed to focus on the outcome, groups translate significantly more letter-strings (t(78)=2.16, p=0.034). However, we find no difference between SC and DYB goals. Furthermore, only two out of 20 groups in SCP achieved the externally set goal and translated more than 13 letter-strings.

It is likely that the time constraint of 45 minutes is one explanation for the observed performance gap, as subjects with a learning goal are nudged to allocate resources towards knowledge acquisition and learning instead of immediate performance. In the long term, this should increase output, as exemplified by the fact that innovative groups ($\overline{x}_{\text{innovative}} = 11.83$) produced on average more outcome that non-innovative groups ($\overline{x}_{\text{non-innovative}} = 8.07$, t(78)=-3.01, p=0.003). If true, we would predict subjects in the learning treatments to spend more time trying to figure out a mechanism, particularly at the beginning of the experiment. We therefore define a new variable called "effort", defined as seconds per trial. As table 3 clearly shows, groups with learning goals invest more time into each trial than those with performance goals (t(78)=-4.83, p<0.000). Furthermore, a specific challenging goal is also significantly more time-consuming (t(78)=-2.04, p=0.045)[10], indicating that an external reference point motivates participants to allocate more time resources in order to achieve the externally set goal.

Figure 3 Number of letter-strings (left) and effort (right) by treatment



The right side of figure 3 illustrates how effort develops with increasing task experience. Performance goals seem to inhibit learning processes in the beginning, as participants speed up the trial-and-error process, whereas participants in learning treatments – particularly those in T4 – allocate more time into the early knowledge accumulation process. Thus, mastering the task and understanding the underlying mechanisms is time consuming, but a prerequisite for being innovative, as exemplified by the high share of innovative groups in the SCL treatments as well as the fact that innovative groups need on average more seconds per trial ($\overline{x}_{\text{innovative}} = 39.35$) than non-innovative groups ($\overline{x}_{\text{non-innovative}} = 11.56$, t(78)=-6.35, p<0.000).

4.3 Chat data

Finally, we use chat data to further parse-out treatment differences and get more insight into how the different goals impact the team innovation process. Table 5 summarises the overall amount of messages sent within groups by treatment (Wald-test: F(3,76)= 2.44, p=0.071). Congruent with the above documented higher effort, subjects in SCL overall send the most messages, indicating that more time is spent on within group communication. The striking difference between SCP and SCL (t(38)=-2.47, p=0.018) suggests that conditional on the goal, groups adopted very different strategies to attain the externally set goal.

Table 5 Descriptive statistics of total messages sent by treatments

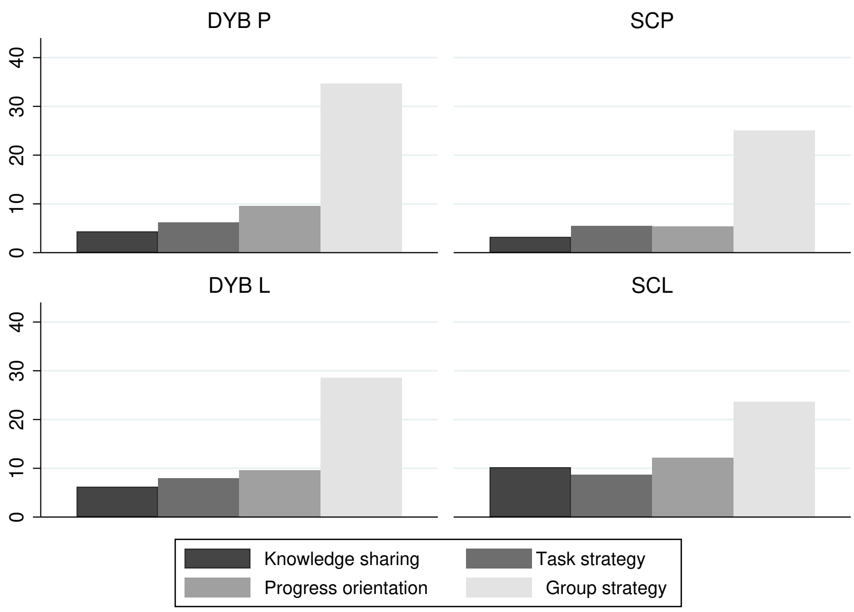
Treatment	N	Mean	SD	Min	Max	Proportion
DYB P	20	129.4	46.6	60	227	26.19%
SCP	20	105.3	36.3	49	176	21.32%
DYB L	20	120.7	28.1	61	178	24.42%
SCL	20	138.8	48.2	77	281	28.08%

[10]When excluding three outliers in treatment 4, the average effort in T4 decreases to 29.79 (44.09). However, differences between learning and performance goals as well as DYB and SC goals remain significant with the same t- and p-values, respectively.

Clustering messages to categories confirms contentual differences between treatments. Our categories of interest are *knowledge sharing, task strategy, progress orientation and group strategy. Knowledge sharing* contains messages with information or suggestions about the cipher (e.g., "always minus 15 in the second round"). It does not contain solutions for a single letter (e.g., "e=44"), because of endogeneity concerns. Single-letter-solution would not only inflate the data, but by design they directly depend on the number of letter-strings translated, as groups can only start translating a new letter-string once all group members successfully completed the current one. The higher the number of translated letter-strings, the higher the number of messages including single-letter-solutions. *Task strategy* includes statements about how to approach or perform the task (e.g., "Let's try trial and error"). *Progress orientation* comprises statements or questions concerning individual or group task progress (e.g., "S is still missing"). Finally, statements assigning specific sub-tasks or questions about the group-directed strategy are classified in category *group strategy* (e.g., "Okay, Then I'll take 71-130"). The average amount of messages matched to these categories is shown by treatment in figure 4.

A separate Wald-test for each category reveals significant differences over all treatments and message types except for *group strategy*[11]. Groups with a learning goal chat more about the cipher or any form of letter-number-translation-pattern (*knowledge sharing*) than groups with a performance goal (t(78)=-3.83, p<0.000). Accordingly, as learning goals encourage innovative behavior, groups who were innovative shared significantly more information about the cipher ($\overline{x}_{\text{innovative}} = 9.5$) than non-innovative groups ($\overline{x}_{\text{non-innovative}} = 5.83$, t(78)=-2.44, p=0.017).

Figure 4 Number of messages per category by treatment



Chatting about task strategy (t(78)=-2.69, p=0.009) or individual/group progress (t(78)=-2.93, p=0.004) significantly differs only between learning and performance goals, but does not significantly differ between DYB and SC goals or innovative and non-innovative groups. Finally, there are no significant differences for *group strategy*.

[11]Separate Wald-test results: *Knowledge sharing*: F(3,76)=7.53, p<0.000; *Task strategy*: F(3,76)=2.57, p=0.060; *Progress orientation*: F(3,76)=6.41, p<0.000; and *group strategy*: F(3,76)=1.03, p=0.383.

As shown in figure 3, groups allocate their time differently depending on the goal. While groups with performance goals primarily focus on translating letter-strings faster, groups with learning goals need more time, especially in the beginning. Regressing chat data categories on effort further supports the assertion that communication impacts the translation process.

Table 6 Random effects GLS regression on effort

Effort					
Variable	(1)	(2)	(3)	(4)	(5)
Learning	16.59***	8.04***	7.91**	7.87**	7.91***
Goal type	6.85**	-1.66	-2.18	-2.23	-1.84
Letter-strings	0.66**	0.73**	0.80**	0.81***	0.88***
Learning*Goal type		17.46***	18.53***	18.47***	18.00***
Messages			-0.02	-0.03	-0.02
Knowledge sharing				4.30	5.06
Task strategy					11.36**
Progress orientation					2.33
Group strategy					-0.83
Constant	-2.30	1.22	3.58	3.59	2.14
Obs.	2573	2573	2573	2573	2573

$^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.10$. Standard errors are clustered on group level.

All coefficients in table 6 for non-chat variables are in line with our previous results. They confirm a significant increase of effort (time invested per trial) when groups receive a learning goal. Additionally, significantly more effort is invested the more letter-strings are translated, congruent with the observation that finding the cipher and acquiring knowledge takes some extra effort. The significant effect of goal type in model (1) is driven by T4, as the coefficient is no longer significant with an interaction term of learning and goal type (1 = SC goal). The total number of messages sent within groups has no effect on effort, as well as *knowledge sharing, progress orientation and group strategy*. Only talking about *task strategy* increases effort. Summing up, table 6 illustrates that difference in effort are primarily due to learning goals, specifically SC learning goals, and groups chatting about how to approach or perform the task. These results shed light on how learning goals foster innovation through group-communication, as subjects allocate more time to spread strategic information and come up with solutions for a complex, non-obvious problem in a non-R&D-like context.

5 Conclusion and discussion

Non-R&D-innovations are a vital, yet still underappreciated and under-explored part of economic progress (Hervas-Oliver et al. 2015; Lee 2015; Lee and Walsh 2016; Rammer et al. 2009). Since they are less formalised and often stem from direct experience, it is difficult to systematically identify their underlying processes (Alhusen et al. 2021; Jensen et al. 2007). This, in turn, inhibits organisations' ability to implement tools and processes that elicit productive or useful innovation outcomes. One approach to capture the learning and knowledge creation processes of non-R&D innovation is the distinction of Jensen et al. (2007). Here, learning by doing, using, and interacting (DUI) describes cumulative learning of sticky knowledge, which contrasts learning by science, technology and innovation (STI), capturing traditional R&D-based, analytical and scientific knowledge. Depending on their specific knowledge environment (Asheim et al. 2011; Asheim and Coenen 2005; Asheim and Hansen 2009), firms that aim for innovation need, and to some extent already do, exhibit different innovation management strategies. For example, firms that lack R&D capabilities instead often compensate through innovation management tools that might include human resource management, team work or an emphasis on organizational innovation (Heidenreich 2009; Moilanen et al. 2014; Rammer et al. 2009). In particular, tools that increase the visibility of problems enable DUI-learning and thereby increase the space for innovation (Lee and Walsh 2016; Lundvall and Johnson 1994; Winter 2003).

This study experimentally examines how goal setting as a low-threshold innovation management tool can promote non-R&D innovation from learning by doing, using, and interacting. We develop a new real-effort task comprising the three core elements of DUI, and apply it to four different kinds of goals that are differentiated by their target (learning vs. performance) and externally set reference point (specific, challenging vs. do-your-best). Our results provide clear evidence that specific, challenging learning goals by far outperform their alternatives in promoting innovation and learning. This supports the hypothesis that often, particularly in complex problem spaces, the most effective course of action is to combine learning frames with performance demands (Masuda et al. 2015b; Miron-Spektor et al. 2018; Shao et al. 2019; Waldman et al. 2019). However, even in combination with a do-your-best frame, learning goals appear somewhat useful in promoting learning and altering the knowledge gaining process by shifting attention to problem-solving. Learning goals make opportunities to optimize solution processes more *visible*, resulting in more groups reducing their input, i.e. trials, or even coming up with the innovative solution, i.e. decoding the cipher. In the beginning, groups with learning goals invest more time into communication and testing, at the cost of short-term output. In the long-run, these investments substantially increase the likelihood of a group being innovative, which strongly increases output.

In contrast, performance goals raise short-term output, but seem to inhibit the learning process in the beginning, as participants speed up the trial-and-error process while allocating less time into the early knowledge accumulation process. Therefore, groups need, on average, more trials, and innovative behavior is very rare. Furthermore, chat data reveals that groups with a performance goal interact less with their group members to e.g. work on task strategies or further the overall progress of the team. Knowledge exchange about the cipher or any form of letter-number-translation-pattern is more scarce, illustrating the potentially detrimental effect of outcome-oriented work on improvement, learning and optimization.

To sum up, our results provide strong evidence that learning goals, and in particular specific, challenging learning goals, support knowledge acquisition and thus innovative behavior within routine, non R&D-like tasks. Learning goals increase the visibility of problems, increase effort and motivate collaborative, strategic team behavior. On the other hand, performance goals effectively increase output at the expense of future knowledge gains. A reasonable inference is that organizations might think about setting goals sequentially or simultaneously (c.f., Miron-Spektor and Beenen 2015). When experience is low and tasks are new, it could be more effective to shift employee resources towards knowledge accumulation. Once the process, task and environment are well understood, a performance goal can help exploiting this newly generated knowledge and shift attention towards output.

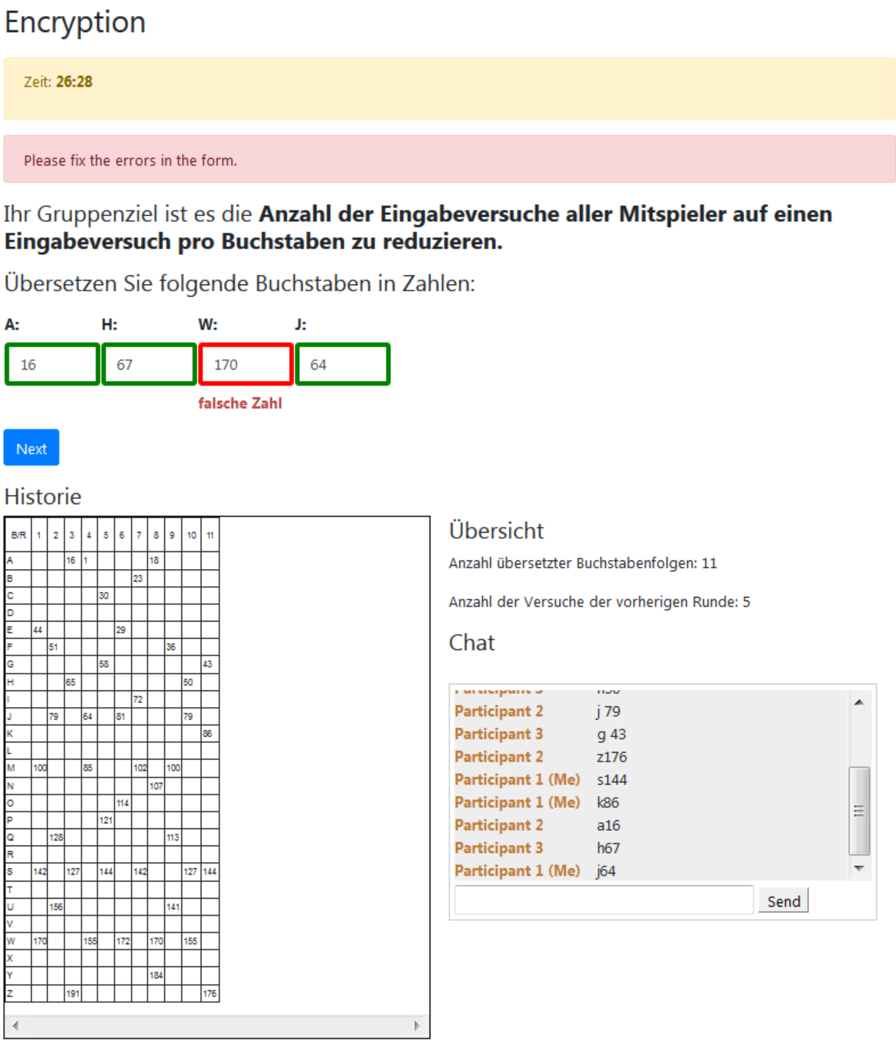
Limitations and future research
This study indicates several follow-up research questions to address limitations and questions that go beyond this experiment. First, as this is a newly developed experimental design, further test are needed to increase robustness and validity. The advantage of our modular construction system is the flexibility to customize many of its features depending on the research questions. Task complexity can be increased (decreased) by longer (shorter) letter-strings or more complex (easier) ciphers. Task duration is adjustable, as well as the number of group members. The option of eliminating or structuring the chat function e.g. through filters enables researchers to better understand the impact of communication on group innovation or vice versa the impact of group structure on innovation relevant knowledge exchange (Charness and Sutter 2012; Meub and Proeger 2017).

Second, the small sample size is a limiting factor. Based on the group structure and the laboratory setting, only 80 independent observations are available. While exploiting the panel-structure and and clustering errors on group level allows for an individual-level analysis with a larger data base, a larger sample size with more diverse populations, experiences and abilities would strengthen results.

Third, various other goals can be tested to contribute to a more differentiated implementation of goal setting as an innovation management tool. As we find specific, challenging learning goals to be most suitable for learning and innovation in non-R&D-like contexts, a longer task duration might reduce this effect as performance outcome should be addressed as soon as skills are acquired. Following Miron-Spektor and Beenen (2015), composing different simultaneous or sequential learning and performance goals might provide more specific insights.

Lastly, the design of our task allows for numerous other innovation research questions to be tested. As markets are dynamic, being innovative once is not sufficient, especially for incremental innovations such as most non-R&D, DUI-mode innovations. A follow-on design could examine how processes change and groups react to changes in the cipher over the course of the experiment. Changes can be both incremental and radical, illustrating different forms of dynamic environments. Furthermore, competition between groups could simulate market competition, and communication between groups could illustrate knowledge and information exchange between firms, as found for example in innovation clusters (Tödtling and Trippl 2005).

Appendix – Screenshot (original German Version)



References

Ackerman, Phillip L., Ruth Kanfer and Maynard Goff (1995). 'Cognitive and Noncognitive Determinants and Consequences of Complex Skill Acquisition'. In: *Journal of Experimental Psychology: Applied* 1 (4), 270–304.

Alhusen, Harm, Tatjana Bennat, Kilian Bizer, Uwe Cantner, Elaine Horstmann, Martin Kalthaus, Till Proeger, Rolf Sternberg and Stefan Töpfer (2021). 'A new measurement conception for the 'doing-using-interacting'mode of innovation'. In: *Research Policy* 50 (4), 104214.

Amabile, Teresa M. (1983). 'The Social Psychology of Creativity: A Componential Conceptualization'. In: *Journal of Personality and Social Psychology* 45 (2), 357–377.

Amara, Nabil, Réjean Landry, Nizar Becheikh and Mathieu Ouimet (2008). 'Learning and novelty of innovation in established manufacturing SMEs'. In: *Technovation* 28 (7), 450–463.

Apanasovich, Natalja, Henar Alcalde Heras and Mario Davide Parrilli (2016). 'The impact of business innovation modes on SME innovation performance in post-Soviet transition economies: The case of Belarus'. In: *Technovation* 57-58, 30–40.

Argote, Linda and Jerry M. Guo (2016). 'Routines and transactive memory systems: Creating, coordinating, retaining, and transferring knowledge in organizations'. In: *Research in Organizational Behavior* 36, 65–84.

Arora, Ashish, Wesley M. Cohen and John P. Walsh (2016). 'The acquisition and commercialization of invention in American manufacturing: Incidence and impact'. In: *Research Policy* 45 (6), 1113–1128.

Arrow, Kenneth J. (1962). 'The Economic Implications of Learning by Doing'. In: *The Review of Economic Studies* 29 (3), 155–173.

Asheim, Bjørn T., Ron Boschma and Philip Cooke (2011). 'Constructing Regional advantage: Platform policies based on related variety and differentiated knowledge bases'. In: *Regional Studies* 45 (7), 893–904.

Asheim, Bjørn T. and Lars Coenen (2005). 'Knowledge bases and regional innovation systems: Comparing Nordic clusters'. In: *Research Policy* 34 (8), 1173–1190.

Asheim, Bjørn T. and Meric S. Gertler (2005). 'The Geography of Innovation: Regional Innovation Systems'. In: *The Oxford Handbook of Innovation.* Ed. by Jan Fagerberg, David C. Mowery and Richard R. Nelson. Oxford University Press, 291–317.

Asheim, Bjørn and Høgni Kalsø Hansen (2009). 'Knowledge bases, talents, and contexts: On the usefulness of the creative class approach in Sweden'. In: *Economic geography* 85 (4), 425–442.

Aslesen, Heidi Wiig, Arne Isaksen and James Karlsen (2012). 'Modes of Innovation and Differentiated Responses to Globalisation-A Case Study of Innovation Modes in the Agder Region, Norway'. In: *Journal of the Knowledge Economy* 3 (4), 389–405.

Berg, Justin M. (2014). 'The primal mark: How the beginning shapes the end in the development of creative ideas'. In: *Organizational Behavior and Human Decision Processes* 125 (1), 1–17.

Brown, John Seely and Paul Duguid (1991). 'Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation'. In: *Organization Science* 2 (1), 40–57.

Brown, Travor C. and Gary P. Latham (2002). 'The Effects of Behavioural Outcome Goals, Learning Goals, and Urging People to do Their Best on an Individual's Teamwork

Behaviour in a Group Problem-Solving Task'. In: *Canadian Journal of Behavioural Science* 34 (4), 276–285.
Brüggemann, Julia, Paolo Crosetto, Lukas Meub and Kilian Bizer (2016). 'Intellectual property rights hinder sequential innovation. Experimental evidence'. In: *Research Policy* 45 (10), 2054–2068.
Cappelen, Ådne, Arvid Raknerud and Marina Rybalka (2012). 'The effects of R&D tax credits on patenting and innovations'. In: *Research Policy* 41 (2), 334–345.
Charness, Gary and Matthias Sutter (2012). 'Groups make better self-interested decisions'. In: *Journal of Economic Perspectives* 26 (3), 157–76.
Chen, Daniel L., Martin Schonger and Chris Wickens (2016). 'oTree-An open-source platform for laboratory, online, and field experiments'. In: *Journal of Behavioral and Experimental Finance* 9, 88–97.
Cohen, Wesley M. (2010). 'Fifty years of empirical studies of innovative activity and performance'. In: *Handbook of the Economics of Innovation* 1, 129–213.
Cyert, Richard M. and James G. March (1963). *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.
Dimos, Christos and Geoff Pugh (2016). 'The effectiveness of R&D subsidies: A meta-regression analysis of the evaluation literature'. In: *Research Policy* 45 (4), 797–815.
Drach-Zahavy, Anat and Miriam Erez (2002). 'Challenge versus threat effects on the goal – performance relationship'. In: *Organizational Behavior and Human Decision Processes* 88, 667–682.
Dweck, Carol S. and Ellen L. Leggett (1988). 'A Social-Cognitive Approach to Motivation and Personality'. In: *Psychological Review* 95 (2), 256–273.
Earley, P. Christopher, Terry Connolly and Göran Ekegren (1989). 'Goals, Strategy-Development, and Task-Performance: Some Limits on the Efficacy of Goal Setting'. In: *Journal of Applied Psychology* 74 (1), 24–33.
Ellström, Per Erik (2010). 'Practice-based innovation: A learning perspective'. In: *Journal of Workplace Learning* 22 (1-2), 27–40.
Feldman, Martha S. and Brian T. Pentland (2003). 'Reconceptualizing Routines Organizational'. In: *Administrative Science Quarterly* 48 (1), 94–118.
Ford, Cameron M. (1996). 'A theory of individual creative action in multiple social domains'. In: *Academy of Management Review* 21 (4), 1112–1142.
Friedman, Shlomit (2013). 'Priming Subconscious Goals'. In: *New Developments in Goal Setting and Task Perforamnce*. Ed. by Edwin A. Locke and Gary P. Latham. Paperback. New: Taylor & Francis. Chap. 33, 549–565.
Grant, Adam M. and James W. Berry (2011). 'the Necessity of Others Is the Mother of Invention: Intrinsic and Prosocial Motivations , Perspective Taking , and Creativity'. In: *The Academy of Management Journal* 54 (1), 73–96.
Greiner, Ben (2015). 'Subject pool recruitment procedures: organizing experiments with ORSEE'. In: *Journal of the Economic Science Association* 1 (1), 114–125.
Guo, Aifang, Jin Chen and Jun Jin (2010). 'An analysis of the complementary innovation mechanism between STI and DUI modes'. In: *International Journal of Learning and Intellectual Capital* 7 (3/4), 265.
Heidenreich, Martin (2009). 'Innovation patterns and location of European low-and medium-technology industries'. In: *Research Policy* 38 (3), 483–494.
Herstad, Sverre J., Tore Sandven and Bernd Ebersberger (2015). 'Recruitment, knowledge integration and modes of innovation'. In: *Research Policy* 44 (1), 138–153.

Hervas-Oliver, Jose-Luis, Francisca Sempere-Ripoll, Carles Boronat-Moll and Ronald Rojas (2015). 'Technological innovation without R&D: unfolding the extra gains of management innovations on technological performance'. In: *Technology Analysis & Strategic Management* 27 (1), 19–38.

Hippel, Eric von (1986). 'Lead Users : A Source of Novel'. In: *Management Science* 32 (7), 791–805.

Hirst, Giles, Daan van Knippenberg and Jing Zhou (2009). 'A Cross-Level Perspective on Employee Creativity: Goal Orientation , Team Learning Behavior , and Individual Creativity'. In: *The Academy of Management Journal* 52 (2), 280–293.

Hirst, Giles, Daan Van Knippenberg, Chin-Hui Chen and Claudia A. Sacramento (2011). 'How Does Bureaucracy Impact Individual Creativity? A Cross-Level Investigation of Team Contextual Influences on Goal'. In: *The Academy of Management Journal* 54 (3), 624–641.

Hoegl, Martin and K. Praveen Parboteeah (2003). 'Goal Setting And Team Performance In Innovative Projects'. In: *Small Group Research* 34 (1), 3–19.

Isaksen, Arne and James Karlsen (2010). 'Different modes of innovation and the challenge of connecting universities and industry: Case studies of two regional industries in Norway'. In: *European Planning Studies* 18 (12), 1993–2008.

Janssen, Onne and Nico W. Van Yperen (2004). 'Employees' Goal Orientations, the Quality of Leader-Member Exchange, and the Outcomes of Job Performance and Job Satisfaction'. In: *The Academy of Management Journal* 47 (3), 368–384.

Jenkins, G. Douglas Jr., Atul Mitra, Nina Gupta and Jason D. Shaw (1998). 'Are Financial Incentives Related to Performance? A Meta-Analytic Review of Empirical Research'. In: *Journal of Applied Psychology* 83 (5), 777–787.

Jensen, Morten Berg, Björn Johnson, Edward Lorenz and Bengt Åke Lundvall (2007). 'Forms of knowledge and modes of innovation'. In: *Research Policy* 36 (5), 680–693.

Kanfer, Ruth and Phillip Ackerman (1989). 'Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition.' In: *Journal of Applied Psychology* 74 (4), 657–690.

Kanfer, Ruth, Phillip L. Ackerman, Todd C. Murtha, Brad Dugdale and Leissa Nelson (1994). 'Goal Setting, Conditions of Practice, and Task Performance: A Resource Allocation Perspective'. In: *Journal of Applied Psychology* 79 (6), 826–835.

Kerr, Steven and Steffen Landauer (2004). 'Using stretch goals to promote organizational effectiveness and personal growth: General Electric and Goldman Sachs.' In: *Academy of Management Executive* 18 (4), 134–138.

Kerr, Steven and Douglas LePelley (2013). 'Stretch Goals: Risks, Possibilities, and Best Practice'. In: *New Developments in Goal Setting and Task Performance*. Ed. by Edwin A. Locke and Gary P. Latham. Paperback. New York: Routledge. Chap. 3, 21–31.

Kesting, Peter and John Parm Ulhøi (2010). 'Employee-driven innovation: Extending the license to foster innovation'. In: *Management Decision* 48 (1), 65–84.

Klein, Howard J., Michael J. Wesson, John R. Hollenbeck, Patrick M. Wright and Richard P. Deshon (2001). 'The assessment of goal commitment: A measurement model meta-analysis'. In: *Organizational Behavior and Human Decision Processes* 85 (1), 32–55.

Kozlowski, Steve W.J. and Bradford S. Bell (2006). 'Disentangling Achievement Orientation and Goal Setting: Effects on Self-Regulatory Processes'. In: *Journal of Applied Psychology* 91 (4), 900–916.

Kramer, William S., Amanda L. Thayer and Eduardo Salas (2013). 'Goal Setting in Teams'. In: *New Developments ind Goal Setting and Task Performance*. Ed. by Edwin A. Locke and Gary P. Latham. Paperback. New York: Routledge. Chap. 19, 287–310.

Latham, Gary P. (2004). 'The motivational benefits of goal-setting'. In: *Academy of Management Executive* 18 (4), 126–129.

Latham, Gary P. and Edwin A. Locke (2007). 'New developments in and directions for goal-setting research'. In: *European Psychologist* 12 (4), 290–300.

Latham, Gary P. and Lise M. Saari (1982). 'The Importance of Union Acceptance for Productivity Improvement Through Goal Setting'. In: *Personnel Psychology* 35 (4), 781–787.

Lee, Thomas W., Edwin A. Locke and Soo H. Phan (1997). 'Explaining the assigned goal-incentive interaction: The role of self-efficacy and personal goals'. In: *Journal of Management* 23 (4), 541–559.

Lee, You-Na (2015). 'Evaluating and extending innovation indicators for innovation policy'. In: *Research Evaluation* 24 (4), 471–488.

Lee, You-Na and John P. Walsh (2016). 'Inventing while you work: Knowledge, non-R&D learning and innovation'. In: *Research Policy* 45 (1), 345–359.

Levitt, Barbara and James G. March (1988). 'Organizational Learning'. In: *Annual Review of Sociology* 14 (1), 319–338.

Locke, Edwin A. (2004). 'Linking goals to monetary incentives.' In: *Academy of Management Executive* 18 (4), 130–133.

Locke, Edwin A. and Gary P. Latham (1990a). *A theory of motivation and task performance*. Ed. by Edwin A. Locke and Gary P. Latham. Englewood Cliffs, NJ: Prentice-Hall, Inc.

— (2002). 'Building a practically useful theory of goal setting and task motivation: A 35-year odyssey'. In: *American Psychologist* 57 (9), 705–717.

— (2006). 'New Directions in Goal-Setting Theory'. In: *Current Directions in Psychological Science* 15 (5), 265–268.

— (2013a). 'Goal Setting Theory, 1990'. In: *New Developments in Goal Setting and Task Performance*. Ed. by Edwin A. Locke and Gary P. Latham. Paperback. New York: Routledge. Chap. 1, 3–15.

— (2013b). 'Goal Setting Theory: The Current State'. In: *New Developments in Goal Setting and Task Performance*. Ed. by Edwin A. Locke and Gary P. Latham. Paperback. New York: Routledge. Chap. 37, 623–630.

— (2013c). *New Developments in Goal Setting and Task Performance*. Paperback. New York: Routledge.

— (1990b). *A theory of goal setting and task performance*. Ed. by Edwin A. Locke and Gary P. Latham. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Lundvall, Bengt-Åke (2016). *The Learning Economy and the Economics of Hope*. Electronic. New York: Anthem Press.

Lundvall, Bengt-Åke and Björn Johnson (1994). 'The learning economy'. In: *Journal of Industry Studies* 1 (2), 23–42.

Malerba, Franco (1992). 'Learning by Firms and Incremental Technical Change'. In: *The Economic Journal* 102 (413), 845–859.

March, James G. and Herbert Simon (1958). *Organizations*. New York: Wiley.

Masuda, Aline D., Edwin A. Locke and Kevin J. Williams (2015a). 'The effects of simultaneous learning and performance goals on performance: An inductive exploration'. In: *Journal of Cognitive Psychology* 27 (1), 37–52.

Masuda, Aline D., Edwin A. Locke and Kevin J. Williams (2015b). 'The effects of simultaneous learning and performance goals on performance: An inductive exploration'. In: *Journal of Cognitive Psychology* 27 (1), 37–52.

Mento, Anthony J., Robert P. Steel and Ronald J. Karren (1987). 'A meta-analytic study of the effects of goal setting on task performance: 1966–1984'. In: *Organizational Behavior and Human Decision Processes* 39 (1), 52–83.

Meub, Lukas and Till Proeger (2017). 'The impact of communication regimes and cognitive abilities on group rationality: Experimental evidence'. In: *Journal of Economic Behavior & Organization* 135, 229–238.

Miron-Spektor, Ella and Gerard Beenen (2015). 'Motivating creativity: The effects of sequential and simultaneous learning and performance achievement goals on product novelty and usefulness'. In: *Organizational Behavior and Human Decision Processes* 127, 53–65.

Miron-Spektor, Ella, Amy Ingram, Joshua Keller, Wendy K. Smith and Marianne W. Lewis (2018). 'Microfoundations of organizational paradox: The problem is how we think about the problem'. In: *Academy of Management Journal* 61 (1), 26–45.

Moilanen, Mikko, Stein Østbye and Kristin Woll (2014). 'Non-R&D SMEs: External knowledge, absorptive capacity and product innovation'. In: *Small Business Economics* 43 (2), 447–462.

Nelson, Richard R. and Sidney G. Winter (1982). *An Evolutionary Theory of Economic Growth*. The Belknap Press of Harvard University Press.

— (2002). 'Evolutionary Theorizing in Economics'. In: *The Journal of Economic Perspectives* 16 (2), 23–46.

Parrilli, Mario Davide and Henar Alcalde Heras (2016). 'STI and DUI innovation modes: Scientific-technological and context-specific nuances'. In: *Research Policy* 45 (4), 747–756.

Pavitt, Keith (1984). 'Sectoral patterns of technical change: towards a taxonomy and a theory'. In: *Research Policy* 13 (6), 343–373.

Poole, Marshall Scott and Jonelle Roth (1989). 'Decision development in small groups IV: A typology of group decision paths'. In: *Human Communication Research* 15 (3), 323–356.

Pruett, Mark and Howard Thomas (2008). 'Experience-based learning in innovation and production'. In: *R&D Management* 38 (2), 141–153.

Rammer, Christian, Dirk Czarnitzki and Alfred Spielkamp (2009). 'Innovation success of non-R&D-performers: substituting technology by management in SMEs'. In: *Small Business Economics* 33 (1), 35–58.

Rosenberg, Nathan (1982). *Inside the Black Box: Technology and Economics*. New York: Cambridge University Press.

Seijts, Gerard H. and Gary P. Latham (2001). 'The Effect of Distal Learning, Outcome, and Proximal Goals on a Moderately Complex Task'. In: *Journal of Organizational Behavior* 22 (3), 291–307.

Seijts, Gerard H., Gary P. Latham and Meredith Woodwark (2013). 'Learning Goals: A Qualitative and Quantitative Review'. In: *New Developments in Goal Setting and Task Performance*. Ed. by Edwin A. Locke and Gary P. Latham. Paperback. New York: Routledge. Chap. 13, 195–212.

Seijts, Gerard H, Gary P. Latham and Glen Whyte (2000). 'Effects of Self- and Group Efficacy on Group Performance in a Mixed-Motive Situation'. In: *Human Performance* 13 (3), 279–298.

Seijts, Gerard and Gary P. Latham (2005). 'Learning versus performance goals: When should each be used?' In: *Academy of Management Executive* 19 (1), 124–131.
Seijts, Gerard, Gary P. Latham, Kevin Tasa and Brandon W. Latham (2004). 'Goal Setting and Goal Orientation : An Integration of Two Different Yet Related Literatures'. In: *The Academy of Management Journal* 47 (2), 227–239.
Shalley, Christina E. and Gamze Koseoglu (2013). 'Goals and Creativity'. In: *New Developments in Goal Setting and Task Performance*. Ed. by Edwin A. Locke and Gary P. Latham. Paperback. New York: Routledge. Chap. 22, 343–356.
Shao, Yan, Bernard A. Nijstad and Susanne Täuber (2019). 'Creativity under workload pressure and integrative complexity: The double-edged sword of paradoxical leadership'. In: *Organizational Behavior and Human Decision Processes* 155, 7–19.
Simmons, Aneika L. and Run Ren (2009). 'The influence of goal orientation and risk on creativity'. In: *Creativity Research Journal* 21 (4), 400–408.
Smith, Vernon L. (1976). 'Experimental Economics: Induced Value Theory'. In: *American Economic Review* 66 (2), 274–279.
Thomä, Jörg (2017). 'DUI mode learning and barriers to innovation—A case from Germany'. In: *Research Policy* 46 (7), 1327–1339.
Tjosvold, Dean and Zi-You Yu (2004). 'Goal Interdependence and Applying Abilities for Team In-Role and Extra-Role Performance in China'. In: *Group Dynamics: Theory, Research, and Practice* 8 (2), 98–111.
Tödtling, Franz and Michaela Trippl (2005). 'One size fits all?: Towards a differentiated regional innovation policy approach'. In: *Research policy* 34 (8), 1203–1219.
VandeWalle, Don (1997). 'Development and Validation of a Work Domain Goal Orientation Instrument'. In: *Educational and Psychological Measurement* 57 (6), 995–1015.
Waldman, David A., Linda L Putnam, Ella Miron-Spektor and Donald Siegel (2019). 'The role of paradox theory in decision making and management research'. In: *Organizational Behavior and Human Decision Processes* 155, 1–6.
Winter, Sidney G. (1984). 'Schumpeterian competition in alternative technological regimes'. In: *Journal of Economic Behavior & Organization* 5 (3-4), 287–320.
— (2003). 'Understanding dynamic capabilities'. In: *Strategic Management Journal* 24 (10), 991–995.
Winters, Dawn C. and Gary P. Latham (1996). 'Effect of Learning Versus Ouctome Goals on a Simple Versus a Complex Task'. In: *Group & Organization Management* 21 (2), 236–250.
Wong, Alfred, Dean Tjosvold and Chunhong Liu (2009). 'Innovation by Teams in Shanghai, China: Cooperative Goals for Group Confidence and Persistence'. In: *British Journal of Management* 20 (2), 238–251.
Yaping, Gong, Jia-Chi Huang and Jiing-Lih Farh (2009). 'Employee Learning Orientation, Transformational Leadership, and Employee Creativity: The Mediating Role of Employee Creative Self-Efficacy'. In: *The Academy of Management Journal* 52 (4), 765–778.

Chapter 4

The Effect of Monetary Incentives and Self-Chosen Payment Schemes on Non-R&D Innovation

The Effect of Monetary Incentives and Self-Chosen Payment Schemes on Non-R&D Innovation

Elaine Horstmann and Kilian Bizer

Abstract

With growing recognition of non-R&D innovations in economic research, the demand for understanding and identifying key factors that encourage or inhibit these innovations is thriving. Firms that lack formal R&D structures often aim to compensate by relying on innovation management tools. Among others, monetary incentives and worker participation are two key elements used to motivate employees, increase performance and generate creative ideas. By applying a new real-effort task based on learning by doing, using, and interacting, as a structural approach to capture the knowledge creation processes behind non-R&D innovations, we experimentally examine the effects of monetary incentives and participation in the compensation decision process on non-R&D innovation. In a 2 (piece-rate vs. fixed payment) x 2 (voting before vs. voting after) design, we find a strong and mostly homogeneous preference for fixed payments. Neither the compensation scheme, nor the opportunity to democratically vote for an incentive structure, facilitates or inhibits innovative behavior. Additionally, we find no evidence for monetary incentives to crowd out creative ideas or innovation.

This work was financially supported by the German Federal Ministry of Education and Research, Grant Number 16IFI005.

1 Introduction

Innovating can be a tricky, strenuous, and exhausting business. But, it is a key factor for firm success and market survival. Consequently, many organizations are interested in creating framework conditions that motivate employees to endure in the face of continuous complexity and persist through the *innovation quest* (Adler and Chen 2011; Brück et al. 2021; Paolillo and Brown 1978; Speckbacher 2017). This study examines how different payment schemes in conjunction with employee self-selection by a democratic voting mechanism affect group innovation and performance.

Most research on innovation processes in economics has focused on innovations that emerge from or in the context of formalized R&D structures. However, a growing strand of economic literature acknowledges and addresses the importance of non-R&D innovations particularly for small and medium sized enterprises (Alhusen et al. 2021; Barge-Gil et al. 2011; Hervas-Oliver et al. 2012, 2011, 2015; Jensen et al. 2007; Lee and Walsh 2016). Several conceptions of behind-the-scenes processes for non-R&D innovation have been proposed, all of which center employees, their work environment, and several human resource as well as accounting management tools (Ellström 2010; Høyrup 2010; Kesting and Ulhøi 2010; Lundvall and Nielsen 2007). One of the most prominent concepts focuses on two different learning and knowledge creation modes behind R&D and non-R&D innovations, namely *learning by doing, using, interacting* and *learning by science, technology, and innovation* (Jensen et al. 2007). This conception creates a natural starting point to experimentally investigate the effect of specific innovation management tools on non-R&D innovations. We therefore apply a new experimental real-effort task based on learning by doing, using, interacting, to examine the effects of two innovation management tools – monetary incentives and worker participation – on non-R&D innovation.

Lee and Walsh (2016) offer a condensed overview of how differences in innovation outcomes of R&D and non-R&D work are based on the nature of knowledge. Knowledge environments differ in terms of R&D valuing abstract knowledge more than practical knowledge, while the routinized, less formalized regime of non-R&D refers to practical, on-the-job and online learning that comprises experimentation and trial-and-error processes, creating knowledge from direct experience and learning by cumulative experience (Brown and Duguid 1991; Jensen et al. 2007; Lundvall 2016; Pavitt 1984; Winter 1984). In contrast, R&D favors science-based innovation emerging from codified scientific and technological knowledge, high skilled academic workforce and patents (Jensen et al. 2007; Parrilli and Heras 2016).

In their seminal paper Jensen et al. (2007) name these opposite, but not mutually exclusive, knowledge creation modes of innovation: learning by doing, using, and interacting (DUI) and learning by science, technology, and innovation (STI). The former primarily describes the knowledge and learning environment of non-R&D innovation, whereas the latter one characterizes the main basis for R&D innovation. The interactive process of DUI-mode innovation includes three key channels of learning, occurring as a by-product of normal jobs especially outside of R&D (Lee and Walsh 2016). First, learning-by-doing, as introduced by Arrow (1962), is described as repeating the same manufacturing operation in order to learn by actually doing and repeating it, and therefore, steadily evolving over time. Thereby, experiential learning and steady development of the production process generates productivity growth (ibid.). Learning-by-doing is a company-internal process, which is related to production activity and daily learning in working (Argote and Guo 2016; Brown and Duguid 1991; Malerba 1992). Accumulating the gained-while-producing experience is essential for optimizing the production process and thus reducing costs or increasing output. In other words, learning-by-doing means, getting better at what you are doing by actually doing it (Lundvall and Johnson 1994). Second, learning-by-using is associated with the consumption or application of a respective good or service (Lundvall 2016; Lundvall and Johnson 1994). This is not restricted to final products, but also involves prototypes. Furthermore, using technologies, machines, equipment or the product/ service itself creates knowledge which is gained through a learning-by-using process (Amara et al. 2008; Rosenberg 1982). Learning-by-using occurs both inside and outside a company (Hippel 1986). Finally, learning-by-interacting is not only the most investigated component (Apanasovich et al. 2016), it is also an integral part of knowledge creation, as "almost all learning is interactive" (Lundvall 2016, p.177). Every form of (in-)formal interaction or communication, both inside the company with colleagues, or outside with external actors such as customers, suppliers or other agents along the firm's supply chain (Aslesen et al. 2012), holds the potential for information and knowledge transfer.

As STI is mainly characterized by formal learning within explicitly supportive structures, like R&D units or business models that evolve around innovation goals, it is by definition embedded in an innovation promoting framework. For DUI innovation, on the other hand, creating supportive structures within organizations is much harder. Past research suggest that, among others, innovation management tools can be a crucial element to compensate for the lack of R&D as a source of innovation (Hervas-Oliver et al. 2015; Rammer et al. 2009). Two examples for such tools are monetary incentive schemes and modern modes of employee participation e.g. through democratic choice.

It is broadly excepted that monetary incentives can be effective in furthering motivation, performance and effort. However, their impact can vary widely (Bonner and Sprinkle 2002; Camerer and Hogarth 1999; Eisenhardt 1988; Jenkins et al. 1998; Smith 1976). Monetary incentives mainly increase motivation, which increases effort and, in turn, leads to increasing performance, even though this relation is not constant per se and results can be mixed (e.g., Bonner and Sprinkle 2002). This exacerbates the problem of identifying fitting compensation schemes. What is more, going beyond effort and using monetary incentives to increase creativity or innovation output is even more ambitious (Byron and Khazanchi 2012; Kaplan and Norton 1996). In our experiment, we use two simple, but contrary compensation schemes: fixed payment and piece-rate payment (Cadsby et al. 2007; Lazear 1986). While the former is tied to the input-side of a production function, e.g. hours worked, piece-rate payment, as one form of pay for performance, is affiliated with the output-side. Under a fixed payment scheme, individuals receive a certain compensation irrespective of their performance, reducing pressure and thereby leaving them more capacities for the multistage process of creativity and innovation (Grabner 2014; Holmstrom and Milgrom 1991; Manso 2011). However, non-R&D and especially DUI-mode innovation emerge as a byproduct of normal, routinized jobs. Therefore, compensation schemes should not distract individuals from producing a proper amount of output while exploring new approaches (Amabile et al. 1996; Shalley et al. 2004). With a piece-rate system, individuals receive a clear incentive to focus on output. The downside is that outcome-oriented schemes might discourage employees from exploring new approaches, thus *only* motivating increases in effort at the detriment of creativity (Ederer and Manso 2013). Consequently, pay for performance is uncommon in departments requiring creativity, e.g. R&D-department, but more popular for routine tasks (Ederer and Manso 2013; Lambert et al. 1993; Sprinkle 2008).

Besides the specific kind of incentive scheme, firms are increasingly experimenting with new participatory mechanisms that involve the employees as stakeholders in determining organisation frameworks. While most remuneration systems are still being assigned in a top-down manner, the positive effects of giving employees more opportunities for self-management and self-determination with regard to long-run effort or performance have been acknowledge from research as well as industry (e.g., Franke et al. 2016; Harrison and Freeman 2004; Semler 1989, 2007; Sliwka 2001). Some companies and start-ups have stretched employees' authorities, experimenting with delegating or democratizing the decision of salary adjustments and job promotions (Brück et al. 2021; Jeworrek and Mertins 2019; Shaw 2021). Prior research on the consequences of delegating salary determination processes has found positive effects on motivation and creativity (Brück et al. 2021; Charness et al. 2012; Harrison and Freeman 2004; Mellizo et al. 2014). Self-determination theory argues (e.g., Gagné and Deci 2005) that self-set salaries strengthen an individual's intrinsic motivation through three basic psychological needs – autonomy, competence, and relatedness – which subsequently increases effort, performance and creative idea generation (Ryan and Deci 2000). Moreover, intrinsic motivation is expected to be an important driver for creativity and thus innovation (Amabile and Gitomer 1984; Bailyn 1985; Paolillo and Brown 1978).

Several studies document a positive effect of involving employee stakeholders by delegating compensation decisions. Examining the effect of self-set wages and randomly assigned wages for performing a complex real-effort laboratory task, Faillo and Piovanelli (2017) find an increase in performance when individuals can self-set their wages and an even larger effect, when these individuals are highly motivated as well. Even though highly motivated participants ask for lower wages on average. Mellizo et al. (2014) confirm that work performance can be sensitive to the remuneration selection process. If groups are enabled to vote for either a tournament or a revenues-sharing scheme, they enhance effort, irrespective of the compensation schemes implemented. Furthermore, delegating the compensation determination process – even to entirely self-set salaries – not only increases effort, but Charness et al. (2012) also find a positive impact for both workers and firms in a gift exchange game, indicating a positive welfare effect for both sides. Moreover, they find evidence that increasing effort is likely driven by the increment of responsibility, rather than positive reciprocity for delegation. Jeworrek and Mertins (2019) confirm the importance of responsibility for successfully implementing such compensation decision processes in a natural field experiment. Additionally, the enhancing effect of self-set salaries or self-sorting into the respective payment schemes has been shown to also be driven by a stronger homogeneity of self-selected individuals when compared to an assigned payment scheme (Eriksson et al. 2009). For example, there is evidence that more productive individuals more often opt for a performance-based compensation, whereas risk-averse individuals prefer a fixed payment scheme (Cadsby et al. 2007; Eriksson et al. 2009).

Finally, fitting remuneration systems need to account for the fact that extrinsic incentives, such as monetary rewards, have the potential to crowd out intrinsic motivation, which is a crucial part of creativity and innovation (Frey and Jegen 2001; Gneezy and Rustichini 2000). Brück et al. (2021) find an enhancing effect of self-set salaries on quantitative output, which does not come at the cost of creativity. Furthermore, they also confirm a positive effect of having an impact on the compensation determination process on the three basic psychological needs of intrinsic motivation. Moreover, Chen et al. (2012) tie the reward system directly to creative outcomes of groups. They find inter-group tournament payments to raise creative group solutions more than group piece-rate payments, with a between-group tournament based on creative performance being more convenient for enhancing creative performance in groups.

Building on the above literature, this study experimentally examines the impact of assigned or democratically voted compensation schemes on non-R&D, DUI-mode innovation. Groups of three are either assigned, or vote for one of two payment schemes: a pay-for-performance based piece rate system or a fixed payment. Voting happens either ex ante, or ex post after outcomes are generated. To depict performance and innovation, we exploit a new real-effort task that represents the three key components of innovation resulting from learning by doing, using, and interacting[1]. Results show strong preferences for a fixed payment scheme, irrespective of voting timing. Moreover, the majority of participants votes homogeneously, in line with the hypothesis that individuals prefer a fixed payment scheme if they are uncertain about the upcoming task. Neither the payment scheme, nor delegating the compensation decision to subjects, impacts innovative performance. We conclude that worker participation has neither a positive, nor a harmful effect for DUI-innovations. Finally, as innovative performance is independent from different assigned or self-selected payment scheme, we do not find a crowding out effect of monetary incentives on creative ideas and innovation.

2 Experimental Design

As a crucial part of empirical economics, experiments allow researchers to separate specific action mechanisms and causally quantify the effect of single interventions on pre-determined outcomes. Experiments allow to exclude or control confounding factors that naturally occur in data. However, in economic innovation research, experiments are still scarce (Brüggemann et al. 2016; Sørensen et al. 2010), especially in non-R&D innovation research. Thus, we develop a new real-effort task, comprising the key elements of innovations emerging from routine-like tasks with learning by doing, using, and interacting. For that purpose, we design an encryption task, where participants translate letter-strings into numbers. Transforming letters into numbers follows a certain encryption mechanisms, i.e. cipher, which represents the innovation. This setup creates a task-based objective measure of innovation. Besides, within our design, it also fulfills the two core criteria for innovation – usefulness and novelty – because knowing the cipher makes the translation process distinctly more efficient and the cipher is a novel task element for participants to detect (Berg 2014; Ford 1996; Miron-Spektor and Beenen 2015). According to the work context of non-R&D innovation, participants are neither assigned to find the cipher or any form of translation pattern, nor is it suggested in the instructions.

Matching participants into groups and providing them with a chat box ensures learning by interacting. The chat box is available during the entire translation task, enabling group members to share information, discuss task strategies, or suggest solutions. Since no specific task approach is provided in the instructions, participants have to experiment and rely on a simple trial-and-error process to discover the correct translations in the first place. Through immediate feedback after entering a number as a translation trial, participants learn by doing, i.e. simply by working on the task. Additionally, a history table summarizes

[1]All supplementary material can be obtained from our online repository: https://researchbox.org/495&PEER_REVIEW_passcode=PMOTYR. It contains the list of letter-strings, its related translation, screen-shots of task- and voting-screens, instructions, post-experimental questionnaire, examples for chat data coding and further results from the analysis excluding outliers.

a group's cumulative correct translations and allows participants to derive task knowledge, which enhances performance and benefits learning. Thus, providing feedback and a tool to accumulate functioning solutions, facilitates the creation and improvement of a group's search pattern. In the best case, they will discover the cipher, i.e. the innovation. Similar to learning in early design stages that involve working with steel and clay prototypes, every idea about the cipher can be applied, tested and reviewed. This is learning by using. Information gained while performing the task – learning by doing – are transformed to a proposed translation solution (equivalent to a clay prototype), which is applied in subsequent trials.

Thus, our comprehensive design translates all three key elements of non-R&D innovation as suggested by the DUI-literature into an experimental task (Brown and Duguid 1991; Jensen et al. 2007). Importantly, as participants are neither framed to be innovative, nor to find the cipher or innovation, it allows us to understand innovation as a byproduct of employees' normal jobs (Lee and Walsh 2016). Finally, this task provide quantifiable measures (e.g., being innovative, number of trials, letter-strings as output and time investments) while still allowing for high controllability of external factors like communication.

2.1 Experimental Task

To match the purpose of our research question, groups of three (random fixed matching) translate letter-strings consisting of three characters into integers for 30 minutes. Letters are randomly drawn from a pool of 13 letters, containing every second letter of the German alphabet (A, C, E, ..., U, W, Y). They have to be translated into integers from the interval [1;150]. The sequence of letter-strings is predetermined and fixed, but unknown to participants, to make results from different treatments and group fully comparable. Every group member has their own screen and the translations progress is independent of the other group members. If an individual inserts all three translation number correctly, they will automatically forwarded to the next letter-string in line. However, an intra-group chat box is available during the entire translation task, enabling group members to share information and solutions if they like to. Furthermore, a history table with all the letters and its corresponding correct translations in the respective, previous rounds is displayed. This enables participants to detect the pattern's characteristics over progressing rounds. Additionally, each individual is shown their accumulated number of correctly translated letter-strings, their and their group's number of trials in the previous round, and the sum of all group members translated letter-strings[2]. Finally, a timer indicates the remaining time. Over all treatments, groups are instructed with the group goal to do their best and translate, in sum, as many letter-strings as possible within the 30 minutes.

Encyrption mechanism

The translation of letters into numbers follows a specific pattern, which is composed by a within- and a between-variation (table 1). This pattern is called a cipher. The relationship between consecutive and alphabetically increasing letters is described by the between-variation, marked by the respective starting number x. This starting number x decreases by -11 from A to Y. Changes within each letter are illustrated by the within-variation and within each single letter the variation is $x \longrightarrow x + 17 \longrightarrow x_1 - 13 \longrightarrow x_2 - 4 \longrightarrow x_3 + 17 \longrightarrow \cdots \longrightarrow x_n - 4$, with the starting number x and x_i-4 = x. Therefore,

[2]For a screenshot please refer to the appendix.

the within-variation repeats itself, if the respective letter has to be translated for the fourth time.

Table 1 Encryption of letters

	x	x+17	x_1-13		x	x+17	x_1-13		x	x+17	x_1-13		x	x+17	x_1-13
A	133	150	137	I	89	106	93	O	56	73	60	U	23	40	27
C	122	139	126	K	78	95	82	Q	45	62	49	W	12	29	16
E	111	128	115	M	67	84	71	S	34	51	38	Y	1	18	5
G	100	117	104												

The starting number $x = x_i - 4$.

Using "A" as an example, the starting number for letter "A" is $x = 133$. With a between-variation of -11 from A to Y, "C" starts with $x = 133 - 11 = 122$ and so on. Applying the within-variation, for "A" we get $133 \longrightarrow 150 \longrightarrow 137 \longrightarrow 133 \longrightarrow 150 \longrightarrow \cdots \longrightarrow 133(forx_{n} - 4)$. In table (2), we see letter "E" and "W" re-start in the 6^{th} letter-string.

Table 2 First ten letter-strings and their translation

No.	Letter-string	Translation	No.	Letter-string	Translation
1	MWE	M=67, W=12, E=111	6	EWK	E=111, W=12, K=78
2	SIW	S=34, I=89, W=29	7	GSC	G=100, S=51, C=139
3	AOE	A=133, O=56, E=128	8	UAM	U=23, A=150, M=71
4	MWC	M=84, W=16, C=122	9	CWI	C=126, W=29, I=106
5	YQE	Y=1, Q=45, E=115	10	OGS	O=73, G=117, S=38

No. refers to the chronological order. The full set of all 40 letter-strings is uploaded in the online repository.

2.2 Treatments

With our four treatments, we examine the impact of two simple and common compensation schemes (salary and piece rate), in conjunction with worker participation regarding their payment, on innovative behavior (see table (3)). These two payment schemes compensate for different production factors, as wage is a compensation especially for the input factor, hours worked, whereas piece rate primarily focuses on outputs. In treatment T1 and T2, we assign either of these to groups. Group members in T1 will be compensated according to their output performance, i.e. piece rate. At that, participants receive 1 token per translated letter-string, with a conversion rate of 1 token = 1 €. An individual translating 8 letter-strings receives 8 token for the translation task. In T2, participants receive a fixed payment, i.e. salary, of 10 token, irrespective of the amount of letter-strings translated.

To implement worker participation, group members in T3 and T4 can vote for either one of these two payment schemes. The payment scheme that receives the majority of votes is implemented. In T3, voting proceeds after reading the translation task instructions and answering the control questions, but prior to the task itself, i.e. participants are informed about the translation task, but are uncertain about their performance. Every group member has one vote and can decide between a pay-for-performance (T3.1) or fixed payment (T3.2). After voting, group members are informed about the voting result and which payment scheme is implemented. Afterwards, the translation task starts. In T4, we conduct the same form for voting procedure, but after groups have finished the translation task, i.e. group members know their and their groups performance, making

Table 3 Treatments

Assigned payment	**Voting for payment**	*Sub-treatments: Self-selection to payment*
Pay for performance (T1): Participants receive 1 token per translated letter-string	*Voting* ***before*** *task (T3)*: Participants vote before the task, whether they would like to be paid by performance or equally	Pay for performance (T3.1) Fixed payment (T3.2)
Fixed Payment (T2): Every participant receives 10 token irrespective of performance	*Voting* ***after*** *task (T4)*: Participants vote after the task, whether they would like to be paid by performance or equally	Pay for performance (T4.1) Fixed payment (T4.2)

fixed payment (T4.2) the dominant strategy if the number of translated letter-strings are below ten and pay-for-performance (T4.1) otherwise. However, we inform participants about the downstream voting procedure as well as as the two payment schemes prior to the translation task.

2.3 Experimental Procedure

The experiment is programmed in oTree (Chen et al. 2016) and took place in the Laboratory for Behavioral Economics at the University of Goettingen in November and December 2019. We recruited 240 subjects using ORSEE (Greiner 2015), resulting in 80 groups, which are equally distributed among the four treatments. The average age is 24 years, 55.08% are female and 33.75% study at the faculty of business and economics. Besides, 29.17% attended a lecture, where economic experiments are part of the syllabus. Before reading the task instructions, participants completed the social value orientation slider measure task (SVOSM) designed by Murphy et al. (2011). We sampled all six primary and all nine secondary slider items.[3]

2.4 Measures

Innovation and learning. We use the number of trials per letter-string as a proxy for learning and a reduction of trials to one per letter-string as a proxy for detecting the cipher. Therefore, in our experiment, detecting the cipher corresponds to creating an innovation. If a participant uses only one trial for at least one letter-string translated, they are considered an innovator. Given the group structure in our design, we count groups with at least one innovator as innovative groups. Among all 240 participants, 46 (19.17%) from 26 groups (32.5%) are classified as innovators.

Output performance. The accumulated number of translated letter-strings within the 30 minutes translation task measures the output performance. Only letter-strings that have been fully translated within the time frame are counted. Furthermore, letter-strings are measured on individual level, but we calculate the within-group mean for group level analyses.

Voting and payment scheme. Treatments differ in terms of the voting option and the compensation scheme. In the analysis, the *voting* variable is 1 for treatments T3 and T4, i.e. participants have the option to vote, and 0 otherwise. Within voting treatments, we distinguish between a nem con vote and a 2/3-majority, where we also check for voting

[3]We share complementary information about the chat data, additional tables, tests and explanations in the online repository.

downs. The *compensation* variable is 1 for pay-for-performance, and 0 for a fixed payment. Finally, we combine both variables to analyse the self-selection into T3.1 and T3.2 or T4.1 and T4.2 respectively.

Effort. To analyze whether participants rely on a quick trial-and-error strategy or take more time to try to find out more efficient ways to master the task or innovate, i.e. find the ciper, we calculate an additional effort variable. Effort is defined as the number of seconds spent translating a letter-string, divided by the number of trials needed for said letter-string.

Questionnaire. Prior to the main translation task, participants complete the *social value orientation* slider measure task (SVOSM) (Murphy et al. 2011). The fist six self-other allocations help to sort social orientations as altruistic, pro-social, individualistic, and competitive. The remaining nine slider items disentangle joint maximization from inequality aversion within pro-social motives. We explore whether SVO attitudes predict behavior under the two voting schemes. The post-experimental questionnaire includes, among others, questions about *task effort* (Chronbach's $\alpha = 0.89$), *individual* and *group performance orientation* (IPO and GPO), cooperative and competitive behavior, adopted from Van Mierlo and Kleingeld (2010), as well as the motivational effect of participating in the compensation decision, and satisfaction with the implemented payment scheme. We calculate an additive, equally-weighted index for *cooperation & GPO* ($\alpha = 0.67$) and *competition & IPO* ($\alpha = 0.63$). To track the *motivational effect* in T1 and T2, we ask on a 5-point Likert-scale (1 = strongly disagree, 5 = strongly agree), whether an opportunity to vote on the compensation system would have increased effort and motivation in the translation task ($\alpha = 0.81$). Accordingly, we ask in T3 and T4, whether participating in the compensation decision motivated them, or if they had preferred not being involved. Finally, participants in T1 and T2 rate their *payment scheme satisfaction* by indicating how satisfied they are with their assigned compensation scheme, i.e. pay-for-performance (T1) or fixed payment (T2), and if they would have preferred a fixed payment (T1) or pay-for-performance (T2) ($\alpha = 0.50$). Participants in T3 and T4 rate their satisfaction with the implemented compensation scheme and if they would have preferred being payed according to a non-elected, i.e. assigned, compensation scheme ($\alpha = 0.86$).

Chat data. To make the chat data usable, we apply a modified version of the decision function coding system developed by Poole and Roth (1989). In a first step, we code the chat data based on five categories with 18 subcategories. We then condense them into four subcategories that determine our chat data analysis: insight sharing, history reference, task strategy and motivation. According to the independent coder method and in order to increase coding reliability, we compare the results from two coders for consistency. The authors of this paper made adjustments if needed. The modified coding scheme and the final categories used, including examples and descriptive statistics, are uploaded to the online repository.

3 Results

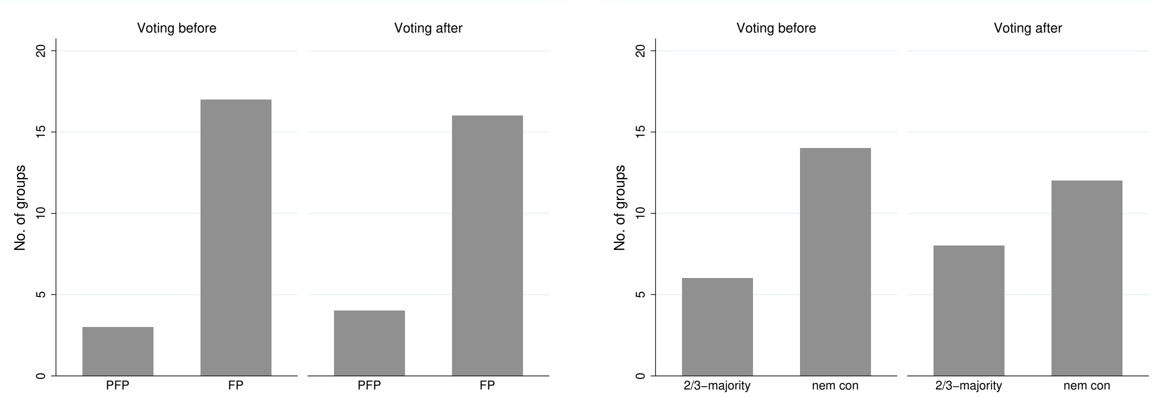
We first report results on voting behavior and specific SVO types. We then focus on the relationship between innovation, learning and our treatment interventions. After reviewing output and effort data, the section concludes with a chat data analysis to provide insights into the knowledge-generating process within the DUI-mode interaction-dimension.

3.1 Voting behavior and SVO types

To date, personality or character traits associated with different self-selection patterns regarding performance-dependent or -independent incentive schemes are largely unknown. What is more, there is little quantitative data on the distribution of respective employee preferences. According to our post-experimental questionnaire, opinions on the desirability of employee-driven salary provision are mixed.[4] However, participants who had the opportunity to vote for their payment scheme indicated a significantly higher task motivation due to voting ($\overline{x} = 3.58$) than participants who did not have the opportunity ($\overline{x} = 2.65, t(238) = -7.70; p < 0.000$[5]). Moreover, they were also more satisfied with their implemented payment scheme than participants ($\overline{x} = 4.25$) who were did not have a choice ($\overline{x} = 3.62, t(238) = 4.54; p < 0.000$).

In the voting treatments, participants have the opportunity to vote for their compensation scheme either prior (T3) or after (T4) the translation task. The left figure (1) shows the number of groups who implemented either a *pay-for-performance* (PFP) or a *fixed payment* (FP) scheme is. A significant majority of groups voted for fixed payment, irrespective of timing ($\chi^2(3) = 40.00; p = 0.000$). Additionally, participants who voted for fixed payment were more satisfied with their choice ($\overline{x} = 4.47$) than participants who voted for PFP ($\overline{x} = 3.21, t(118) = -5.00; p < 0.000$)[6]. Voting within groups was mostly homogeneous (nem con) as depicted in the right figure (1), only a minority of groups enforced the payment scheme with a 2/3-majority ($\chi^2(3) = 10.73; p = 0.013$).

Figure 1 Voting behavior: votes for payment scheme (left) and homogeneity of votes (right) by treatment



As social preferences as well as performance might impact the voting behavior, we use the SVO slider task to approximate social preferences. Table (4) presents the distribution of SVO types over treatments[7]. Overall, about 2/3 of participants are classified as *pro-social*, while 1/3 fall into the category of *individualism*. Due to the low number of cases, we disregard the two more extreme categories *altruism* and *competitiveness*. The left section

[4]Participants were asked if they would like to work for a company in which employees discuss and vote on each other's salaries or companies that hand over the responsibility of determining salaries to their employees ($\overline{x} = 3.03$ on a 5-point Likert scale).

[5]All p-values are from two-sided t-test.

[6]Participants in the assigned treatments are more satisfied with a fixed ($\overline{x} = 3.8$) instead of a pay-for-performance compensation, as well ($\overline{x} = 3.44, t(118) = -1.98; p = 0.050$).

[7]Table (4) displays column percentages. One observation T1 is missing, as the mean allocation for self is zero and the ratio of mean allocations between self and other cannot be calculated

of table (4) shows an almost uniform distribution with no significant differences among all four treatments ($\chi^2(9) = 6.14; p = 0.726$), indicating no bias for further analyses. A sub-analysis of *pro-social* participants reveals no significant distributional difference of both *inequality averse* and *joint gain maximizing* preferences among all treatments ($\chi^2(6) = 5.27; p = 0.509$) as well.

Table 4 Distribution of SVO types by treatments and voting behavior

	Treatment						Voting behavior				
SVO type	*PFP (T1)*	*FP (T2)*	*VB (T3)*	*VA (T4)*	**Total**	**SVO type**	*VB: PFP (T3.1)*	*VB: FP (T3.2)*	*VA: PFP (T4.1)*	*VA: FP (T4.2)*	**Total**
Altruism	0 0.00%	1 1.67%	0 0.00%	1 1.67%	2 0.84%	Altruism	0 0.00%	0 0.00%	1 7.14%	0 0.00%	1 0.83%
Prosociality	38 64.41%	38 63.33%	42 70.00%	36 60.00%	154 64.44%	Prosociality	5 55.56%	37 72.55%	7 50.00%	29 63.04%	78 65.00%
Individualism	20 33.90%	21 35.00%	18 30.00%	23 38.33%	82 34.31%	Individualism	4 44.44%	14 27.45%	6 42.86%	17 36.96%	41 34.17%
Competitiveness	1 1.69%	0 0.00%	0 0.00%	0 0.00%	1 0.42%	Competitiveness	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%
Total	59	60	60	60	239	Total	9	51	14	46	120

We first check for an impact of social preferences on voting behavior, especially if participants vote before the main task (see right section of table 4). In T3, the majority of participants voted for a fixed payment, with no significant difference among SVO types ($\chi^2(1) = 1.05; p = 0.305$). The same holds true for T4, even though results are less strong ($\chi^2(1) = 3.69; p = 0.158$). However, we expect voting behavior in T4 to be primarily determined by participants own performance. Based on the payment scheme structures, participants with more than ten translated letter-strings should vote for *pay-for-performance* compensation, those with fewer than ten should vote for a *fixed payment* and those with exactly ten should be indifferent. Five participants translated ten letter-strings (table (5)) and four of them voted for a fixed payment. In the other two groups, the majority voted according to the economic rationale. Most participants translated fewer than ten letter-strings and accordingly voted for fixed payment, which explains the large share of groups receiving a fixed payment in T4 (see figure 1).

Table 5 Economic rationale of voting behavior in T4

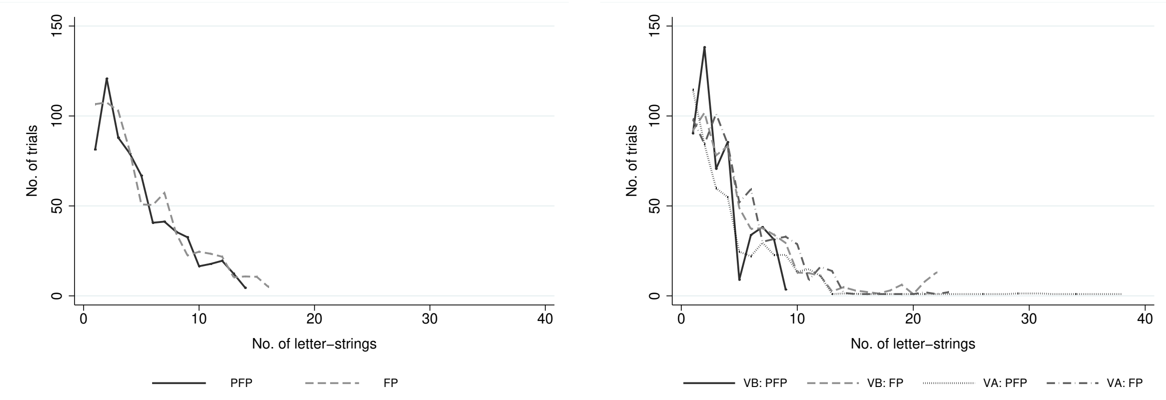
	No. participants	**Vote for PFP**	**Vote for FP**
<10 letter-strings	43	5	38
=10 letter-strings	5	1	4
>10 letter-strings	12	8	4
Total	60	14	46

Overall, SVO types are approximately uniformly distributed among treatments, but due to small and insignificant differences, they cannot explain voting behavior in T3. Voting behavior in T4 is mainly driven by task performance, as we find the majority voting in line with their performance output.

3.2 Innovation and learning performance

The number of trials per letter-string indicate innovation and learning performance. With increasing task experience and cumulative learning, participants need fewer trials per letter-strings, as shown in figure (2). The left section of this figure illustrates learning performance between *pay-for-performance* and *fixed payment* if payment schemes are assigned. Additionally, the right section depicts the four endogenous sub-treatments for our voting treatments. There are no noticeable differences.

Figure 2 Number of trials by assigned (left) and chosen (right) payment scheme



In line with figure (2), the number of trials in table (6) show no real group-level differences between treatments. In order to test for treatment effects, we run a linear regression of the group-level average number of trials on a categorical treatment variable excluding the intercept. The subsequent Wald-test accepts the hypothesis of all four treatments being equal, indicating no significant treatment effects (F(3,76)=1.62; p=0.192). Therefore, test on group-level reveal no significant effects for our main treatment interventions: the payment scheme ($WRS^8 : z = -1.27; p = 0.22$) and the voting opportunity ($WRS : z = 1.16; p = 0.16$). Furthermore, differences between single treatments are barely significant only between T1 and T4.1 ($WRS : z = 1.70; p = 0.09$), which is driven by two outlier groups.

Table 6 Average trials, number of letter-strings and effort per group by treatments

Variable	Treatment				Voting			
	PFP (1)	FP (2)	VB (3)	VA (4)	VB: PFP (3.1)	VB: FP (3.2)	VA: PFP (4.1)	VA: FP (4.2)
Average trials	231.04	270.62	224.20	209.40	217.28	225.42	138.88	227.03
	(70.69)	(98.36)	(93.77)	(101.79)	(103.56)	(95.34)	(111.22)	(94.83)
Number of letter-strings	7.15	6.58	7.17	9.08	6	7.37	16.83	7.14
	(2.70)	(3.72)	(4.98)	(7.28)	(3.00)	(5.29)	(13.26)	(3.46)
Effort	7.96	5.42	17.09	7.17	4.67	19.29	4.55	7.82
	(6.79)	(2.68)	(50.70)	(10.84)	(1.71)	(54.94)	(2.07)	(12.08)
Obs	20	20	20	20	3	17	4	16

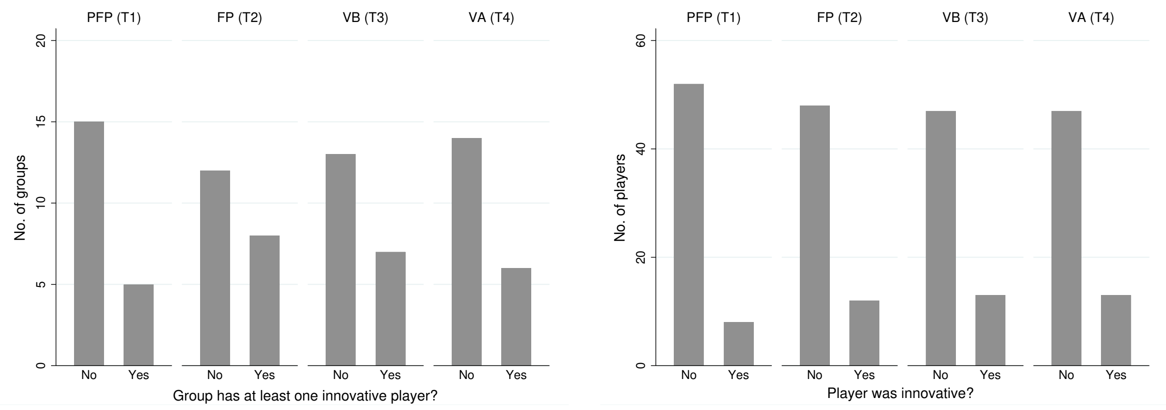
Standard deviation in parentheses.

We stick with the group structure for now and check whether the payment scheme or voting opportunity affect the number of groups with at least one innovative player. As shown in figure (3), only the minority of groups, and therefore players, are innovative, i.e. reducing their number of trials to only one at least once. Again, we find no significant treatment effect on both group ($\chi^2(3) = 1.14; p = 0.768$) and individual level ($\chi^2(3) = 1.83; p = 0.609$).

As our experimental design allows participants to work at their own pace and style, the chat box is the only design element constraining the independence of individual observations. Considering that, we run random effects GLS and logit regressions with standard errors clustered on group-level to analyse individual behavior, taking the group

[8]Unless specified otherwise, we run Wilcoxon-ranksum-tests on group-level.

Figure 3 Number of innovative groups (left) and innovative players (right)



structure into account. In table (7) we use the full sample. The first five columns report the coefficients for RE GLS regressions of the number of individual trials. If individual's trials are regressed on the treatment interventions only, we find participants in voting treatments to need significantly fewer trials and those with PFP to use significantly more trials (model 1). The effect of voting disappears with additional explanatory variables and the effect for compensation even reverses. The step-wise expansion of our regression model reveals no effect for voting, but participants who choose their payment scheme and self-select into *pay-for-performance* need significantly more trials, as shown by the interaction term. A partial explanation is the significant positive effect for innovative letter-strings, i.e. letter-strings translated with only one trial, as participants who self-select themselves into PFP translate significantly more innovative letter-strings. Controlling for the total number of letter-strings or regressing the number of individual trials on the voting sub-sample (see table (8), right section) dissolves this effect. This, again, confirms the lack of differences within voting treatments in table (6) and figure (2). Finally, participants of two groups in T 4.1 translated considerably more letter-strings than all other groups. Excluding these outlying observations from the regression (model 5) of table (7), we find no significant interaction term anymore. Therefore, we conclude that voting in combination with our two payment schemes has neither an effect on the number of trials, nor on innovative behavior.

Table 7 Random effects GLS and logit regression of the number of individual trials

Full sample	**No. of individual trials**					**No. of innovative letter-strings**				
	RE					*Logit*				
Variable	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Voting	-11.94*	-9.75	-7.00	-6.92	-6.54	0.92	-0.17	-0.57	-0.43	-0.48
Compensation	11.93*	-9.62	-8.25	-8.90*	-8.50*	0.86	-0.51	-0.25	-0.00	0.07
Voting*compensation		-7.13	15.92	16.85*	15.82*		2.87*	0.40	0.11	0.40
Letter-strings			-4.64***	-4.35***	-4.36***			0.21***	0.24***	0.22***
Effort				-0.51**	-0.51**				0.14***	0.14***
Cooperation & GPO					6.59**					-0.62
Task interest					-1.05					0.42
Constant	80.93***	79.58***	118.75***	120.62***	94.69***	-5.31***	-4.48***	-6.00***	-8.15***	-6.45**
Obs.	2001	1799	1799	1799	1799	1806	1806	1799	1799	1799

***$p < 0.01$,** $p < 0.05$,* $p < 0.10$. Standard errors are clustered on group level. *Voting*compensation* is 1 if the group is in T3 or T4 and pay-for-performance got the majority of votes.

Second, we find participants who are compensated based on their output to use significantly fewer trials (table (7), model 5), but this effect is driven by groups in T2 who have been assigned to it (see table (8), model 4). Also, the payment scheme has no effect on innovation attainment.

Accumulated experience and task knowledge, illustrated by the number of translated letter-strings, consistently has a significant effect on trials and innovation (see tables (7) and (8)). With accumulated experience, participants need fewer trials and are more likely to be innovative. The effect does not differ between sub-samples. The same holds true for effort, i.e. the number of seconds per trial. Increasing effort reduces the number of trials and raises the odds for innovative behavior. Both effects are stable, and hold even when excluding the two outlier groups in T4.1.

Table 8 Random effects GLS regression on the number of individual trials by voting

No. of individual trials									
Sub-samples	*Assigned payment*				*Voted payment*				
Variable	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(5)
Compensation	-8.88	-7.51	-9.01**	-8.35**	-16.02	-13.31	6.22	6.40	6.15
Over voted+						-14.13	-13.79**	-13.78**	-14.05**
Letter-strings		-7.01***	-6.23***	-6.37***			-4.01***	-3.83***	-3.79***
Effort			-1.18***	-1.15***				-0.35**	-0.35**
Cooperation & GPO				8.14**					2.61
Task interest				0.22					-1.38
Constant	78.93***	135.62***	138.97***	102.15***	70.31***	71.46***	107.97***	109.45***	102.59***
Obs.	927	824	824	824	1074	1074	975	975	975

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors are clustered on group level.

+ *over voted* is 1 if the player is over voted by its group members and 0 otherwise.

Finally, we use indicators for task interest as well as cooperation and group performance orientation from our post-experimental questionnaire. We find a significant and positive effect of cooperative behavior on the number of trials, indicating cooperating and helping comes at the cost of more time-consuming experimentation. The effect is significant only for treatments with assigned payment schemes. However, it does not impact innovative behavior. At last, being interest in the translation task does not translate into innovative behavior or input reduction.

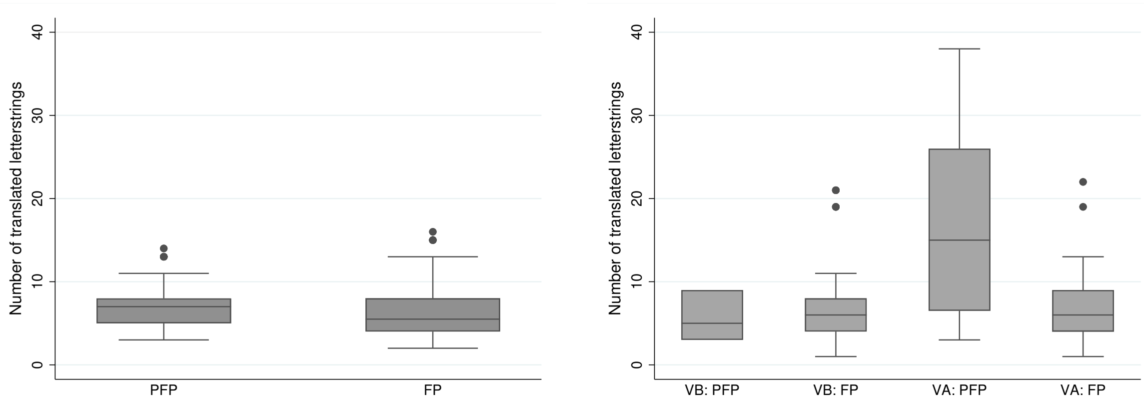
Due to the endogenous structure of payment schemes in voting treatments, we include a variable for over-voting. In table (8), participants who are getting over-voted use significantly fewer trials, possibly indicating some form of decreased motivation. However, the effect vanishes once the two outlier groups in T4.1 are excluded.

In sum, providing participants with a voting opportunity neither reduces trials, nor does it promote innovative behavior. Paying participants a fixed amount increases their input, i.e. trials, but uncoupling the payment from performance does not cause a detachment from output in favor of searching for innovative input solutions.

3.3 Output performance and effort

To measure output performance, we use the number of letter-strings individuals translate within the given 30 minutes. Figure (4, left section) shows no differences in the number of letter-strings between assigned payment schemes. Accordingly, the tobit regression in table (9) displays no significant effect for compensation. In the right section of figure (4), the number of letter-strings are outlined conditional on the sub-treatments among treatments with voting. Ex-post voting for *pay-for-performance* (T4.1) is a clear outlier, while there are no differences between the other three sub-treatments. However, the regression coefficient for interacting voting and compensation is insignificant (table (9)). The salient box plot difference hinges on the two outlier groups. All members of these groups were innovative players, thus translating significantly more letter-strings than non-innovative players (table (9)), showing the great utility of innovative behavior and explaining the distribution for T4.1 in figure (4).

Figure 4 Number of letter-strings assigned payment scheme (left) and chosen payment scheme (right)



Effort, measured in seconds per trials, can yield information about how participants approached the translations. If they need more time per trial, they might spend time on trying to find the cipher or at least trying to organize the translation task more efficiently. On the other hand, if they invest less time, they probably go with the quick and easy trial-and-error approach. In figure (5), we see a roughly increasing trend of effort, suggesting participants re-allocate their time to find the cipher, i.e. the innovation, or optimize their translation strategy. However, there are no differences between treatments, neither for those with assigned nor with self-chosen payment schemes. In line with figure (5), there are not significant treatment effects in the random effects GLS regression of effort in table (9). Again, participants, who are innovative invest more effort in translating letter-strings, implying spending more time on finding the innovation. Results are robust towards excluding the two outlier groups in T4.1.

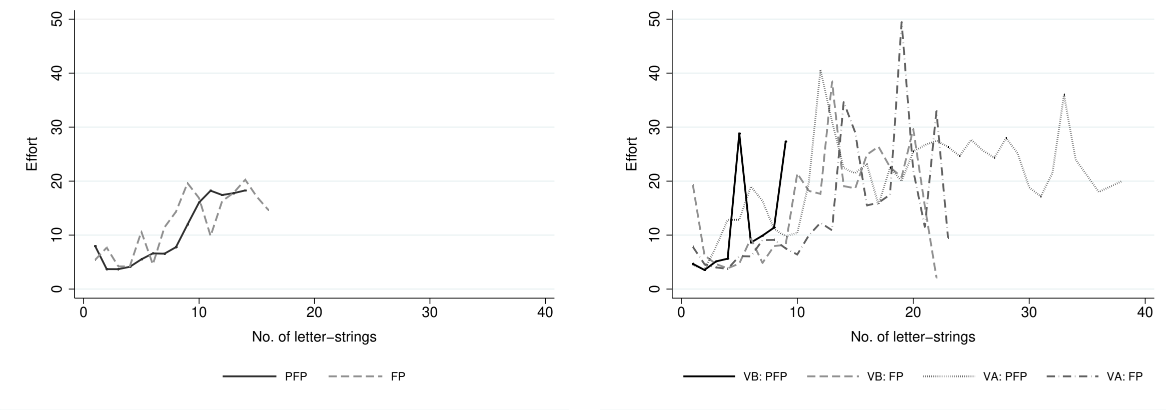
Table 9 Tobit regression on letter-strings and random effects GLS regression on effort

Full sample	**Letter-strings**					**Effort**					
	Tobit					*RE*					
Variable	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)	(6)
Voting	1.26			1.48	0.88	1.96			0.63	0.86	0.85
Compensation		1.45		1.79	1.03		-0.39		-0.53	-0.25	0.12
Innovative player			7.43***	7.24***	7.04***			15.29***	15.25***	15.33***	16.85***
Voting*compensation					2.08					-0.74	-0.02
Letter-strings											-0.20
Constant	6.87***	7.01***	6.07***	4.77***	5.17***	7.14***	8.25***	4.78***	4.66***	4.50***	5.31***
Obs.	240	240	240	240	240	2001	2001	2001	2001	2001	1799

$^{***}p < 0.01,^{**}p < 0.05,^{*}p < 0.10$. Standard errors are clustered on group level. *Tobit:* lower limit = 0 & upper limit = 40. *Innovative player* is 1 if player was innovative, otherwise 0.

Finally, we find no effect of our treatment interventions on output performance and effort. Not even paying participants based on their output performance enables them to translate more letter-strings or to increase effort to optimize or innovative their translation strategy in order to generate more output.

Figure 5 Effort with assigned payment scheme (left) and chosen payment scheme (right)



3.4 Chat data

Chat data allows us to gain insight into a group's innovation process and increase our understanding of how individuals and groups approach the task and work on innovations. Table (10) summarizes group-level descriptive statistics for all treatments and sub-treatments (Wald-test[9]: F(3,76)= 2.13, p=0.102). Most messages are sent when participants vote post-translation task, and the frequency of messages is almost the same among the other treatments. The significant difference between voting before and after the task (t(34)=-1.95, p=0.06) emerges from groups who vote for pay-for-performance ex post, including the two outlier groups[10]. Innovative groups, i.e. those with at least one innovative player, send on average 21.39 more massages than non-innovative groups (t(78)=-3.47, p<0.000). To find out what groups chat about, we apply a modified version of the decision function coding system by Poole and Roth (1989).

Table 10 Descriptive statistics of total messages sent in groups by treatments

Treatment	N	Mean	SD	Min	Max	Proportion
Pay-for-performance	20	60.3	24.04	29	124	23.12%
Fixed payment	20	62.55	22.49	22	98	23.98%
Voting before	20	59.7	27.02	12	126	22.89%
Voting after	20	78.3	33.09	39	188	30.02%
Voting before: PFP	3	46.67	15.14	36	64	2.68%
Voting before: FP	17	62	28.30	12	126	20.20%
Voting after: PFP	4	120	46.51	89	188	9.20%
Voting after: FP	16	67.87	19.36	39	116	20.82%

We cluster messages into categories to ascertain contentual differences between treatments. The four categories of interest are *insight sharing*, *history reference*, *task strategy* and *motivation*. Messages about the cipher and its characteristics (e.g., "the first number +17= second number") are assembled within the first category. When participants make statements that refer to the history table (e.g., "Now we have all numbers at least once. This means that we can now always solve this according to the same scheme."), these

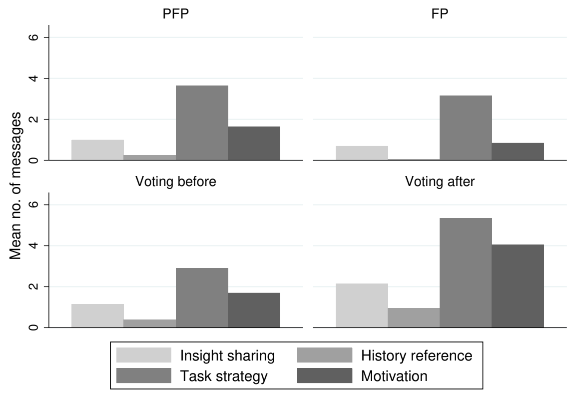
[9]All chat data tests are run on group-level.

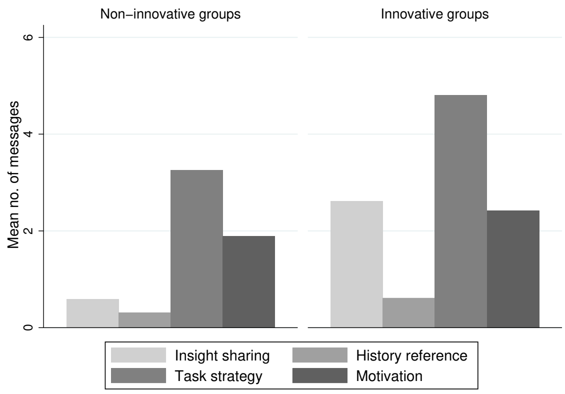
[10]When excluding the outliers significant differences vanish (t(36)=-1.48; p=0.147)

fall into the category *history reference*. This subcategory of sharing insights allows us to learn about the importance of the history table and if group members share how they got insights, or just the insight. General and specific task relevant statements about how to approach the task (e.g., "one does 1-50, another 50-100, another 100-150") are summarized in *task strategy*. Finally, *motivation* pools messages of team motivation or showing appreciation for each others input (e.g., "let's got" or "good hint"). The total number of messages within these categories by treatment are illustrated in figure (6).

Differences over all treatments and message types are significant, except for insight sharing (F(3,76)=1.54; p=0.210)[11]. According to figure (6), groups in voting treatments, and especially in T4, seem to share more insights about the cipher as groups in non-voting treatments do. However, the difference is not significant (t(78)=-1.58, p= 0.11). The other categories imply the same impression. If groups have the opportunity to vote, they refer more often to the history table (t(78)=-2.30, p<0.00) and support each other significantly more frequently (t(78)=-2.58, p=0.01). Differences in the number of messages containing task strategic information are not significant (t(78)=-1.00, p=0.32). Moreover, we find no significant differences within categories for different payment schemes.

Figure 6 Mean no. of messages by treatment (left) and by innovative groups (right)





The right bar chart compares the four message categories between innovative and non-innovative groups. Within every category, the mean number of messages is larger for innovative groups. Innovative groups send on average 2.02 more messages containing insight information than non innovative groups (t(78)=-4.05, p<0.00), but the do not refer significantly more often to the history table (t(78)=-1.54, p=0.13). Moreover, they discuss task strategies more frequently (t(78)=-2.05, p=0.04), but do not motivate each other more often (t(78)=-0.77, p=0.45).

Finally, as effort indicates how much time participants spend on every trial, we regress effort on our chat data categories to gain insight about how participants allocate their time. Table (11) replicates the results for our non-chat variable as in table (9), but all chat data categories, including the total number of messages sent by groups, have no impact on effort. Only innovative players allocate more time per trial, what probably allows them to discover the innovation in the first place.

In conclusion, chat data support results from our previous analyses. Innovative players approach the task differently than non-innovative players, as they discuss insights about the cipher more frequently, making their innovation knowledge transferable. Reviewing task

[11] Wald-test for respective categories: *history reference*: F(3,76)=5.09, p=0.003); *task strategy*: F(3,76)=2.47, p=0.068; *motivation*: F(3,76)=5.21, p=0.002.

Table 11 Random effects GLS regression of effort controlling for chat data

Full sample	Effort			
Variable	(1)	(2)	(3)	(4)
Voting	0.57	0.51	0.13	-0.19
Compensation	-0.52	-0.57	-0.99	-1.20
Innovative player	14.57***	14.54***	14.05***	14.08***
Messages		0.02	0.01	0.00
Insight sharing			0.33	0.21
History reference				-0.35
Task strategy				-0.04
Motivation				0.35
Constant	4.77***	4.43***	4.63***	4.73***
Obs.	2030	2030	2030	2030

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Standard errors are clustered on group level.

strategies more extensively potentially increases the efficacy of within-group collaboration, emphasizing the interaction part of DUI-mode innovation.

4 Conclusion

Monetary incentives are broadly recognized to affect motivation, effort, and performance, with a long-standing literature pointing to heterogeneous and context-depending effects (Bonner and Sprinkle 2002; Camerer and Hogarth 1999; Eisenhardt 1988). Thus, creating fitting compensation schemes to enhance any aspect of worker performance such as effort or creativity is essential, yet challenging (Byron and Khazanchi 2012; Kaplan and Norton 1996). Existing evidence suggests fixed payment systems to be advantageous for innovation, as lower performance pressure allows for more capacities to endure the complex, multistage process of creativity (Grabner 2014; Holmstrom and Milgrom 1991; Manso 2011). Some researchers even found performance-based reimbursement to reduce innovation (Ederer and Manso 2013).

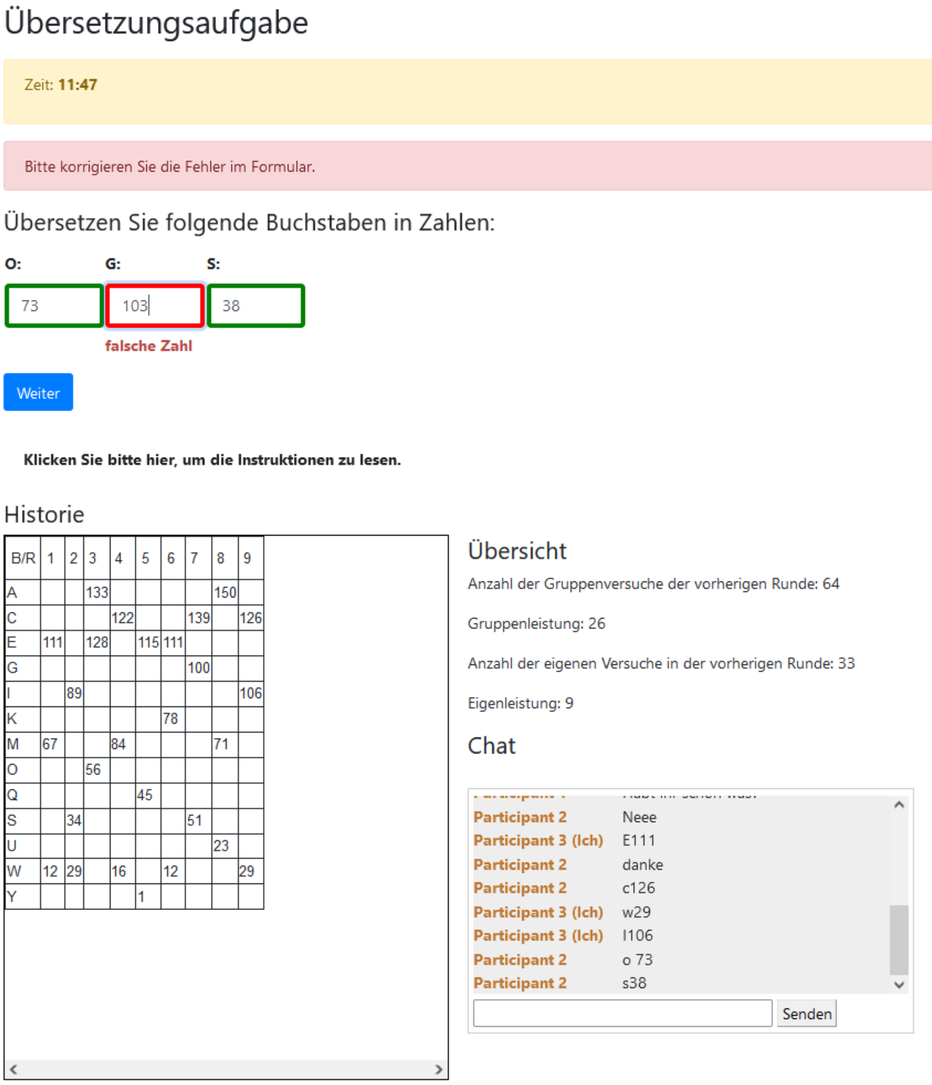
To avoid the possible detrimental effects of monetary incentives, researchers and practitioners have proposed to increase employee involvement in the compensation decision process. This could, for example, include self-set wages or self-chosen payment schemes (e.g., Brück et al. 2021; Charness et al. 2012; Jeworrek and Mertins 2019; Mellizo et al. 2014). Worker participation has been found to foster intrinsic motivation, which is expected to be a crucial prerequisite for creativity and innovation (Amabile and Gitomer 1984; Bailyn 1985; Gagné and Deci 2005; Paolillo and Brown 1978). Building on prior research, this paper examines the effects of two common monetary incentive schemes – piece-rate, as a form of pay for performance, and fixed rate –, in combination with worker participation via democratic votes on non-R&D innovation.

To do that, we exploit a laboratory real-effort task that represents the key elements of learning by doing, using and interacting (Jensen et al. 2007) as a form of non-R&D innovation. Results show that for non-R&D innovation, pay for performance has no significant innovation-reducing effect compared to a fixed payment scheme (c.f., Ederer and Manso 2013). Thus, for innovation elicited through routine-like tasks, monetary incentives appear to be less decisive.

Furthermore, these results are robust against employee self-selection through a democratic vote within work groups. Voting before the main task causes the majority to choose a fixed payment, which is independent of social preferences, but probably affected by ambiguity about the subsequent task. Voting behavior after the translation task is mostly in line with the economic rationale to maximize the payoff. There are no substantial differences in performance. Therefore, neither involving participants prior the task, nor holding out the prospect to decide about the compensation schemes after the task, affects innovative behavior. While innovative players outperform their counterparts in terms of input, i.e. trials per letter-string, output and effort, they are equally distributed among treatments. Chat data analysis also reveals differences only for innovative players vs. non-innovative players, as their communication strategy significantly differs based on frequencies for innovation knowledge transfer and reviewing task strategies.

Contributing to the ambivalent state of the current literature, we conclude that independent of the voting-timing, participants do not reciprocate for the compensation delegation. Due to the large share of voting for fixed compensation prior the task, responsibility, as discussed in Charness et al. (2012) and Jeworrek and Mertins (2019), might be insufficient for innovative behavior or increasing output performance, as the payment is decoupled from performance. As neither payment schemes nor self-selected compensation schemes are fostering or inhibiting non-R&D innovation, firms should allocate more resources to other innovation management tools such as goal setting, team work, innovation networks, or error tolerance.

Appendix A - Screenshot (original German Version)



Übersetzungsaufgabe

Zeit: **11:47**

Bitte korrigieren Sie die Fehler im Formular.

Übersetzen Sie folgende Buchstaben in Zahlen:

O:	G:	S:
73	103	38
	falsche Zahl	

Weiter

Klicken Sie bitte hier, um die Instruktionen zu lesen.

Historie

B/R	1	2	3	4	5	6	7	8	9
A			133					150	
C				122			139		126
E	111		128		115	111			
G							100		
I		89							106
K						78			
M	67			84				71	
O			56						
Q					45				
S		34					51		
U								23	
W	12	29		16		12			29
Y					1				

Übersicht

Anzahl der Gruppenversuche der vorherigen Runde: 64

Gruppenleistung: 26

Anzahl der eigenen Versuche in der vorherigen Runde: 33

Eigenleistung: 9

Chat

Participant 2	Neee
Participant 3 (Ich)	E111
Participant 2	danke
Participant 2	c126
Participant 3 (Ich)	w29
Participant 3 (Ich)	I106
Participant 2	o 73
Participant 2	s38

Senden

References

Adler, Paul S. and Clara Xiaoling Chen (2011). ‘Combining creativity and control: Understanding individual motivation in large-scale collaborative creativity’. In: *Accounting, Organizations and Society* 36 (2), 63–85.

Alhusen, Harm, Tatjana Bennat, Kilian Bizer, Uwe Cantner, Elaine Horstmann, Martin Kalthaus, Till Proeger, Rolf Sternberg and Stefan Töpfer (2021). ‘A new measurement conception for the ‘doing-using-interacting’mode of innovation’. In: *Research Policy* 50 (4), 104214.

Amabile, Teresa M., Regina Conti, Heather Coon, Jeffrey Lazenby and Michael Herron (1996). ‘Assessing the work environment for creativity’. In: *Academy of Management Journal* 39 (5), 1154–1184.

Amabile, Teresa M. and Judith Gitomer (1984). ‘Children’s artistic creativity: Effects of choice in task materials’. In: *Personality and Social Psychology Bulletin* 10 (2), 209–215.

Amara, Nabil, Réjean Landry, Nizar Becheikh and Mathieu Ouimet (2008). ‘Learning and novelty of innovation in established manufacturing SMEs’. In: *Technovation* 28 (7), 450–463.

Apanasovich, Natalja, Henar Alcalde Heras and Mario Davide Parrilli (2016). ‘The impact of business innovation modes on SME innovation performance in post-Soviet transition economies: The case of Belarus’. In: *Technovation* 57-58, 30–40.

Argote, Linda and Jerry M. Guo (2016). ‘Routines and transactive memory systems: Creating, coordinating, retaining, and transferring knowledge in organizations’. In: *Research in Organizational Behavior* 36, 65–84.

Arrow, Kenneth J. (1962). ‘The Economic Implications of Learning by Doing’. In: *The Review of Economic Studies* 29 (3), 155–173.

Aslesen, Heidi Wiig, Arne Isaksen and James Karlsen (2012). ‘Modes of Innovation and Differentiated Responses to Globalisation-A Case Study of Innovation Modes in the Agder Region, Norway’. In: *Journal of the Knowledge Economy* 3 (4), 389–405.

Bailyn, Lotte (1985). ‘Autonomy in the industrial R&D lab’. In: *Human Resource Management* 24 (2), 129–146.

Barge-Gil, Andres, Maria Jesus Nieto and Lluis Santamaria (2011). ‘Hidden innovators: The role of non-R&D activities’. In: *Technology Analysis & Strategic Management* 23 (4), 415–432.

Berg, Justin M. (2014). ‘The primal mark: How the beginning shapes the end in the development of creative ideas’. In: *Organizational Behavior and Human Decision Processes* 125 (1), 1–17.

Bonner, Sarah E. and Geoffrey B. Sprinkle (2002). ‘The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research’. In: *Accounting, Organizations and Society* 27 (4-5), 303–345.

Brown, John Seely and Paul Duguid (1991). ‘Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation’. In: *Organization Science* 2 (1), 40–57.

Brück, Christian, Thorsten Knauer, Harald Meier and Anja Schwering (2021). ‘Self-set salaries and creativity’. In: *Journal of Business Economics* 91 (1), 91–121.

Brüggemann, Julia, Paolo Crosetto, Lukas Meub and Kilian Bizer (2016). ‘Intellectual property rights hinder sequential innovation. Experimental evidence’. In: *Research Policy* 45 (10), 2054–2068.

Byron, Kris and Shalini Khazanchi (2012). 'Rewards and creative performance: a meta-analytic test of theoretically derived hypotheses'. In: *Psychological Bulletin* 138 (4), 809.

Cadsby, C. Bram, Song Fei and Francis Tapon (2007). 'Sorting and incentive effects of pay for performance: An experimental investigation'. In: *Academy of Management Journal* 50 (2), 387–405.

Camerer, Colin F. and Robin M. Hogarth (1999). 'The effects of financial incentives in experiments: A review and capital-labor-production framework'. In: *Journal of Risk and Uncertainty* 19 (1), 7–42.

Charness, Gary, Ramón Cobo-Reyes, Natalia Jiménez, Juan A. Lacomba and Francisco Lagos (2012). 'The hidden advantage of delegation: Pareto improvements in a gift exchange game'. In: *American Economic Review* 102 (5), 2358–79.

Chen, Clara Xiaoling, Michael G. Williamson and Flora H. Zhou (2012). 'Reward system design and group creativity: An experimental investigation'. In: *The Accounting Review* 87 (6), 1885–1911.

Chen, Daniel L., Martin Schonger and Chris Wickens (2016). 'oTree-An open-source platform for laboratory, online, and field experiments'. In: *Journal of Behavioral and Experimental Finance* 9, 88–97.

Ederer, Florian and Gustavo Manso (2013). 'Is pay for performance detrimental to innovation?' In: *Management Science* 59 (7), 1496–1513.

Eisenhardt, Kathleen M. (1988). 'Agency- and Institutional-Theory Explanations: The Case of Retail Sales Compensation'. In: *Academy of Management Journal* 31 (3), 488–511.

Ellström, Per Erik (2010). 'Practice-based innovation: A learning perspective'. In: *Journal of Workplace Learning* 22 (1-2), 27–40.

Eriksson, Tor, Sabrina Teyssier and Marie Claire Villeval (2009). 'Self-selection and the efficiency of tournaments'. In: *Economic Inquiry* 47 (3), 530–548.

Faillo, Marco and Costanza Piovanelli (2017). 'Wage delegation and intrinsic motivation: an experimental study'. In: *CEEL working papers* No. 1701.

Ford, Cameron M. (1996). 'A theory of individual creative action in multiple social domains'. In: *Academy of Management Review* 21 (4), 1112–1142.

Franke, Jörg, Ruslan Gurtoviy and Vanessa Mertins (2016). 'Workers' participation in wage setting: A gift-exchange experiment'. In: *Journal of Economic Psychology* 56, 151–162.

Frey, Bruno S. and Reto Jegen (2001). 'Motivation crowding theory'. In: *Journal of economic surveys* 15 (5), 589–611.

Gagné, Marylène and Edward L. Deci (2005). 'Self-determination theory and work motivation'. In: *Journal of Organizational Behavior* 26 (4), 331–362.

Gneezy, Uri and Aldo Rustichini (2000). 'Pay enough or don't pay at all'. In: *The Quarterly Journal of Economics* 115 (3), 791–810.

Grabner, Isabella (2014). 'Incentive system design in creativity-dependent firms'. In: *The Accounting Review* 89 (5), 1729–1750.

Greiner, Ben (2015). 'Subject pool recruitment procedures: organizing experiments with ORSEE'. In: *Journal of the Economic Science Association* 1 (1), 114–125.

Harrison, Jeffrey S. and R. Edward Freeman (2004). 'Is organizational democracy worth the effort?' In: *The Academy of Management Executive* 18 (3), 49–53.

Hervas-Oliver, Jose-Luis, José Albors-Garrigos and Juan-Jose Baixauli (2012). 'Beyond R&D activities: the determinants of firms' absorptive capacity explaining the access to scientific institutes in low-medium-tech contexts'. In: *Economics of Innovation and New Technology* 21 (1), 55–81.

Hervas-Oliver, Jose-Luis, Jose Albors Garrigos and Ignacio Gil-Pechuan (2011). 'Making sense of innovation by R&D and non-R&D innovators in low technology contexts: A forgotten lesson for policymakers'. In: *Technovation* 31 (9), 427–446.

Hervas-Oliver, Jose-Luis, Francisca Sempere-Ripoll, Carles Boronat-Moll and Ronald Rojas (2015). 'Technological innovation without R&D: unfolding the extra gains of management innovations on technological performance'. In: *Technology Analysis & Strategic Management* 27 (1), 19–38.

Hippel, Eric von (1986). 'Lead Users : A Source of Novel'. In: *Management Science* 32 (7), 791–805.

Holmstrom, Bengt and Paul Milgrom (1991). 'Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design'. In: *Journal of Law, Economics, & Organization* 7, 24–52.

Høyrup, Steen (2010). 'Employee-driven innovation and workplace learning: Basic concepts, approaches and themes'. In: *Transfer: European Review of Labour and Research* 16 (2), 143–154.

Jenkins, G. Douglas Jr., Atul Mitra, Nina Gupta and Jason D. Shaw (1998). 'Are financial incentives related to performance? A meta-analytic review of empirical research'. In: *Journal of Applied Psychology* 83 (5), 777.

Jensen, Morten Berg, Björn Johnson, Edward Lorenz and Bengt Åke Lundvall (2007). 'Forms of knowledge and modes of innovation'. In: *Research Policy* 36 (5), 680–693.

Jeworrek, Sabrina and Vanessa Mertins (2019). 'Wage delegation in the field'. In: *Journal of Economics & Management Strategy* 28 (4), 656–669.

Kaplan, Robert S. and David P. Norton (1996). 'Linking the balanced scorecard to strategy'. In: *California Management Review* 39 (1), 53–79.

Kesting, Peter and John Parm Ulhøi (2010). 'Employee-driven innovation: Extending the license to foster innovation'. In: *Management Decision* 48 (1), 65–84.

Lambert, Richard A., David F. Larcker and Keith Weigelt (1993). 'The structure of organizational incentives'. In: *Administrative Science Quarterly* 38 (3), 438–461.

Lazear, Edward P (1986). 'Salaries and Piece Rates'. In: *The Journal of Business* 59 (3), 405–431.

Lee, You-Na and John P. Walsh (2016). 'Inventing while you work: Knowledge, non-R&D learning and innovation'. In: *Research Policy* 45 (1), 345–359.

Lundvall, Bengt-Åke (2016). *The Learning Economy and the Economics of Hope*. Electronic. New York: Anthem Press.

Lundvall, Bengt-Åke and Björn Johnson (1994). 'The learning economy'. In: *Journal of Industry Studies* 1 (2), 23–42.

Lundvall, Bengt-Åke and Peter Nielsen (2007). 'Knowledge management and innovation performance'. In: *International Journal of Manpower* 28 (3-4), 207–223.

Malerba, Franco (1992). 'Learning by Firms and Incremental Technical Change'. In: *The Economic Journal* 102 (413), 845–859.

Manso, Gustavo (2011). 'Motivating innovation'. In: *The Journal of Finance* 66 (5), 1823–1860.

Mellizo, Philip, Jeffrey Carpenter and Peter Hans Matthews (2014). 'Workplace democracy in the lab'. In: *Industrial Relations Journal* 45 (4), 313–328.

Miron-Spektor, Ella and Gerard Beenen (2015). 'Motivating creativity: The effects of sequential and simultaneous learning and performance achievement goals on product

novelty and usefulness'. In: *Organizational Behavior and Human Decision Processes* 127, 53–65.
Murphy, Ryan O., Kurt A. Ackermann and Michel Handgraaf (2011). 'Measuring social value orientation'. In: *Judgment and Decision Making* 6 (8), 771–781.
Paolillo, Joseph G. and Warren B. Brown (1978). 'How organizational factors affect R&D innovation'. In: *Research management* 21 (2), 12–15.
Parrilli, Mario Davide and Henar Alcalde Heras (2016). 'STI and DUI innovation modes: Scientific-technological and context-specific nuances'. In: *Research Policy* 45 (4), 747–756.
Pavitt, Keith (1984). 'Sectoral patterns of technical change: towards a taxonomy and a theory'. In: *Research Policy* 13 (6), 343–373.
Poole, Marshall Scott and Jonelle Roth (1989). 'Decision development in small groups IV: A typology of group decision paths'. In: *Human Communication Research* 15 (3), 323–356.
Rammer, Christian, Dirk Czarnitzki and Alfred Spielkamp (2009). 'Innovation success of non-R&D-performers: substituting technology by management in SMEs'. In: *Small Business Economics* 33 (1), 35–58.
Rosenberg, Nathan (1982). *Inside the Black Box: Technology and Economics*. New York: Cambridge University Press.
Ryan, Richard M. and Edward L. Deci (2000). 'Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being'. In: *American Psychologist* 55 (1), 68.
Semler, Ricardo (1989). 'Managing without managers'. In: *Harvard Business Review* 67 (5), 76–84.
— (2007). 'Out of this world: Doing things the Semco way'. In: *Global Business and Organizational Excellence* 26 (5), 13–21.
Shalley, Christina E., Jing Zhou and Greg R. Oldham (2004). 'The effects of personal and contextual characteristics on creativity: Where should we go from here?' In: *Journal of Management* 30 (6), 933–958.
Shaw, Dougal (2021). 'The company where colleagues decide your salary'. In: *BBC News, May 5 2021*. Retrieved November 6, 2021 from https://www.bbc.com/news/business-56915767.
Sliwka, Dirk (2001). 'On the costs and benefits of delegation in organizations'. In: *Journal of Institutional and Theoretical Economics* 157 (4), 568–590.
Smith, Vernon L. (1976). 'Experimental economics: Induced value theory'. In: *The American Economic Review* 66 (2), 274–279.
Sørensen, Flemming, Jan Mattsson and Jon Sundbo (2010). 'Experimental methods in innovation research'. In: *Research Policy* 39 (3), 313–322.
Speckbacher, Gerhard (2017). 'Creativity research in management accounting: A commentary'. In: *Journal of Management Accounting Research* 29 (3), 49–54.
Sprinkle, Geoffrey B. (2008). 'Discussion of Measuring and motivating quantity, creativity, or both'. In: *Journal of Accounting Research* 46 (2), 375–382.
Van Mierlo, Heleen and Ad Kleingeld (2010). 'Goals, strategies, and group performance: Some limits of goal setting in groups'. In: *Small Group Research* 41 (5), 524–555.
Winter, Sidney G (1984). 'Schumpeterian competition in alternative technological regimes'. In: *Journal of Economic Behavior & Organization* 5 (3-4), 287–320.

Chapter 5

Experience Dominates Choice Bracketing in Risky Decisions under Incomplete Information

Experience Dominates Choice Bracketing in Risky Decisions under Incomplete Information

Alexander Erlei and Elaine Horstmann

Abstract

This article investigates the influence of choice bracketing on risk-behavior when the objective probabilities are unknown. Participants first learn about their environment using stochastic descriptions, sampled experience, or a combination, and then decide how to invest three points across two risky and one safe option. Depending on treatment, investments are structured sequentially or simultaneously. Results from five experiments show that broad choice bracketing leads to less risky and more diverse investments when subjects decide from descriptions, which dissolves once sampled experience becomes the primary learning mode. Sampling induces more risk-averse behavior through subjects disproportionally shunning high-risk high-reward options after experiencing a relatively bad mean. We also find evidence that people care more about satisfying a certain mean payoff after gathering information through experience. Finally, deciding against exploration decreases risk-taking.

This work was financially supported by the German Federal Ministry of Education and Research, Grant Number 16IFI005.

1 Introduction

The proper allocation of scarce resources in risky and uncertain environments lies at the core of many fundamental economic problems. To learn about the probabilistic consequences attached to each potential action, people rely on descriptive information, personal experience, or a combination. In principle, both modes can convey the exact same knowledge. However, years of behavioral research have documented a robust description-experience gap (Hertwig et al. 2004; Wulff et al. 2018), showing that decision-makers make different choices depending on the learning mode.

In this paper, we expand on the existing literature by embedding choice bracketing (Read et al. 1999) into experimental setups that differentiate between learning from stochastic descriptions and learning from experience. Subjects first gather information about their uncertain environment, and then choose between different risky options in a sequential or simultaneous manner, the latter reflecting the bracketed choice architecture of an ambidextrous decision unit (O'Reilly III and Tushman 2013). Rather than evaluating options one by one, their potential consequences are assessed together, creating new dependencies between them (Moher and Koehler 2010). Research on consumer behavior shows that this kind of choice bracketing can be a powerful determinant of human behavior, resulting in more diverse outcomes (Read and Loewenstein 1995), as well as increased expected value maximization when people decide from experience (Hadar et al. 2021) or descriptions (Webb and Shu 2017).

However, to this date, there has been no effort to analyze the effect of choice bracketing on risk-behavior conditional on the learning mode. Additionally, most research on bracketing as well as risky choices relies on known, objective probabilities (Hertwig and Wulff 2021). In reality, it is much more likely that decision-makers have some prior knowledge about their actions, while still dealing with uncertainty regarding the particular outcome distributions. This is true for both experience and descriptions. Examples reach from investment portfolios to resource allocations within organizations that affect e.g. investments in R&D.

This paper therefore analyzes the effect of choice bracketing when decision-makers face three initially uncertain options for which they receive incomplete information. We differentiate between a low-risk, a high-risk and a safe option, replicating a common investment dilemma. Over five studies, we gradually increase the importance of experience, documenting how broad choice bracketing initially leads to less risky and more diverse choices, which dissolves once subjects decide primarily from sampled experience. This effect does not appear to be driven by sampling bias or skewed subject beliefs. Rather, peoples preferences change towards securing higher mean payoffs. We also find evidence that people become generally more risk-averse after deciding not to engage in costly exploration.[1]

2 Related Literature

2.1 Description-Experience gap in risky choices

In real-world decisions, people broadly draw on two sources to learn about the risk associated with their decision alternatives. One, stated information where aggregate

[1]We provide the material, data and pre-registrations via an online repository: `https://osf.io/xfu5k/?view_only=debf4303f7fb4a8aa57c5c4bcc8e7abb`.

statistics are represented in some descriptive way, or two, personal experience. Imagine an investor pondering how to spend a certain budget. If the investor relies on descriptions, they might look at different investment fund reports and allocate money 7 according to the provided probabilities and outcomes. Alternatively, the investor could draw from personal experience and private information, investing for example into early-stage startups based on their impression of the people behind the project. Sometimes, the investor will have access to both learning modes, weighing the possibly diverging information signals. Throughout life, people make countless such economic or social choices, and the way individuals learn about risk crucially determines their success. Thus, understanding not only what drives human choices under risk, but also which interventions facilitate more or less risk-seeking behavior, can be a powerful asset in predicting people's decisions and constructing useful choice architectures.

A long-standing experimental literature shows that people tend to make systematically different choices conditional on whether they learn about their alternatives through description or experience, even when the transmitted information are virtually the same (Erev et al. 2010; Hertwig et al. 2004; Hertwig and Erev 2009; Ungemach et al. 2009; Wulff et al. 2018). In particular, individuals tend to overweigh rare events when learning from descriptions (Tversky and Kahneman 1992), but underweigh rare events when learning from experience (Erev and Barron 2005; Weber et al. 2004; Yechiam and Busemeyer 2006). Furthermore, experience through consequential (i.e. costly) sampling can lead risk-neutral agents to prematurely forego on average superior actions with a positively skewed or symmetric outcome distribution (Denrell 2007). In other words, experience tends to disfavor choices that are often underwhelming, but sometimes result in disproportionately high payoffs – a feature that characterizes many scenarios where individuals strive for excellence or innovation. Instead of striving for the maximum, people choose options that provide good outcomes most of the time (Erev and Haruvy 2016).

Of course, the reverse is also true. When an option provides a good outcome most of the time, with the rare risk of some bad or catastrophic result, decisions based on learning from experience are more robust to negative outliers. Decisions based on learning from descriptions will tend to overweigh the associated risks, e.g. in the case of vaccines or terrorism (Yechiam et al. 2005).

The reasons for this gap are subject to continuous debate. A lot evidence points to the fact that information gained through inconsequential experience exhibit sampling bias, which induces different probability estimates and, consequently, gaps in behavior (Camilleri and Newell 2011a). In some studies, these information asymmetries appear to fully explain choice differences (e.g. Aydogan and Gao 2020; Fox and Hadar 2006; Rakow et al. 2008), in others, sampling bias only attenuates the gap while differences of varying degrees still exist (e.g. Camilleri and Newell 2011b; Frey et al. 2015; Hau et al. 2010; Ungemach et al. 2009). Other research suggests that this relationship might be partially mediated by the degree of ambiguity or uncertainty, as decision-makers have less access to objective probabilities which leads to regression and thereby stronger weighting of rare events (Abdellaoui et al. 2011; Ert and Trautmann 2014). Glöckner et al. (2016) use cumulative prospect theory (CPT) to model choices from experience based on subjects' subjective probability estimates (as opposed to objective estimates). Their analysis suggests that people who learn from experience sometimes exhibit lower sensitivity to probabilities, i.e. *over-weigh* rare events compared to decisions from descriptions. In particular, removing certain alternatives leads to a reversal of the description-experience gap. Similarly, Kellen et al. (2016) document lower probability sensitivity and higher outcome sensitivity for decisions from experience, suggesting a reversed-S-shaped probability function with greater linearity in outcomes. Incorporating prior beliefs and Bayesian updating even dissolves the description-experience gap for choice environments that include safe options (Aydogan 2021). Thus, while outcome differences often show decisions from experience to under-weigh rare events, incorporating subjectivity in probability representation suggests that this is not necessarily because people are more sensitivity to probabilities. Contrary, there might even be contexts in which they are less sensitive, over-weighing rare events to a larger degree, thereby reversing the description-experience gap. The heterogeneity in experimental results, methods and interpretations highlights the complexity inherent to human learning and probabilistic inference, emphasizing the need to analyze behavioral interventions conditional on the method of information acquisition and processing. Furthermore, the vast majority of studies relies on binary two-outcome gambles or binary gambles that involve one risky and one safe option. Under decision environments that engender more complexity, analyses become more complicated and popular models of decision-making like CPT lose a lot of explanatory power (e.g. Birnbaum 2008). Given that learning from experience inevitably suffers from more noisy subjective probability representation, a larger number of decision alternatives with more heterogeneity in outcomes likely increases the importance of ambiguity attitudes and intuition. This, in turn, possibly accentuates the differential effect of the learning mode on risky choices. At the same time, it is a more accurate representation of many real-world decision environments. Thus, analyzing behavioral interventions and decision making in relation to the learning mode becomes even more relevant.

While a lot of the literature clearly distinguishes between the two modes of learning, many risky real-life choices fall between these two extremes. People have access to some descriptions, while simultaneously drawing on personal experience. The investor looking for a suitable start-up will not decide on gut feeling, personality assessment or learned heuristics alone. Instead, they combine such existing, possibly internalized knowledge with stated information like budget plans or revenue forecasts. The relative effect of both learning modes on decisions is not easy to predict, although many studies suggest experience to be the dominant factor in most situations where people have to combine and weigh information signals from the two sources (Erev et al. 2017; Jessup et al. 2008; Lejarraga and Gonzalez 2011; Newell et al. 2016). This appears to be moderated by the specific entrance point of descriptions. At the beginning of learning, information from descriptions have more influence, both because the information receive more weight and because they shape future choices, e.g. in subsequent learning from experience (Barron et al. 2008; Hertwig and Wulff 2021). Furthermore, there is evidence that the complexity of descriptions correlates negatively with their impact on decision-making (Lejarraga 2010). However, there still is much to learn about the interaction between the two learning modes, e.g. regarding the decision domain, the specificities of descriptions, the scope of experience, or the interplay with other elements of choice architecture.

2.2 Choice Bracketing

Choice Bracketing broadly refers to the way different options are presented to a decision-maker (Haisley et al. 2008; Moher and Koehler 2010; Read et al. 1999). It thus constitutes a feature of an individual's choice architecture. People can bracket their choices broadly or narrowly. An investor assessing their different options simultaneously by e.g. spreading 20 different reports across their desk presumably assess the consequences of all potentialities together. This is an example of broad bracketing. Contrary, the investor who goes from one investment to the next sequentially, assessing each option in isolation, follows narrow bracketing.

In their seminal paper, Read et al. (1999) propose that broad bracketing generally leads to higher utility and more expected value maximization, as people are able to integrate more information into their decision-making. For example, Brown et al. (2021) show how inducing people to think jointly about annuitization and retirement spending increases rational valuation. Broad bracketing also increases self-control and reduces cognitive effort required to simultaneously process and memorize different possibilities and their consequences, as the choice environment takes up some of the load (Koch and Nafziger 2019). On the other hand, broad bracketing can increase diversification (Read and Loewenstein 1995), which might hurt utility if one option dominates the other alternatives and more diverse investments or allocations therefore simply decrease output.

Webb and Shu (2017) confirm experimentally that broad decision frames affect risk-preferences and improve choices under risk by increasing the importance of potential losses. Haisley et al. (2008) find people to be more likely to buy lottery tickets when making several sequential rather than only one choice. Thus, broad bracketing decreased people's willingness to invest into high-risk high-reward options with a low expected value. Other research on myopic loss aversion is largely in line with these results, showing that people are usually more willing to tolerate risks with narrow bracketing and therefore a stronger focus on the short-term (Benartzi and Thaler 1999; Looney and Hardin 2009; Thaler et al. 1997), except for Hardin and Looney (2012), who find broad bracketing to increase risk-taking.

Given the behavioral consequences of choice bracketing, and recent estimates that a large majority of people does not engage in broad bracketing (Ellis and Freeman 2020; Rabin and Weizsäcker 2009), analyzing the effects of choice environments that encourage individuals to bracket their available risky options could prove highly valuable as a cost-effective measure to guide people's risk behavior. However, to date, there is little evidence how choice bracketing affects risk behavior under incomplete information and more than one risky option (see e.g. Birnbaum (2008) for a review on multiple outcome lotteries). Furthermore, it is likely that an individual's learning mode has the potential to mediate the effects of choice bracketing. The vast majority of research relies on descriptions, where subjects need to invest disproportionally less resources in order to learn about their decision alternatives. We only know about one study testing the robustness of choice bracketing against experienced-based decisions. Chaudhry et al. (2020) find that the effect of broad decision frames does not disappear when subjects make repeated choices with feedback in-between. However, their design brackets probability and outcome information, rather than the options themselves. Subjects also do not learn from experience, but experience by making multiple decisions based on cumulative descriptions. Thus, we still know very little about how experience as a learning mode affects choice bracketing, particularly for problem bracketing as opposed to outcome bracketing.

There are several reasons to suspect why the learning mode might affect the degree to which choice bracketing translates into risk-behavior. For instance, learning through sampled experience requires effort, which could increase people's attachment to the gathered information. If — as suggested by the literature — information gathered from experience loom larger than those received through descriptions (Meyer and Kunreuther 2017; Rakow et al. 2008), any resulting preferences might be much less malleable (Steinhart et al. 2013; Wachinger et al. 2013). Experience could shift people's attention to recently observed small samples, overriding prior aggregated information. This sampling bias would tend to disfavor heavy-tailed, high-risk option, leading to less risk-seeking behavior. Simultaneously, continuous sampling in environments that establish safe alternatives as a natural reference point should disfavor positively skewed distributions, again discouraging high-risk actions. In general, in so far as learned experience induces relative risk aversion (Lejarraga et al. 2016), it could also crowd-out the effect of broad choice bracketing by naturally leading to more diversification. On the other hand, if recent results on the over-weighing of rare events in learning from experience hold in gambles with three options (including a safe option) and a multitude of outcomes, experience should c.p. increase high-risk choices. In this case, broad choice bracketing could be especially influential in regulating the relatively low probability sensitivity of decisions from experience. For both scenarios, effects should be particularly pronounced when people choose between positively skewed and symmetric (or negatively skewed) risky options. Intriguingly, this setup describes very common investment and innovation contexts, where individuals and organizations decide whether they prefer relatively small, but safe gains, or are willing to incur potentially high losses for big payoffs. We replicate this exact scenario across five experiments.

3 Experiment 1 (E1): Description

The experiments were conducted on Amazon Mechanical Turk (MTurk) using oTree (Chen et al. 2016). We restricted the sample to workers who successfully completed at least 100 HITs with a minimum approval rate of 90. Throughout all five experiments, we asked subjects to allocate three points over three options: low-risk, high-risk, safe. The three options never changed. The low-risk option mimes a normal distribution drawn from a relatively narrow interval ($E(x_1) = 6.5, \sigma^2 = 3.25$). The high-risk option resembles a heavy-tailed distribution drawn from a relatively broad interval ($E(x_2) = 7, \sigma^2 = 145.7$).[2] The third, safe option always returns a value of six. Experiments 1, 3, 3b and 4 were preregistered.

3.1 Procedure and Participants

Subjects first indicated their current device and completed an attention check. Those who failed the attention check, or did not use a PC/laptop, were not allowed to participate in the experiment. The same protocol applied to Experiments 2 – 5.

In Experiment 1 (*Description*), subjects then read through the instructions. They learned that their task was to allocate three points across three options. For each point allocated, a coin would be drawn from the respective option. To learn about the contents of the three options, they would first be able to observe summary statistics from 100 draws per option. After completing four comprehension questions, subjects proceeded to said summary screen. If a participant failed to answer all questions correctly within two trials, they were dropped from the experiment.

Summary statistics were fixed for all subjects and included an options mean ($\bar{x}_1 = 6.51$; $\bar{x}_2 = 5.79$; $\bar{x}_{safe} = 6$), standard deviation ($\sigma_{x_1} = 1.79$; $\sigma_{x_2} = 10.53$; $\sigma_{x_{safe}} = 0$), lowest draw ($x_1 = 3$, $x_2 = 0$, $x_{safe} = 6$) and highest draw ($x_1 = 10$, $x_2 = 40$, $x_{safe} = 6$). Thus, in the pure description paradigm, the fact that the riskiest option returned the highest mean on

[2] Option x_1 contained eight coins (3, 4, 5, 6, 7, 8, 9, 10) with corresponding probabilities of (0.05, 0.1, 0.15, 0.2, 0.2, 0.15, 0.1, 0.05). Option x_2 contained eight coins (0, 1, 2, 3, 20, 25, 35, 40) with corresponding probabilities of (0.25, 0.25, 0.15, 0.15, 0.05, 0.05, 0.05, 0.05).

average was obscured. We did this for three main reasons. One, given what we know about decisions from descriptions, an option that has the highest upside as well as the highest mean could well have dominated the other two options. Two, we expected potentially weaker effects of choice bracketing on risky choices when the high-risk option maximizes average returns, since broad bracketing has been repeatedly associated with expected-value maximizing behavior. Three, relatively small samples naturally bias the experienced mean return of options with positively skewed outcome distributions downwards. By providing a smaller descriptive mean, we aimed to increase comparability between descriptions any experience. To make sure that subjects understood the information, we provided explanations for all four indicators on the top of the page.

After observing the summary statistics, subjects in *Description* proceeded to the main decision and used three points to freely draw from the options (presented as money bags). Here lies the main intervention of this paper: subjects in *sequential* made three consecutive decisions without intermediary feedback, each time investing one point into one option. Subjects in *simultaneous* made one decision on how to allocate all three points.

The experiment concluded with a questionnaire that included a battery of demographic questions, the Elaboration on Potential Outcomes (EPO) inventory (Nenkov et al. 2008), and a question capturing risk attitudes (Dohmen et al. 2011). Finally, subjects were informed about their draws and paid accordingly. We used ECU during the experiment, where 1 ECU equaled 10 cents. Subjects also earned a fixed reward of \$0.5. We gathered data until 125 independent observations per treatment, resulting in a final sample of 250 (43.6% female) participants.

3.2 Results Description

Figure 1 compares the average allocation of points over the three options between *simultaneous* and *sequential*.

Subjects who made three consecutive decisions were significantly more likely to choose the heavy-tailed high-risk option with the ostensibly lower mean ($t(248) = 2.13$, $p = 0.03$, $d = 0.27$). This was primarily because subjects in *simultaneous* invested more diversely. An ordered logistic regression over a categorical diversification variable ($1 =$ no, $3 =$ full diversification) revealed a significant treatment dummy (Odds Ratio (OR): $= 0.56$, $p = 0.016$). Whereas almost 50% in *sequential* chose to allocate all three points into one option (65% high-risk), only 36% did so in *simultaneous* (51% high-risk). Ordered logistic regressions for all three options further revealed a significant effect of self-reported risk-attitudes on choosing the high-risk option. There are no gender differences (see Table 1). The results are consistent with prior evidence showing that broad bracketing can increase expected value maximization, or alternately discipline risk-seeking behavior.

Result 1
When making decisions from stochastic descriptions, subjects who decide simultaneously and thus bracket their choices broadly invest less in a high-risk high-reward option.

Result 2
Most subjects exhibit risk-seeking behavior, only a minority opts for the option that guarantees a safe return.

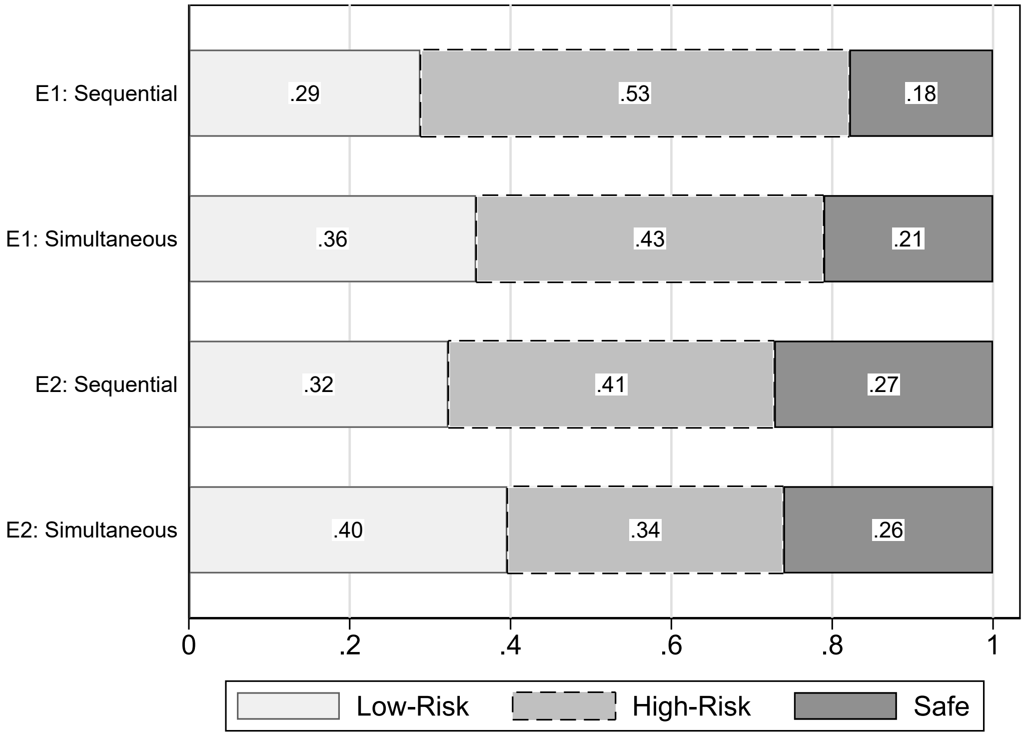


Figure 1 Mean share of points invested into the three options for Experiment 1 (E1) and Experiment 2 (E2).

Table 1 Ordered Logistic Regressions on the share of points (0 – 3) invested into each of the three respective investment options.

	Low-Risk		High-Risk		Safe	
	OR	*SD*	*OR*	*SD*	*OR*	*SD*
sequential	0.59**	0.14	1.71**	0.39	0.72	0.19
risk	0.92	0.05	1.10**	0.05	1.00	0.05
EPO (evaluation)	1.05	0.14	0.82	0.11	1.00	0.15
EPO (positive)	1.02	0.12	1.13	0.13	0.90	0.11
EPO (negative)	0.99	0.09	0.97	0.09	1.12	0.11
female	0.86	0.21	1.22	0.29	0.99	0.26
N		250		250		250

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4 Experiment 2 (E2): Voluntary Sampling

4.1 Procedure and Participants

Experiment 2 (*Voluntary Sampling*) allowed subjects to decide whether they wanted to additionally sample through all three options. Apart from that, it was analogous to E1. Participants chose whether to sample via a button after the summary screen. Sampling was costly, with an initial fee of 0.2 ECU that increased by the same amount after every fifteenth point allocated. Mirroring real-life opportunity costs, the primary intention was to induce a reflective sampling decision, and consequently re-affirm the effect of choice bracketing on risky choices when the decision to solely rely on descriptions is voluntary. Feedback was provided after every round. Subjects autonomously chose when to terminate sampling and proceed to the main investment decision. Allowing subjects to self-select into sampling provided us with two major advantages. First, subjects who opted against sampling offered a quasi-control group for *Description* to analyze whether deciding against costly sampling, i.e. learning from experience, affects risk-behavior when learning from descriptions. And, consequently, if the effect of choice bracketing still holds. This scenario is analogous to the investor foregoing their own search and evaluation efforts in favor of solely relying on fund brochures. Second, we get a first approximation of how learning from experience might affect the influence of choice bracketing on risky choices. However, since sampling was costly, and participants received statistical information from 100 draws per option, we did not expect a lot of sampling, severely inhibiting our ability to draw conclusive inferences.

Like in E1, we used ECU, where 1 ECU equaled 15 cents. Subjects earned a fixed reward of \$0.5. We gathered data until 125 independent observations per treatment, resulting in a final sample of 258 (46.12% female) participants. Subjects who failed to answer four comprehension questions twice were not allowed to participate in the experiment.

4.2 Results Voluntary Sampling

First, we find that general allocation patterns closely resemble those from *Description* (see Figure 1). Across all observations, subjects in *sequential* were more likely to choose the high-risk option. However, overall shares are significantly lower, with a 13 percentage point drop between the *sequential* treatments (t(248) = 2.47, p = 0.01, d = 0.31) and a 9 percentage point drop for *simultaneous* (t(256) = 1.95, p = 0.05, d = 0.24).

Differentiating subjects conditional on whether they decided to sample (N = 91) reveals that subjects in *Voluntary Sampling* who purely learned from descriptions (N = 167) invested significantly less into the high-risk option than subjects in *Description* (*sequential*: 0.53 vs. 0.41, t(210) = -2.24, p = 0.03, d = 0.3; *simultaneous*: 0.43 vs. 0.31, t(203) = -2.32, p = 0.02, d = 0.33). Thus, rejecting the opportunity to sample and therefore explore an uncertain environment appears to reduce risk-taking. For this subset, we also replicate the negative broad bracketing effect on choosing the high-risk option (0.41 vs. 0.31; t(165) = 1.66, p = 0.09, d = 0.26).

As expected, most subjects opted against sampling, and those who sampled drew on average 16.35 and 7.52 times in *simultaneous* (N = 53) and *sequential* (N = 38) respectively. Thus, learning from experience was very limited, with presumably at best very small as well as treatment-dependent effects. In both conditions, the low-risk (*simultaneous*: 6.2 draws; *sequential*: 2.8 draws) and the high-risk option (*simultaneous*: 7 draws; *sequential*: 2.9 draws) dominated sampling choices. Looking at subjects' final investment decisions, we do not document a choice bracketing effect (0.41 vs. 0.39). For both decision structures, sampling tended to increase the share of points allocated to the low-risk option while decreasing the share used for the safe option. However, differences are small, possibly subject to selection bias, and due to the small samples and little sampling, we refer to Experiments 3 – 5 for an in-depth analysis of experience learning and choice bracketing.

Overall, these results support the claim from E1 *Description* that broad bracketing decreases investments into high-risk actions and further suggest that individuals who reject the possibility of sampling become less willing to use high-risk options. Instead, they opt for safe and less volatile choices. While this effect could be driven by a selection effect, we find no quantifiable differences between subjects who sampled and subjects who did not sample. In particular, there are no differences in stated risk preferences ($t(254) = 0.53$, $p = 0.59$), the EPO inventory ($t(254) = 0.18$, $p = 0.86$) or answer length in an open-ended post-experimental question ($t(254) = 1.35$, $p = 0.18$) as a proxy for effort. Regressing the share of points invested into the high-risk option on our treatment dummy, an experiment dummy, risk attitudes and the EPO inventory using the subsample that decided against sampling reveals a strong negative effect of *Voluntary Sampling* on investments into the high-risk option (OR: 0.56, $p = 0.002$). For the minority that sampled, differences for the high-risk option disappear, indicating that learning from experience might erode the effect of choice bracketing. We will test this intuition in Experiments 3, 4 and 5.

Result 3
Deciding against exploration decreases subjects' willingness to invest in a high-risk option.

5 Experiment 3a (E3a): Forced Sampling

5.1 Procedure and Participants

In Experiment 3 (*Forced Sampling*), subjects first made an incentivized allocation decision analogous to *Description* based on summary information from 20 draws. This smaller sampled ensured that subjects received more information signals through sampling than from descriptions. Provided information closely replicated the values from the first two experiments. They then sampled through the options 100 times by freely clicking on them one by one, receiving feedback on their draw every round. Finally, subjects made another incentivized allocation decision, analogous to the first one. We disclosed all draws and payoffs only at the end of the experiment. Subjects in this setup acquired information for the second investment decision primarily through their own sampling, which also increased the likelihood of experiencing the "true", higher mean for the high-risk option. Hence, we expected choice bracketing to be less impactful.

We used ECU, where 1 ECU equaled 5 cents. Subjects earned a fixed reward of \$0.5. We gathered data until 125 independent observations per treatment, resulting in a final sample of 250 (50.8% female) participants. Subjects who failed to answer five comprehension questions twice were not allowed to participate in the experiment.

5.2 Results Forced Sampling

For the first decision based on summary statistics from 20 draws, results are directionally similar to the first two experiments, but not significant (see Figure 2).

Overall, investments are more equally distributed, which could be influenced (1) by the increased uncertainty due to the 20-draw sample and (2) different risk preferences induced by the availability of a second decision. After 100 sampling rounds with participants choosing the low-risk (high-risk) option on average 44 (39) times (no treatment differences), subjects in both conditions invested significantly more into the low-risk option (*sequential*: $t(248) = 3.09$, $p = 0.00$, $d = 0.39$; *simultaneous*: $t(248) = 2.57$, $p = 0.01$, $d = 0.32$). In

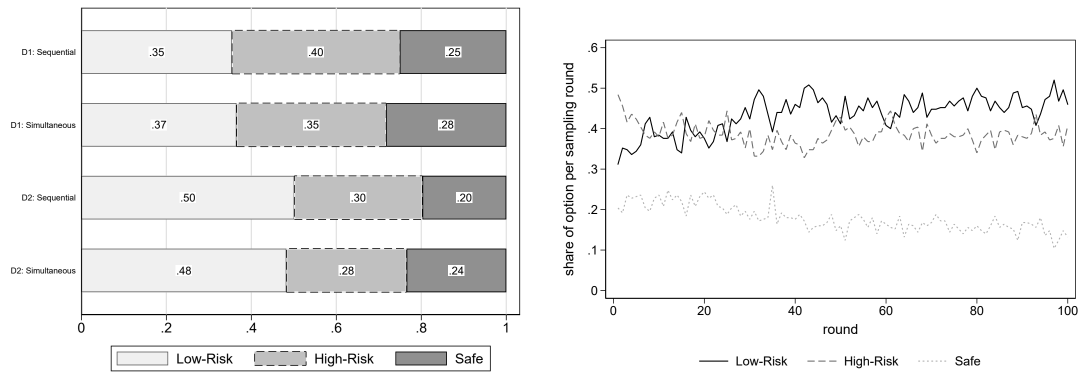


Figure 2 Left: Mean share of points invested into each of the three options during the first (D1) and second (D2) decision of Experiment 3. Right: Average change from the first to the second investment decision when subjects experienced a higher mean for the low-risk ($x_{1,exp}$) or the high risk ($x_{2,exp}$) option.

sequential, but not *simultaneous*, subjects also invested significantly less in the high-risk option compared to the first decision (t(248) = -1.94, p = 0.05, d = 0.25).

To control for the primary learning mode, we asked subjects after the experiment whether "the provided summary information" or their "own sampling experience" had more influence on their second investment decision. Out of 250 subjects, 162 (65%) chose their sampled experience. For this subsample, we find a significant negative effect of choice bracketing on choosing the high-risk option for the first decision (0.46 vs 0.36, t(160) = 1.79, p = 0.08, d = 0.28), but not the second decision (0.31 vs. 0.25, t(160) = 1.12, p = 0.26, d = 0.18). While subjects in *sequential* still tended to invest more into the high-risk option, the low-risk option dominated overall allocations (*sequential*: 0.53, *simultaneous*: 0.5). This is consistent with our interpretation of experience crowding-out the diversifying effect of choice bracketing by discouraging investments into heavy-tailed high-risk options (see Figure 10).

Splitting the data conditional on whether subjects sampled a higher mean for the high-risk than for the low-risk option (N = 133) reveals two things.[3] First, the share of points invested into the high-risk option remains largely constant when its experienced mean lies above the low-risk option. Second, when subjects sampled a higher mean for the low-risk option (analogous to the descriptions), investments into the high-risk option collapsed by up to 20 percentage points (see Figure 3). These results are not in line with over-weighting of rare events in decisions from experience. However, we also rely on objective inferences, which could skew the results.

Regressing the share of points allocated to either option on our main explanatory variables and sampling experience suggests some trends in how subjects respond to feedback from different payoff distributions (see Table 2). First, the sampled mean (weighted for recency) has a substantially larger effect on investments into the low-risk compared to the high-risk option. Second, the continuous difference in sampled means between the two risky options appears to be more important for investments into the high-risk option. This could point to subject difficulties in inferring incremental changes in expected returns for

[3]We focus on the two risky options because 92% of subjects sampled a higher mean for the low-risk than the safe option and there were no additional information gains from sampling the safe option. Only 4% of participants experienced the safe option as the expected-value maximizing option, and 70% of those almost exclusively sampled the safe option, indicating little regard for sampling.

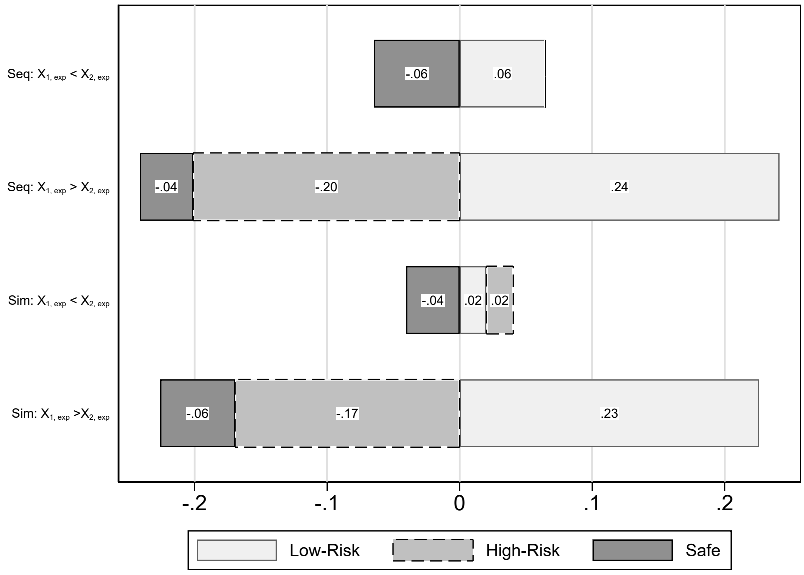


Figure 3 Average change from the first to the second investment decision when subjects experienced a higher mean for the low-risk ($x_{1,exp}$) or the high risk ($x_{2,exp}$) option.

Table 2 Ordered Logistic Regressions (odds ratios) on the share of points $(0-3)$ invested into each of the three respective investment options during the second investment decision.

	Low-Risk		High-Risk		Safe		Low-Risk		High-Risk		Safe	
	OR	*SD*	*OR*	*SD*	*OR*	*SD*	*OR*	*SD*	*OR*	*SD*	*OR*	*SD*
sequential	1.11	0.26	0.99	0.25	0.79	0.20	1.12	0.33	1.29	0.42	0.49**	0.16
risk	0.92	0.05	1.17***	0.07	0.93	0.05	0.92	0.06	1.28**	0.08	0.90	0.06
EPO	1.01	0.14	0.98	0.01	1.00	0.01	1.02	0.02	0.98	0.02	1.01	0.02
female	0.95	0.23	1.19	0.30	1.19	0.31	0.86	0.26	1.19	0.39	1.25*	0.42
$\overline{\mathrm{X}}_{\alpha,\mathrm{exp}}$	2.04***	0.41	1.13***	0.05			2.45***	0.66	1.20***	0.07		
$\overline{\mathrm{X}}_{2,\mathrm{exp}} - \overline{\mathrm{X}}_{1,\mathrm{exp}}$	0.84***	0.04	1.23***	0.08			0.79***	0.05	1.28***	0.11		
$\overline{\mathrm{X}}_{1,\mathrm{exp}} - \overline{\mathrm{X}}_{\mathrm{safe}}$					0.54***	0.09					0.36***	0.09
$\overline{\mathrm{X}}_{2,\mathrm{exp}} - \overline{\mathrm{X}}_{\mathrm{safe}}$					0.91**	0.04					0.91	0.06
N		250		250		250		162		162		162

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Left: Full sample. Right: Sub-sample of subjects who stated sampled experience as their primary information source. We used weighted sampled means following Plonsky et al. (2015) and tested $\alpha \in [1, 0.75, 0.5, 0.1]$. The weighing function is: $x_{t,exp} = (1 - \frac{1}{t^\alpha})x_{t-1,exp} + \frac{1}{t^\alpha} * V_t$, where V_t is the drawn coin in round t and $\alpha \in [0, 1]$ is a free parameter. Differences are generally pretty small. For the experienced mean $\overline{\mathrm{X}}_{\alpha,\mathrm{exp}}$, $\alpha = 0.5$ appears to perform best for the low-risk option (Hosmer and Lemeshow's goodness-of-fit test slightly favors $\alpha = 0.1$, but pseudo R^2 suffers substantially, and differences are small) as well as for the high-risk option. For the difference in experienced means between the two risky options, equal weighting of sampled experience ($\alpha = 1$) performs best. When regressing onto safe investments, we used $\alpha = 1$ for the differences in sampled means with the high-risk and $\alpha = 0.5$ with the low-risk option.

the high-risk option. Instead, reference points become more important.

Figure 4 shows kernel density estimates for the difference in experienced means between the two risky options conditional on subjects investing more into the high-risk or the

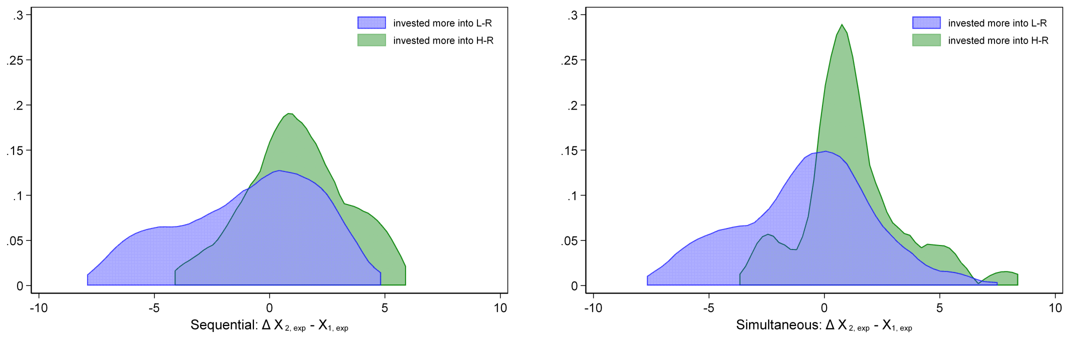


Figure 4 Kernel density estimates for the difference in sampled means between the two risky options for subjects who invested more into the low-risk (blue) or high-risk (green) option during the second decision.

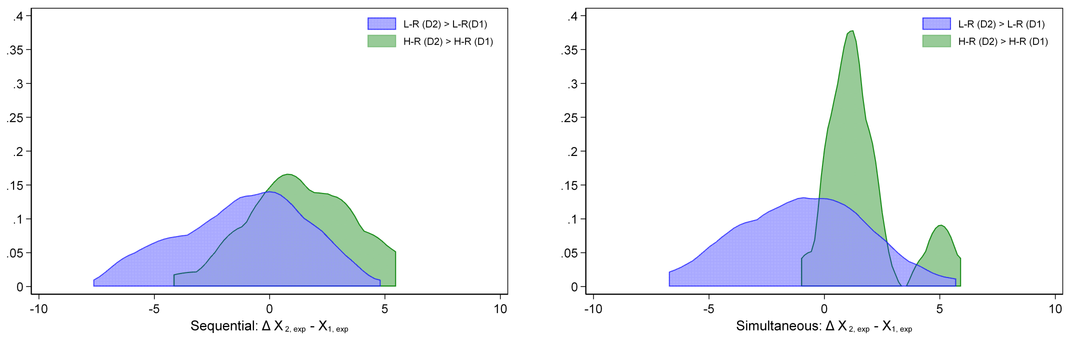


Figure 5 Subsample of subjects who stated that their sampled experience was their primary information source. Kernel density estimates for the difference in sampled means between the two risky options for subjects who increased investments into the low-risk (blue) or high-risk (green) option from D1 to D2.

low-risk option after sampling.[4]

In accordance with the results above, a large majority of subjects who invested more into the high-risk than the low-risk option also experienced a larger mean for the high-risk option. For the low-risk option, the slope is substantially broader, underscoring how experience biases decision-making towards more negatively skewed distributions. Between treatments, there are small, but noticeable differences. For both decision types, functions in *sequential* are slightly denser for values smaller than zero, i.e. those participants who sampled a higher mean for the low-risk option. When deciding sequentially, subjects appear to be somewhat more willing to accept worse experiences for the high-risk option.

Restricting the sample to subjects indicating experience as their primary information source accentuates these differences. Subjects in *sequential* needed less favorable sampling in order to either (1) increase their investment into the high-risk option between the two decisions (Figure 5) or (2) invest more into the high-risk than the low-risk option after sampling (Figure 12). Thus, while there remains some evidence that choice bracketing might influence risk-behavior when subjects primarily learn from experience, the overall effects appear to be, at best, small.

[4]Overall, there were no sampling differences between treatments (see Figure 11).

Result 4
The effect of choice bracketing on risky choices decreases with the introduction of experience as the primary learning mode.

6 Experiment 3b (E3b): Forced Sampling 2

6.1 Procedure and Participants

Experiment 3b adjusts two core features of E3a to test the following pre-registered hypothesis: Choice Bracketing does not affect risky choices under incomplete information when people primarily learn from experience.

First, the summary information for the first investment decision stated a mean of 7.08 ECU for the high-risk option, which is line with its average return over a large sample. Thus, subjects saw the following summary statistics based on 20 draws: mean ($\bar{x}_1 = 6.51$; $\bar{x}_2 = 7.08$; $\bar{x}_{safe} = 6$), standard deviation ($\sigma_{x_1} = 1.79$; $\sigma_{x_2} = 12.71$; $\sigma_{x_{safe}} = 0$), lowest draw ($x_1 = 3$, $x_2 = 0$, $x_{safe} = 6$) and highest draw ($x_1 = 10$, $x_2 = 40$, $x_{safe} = 6$). By providing subjects with information about the correct mean for the high-risk option, we wanted to control for the possibility that the previous descriptions biased risky choices after sampling. For instance, subjects might trust their own sampling experience less when it conflicts with descriptions, even if those are based on 20 draws.

Second, subjects only learned about the second investment decision and the sampling phase after completing the first investment decision. Subjects learned in the instructions that there would be a second stage, but no further information were provided. One reason why the effect of broad bracketing in E3 *Forced Sampling* was more ambivalent could be that subjects in both treatments took the possibility of a second round into account, consequently aligning the results.

We pre-registered to gather data until 150 independent observations per treatment. This resulted in a final sample of 301 observations (44.9% female). Throughout the experiment we used ECU, where 1 ECU equaled 10 cents. Subjects earned a fixed reward of \$0.5. Subjects who failed to answer five comprehension questions twice were not allowed to participate in the experiment.

6.2 Results Forced Sampling 2

Figure 6 shows subjects' allocation of points for both investment decisions (left) and the sampling phase (right). In line with the other experiments, we document a significant choice bracketing effect on risky choices when subjects learn from descriptions. Inducing broad bracketing through simultaneous choices leads to a significant increase in safe investments (0.191 vs 0.126, $t(299) = 2.08$, $p = 0.039$, $d = 0.24$) at the expense of high-risk investments (0.538 vs 0.624, $t(299) = 2.01$, $p = 0.045$, $d = 0.23$). Like in E1 *Description*, investments under *simultaneous* were significantly more diverse ($OR = 0.46$, $p = 0.00$). This result suggests that the effect of broad bracketing on expected value maximization might be dependent on the risk profile of the involved decision alternatives.

While the effect on high-risk investments is smaller than in E1 and only borderline significant, it is essentially the fourth replication of our main effect following a very soft intervention. Furthermore, the provided descriptions were based on a small sample, and showed the high-risk option to maximize expected income. Thus, we'd expect less deterrence from high-risk investments due to broad bracketing.

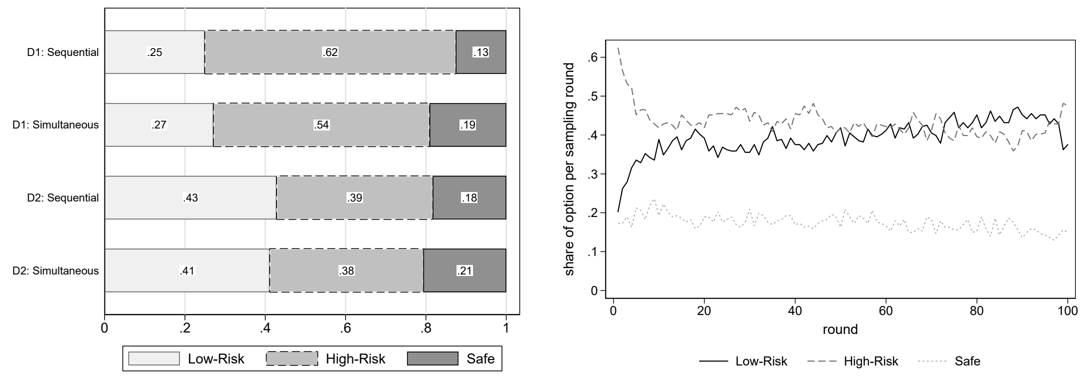


Figure 6 Left: overview subject choices over 100 sampling rounds. Right: mean share of points invested into each of the three options during the first (D1) and second (D2) decision of Experiment 3b.

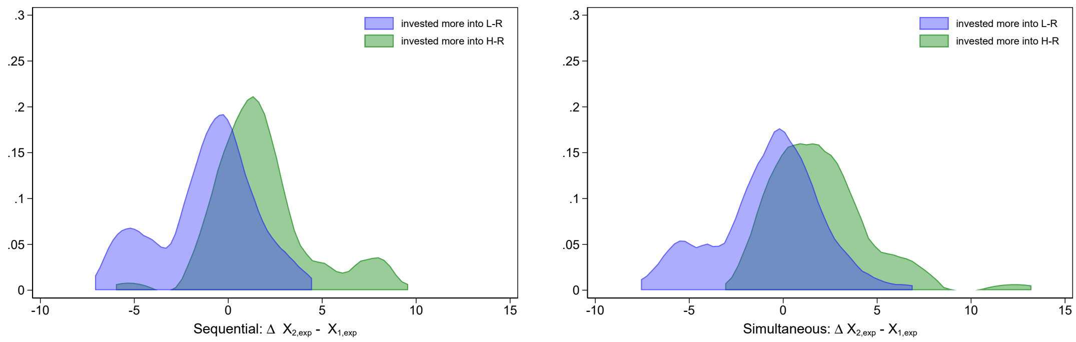


Figure 7 Kernel density estimates for the difference in sampled means between the two risky options for subjects who invested more into the low-risk (blue) or high-risk (green) option during the second decision.

After sampling, the effect of choice bracketing on risky choices largely vanishes (D2). There are virtually no differences in aggregate investments between *sequential* and *simultaneous*. Subjects in *sequential* still invested less diversely (OR = 0.63, p = 0.04), but the effect is substantially smaller than in D1. As shown on the right side of Figure 6, subjects' predominantly sampled through the high-risk (43.2) and the low-risk (39.3) option, experiencing the different payoff distributions and altering their investments in favor of the low-risk option – as predicted by theory and prior research. This pattern was mainly driven by subjects who sampled a lower mean for the high-risk than for the low-risk (*simultaneous*: $0.49 \longrightarrow 0.23$, *sequential*: $0.55 \longrightarrow 0.2$), but also present for those sampling a higher average mean for the high-risk option (*simultaneous*: $0.57 \longrightarrow 0.48$, *sequential*: $0.69 \longrightarrow 0.55$). Furthermore, Figure 7 shows that subjects in both treatments had roughly the same sampling experience conditional on the relation of points allocated between the two risky options during D2. Thus, E3b underscores the results of E3a, showing how learning from experience erodes the effect of choice bracketing on risky choices under incomplete information.

To analyze how sampling changed subjects' investment behavior, we run ordered logistic regressions on the share of points invested into each respective option (Table 3).[5] Sampling

[5]Here, we omit first-stage investments because (1) participants did not receive any information about their draws before the experiment concluded and (2) including first-stage investments as another

a higher mean has a large and significant effect for the low-risk option, while the effect for the high-risk option is only significant on the 10% level and much smaller. Instead, risk-preferences and the option's performance relative to the low-risk option appear to explain behavior. In particular and in line with descriptive tendencies in E3a *Forced Sampling*, sampling a relatively worse mean has a stronger negative effect for the high-risk ($\overline{X}_{1,exp} - \overline{X}_{2,exp}$: OR = 0.63, p < 0.00, CI (95%): [0.56, 0.71]) than the low-risk option ($\overline{X}_{2,exp} - \overline{X}_{1,exp}$: OR = 0.79, p < 0.00, CI (95%): [0.73, 0.86]) using the same ordered logit regressions.

Table 3 Ordered Logistic Regressions (odds ratios) on the share of points (0 – 3) invested into each of the three respective investment options during the second investment decision.

	Low-Risk		High-Risk		Safe		Low-Risk		High-Risk		Safe	
	OR	*SD*	*OR*	*SD*	*OR*	*SD*	*OR*	*SD*	*OR*	*SD*	*OR*	*SD*
sequential	0.89	0.19	1.17	0.27	0.75	0.18	1.06	0.29	0.93	0.28	0.75	0.23
risk	0.97	0.05	1.17***	0.06	0.92	0.04	0.98	0.05	1.25***	0.08	0.84**	0.06
EPO	1.03	0.14	0.99	0.14	0.99	0.01	1.04**	0.02	0.96**	0.02	0.99	0.02
female	0.81	0.18	0.86	0.20	1.84**	0.46	0.82	0.23	1.05	0.31	1.71*	0.55
$\overline{X}_{\alpha,exp}$	1.30***	0.09	0.93*	0.04			1.25**	0.12	0.94	0.05		
$\overline{X}_{2,exp} - \overline{X}_{1,exp}$	0.79***	0.03	1.59***	0.09			0.77***	0.04	1.66***	0.15		
$\overline{X}_{1,exp} - \overline{X}_{safe}$					1.00	0.09					0.85	0.14
$\overline{X}_{2,exp} - \overline{X}_{safe}$					0.91**	0.04					0.89*	0.06
N		301		301		301		191		191		191

Note. * p < 0.10, ** p < 0.05, *** p < 0.01. Left: Full sample. Right: Sub-sample of subjects who stated sampled experience as their primary information source. We used weighted sampled means following Plonsky et al. (2015) and tested $\alpha \in [1, 0.75, 0.5, 0.1]$. For the experienced mean $\overline{X}_{\alpha,exp}$, $\alpha = 0.1$ performs best (Hosmer and Lemeshow's goodness-of-fit test) for the low-risk option, and $\alpha = 0.5$ performs best for the high-risk option, although differences are small. For the difference in experienced means between the two risky options, equal weighting of sampled experience ($\alpha = 1$) performs best. When regressing onto safe investments, we used ($\alpha = 1$) for the differences in sampled means with both risky options. Reducing α hurts model fit.

Overall, results from E3b support the assertion that subjects are more sensitive to relatively bad outcomes from a high-risk actions and confirm that the sampled mean is less important for options with relatively heavy-tailed distributions, possibly because it is harder to accurately infer or intuit their average return. For the safe option, only the difference in sampled means compared to the high-risk option significantly explains behavior. Hence, subjects appear to rely more on reference points when choosing over high-risk options with high variability, comparing them to their alternatives and thereby inferring desirability. Given the relatively small average experience sample for each option, subjects experienced volatile returns from the high-risk option, possibly exacerbating issues regarding the inference of incremental differences in experienced means. As a response, they opt for the high-risk option only once they can clearly establish its superiority.

Restricting the regression analysis to subjects who indicated sampled experience as their primary information source for D2 (N = 191) largely replicates the results for the full sample (see columns 5 – 7 in Table 3). Notably, the elaboration on potential outcomes (EPO) scale appears to have a significant effect on main investments for subjects who

explanatory variable does not change much for the other coefficients.

indicated a stronger emphasis on experience. The EPO scale is positively associated with evaluating the positive and negative consequences of one's behavior as well as stronger self-regulation and control. These traits are closely related to the purported effects of broad bracketing. Indeed, our regression analysis shows that subjects with higher EPO scores tend to forego large but more uncertain gains from the high-risk option in favor of the low-risk option – much like subjects who allocate their points simultaneously. One potential interpretation is that under description-learning, choice bracketing crowds-out the effects of personality traits with correlated behavioral consequences. However, when the effects of choice bracketing disappear due to experience-based learning, people's character traits re-establish explanatory power. A second interpretation is that under experience, people are more focused on securing a certain mean payoff, and thus punish unfavorable draws or blanks more harshly. If experience decreases peoples' willingness to forego relatively safe payoffs – for example because the frequent experience of low-value draws increases the salience of that possibility –, it also naturally limits the influence of choice bracketing by biasing peoples choices towards satisficing behavior. Experiment 4 will test whether the effect of EPO on risky choices from experience replicates.

Result 5
When making decisions from stochastic descriptions, subjects who decide simultaneously and thus bracket their choices broadly invest less into a high-risk high-reward option even when that option maximizes expected value.

Result 6
Learning from experience increases investments into a low-risk option at the expense of a high-risk option. This effect is partially driven by subjects disproportionately shunning the high-risk option after experiencing a lower average mean compared to the low-risk option.

Result 7
We replicate the result from E3 Forced Sampling: The effect of choice bracketing on risky choices decreases with the introduction of experience as the primary learning mode.

7 Experiment 4 (E4): Pure Sampling

7.1 Procedure and Participants

Experiment 4 consisted of two stages. First, subjects gained information about the options solely from sampling through them 100 times. There were no descriptive information about the options. Subjects learned that after sampling, they would draw from the options three times, earning money according to the value of the coins. They were not informed about the second stage.

Second, we elicited incentivized subject beliefs about the three options. After completing their main investment decision, subjects were asked to rank the three options according to four metrics: (1) mean, (2) standard deviation, (3) lowest draw and (4) highest draw. We provided explanations for these four metrics both in the instructions for the elicitation task (after the main investments), and on the decision page. Subjects could earn an additional 6 ECU if they correctly ranked all three options according to all four metrics. If they correctly ranked all three options according to three metrics, they received 3 ECU.

Otherwise, they did not earn any bonus money.

Additionally, we asked subjects to estimate the *true value* of each option via a slider. Per option, participants earned additional 3 ECU if their estimation was within 5% of the true value, 2 ECU if it was within 10% of the true value, and 0.5 ECU if it was within 15% of the true value. Otherwise, they did not earn any bonus money. Thus, subjects could earn up to 9 ECU in additional bonus money for correctly estimating the true mean of all three options, and up to 6 ECU for correctly ordering the three options with regard to our four main metrics in this study. Subjects first completed the ranking elicitation, and then the mean estimation on another page. All payoffs and draws from all tasks within this study were only revealed after the post-experimental questionnaire.

We pre-registered to gather data until 125 independent observations per treatment. This resulted in a final sample of 254 observations (41.3% female). Throughout the experiment we used ECU, where 1 ECU equaled 10 cents. Subjects earned a fixed reward of $0.5. Subjects who failed to answer three comprehension questions twice were not allowed to participate in the experiment.

7.2 Results Sampling

Figure 8 shows subjects' sampling behavior (left) as well as their main decision investments afterwards (right). As expected, sampling into the two risky options increased over time, although the safe option remained fairly relevant throughout. On average, subjects sampled a mean return of 6.4 for the low-risk, 6.05 for the high-risk and 6 for the safe option (no treatment differences).

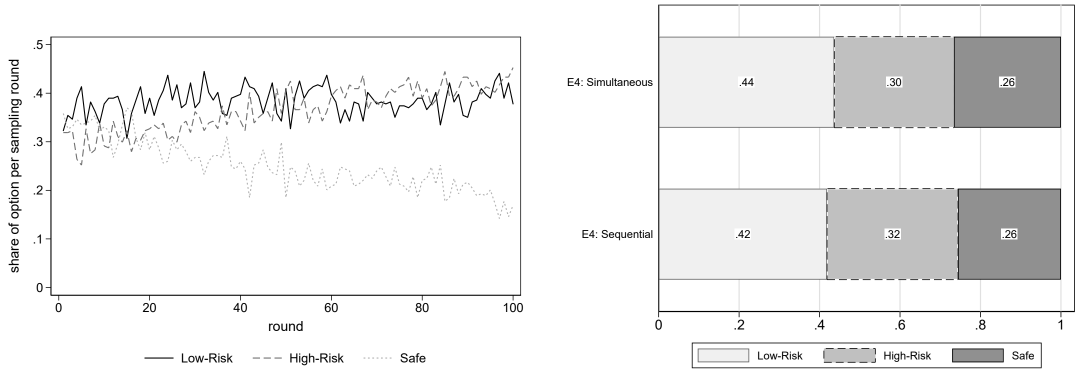


Figure 8 Left: overview subject choices over 100 sampling rounds. Right: mean share of points invested into the three options.

In accordance with our hypothesis and results from E3a and E3b, broad choice bracketing through simultaneous choice did not significantly affect aggregate main investments for any option. Restricting the sample to subjects experiencing a larger mean for the high-risk option does not change that result (*sequential*/*simultaneous* LR: 0.33/0.32; HR: 0.49/0.44; Safe: 0.18/0.24), confirming that this is not due to subjects sampling disproportionally bad means for the heavy-tailed high-risk option. This affirms one main insight from this paper: In our experimental setting, the effect of choice bracketing on risky choices largely vanishes once people learn from experience. As shown in Figure 8, that holds true even for the density estimates of the difference in experienced (sampled) means between the two risky options conditional on subjects' investments. For example, Figure 4 showed that when risky choices are made under both sampled experience and descriptions, people in

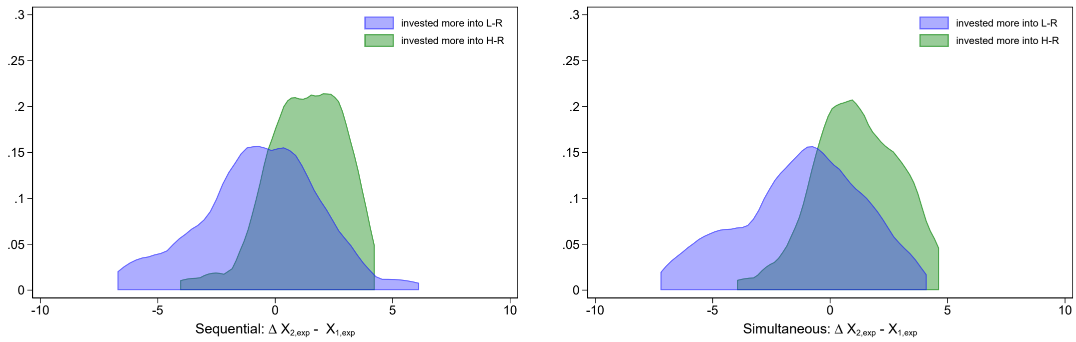


Figure 9 Kernel density estimates for the difference in sampled means between the two risky options for subjects who invested more into the low-risk (blue) or high-risk (green) option.

sequential still appeared to be slightly more willing to accept worse experiences for the high-risk option. For individuals who purely learn from sampling, however, we do not document these differences, further highlighting the erosion of choice bracketing effects. Similar to E3b *Forced Sampling*, there do appear to be residual effects of our decision frame on the distribution of investments. Specifically, an ordered logistic regression of the treatment dummy on the diversification variable confirms that subjects in *simultaneous* invested more diversely (OR = 2.41, $p < 0.00$). Pairwise chi-square tests show that investment distributions significantly differed for the low-risk option ($\tilde{\chi}^2$ (3) = 16.51, $p < 0.00$), but not the high-risk ($\tilde{\chi}^2$ (3) = 5.42, p = 0.14) or the safe option ($\tilde{\chi}^2$ (3) = 7.62, p = 0.06).

Ordered logistic regression results (Table 4) reveal some generally consistent behavioral patterns. As expected, risk-seeking attitudes positively affect investments into the high-risk at the expense of investments into the low-risk option. Contrary to decisions from descriptions, the EPO scale again exhibits significant explanatory power. In particular, people who are focused on future outcomes seem to be more inclined to use the safe option. Regarding sampling, we document two interesting main findings. First, differences in sampled means are significant and strong predictors for investment across all three options, whereas the sampled mean per se has no significant effect for investments into the high-risk option. Like before, subjects appear more sensitive to relatively bad draws from the high-risk option ($\overline{X}_{1,exp} - \overline{X}_{2,exp}$: OR = 0.71, $p < 0.00$, CI (95%): [0.61, 0.81]) than the low-risk option ($\overline{X}_{2,exp} - \overline{X}_{1,exp}$: 0.82, $p < 0.00$, CI (95%): [0.75, 0.91]), although differences in odds ratios are not significant on the 5%-level (z = 1.81, p = 0.07). Second, for the low-risk option, focusing on smaller, more recent samples tends to increase model fit, whereas for the high-risk option, more distant experiences still matter. One potential interpretation is that the more extreme values from the high-risk option are more salient and thus have a more durable effect on decision-making.

While the regression results confirm that subjects utilized experiential learning to inform their investment decision, there is still some uncertainty about the effectiveness and accuracy with which subjects process the sampling data. In order to approximate that, we elicited incentivized beliefs for all three options. We first exclude participants who either indicated an expected mean of zero for any option, as well as those who indicated expected means of higher than 20 for the low-risk, and higher than 10 for the safe option (N = 50). This conservative approach guarantees that subjects with especially negligent answers do not bias the analyses. As there were no coins with a value higher than 10

Table 4 Ordered Logistic Regressions (odds ratios) on the share of points (0 – 3) invested into each of the three respective investment options.

	Low-Risk		High-Risk		Safe	
	OR	*SD*	*OR*	*SD*	*OR*	*SD*
sequential	0.96	0.23	0.96	0.24	0.87	0.22
risk	0.91**	0.04	1.27***	0.07	0.95	0.47
EPO	0.97**	0.13	0.97*	0.15	1.07***	0.02
female	1.03	0.24	1.06	0.27	1.22	0.31
$\overline{\mathrm{X}}_{\alpha,\mathrm{exp}}$	1.47***	0.12	1.05	0.05		
$\overline{\mathrm{X}}_{2,\mathrm{exp}} - \overline{\mathrm{X}}_{1,\mathrm{exp}}$	0.82***	0.04	1.42***	0.10		
$\overline{\mathrm{X}}_{1,\mathrm{exp}} - \overline{\mathrm{X}}_{\mathrm{safe}}$					0.80***	0.07
$\overline{\mathrm{X}}_{2,\mathrm{exp}} - \overline{\mathrm{X}}_{\mathrm{safe}}$					0.83***	0.04
N		254		254		254

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We used weighted sampled means following Plonsky et al. (2015) and tested $\alpha \in [1, 0.75, 0.5, 0.1]$. For the experienced mean $\overline{\mathrm{X}}_{\alpha,\mathrm{exp}}$, $\alpha = 0.1$ performs best for the low-risk option, and $\alpha = 0.5$ performs best for the high-risk option, although differences are small. For the difference in experienced means between the two risky options, equal weighting of sampled experience ($\alpha = 1$) performs best. When regressing investments into the safe option, we used ($\alpha = 0.5$) for the differences in sampled means with the low-risk, and ($\alpha = 1$) for differences in sampled means with the high-risk option. Reducing α for the high-risk option hurts model fit.

in the low-risk option, and no values that differed from 6 in the safe option, it does not make sense for any subject to estimate means above 20 or 10 respectively.[6] Since subjects were incentivized to guess the "true" mean, it is still possible that they thought an option might contain coins they did not yet sample, which is why we do not drop participants who e.g. estimated a mean of 10 for the low-risk option.

Table 5 summarizes the belief elicitation results. There are no relevant treatment differences. Overall, subject beliefs are relatively accurate. For the high-risk and the safe option, median mean beliefs from the slider estimates coincide with the option's true mean, and for the low-risk option, they are not far off either. Subjects on average slightly overestimate average returns for the low-risk and the high-risk option, although results vary widely for the high-risk option.

Asking subjects to rank the different options according to their average return reveals that most people expected the low-risk option to maximize expected value. For the high-risk option, estimations are much more polarized, with roughly one-third expecting maximization, and over 50% judging the option to return the lowest return on average.

[6] Except for one subject who did not sample the low-risk option, every participant sampled each option at least once.

Table 5 Belief Elicitation Results

	Low-Risk			High-Risk			Safe		
Est. Mean (SD)	7.06 (1.9)			8.11 (6.3)			5.88 (1.2)		
Est. Median	7			7			6		
	Low-Risk			High-Risk			Safe		
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
Ranked Mean %	45.5	41.2	12.6	33.9	13.1	53.0	20.7	44.9	34.3
Ranked SD %	15.0	70.0	15.0	75.7	12.4	11.9	9.3	17.6	73.1
Ranked HD %	11.7	72.9	15.4	82.9	3.7	13.3	5.3	23.4	71.3
Ranked LD %	17.9	70.9	11.2	51.0	4.6	44.4	37.8	24.5	37.8

This highlights the proposed difficulty to infer representative probability distributions for options with heterogeneous, strongly skewed outcomes. For the standard deviation, the majority of subjects answered in line with the underlying distribution. Most people indicated that the high-risk option had the highest standard deviation (76%), followed by the low-risk option (70%) and the safe option (73%). Given that the standard deviation is a relatively complicated measure which most lay people do not regularly – if at all – engage with, the high percentages point to a relatively attentive sample with an effective learning period.

Including subject beliefs in the regression analysis points to one possibly distinct difference between risky choices from descriptions and risky choices from experience: people care relatively more about their average as compared to their maximum payoff when deciding from experience. Controlling for estimated mean differences, the relative ranking of a risky option regarding its mean has a large and significant effect on investments, as subjects choose to invest far less points into an option when they do not consider it to maximize average returns (see Table 6).

These results somewhat contrast E1 and E2, where many subjects chose the high-risk option despite its ostensibly lower mean. This was also where choice bracketing had the most pronounced effect, as people in *sequential* were significantly more likely to "chase" larger returns. In combination with the fact that EPO attitudes, which measure a person's propensity to regulate themselves while considering future outcomes, become relevant under pure sampling (as they were for subjects who indicated sampled experience as their primary information source in E3b), E4 suggests that one reason why choice bracketing does not influence behavior under experience might be a lack of Upper Confidence Bound strategies. While under descriptions, broad bracketing through simultaneous choice disciplines some people towards securing relatively safe payoffs, most people who learn from experience already adopt such a pattern.[7]

Finally, when controlling for subject beliefs, choice bracketing still does not significantly explain behavior, suggesting that our results are not due to sampling bias. Interestingly, if anything, the effect seems to be *reversed*, with subjects in *sequential* investing *less* into

[7] Of note, controlling for subject beliefs does not change the apparent tendency of subjects to "punish" relatively worse experiences of high-risk options more intensely ($\overline{\mathrm{X}}_{1,\mathrm{exp}} - \overline{\mathrm{X}}_{2,\mathrm{exp}}$: OR = 0.70, $p < 0.00$, CI (95%): [0.58, 0.86]) than those of low-risk options ($\overline{\mathrm{X}}_{2,\mathrm{exp}} - \overline{\mathrm{X}}_{1,\mathrm{exp}}$: 0.88, $p < 0.00$, CI (95%): [0.77, 1.00] $\longrightarrow$ z = 1.83, p = 0.07)

Table 6 Ordered Logistic Regressions (odds ratios) on the share of points (0 – 3) invested into each of the three respective investment options.

	Low-Risk		High-Risk		Safe	
	OR	*SD*	*OR*	*SD*	*OR*	*SD*
sequential	1.13	0.31	0.76	0.23	1.19	0.39
risk	0.98	0.06	1.36***	0.09	0.79***	0.05
EPO	0.97**	0.02	0.99	0.02	1.07***	0.02
female	1.23	0.36	1.12	0.36	0.97	0.32
$\overline{\mathrm{X}}_{\alpha,\mathrm{exp}}$	1.54***	0.16	1.03	0.05		
$\overline{\mathrm{X}}_{2,\mathrm{exp}} - \overline{\mathrm{X}}_{1,\mathrm{exp}}$	0.88*	0.06	1.42***	0.14		
ranked mean 2	0.40***	0.13	0.17***	0.09	0.31***	0.13
ranked mean 3	0.12***	0.06	0.31***	0.12	0.19***	0.09
$\overline{\mathrm{X}}_{2,\mathrm{est}} - \overline{\mathrm{X}}_{1,\mathrm{est}}$	0.98	0.02	1.00	0.07		
$\overline{\mathrm{X}}_{1,\mathrm{est}} - \overline{\mathrm{X}}_{\mathrm{safe,est}}$	0.93	0.07			1.05	0.08
$\overline{\mathrm{X}}_{2,\mathrm{est}} - \overline{\mathrm{X}}_{\mathrm{safe,est}}$			1.00	0.07	0.96	0.03
$\overline{\mathrm{X}}_{1,\mathrm{exp}} - \overline{\mathrm{X}}_{\mathrm{safe}}$					0.84*	0.09
$\overline{\mathrm{X}}_{2,\mathrm{exp}} - \overline{\mathrm{X}}_{\mathrm{safe}}$					0.81***	0.07
N		198		198		198

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We used the same specifications as in Table 4. $\overline{\mathrm{X}}_{\mathrm{est}}$ refers to mean values calculated from slider estimations

the high-risk option. Choosing a more exclusive sample by dropping all subjects who estimated a mean return of less than 5 (N = 19) or more than 7 (N = 10) for the safe option[8] and re-running the regression from Table 6 even indicates a borderline significant negative effect of *sequential* choice on investments into the high-risk option (OR: 0.53, p = 0.068; see Table 7). This rather puzzling result might be explained by the diversifying effect of a *simultaneous* decision frame. However, since the effect only appears in that sub-sample and is not significant for conventional p-value thresholds, we caution not to over-interpret it and present it as a potential starting point for future inquiry.

Result 8
When subjects learn purely from experience, broad bracketing does not inhibit investments into high-risk high-reward options.

8 Conclusion

Five studies show that the effect of choice bracketing on risk behavior is dependent on the learning mode. When people learn from stochastic descriptions, broad bracketing

[8]The only reason attentive subjects would deviate from a prediction of 6 is uncertainty induced by expectations of unexplored coins. Restricting the sample around the safe option thus has a small likelihood of excluding attentive subjects, particularly if the deviations from 6 are relatively large.

reduces risk-taking. Compared to a sequential decision frame, simultaneous choice deters investments into a high-risk high-reward option in favor of safe and less volatile returns.

Introducing experience as the primary learning mode reduces that effect. Instead, subjects appear to behave more risk-averse by disproportionally investing into the low-risk option. Sampling a lower mean payoff strongly reduces investments for the high-risk, but not the low-risk option, which offsets the diversifying effect of broad bracketing. This supports the interpretation that learning from experience crowds-out the effect of choice bracketing as it biases people towards options with more negatively skewed outcome distributions that minimize immediate regret. While a narrow decision frame might still somewhat favor options that promise potentially large rewards (see e.g. the "big eyes effect" (Grosskopf et al. 2006)), its effect is partially overshadowed by strong experience-induced preferences for options that return good outcomes most of the time. Thus, in our decision environment, the effects of broad bracketing and experience are positively correlated, decreasing the differential impact of simultaneous choice. Contrary and in line with much of the literature on the description-experience gap, people who learned from descriptions generally favored the high-risk option despite its ostensibly lower mean. This naturally leaves more room for the impact of broad bracketing to unfold.

Introducing experience as the sole learning mode supports that conjunction. When people receive no descriptive information and purely learn through sampling, the negative effect of broad bracketing on high-risk investments vanishes. Results from incentivized belief elicitation suggest that this is not because subject beliefs are, on average, fundamentally biased. While many participants severely under-estimate mean returns from the high-risk option compared to the two alternatives, a sizeable share also thought that it would maximize mean returns, while subjects that learned from descriptions mostly saw the high-risk option to have the lowest mean. Mean estimates are also remarkably close to the option's true values. Thus, despite relatively accurate beliefs and a large percentage of people thinking the high-risk option not only to have the highest upside, but to exhibit the largest mean returns, they invested less than participants whose only information source stated the high-risk option to have the lowest average return. These results, in combination with regression analyses showing a subject's subjective mean ranking to have strong explanatory power for investments, suggest that people who learn from experience have stronger preferences to secure a certain mean payoff. Furthermore, for subjects in E3b *Forced Sampling* who indicated to primarily rely on their sampled experience as well as for subjects in E4 *Pure Sampling*, the EPO scale significantly explains behavior. In all other cases, it has no explanatory power. This is more evidence that in learning from experience, people's tendencies to self-regulate themselves towards relatively secure payoffs are more important determinants for risky choices. Thus, we propose that people who learn from experience are already biased towards the behavioral patterns often associated with broad choice bracketing: less uncertainty-seeking, more self-control, less seeking of high-risk high-reward gains.

This interpretation is consistent with the proposition that ambiguity, which surrounds at least two of our options in the sampling paradigm, induces anticipation of regret because bad outcomes can be attributed to the decision-maker (Heath and Tversky 1991). Such mechanisms might be particularly pronounced in environments where uncertain, ambiguous options are constantly compared to a safe reference point (Glöckner et al. 2016), leading to loss-aversion for those that regularly perform below the comparison point, e.g. in our study the high-risk option. On the other hand, for subjects who actually sampled a higher mean, recent findings would suggest that decisions from experience actually exhibit *more* sensitivity to probabilities, and therefore a potential over-weighting of rare events compared to decisions from descriptions. Hence, people who experience larger average returns due to rare positive events should – conditional on preferences being exogenous – invest more into the high-risk option. In turn, the effect of choice bracketing on risky choices might re-appear, as subjects exhibit more risk-seeking patterns. However, we do not find a choice bracketing effect for subjects who either indicated larger estimated means, or sampled a larger mean for the high-risk option. One probable reason is the safe alternative. While subjects in the pure sampling paradigm do not immediately know that the option always returns a certain value, they do with a high probability in experiments 3a and 3b. Moreover, sampling investments into the safe alternative decrease over time across all sampling studies, suggesting increasing confidence into the robustness of the safe option. As Glöckner et al. (2016) show, decisions from experience do show under-weighting of rare events when the choice environment includes a safe alternative. Another reason might be that higher sensitivity to probabilities in decisions from experience is largely predicated on cumulative prospect theory, which has ambivalent use in environments with more than two choice alternatives that comprise multiple outcomes (Birnbaum 2008). Thus, results are still consistent with subjects shunning high-risk high-reward options when information is gathered through sampling, thereby obviating the need for broad bracketing.

Importantly, it might not always be desirable to motivate low-risk at the expense of high-risk choices. Experiment 3b showed that subjects who invested sequentially were more likely to invest into the expected-value-maximizing high-risk option. This suggests that the effect of broad bracketing on maximization behavior (Haisley et al. 2008) might depend on the risks involved in the decision environment. In line with Read et al. (2001), our results suggest that the diversifying effect of broad bracketing through simultaneous choice can be detrimental to decision quality, for example if (1) a high-risk option maximizes expected values, (2) people prefer low-variance outcomes or (3) one option dominates all other alternatives. Particularly in the context of innovations, betting on potentially disruptive technologies can lead to fundamental breakthroughs with large positive externalities. Even in competitive market environments, some firms may want, or need, to create substantial new advantages in order to keep up. Therefore, both decision frames may prove useful in assisting decision-making, depending on the individual or organizational goals.

Beyond that, we find that people who actively decide against sampling, i.e. exploration, are subsequently less likely to exhibit risk-seeking behavior. Intuitively, rejecting the opportunity to reduce uncertainty increases the anticipated regret caused by a bad, risky outcome. Here, we see avenues for future research. More broadly, additional research is needed to understand how choice bracketing interacts with the peculiarities of uncertain environments. This could include among others the influence of a safe alternative or options with different underlying distributions.

Our results can be applied to a vast array of real-world problems, ranging from financial and organizational decision-making to simple consumer behavior like listening to music and compiling playlists. Since many important economic decisions are fundamentally derived from descriptions rather than experiences, we expect the behavioral consequences of choice bracketing to be widespread and of significant importance.

Appendix

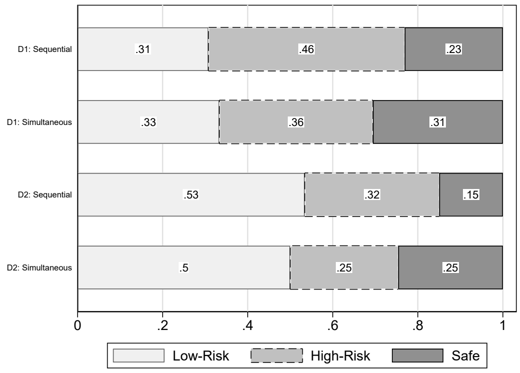


Figure 10 Experiment 3: Subsample of subjects who stated sampled experience as their primary information mode. Mean share of points invested into each of the three options during the first (D1) and second (D2) decision.

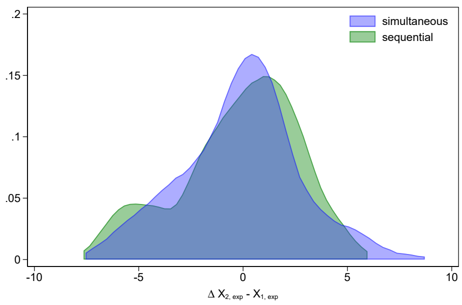


Figure 11 Kernel density estimates for the difference in sampled means between the treatments (X_2= high-risk option, X_1= low-risk option).

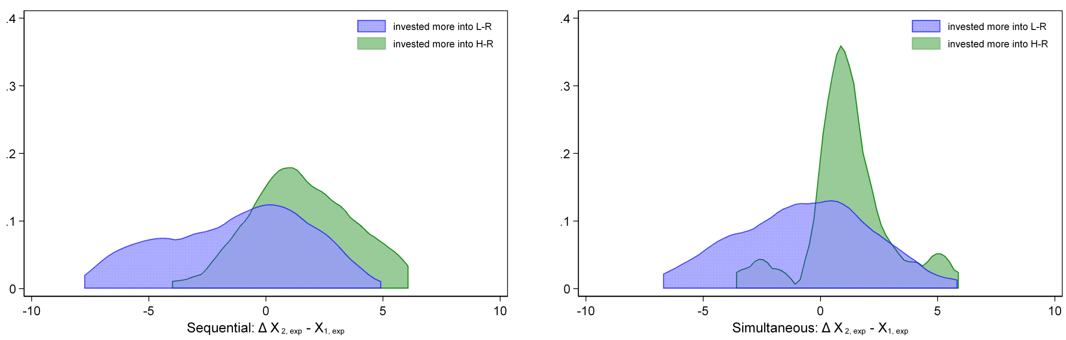


Figure 12 Subsample of subjects who stated that their sampled experience was their primary information source. Kernel density estimates for the difference in sampled means between the two risky options for subjects who increased investments into the low-risk (blue) or high-risk (green) option from D1 to D2.

Table 7 Ordered Logistic Regressions (odds ratios) on the share of points (0 – 3) invested into each of the three respective investment options.

	Low-Risk		High-Risk		Safe	
	OR	*SD*	*OR*	*SD*	*OR*	*SD*
sequential	1.35	0.42	0.53*	0.18	1.09	0.38
risk	0.99	0.07	1.42***	0.11	0.75***	0.06
EPO	0.96*	0.02	0.99	0.02	1.06**	0.03
female	0.93	0.29	1.49	0.52	0.93	0.33
$\overline{\mathrm{X}}_{\alpha,\mathrm{exp}}$	1.65***	0.19	1.02	0.06		
$\overline{\mathrm{X}}_{2,\mathrm{exp}} - \overline{\mathrm{X}}_{1,\mathrm{exp}}$	0.83**	0.06	1.47***	0.17		
ranked mean 2	0.43**	0.15	0.24**	0.13	0.34**	0.15
ranked mean 3	0.15***	0.09	0.31***	0.14	0.24***	0.13
$\overline{\mathrm{X}}_{2,\mathrm{est}} - \overline{\mathrm{X}}_{1,\mathrm{est}}$	0.98	0.03	1.06	0.11		
$\overline{\mathrm{X}}_{1,\mathrm{est}} - \overline{\mathrm{X}}_{\mathrm{safe,est}}$	0.82	0.09			1.14	0.13
$\overline{\mathrm{X}}_{2,\mathrm{est}} - \overline{\mathrm{X}}_{\mathrm{safe,est}}$			1.03	0.12	0.96	0.04
$\overline{\mathrm{X}}_{1,\mathrm{exp}} - \overline{\mathrm{X}}_{\mathrm{safe}}$					0.84	0.09
$\overline{\mathrm{X}}_{2,\mathrm{exp}} - \overline{\mathrm{X}}_{\mathrm{safe}}$					0.82**	0.07
N		171		171		171

Note. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We used the same specifications as in Table 4.

References

Abdellaoui, Mohammed, Aurélien Baillon, Laetitia Placido and Peter P. Wakker (2011). 'The rich domain of uncertainty: Source functions and their experimental implementation'. In: *American Economic Review* 101 (2), 695–723.

Aydogan, Ilke (2021). 'Prior Beliefs and Ambiguity Attitudes in Decision from Experience'. In: *Management Science* 67 (11), 6934–6945.

Aydogan, Ilke and Yu Gao (2020). 'Experience and rationality under risk: re-examining the impact of sampling experience'. In: *Experimental Economics* 23, 1–29.

Barron, Greg, Stephen Leider and Jennifer Stack (2008). 'The effect of safe experience on a warnings' impact: Sex, drugs, and rock-n-roll'. In: *Organizational Behavior and Human Decision Processes* 106 (2), 125–142.

Benartzi, Shlomo and Richard H. Thaler (1999). 'Risk aversion or myopia? Choices in repeated gambles and retirement investments'. In: *Management Science* 45 (3), 364–381.

Birnbaum, Michael H. (2008). 'New paradoxes of risky decision making.' In: *Psychological Review* 115 (2), 463.

Brown, Jeffrey R., Arie Kapteyn, Erzo F.P. Luttmer, Olivia S. Mitchell and Anya Samek (2021). 'Behavioral impediments to valuing annuities: complexity and choice bracketing'. In: *Review of Economics and Statistics* 103 (3), 533–546.

Camilleri, Adrian R. and Ben R. Newell (2011a). 'Description-and experience-based choice: Does equivalent information equal equivalent choice?' In: *Acta Psychologica* 136 (3), 276–284.

— (2011b). 'When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms'. In: *Psychonomic Bulletin & Review* 18 (2), 377–384.

Chaudhry, Shereen J., Michael Hand and Howard Kunreuther (2020). 'Broad bracketing for low probability events'. In: *Journal of Risk and Uncertainty* 61 (3), 211–244.

Chen, Daniel L., Martin Schonger and Chris Wickens (2016). 'oTree—An open-source platform for laboratory, online, and field experiments'. In: *Journal of Behavioral and Experimental Finance* 9, 88–97.

Denrell, Jerker (2007). 'Adaptive learning and risk taking.' In: *Psychological Review* 114 (1), 177.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp and Gert G. Wagner (2011). 'Individual risk attitudes: Measurement, determinants, and behavioral consequences'. In: *Journal of the European Economic Association* 9 (3), 522–550.

Ellis, Andrew and David J. Freeman (2020). 'Revealing choice bracketing'. In: *arXiv preprint arXiv:2006.14869*.

Erev, Ido and Greg Barron (2005). 'On adaptation, maximization, and reinforcement learning among cognitive strategies.' In: *Psychological Review* 112 (4), 912.

Erev, Ido, Eyal Ert, Ori Plonsky, Doron Cohen and Oded Cohen (2017). 'From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience.' In: *Psychological Review* 124 (4), 369.

Erev, Ido and Ernan Haruvy (2016). 'Learning and the Economics of Small Decisions'. In: *The Handbook of Experimental Economics, Volume 2*. Princeton University Press, 638–716.

Erev, Ido et al. (2010). 'A choice prediction competition: Choices from experience and from description'. In: *Journal of Behavioral Decision Making* 23 (1), 15–47.

Ert, Eyal and Stefan T. Trautmann (2014). 'Sampling experience reverses preferences for ambiguity'. In: *Journal of Risk and Uncertainty* 49 (1), 31–42.
Fox, Craig R. and Liat Hadar (2006). '"Decisions from experience"= sampling error+ prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004)'. In: *Judgment and Decision Making* 1 (2), 159.
Frey, Renato, Rui Mata and Ralph Hertwig (2015). 'The role of cognitive abilities in decisions from experience: Age differences emerge as a function of choice set size'. In: *Cognition* 142, 60–80.
Glöckner, Andreas, Benjamin E. Hilbig, Felix Henninger and Susann Fiedler (2016). 'The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice.' In: *Journal of Experimental Psychology: General* 145 (4), 486.
Grosskopf, Brit, Ido Erev and Eldad Yechiam (2006). 'Foregone with the wind: Indirect payoff information and its implications for choice'. In: *International Journal of Game Theory* 34 (2), 285–302.
Hadar, Liat, Shai Danziger and Vicki G. Morwitz (2021). 'Choice bracketing and experience-based choice'. In: *Journal of Behavioral Decision Making*.
Haisley, Emily, Romel Mostafa and George Loewenstein (2008). 'Myopic risk-seeking: The impact of narrow decision bracketing on lottery play'. In: *Journal of Risk and Uncertainty* 37 (1), 57–75.
Hardin, Andrew M. and Clayton Arlen Looney (2012). 'Myopic loss aversion: Demystifying the key factors influencing decision problem framing'. In: *Organizational Behavior and Human Decision Processes* 117 (2), 311–331.
Hau, Robin, Timothy J. Pleskac and Ralph Hertwig (2010). 'Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities'. In: *Journal of Behavioral Decision Making* 23 (1), 48–68.
Heath, Chip and Amos Tversky (1991). 'Preference and belief: Ambiguity and competence in choice under uncertainty'. In: *Journal of Risk and Uncertainty* 4 (1), 5–28.
Hertwig, Ralph, Greg Barron, Elke U. Weber and Ido Erev (2004). 'Decisions from experience and the effect of rare events in risky choice'. In: *Psychological Science* 15 (8), 534–539.
Hertwig, Ralph and Ido Erev (2009). 'The description–experience gap in risky choice'. In: *Trends in Cognitive Sciences* 13 (12), 517–523.
Hertwig, Ralph and Dirk U Wulff (2021). 'A description-experience framework of the dynamic response to risk'. In: *Perspectives on Psychological Science*.
Jessup, Ryan K., Anthony J. Bishara and Jerome R. Busemeyer (2008). 'Feedback produces divergence from prospect theory in descriptive choice'. In: *Psychological Science* 19 (10), 1015–1022.
Kellen, David, Thorsten Pachur and Ralph Hertwig (2016). 'How (in) variant are subjective representations of described and experienced risk and rewards?' In: *Cognition* 157, 126–138.
Koch, Alexander K. and Julia Nafziger (2019). 'Correlates of narrow bracketing'. In: *The Scandinavian Journal of Economics* 121 (4), 1441–1472.
Lejarraga, Tomás (2010). 'When experience is better than description: Time delays and complexity'. In: *Journal of Behavioral Decision Making* 23 (1), 100–116.
Lejarraga, Tomás and Cleotilde Gonzalez (2011). 'Effects of feedback and complexity on repeated decisions from description'. In: *Organizational Behavior and Human Decision Processes* 116 (2), 286–295.

Lejarraga, Tomás, Jan K. Woike and Ralph Hertwig (2016). 'Description and experience: How experimental investors learn about booms and busts affects their financial risk taking'. In: *Cognition* 157, 365–383.

Looney, Clayton Arlen and Andrew M. Hardin (2009). 'Decision support for retirement portfolio management: Overcoming myopic loss aversion via technology design'. In: *Management Science* 55 (10), 1688–1703.

Meyer, Robert and Howard Kunreuther (2017). *The ostrich paradox: Why we underprepare for disasters.* Wharton School Press.

Moher, Ester and Derek J. Koehler (2010). 'Bracketing effects on risk tolerance: Generalizability and underlying mechanisms'. In: *Judgment and Decision Making* 5 (5), 339.

Nenkov, Gergana Y., J. Jeffrey Inman and John Hulland (2008). 'Considering the future: The conceptualization and measurement of elaboration on potential outcomes'. In: *Journal of Consumer Research* 35 (1), 126–141.

Newell, Ben R., Tim Rakow, Eldad Yechiam and Michael Sambur (2016). 'Rare disaster information can increase risk-taking'. In: *Nature Climate Change* 6 (2), 158–161.

O'Reilly III, Charles A. and Michael L. Tushman (2013). 'Organizational ambidexterity: Past, present, and future'. In: *Academy of Management Perspectives* 27 (4), 324–338.

Plonsky, Ori, Kinneret Teodorescu and Ido Erev (2015). 'Reliance on small samples, the wavy recency effect, and similarity-based learning.' In: *Psychological Review* 122 (4), 621.

Rabin, Matthew and Georg Weizsäcker (2009). 'Narrow bracketing and dominated choices'. In: *American Economic Review* 99 (4), 1508–43.

Rakow, Tim, Kali A Demes and Ben R Newell (2008). 'Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice'. In: *Organizational Behavior and Human Decision Processes* 106 (2), 168–179.

Read, Daniel, Gerrit Antonides, Laura Van den Ouden and Harry Trienekens (2001). 'Which is better: Simultaneous or sequential choice?' In: *Organizational Behavior and Human Decision Processes* 84 (1), 54–70.

Read, Daniel and George Loewenstein (1995). 'Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices.' In: *Journal of Experimental Psychology: Applied* 1 (1), 34.

Read, Daniel, George Loewenstein and Matthew Rabin (1999). 'Choice Bracketing'. In: *Journal of Risk and Uncertainty* 19 (1-3), 171–197.

Steinhart, Yael, Ziv Carmon and Yaacov Trope (2013). 'Warnings of adverse side effects can backfire over time'. In: *Psychological Science* 24 (9), 1842–1847.

Thaler, Richard H., Amos Tversky, Daniel Kahneman and Alan Schwartz (1997). 'The effect of myopia and loss aversion on risk taking: An experimental test'. In: *The Quarterly Journal of Economics* 112 (2), 647–661.

Tversky, Amos and Daniel Kahneman (1992). 'Advances in prospect theory: Cumulative representation of uncertainty'. In: *Journal of Risk and Uncertainty* 5 (4), 297–323.

Ungemach, Christoph, Nick Chater and Neil Stewart (2009). 'Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)?' In: *Psychological Science* 20 (4), 473–479.

Wachinger, Gisela, Ortwin Renn, Chloe Begg and Christian Kuhlicke (2013). 'The risk perception paradox—implications for governance and communication of natural hazards'. In: *Risk Analysis* 33 (6), 1049–1065.

Webb, Elizabeth C. and Suzanne B. Shu (2017). 'Is broad bracketing always better? How broad decision framing leads to more optimal preferences over repeated gambles'. In: *Judgment and Decision Making* 12 (4), 382–395.
Weber, Elke U., Sharoni Shafir and Ann-Renee Blais (2004). 'Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation.' In: *Psychological Review* 111 (2), 430.
Wulff, Dirk U., Max Mergenthaler-Canseco and Ralph Hertwig (2018). 'A meta-analytic review of two modes of learning and the description-experience gap.' In: *Psychological Bulletin* 144 (2), 140.
Yechiam, Eldad, Greg Barron and Ido Erev (2005). 'The role of personal experience in contributing to different patterns of response to rare terrorist attacks'. In: *Journal of Conflict Resolution* 49 (3), 430–439.
Yechiam, Eldad and Jerome R. Busemeyer (2006). 'The effect of foregone payoffs on underweighting small probability events'. In: *Journal of Behavioral Decision Making* 19 (1), 1–16.

Chapter 6

Contests for Social Dilemmas: Unequal Treatment Erodes the Positive Effect of Group Identity on Cooperation

Contests for Social Dilemmas: Unequal Treatment Erodes the Positive Effect of Group Identity on Cooperation

Elaine Horstmann, Ann-Kathrin Blankenberg, Alexander Erlei, Tim Schneider

Abstract

Group contests have long been one of the primary mechanisms organizations use to increase cooperation and team performance. However, little is known about their interaction with various social intricacies arising within competing work groups. This article exploits the controlled environment of a laboratory public good game to investigate how between-group competition using (1) egalitarian or (2) proportional prize sharing interacts with differing levels of group identity. We theoretically predict and experimentally confirm strictly positive effects of identity on group contributions for a contest with egalitarian prize sharing. For proportional prize sharing, we predict inequity aversion among individual members to decrease and possibly invert the positive effect of group identity on cooperation. In line with that, experimental results show that institutionalizing proportional prize sharing removes the effect of group identity on cooperation. Overall, combining group identity with egalitarian prize sharing maximizes cooperation.

This work was supported by the Federal Ministry of Education and Research (BMBF) [reference number: 01UO1506B/OMAHETI - "The Language of Objects - Material Culture in Context with Social Developments"

1 Introduction

For most people, going to work is not uniquely defined by their financial prospects, but embedded in a larger social context that includes manifold interactions, relations and identities. While often useful and productive, these social intangibles also create new challenges for a firm's institutional framework. For example, establishing organizational units, like teams or work groups, forces individual employees to reconcile different personal and group level interests, challenging them to manage multi-level interactions. Beyond their own pecuniary interests, individuals are tasked to resolve e.g. social and monetary within-group incentives. In turn, firms need to establish fitting monetary incentive-schemes that help their employees cope with the concurrent complexity of social and organizational structures. Here, understanding and leveraging group identity as an essential driver of human decision making (Akerlof and Kranton 2005, 2008) can open up new and revised instruments to further cooperative, productive group behavior.

One of the most prominent organizational tools to foster group performance are contests. Here, different work teams compete to achieve a pre-determined goal that is rewarded with some commonly known prize. Typical examples include R&D races, hackathons or collaborative research projects. These settings imply a complex behavioral frame. Teamwork introduces a classic social dilemma, where individuals try to find a trade-off between the social (contribute) and their individual (free-riding) optimum. Three mechanisms that have often been shown to reduce free-riding are (1) reciprocity due to repeated interactions with the same team members (Andreoni and Croson 2008; Chaudhuri 2011; Keser and Van Winden 2000; Rand and Nowak 2013), (2) competition between teams (Abbink et al. 2012; Puurtinen and Mappes 2009) and (3) group identity (Akerlof and Kranton 2005; Eckel and Grossman 2005). Importantly, the effect of competition is determined by the way it is implemented (Abbink et al. 2010; Tan and Bolle 2007) and particularly how team members benefit from winning (Gunnthorsdottir and Rapoport 2006; Kugler et al. 2010). Here, research predominantly consents that group-level incentives should be adjusted according to individual performance levels.

We argue that under group identity, this might not be optimal anymore. That is because group identity introduces new social conflicts. For instance, group identity impacts inequity aversion (Fehr and Schmidt 1999) among team members, leading to less envy under comparatively lower payoffs (Chen and Li 2009; Kim and Glomb 2014) and more guilt under comparatively higher payoffs (Jensen and Kozlovskaya 2016; Morell 2019). This changes the requirements for an effective monetary contest scheme, as social preferences become increasingly pivotal. Particularly when group contributions are very asymmetric and tensions between group members can arise, resolving this multi-level incentive structure becomes challenging.

In this paper, we therefore interact two distinct mechanisms designed to foster cooperation within organizational units. We explore how (artificial) group identity influences the willingness of individuals to cooperate in a standard public good game embedded in a contest situation with either equal or proportional prize sharing. The faintest form of group identity (low identity) will be introduced by simple partners-matching to consider repeated interaction among the same individuals. To strengthen group identity (high identity), we follow (Eckel and Grossman 2005) and exploit a pre-experimental puzzle task as well as team color tags. Our design thus captures different monetary and social incentives in simultaneous within- and between-group competitions. To the best of our knowledge, the interplay of specific prize sharing rules with the identities of competing groups has not yet been empirically examined.

Contests

Prior research shows that cooperation within groups can be exogenously encouraged by creating a contest situation between groups, as it forces the rational self-interested individual to cooperate with one's group members to win the conflict (Bornstein and Erev 1994; Sherif et al. 1961). Findings from recent experiments indicate a positive effect of intergroup competition (Ahn et al. 2011; Burton-Chellew et al. 2010; Katz et al. 1990; Kugler et al. 2010; Tan and Bolle 2007) as well as pseudo-competition (Burton-Chellew and West 2012) on the willingness to cooperate within groups. Among these, some studies focus on group performance and monetary incentives of different prize sharing rules. They investigate either equal prize sharing (Abbink et al. 2010; Cason et al. 2012; Sheremeta and Zhang 2010), compare equal or proportional prize sharing according to individual effort (Gunnthorsdottir and Rapoport 2006; Ke et al. 2013; Kugler et al. 2010) or examine nested contests, by which members of the victorious group compete for the prize (Ke et al. 2015). Overall, results consistently indicate higher levels of cooperation for proportional than equal prize sharing, even though cooperation under equal prize sharing is still above the predicted theoretical equilibrium (Hoffmann and Thommes 2020; Kugler et al. 2010; Sheremeta 2018).

Group Identity

When interacting in teams, group identity becomes a decisive factor. Early work rooted in social psychology points out that categorizing individuals in groups (Tajfel et al. 1971; Tajfel and Turner 1979), identifying in- and out-group individuals (Stets and Burke 2000; Tajfel 1974) and group salience by comparison or competition (Tajfel 1974, 1978) elicits in-group favoritism. Since then, group identity has also become a well established factor in economics (Akerlof 2002; Akerlof and Kranton 2000, 2005, 2010; Ashforth et al. 2011). Prior studies consistently confirm a positive effect on e.g. work motivation, performance, cooperation, and knowledge sharing (Ashforth and Mael 1989; Ashforth et al. 2011; Chen and Li 2009; Eckel and Grossman 2005; Goette et al. 2012; Solow and Kirkwood 2002; Van Knippenberg 2000; Weng and Carlsson 2015). Identity-homogeneous groups are more likely to reveal less negative reciprocity when group members deviate (Bicskei et al. 2016). Field experiments show that *real identity* reduces free-riding (Chowdhury et al. 2016) and competition among workers (Kato and Shu 2016). In particular, Kato and Shu (2016) show that workers only compete against those with a different social identity but not against their in-group co-workers.

Group Identity and Contest Prize Sharing

There is good reason to believe that group identity should behaviorally interact with different group-level incentive schemes. For instance, imagine a tight-nit work group with strong social cohesion whose members aim to achieve their goals collectively. In a world without identity, establishing a between-group competition that institutionalizes proportional prize sharing or proportional chances of winning introduces the strongest incentive for individuals to contribute. The more they contribute, the higher likelihood of receiving the prize, the higher their individual payout from said prize. However, a strong group identity might *disincentivize* group members from contributing too much. That is because very high contributions will be further removed from the group median, which induces disutility due to inequity aversion (Fehr and Schmidt 1999; Rey-Biel 2008). In particular, disproportionately high-contributing individuals who receive more than their colleagues will suffer feelings of guilt, which is strongly accentuated by a shared group identity (Everett et al. 2015; Güth et al. 2009; Morell 2019; Ockenfels and Werner 2014). Therefore, any prize sharing mechanism that disproportionally favors an individual over their group members will be increasingly costly for groups that share a strong identity.

Second, Fershtman and Gneezy (2001) suggest that people often opt for self-interested actions when they are socially acceptable. In such cases, they even increase their effort to earn money at the detriment of other people. We argue that in a contest, it is socially acceptable to increase one's own contributions in order to earn a higher share of the prize because it simultaneously increases the likelihood that group members earn additional money. That is, irrespective of identity, individuals will tend to contribute relatively high amounts. However, while group identity does neither change the competitive drive induced by contests, nor entirely remove the social excuse to contribute disproportional amounts because it also benefits the group, we hypothesize that under proportional prize sharing, group identity does change the prevailing social norm as well as the social and moral costs of not adhering to that norm (Christensen et al. 2004). Individuals are expected to (and do) care about the payoff of their group (Chen and Li 2009; Goette et al. 2006; Morita and Servátka 2013; Müller 2019), which partially conflicts with their own effort-level. Hence, without identity, people with proportional prize sharing will exert additional effort to secure the prize for themselves, which changes once their attachment to the group increases.

For egalitarian prize sharing, no such friction exists. Here, winning the prize does not increase inequality as everybody receives the same amount. In that case, group identity should increase group contributions unequivocally, as both the individual and their group stand to gain more money by winning the prize. Thus, it is likely that group identity increases cooperative behavior only so far as the additional monetary gains do not induce social frictions or disutility through increased inequity. Whereas the current literature largely favors proportional prize schemes, taking group identity into account could – at least to some extent – shift the balance in favor of equal prize sharing mechanisms.

For organizations, this could be an important insight, as egalitarian incentive mechanisms are inherently less competitive than proportional ones. This potentially reduces frictions within work groups (Bornstein and Gneezy 2002). Under some circumstances, within-team competition can reduce learning intentions (Heidemeier and Bittner 2012), knowledge sharing (Bartel et al. 2017; He et al. 2014), or helping efforts (Drago and Turnbull 1991). Other research finds negative effects of competition and competitive work climates on stress (Fletcher et al. 2008), workaholism (Keller et al. 2016), as well as substantial individual-level heterogeneity in how workers respond to competition (Jones et al. 2017). Therefore, providing managers and human resource personnel with additional information to leverage the benefits from between-group competition while avoiding within-group frictions enhances an organization's ability to build fitting institutional frameworks that help individuals resolve conflicts and ensure smooth operations and business goal attainment. But also beyond organizations, political architects who aim to use contests as an institutional driver of e.g. innovations (Halac et al. 2017) may benefit from considering the social nature of the groups and teams involved.

In line with our prediction, experimental results indicate a significant and positive effect of increased group identity on an individual's willingness to cooperate with egalitarian prize sharing only. In high-identity treatments, cooperation levels between egalitarian and proportional prize schemes are nearly identical. Overall, individuals contribute the most under egalitarian prize sharing with high group identity. Moreover, we find perfect cooperative behavior significantly more often in high identity treatments.[1]

2 Theoretical Framework

Following Gunnthorsdottir and Rapoport (2006), we introduce a market with n groups ($n \geq 2$) competing for an exogenous prize $S > 0$. Let m_k be the number of symmetric players in group k with $k = \{1, ..., n\}$ and $m_k = \{2, ..., K\}$. In total, there are $\sum_{k=1}^{n} m_k = N$ symmetric players in the market. Each player i, where $i = \{1, ..., N\}$, receives an endowment $e > 0$, which can be invested either in a public good or kept for themself. We assume the strategy space to be continuous, implying that individual i can contribute any share of their endowment e to the public good. We denote an individual's contribution to their group's public good by x_{ik} ($0 \leq x_{ik} \leq e$), the group's total contribution by ($X_k = \sum_{i=1}^{m_k} x_{ik}$) and the overall contributions in the market of all N players by X ($X = \sum_{i=1}^{N} X_k$).

Given the usual within-group conflict in standard public good games between investing or keeping their endowment, individuals' dominant strategy is to free-ride on other members, reducing the overall equilibrium contribution to zero. However, by introducing an exogenous prize, for which groups compete, an individual's incentive structure changes. This stems from the probability of winning being determined by group members' ability to cooperate, i.e. to generate a greater public good in comparison to other groups. Depending on prize value and its distribution, restraining from investing is no longer a dominant strategy (Gunnthorsdottir and Rapoport 2006).

Besides their pecuniary interests, individuals perceive themselves as parts of units, such as groups, rather than solely as independent entities. Therefore, an individual's public good contribution is crucially affected by their group identity. Accordingly, we extend our framework by incorporating individual valuation for cooperation with in-group members into their payoff welfare function. Further, we argue that group identity introduces concerns about inequity regarding the contest prize. Depending on the payment scheme, we model and predict differential effects of group identity on cooperation.

The within-group conflict
In accordance with public goods literature (Ledyard 1995; Zelmer 2003), within each group, the investments in the public good, X_k, are uniformly multiplied by the factor $1 < t < m_k$ and equally distributed among all members, leading to an actual payoff in group k from their public good of $\frac{tX_k}{m_k}$. The share of endowment not invested $(e - x_{ik})$ converts directly into the payoff function. Therefore, an individual's payoff from the within-group conflict is given by

$$\pi_{ik}^{in-group} = (e - x_{ik}) + \frac{tX_k}{m_k}. \tag{1}$$

[1] We provide access to the data via the following online repository: https://osf.io/bacvt/?view_only=4f84d243fbef4f759ef7fe06a463ccd3.

The between-group conflict
Like Gunnthorsdottir and Rapoport (2006), we extend the standard public good game with between-group competition by introducing an exogenous, probabilistic prize S for which all n groups compete. A group's probability of winning the prize depends on the group's total contribution relative to overall contributions in the market. A probabilistic contest success function models real-life circumstances more appropriately, as outstanding expenditures not necessarily ensure winning the competition (Gunnthorsdottir and Rapoport 2006). Let the probability of group k winning the prize denoted by

$$\Theta_k = \frac{X_k}{X}. \tag{2}$$

If group k wins the prize, it will be distributed among members according to the prize sharing function f_k given by

$$f_k = \frac{x_{ik}^c}{\sum_{i=1}^{m_k} x_{ik}^c}, \tag{3}$$

with $0 \leq c \leq \infty$. Notice that parameter c determines the type of prize sharing rule. For example, if $c = 0$, all members of the winning group k receive an equal share of the prize, $\frac{S}{m_k}$, which denotes completely egalitarian prize sharing. In contrast, if $c = 1$, each member of the winning group receives a share of the prize proportional to their individual investment to the public good, $\frac{x_{ik}}{X_k}$, which denotes completely proportional prize sharing. An individual's payoff from the between-group competition is given by

$$\pi_{ik}^{BG} = \Theta_k S f_k = \left(\frac{X_k}{X}\right) S \left(\frac{x_{ik}^c}{\sum_{i=1}^{m_k} x_{ik}^c}\right). \tag{4}$$

Payoff structure
An individual's overall expected payoff depends on the outcomes of both the within-group and the between-group conflict. By combining equation (1) and (4), we derive the expected payoff for individual i as a member of group k, given by

$$\pi_{ik} = (e - x_{ik}) + \frac{tX_k}{m_k} + \left(\frac{X_k}{X}\right) S \left(\frac{x_{ik}^c}{\sum_{i=1}^{m_k} x_{ik}^c}\right). \tag{5}$$

Equilibria without group identity
In our set-up, individuals are confronted with diverging investment motives. On the one hand, they have an incentive to keep their endowment for themselves. Given the within-group conflict, this has a higher expected payoff than investing in the public good. On the other hand, investments increase their probability of winning the prize in the between-group conflict.

Lemma 1: *In the unique Nash-equilibrium with symmetric players without group identity, individuals invest* $x_{ik}^* = S\left(\frac{n-1+cn(m_k-1)}{[nm_k]^2(1-\frac{t}{m_k})}\right)$ *in their public good.*

Proof: To determine an individual's equilibrium behavior, we derive the first order condition of equation (5)

$$\frac{\partial \pi_{ik}}{\partial x_{ik}} = \frac{t}{m_k} - 1 + S\left(\frac{x_{ik}^c(X - X_k)}{X^2 \sum_{i=1}^{m_k} x_{ik}^c} + \frac{cX_k(x_{ik}^{c-1}(\sum_{i=1}^{m_k} x_{ik}^c) - x_{ik}^c(\sum_{i=1}^{m_k} x_{ik}^{c-1}))}{X \sum_{i=1}^{m_k} x_{ik}^{2c}}\right). \tag{6}$$

As we assume symmetric players by rearranging and solving for x_{ik}, we obtain[2]

$$x_{ik}^* = S\left(\frac{n-1+cn(m_k-1)}{[nm_k]^2(1-\frac{t}{m_k})}\right). \tag{7}$$

Remark: Due to symmetry in players, this strategy is the unique Nash-equilibrium. However, this strategy is not dominant and depends on individual expectations about other players' investment strategies. Moreover, we exclude the case in which the expected payoff from between-group conflict is excessively high, indicating an employment of the dominant strategy where individuals invest their full endowment. We assume

$$S \leq \frac{e(nm_k)^2(1-\frac{t}{m_k})}{n-1+cn(m_k-1)}. \tag{8}$$

As outlined by equation (7), individual equilibrium investments increase with the value of c, the multiplier t and the value of the prize S. In this equilibrium, each individual receives an expected payoff of

$$\pi^* = e + S\left[\left(\frac{n-1+cn(m_k-1)}{[nm_k]^2(1-\frac{t}{m_k})}\right)(t-1)+\frac{1}{nm_k}\right]. \tag{9}$$

All terms in the bracket are strictly positive, since $t > 1$ and $\frac{t}{m_k} < 1$. Individuals are better off by playing the equilibrium strategy x_{ik}^* rather than free-riding, as in classic public good games. Consequently, introducing between-group competition moderates the within-group conflict and increases cooperation.

Equilibria with group identity

In the next step, we introduce group identity as a new variable. So far, individuals have adjusted their behavior solely based on pecuniary incentives. However, individual decision making is influenced by a person's sense of self, more specifically by the degree of identification with one's membership in groups. Therefore, investment decisions can be expected to depend on in-group attachments as high attachments will be accompanied by an urge for cooperation. In accordance with prior studies on identity (Akerlof and Kranton 2005; Ploner and Regner 2013), we first assume that individuals suffer from utility losses by exhibiting less cooperative behavior towards fellow group members.

As described by Akerlof and Kranton (2005), as well as Huettel and Kranton (2012), this effect depends on the situational context. If individuals do not feel attached to their group, we cannot expect a loss in utility from cooperative behavior. To model this, we introduce z_{ik} ($0 \leq z_{ik}$) as the degree to which an individual i from a group k identifies with their group. The effect of z_{ik} on the disutility caused by withholding parts of the endowment is decreasing in group size using $e^{-\delta_i(m_k-2)}$, where δ_i measures an individual's sensitivity to group size. That is because in public good dilemmas, the size of the group has been shown to correlate negatively with individual cooperation (Brewer and Kramer 1986).[3]

[2] For detailed calculations see Appendix B.

[3] We are aware that our assumption of full contribution as the social norm is quite strict. It might well be the case that a norm, denominated as w, is expected to be rather $x_{ik}^* \leq w \leq e$ instead of $w = e$. However, individuals might equally profit from a utility gain due to an overcontribution with $x_{ik} > w$ as compared to suffering from a utility loss by an underprovision with $x_{ik} < w$. Following this, assuming $x_{ik}^* \leq w \leq e$ instead of $w = e$ would not change the general argument of our analysis.

An individual's expected utility function, formerly outlined by equation (5), changes to

$$\pi_{ik}^{ID} = \left(1 - \frac{z_{ik}}{e^{-\lambda_i(m_k-2)}}\right)(e - x_{ik}) + \frac{tX_k}{m_k} + \left(\frac{X_k}{X}\right) S \left(\frac{x_{ik}^c}{\sum_{i=1}^{m_k} x_{ik}^c}\right). \tag{10}$$

The more an individual identifies with their group, associated with higher values of z_{ik}, the greater their loss in utility as a result of less cooperation. In contrast, individuals with no group attachment, i.e. $z_{ik} = 0$, will base their investment decisions purely on pecuniary aspects.[4] For simplification and because in our experiment, subjects played in small groups of four, we assume λ_i to be zero:

$$\pi_{ik}^{ID} = (1 - z_{ik})(e - x_{ik}) + \frac{tX_k}{m_k} + \left(\frac{X_k}{X}\right) S \left(\frac{x_{ik}^c}{\sum_{i=1}^{m_k} x_{ik}^c}\right). \tag{11}$$

Lemma 2: *With group identity and symmetric players, an individual's optimal investment strategy depends on their intensity of in-group attachment z_{ik}:*

(1) *if $z_{ik} < 1 - \frac{t}{m_k}$, individuals invest $x_{ik1}^* = S\left(\frac{n-1+cn(m_k-1)}{[nm_k]^2(1-\frac{t}{m_k}-z_{ik})}\right)$,*

(2) *if $z_{ik} \geq 1 - \frac{t}{m_k}$, individuals invest their full endowment, $x_{ik2}^* = e$.*

Proof: In order to determine optimal behavior, we derive the first order condition of equation (11), given by

$$\frac{\partial \pi_{ik}^{ID}}{\partial x_{ik}} = z_{ik} - 1 + \frac{t}{m_k} + S\left(\frac{x_{ik}^c(X - X_k)}{X^2 \sum_{i=1}^{m_k} x_{ik}^c} + \frac{cX_k(x_{ik}^{c-1}(\sum_{i=1}^{m_k} x_{ik}^c) - x_{ik}^c(\sum_{i=1}^{m_k} x_{ik}^{c-1}))}{X \sum_{i=1}^{m_k} x_{ik}^{2c}}\right). \tag{12}$$

However, note that by introducing z_{ik}, incentives for investments in the public good have changed. Even while assuming that (8) still holds, it might become an individual's dominant strategy to invest their full endowment depending on the actual value of z_{ik}. We need to distinguish between two cases:

Case 1: $z_{ik} < 1 - \frac{t}{m_k}$
By determining the symmetric Nash equilibrium from equation (12), we derive individual optimal investment strategy

$$x_{ik_1}^* = S\left(\frac{n - 1 + cn(m_k - 1)}{[nm_k]^2(1 - \frac{t}{m_k} - z_{ik})}\right). \tag{13}$$

Case 2: $z_{ik} \geq 1 - \frac{t}{m_k}$
In this case, it is the strict dominant strategy for individuals to invest their full endowment. Since the expected payoff for deviating from full cooperation is negative, we derive

$$x_{ik_2}^* = e. \tag{14}$$

[4] One could argue that people with high group identity also derive greater pleasure, excitement or utility from winning the contest prize S (see e.g., Babcock et al. 2015; Charness and Holder 2019). We refrain from modelling such an additional positive effect since (1) it does not change any fundamental predictions of this model and (2) there is little to no evidence linking group identity to affective or emotional reactions following a successful group competition.

Remark: In both the cases, we observe strictly higher investments in public goods with increased group identity compared to the case of low identity, as long as $z_{ik} > 0$. Notice, a higher number of in-group members m_k, necessitates either higher individual in-group attachment z_{ik} or a larger multiplier t for the public good, in order to attract an individual's full cooperation. This constitutes the circumstance of increasing difficulties to sustain stable cooperation with increasing group size.

Since stable cooperation leads to higher expected payoffs, overall welfare also increases. Absent any considerations regarding within-group equality, higher group attachment leads to expecting more cooperation within the group by higher investments in the public goods.

Egalitarian and Proportional Prize Sharing

We hypothesize that identity has a greater influence on group contributions under egalitarian prize sharing. The first argument concerns differential group salience in an individual's expected payoff function conditional on the sharing rule. Remember, the influence of the prize sharing rule is represented by c as determined through the function $f_k = \frac{x_{ik}^c}{\sum_{i=1}^{m_k} x_{ik}^c}$. With proportional prize sharing and $c = 1$, the expected payoff from (10) simplifies to

$$\pi_{ik}^{ID} = (1 - z_{ik})(e - x_{ik}) + \frac{tX_k}{m_k} + \frac{x_{ik}}{X}S. \tag{15}$$

With egalitarian prize sharing and $c = 0$, the expected payoff follows

$$\pi_{ik}^{ID} = (1 - z_{ik})(e - x_{ik}) + \frac{tX_k}{m_k} + \frac{X_k}{X}\frac{S}{m_k}. \tag{16}$$

Whereas the expected payoff for the prize in (15) is similar to an individual contest with N players and dependent on an individual's contribution x_{ik}, the expected prize under egalitarian prize sharing (16) is increasing in group-expenditure X_k. Since group attachment is expected to function on the group-expenditure level, this makes a clear argument for increased group identity only affecting cooperation when the prize is shared equally amongst all group members.

Inequity Aversion

Thus far, the model abstracted from inequity considerations on group level. However, the level of inequity between group members is crucially affected by the prize scheme. Furthermore, higher group identity should increase the degree of disutility players experience from earning more than their group members. Increasing group attachment under proportional prize sharing then increases (or introduces) that conflict, whereas group attachment under egalitarian prize sharing has no effect on group inequality. This is the second argument for a positive effect of identity on cooperation that is limited to egalitarian prize schemes. To introduce the disutility a player receives when earning more at the expense of their group (i.e. feeling guilty), we extend (4), the expected payoff from the between-group competition. We use $U(\pi_{ik}^{BG})$ to model individual i's utility from receiving π_{ik}^{BG}:

$$\begin{aligned} U(\pi_{ik}^{BG}) = \pi_{ik}^{BG} - c\Bigg(&(1 - z_{ik})\frac{\alpha_i}{m_k - 1}\sum_{j \neq i} max[\pi_{jk}^{BG} - \pi_{ik}^{BG}, 0] - \\ &(1 + z_{ik})\frac{\beta_i}{m_k - 1}\sum_{j \neq i} max[\pi_{ik}^{BG} - \pi_{jk}^{BG}, 0]\Bigg), \end{aligned} \tag{17}$$

where $0 \leq \alpha_i$ and $0 \leq \beta_i < 1$. An individual's expected utility function changes to:

$$\pi_{ik}^{IE} = (1 - z_{ik})(e - x_{ik}) + \frac{tX_k}{m_k} + \left(\frac{X_k}{X}\right) S \left(\frac{x_{ik}^c}{\sum_{i=1}^{m_k} x_{ik}^c}\right) - c\Bigg((1 - z_{ik})$$
$$(\alpha_i \frac{1}{m_k - 1} \sum_{j \neq i} max[\pi_{jk}^{BG} - \pi_{ik}^{BG}, 0]) + (1 + z_{ik})\beta_i \frac{1}{m_k - 1} \sum_{j \neq i} max[\pi_{ik}^{BG} - \pi_{jk}^{BG}, 0]\Bigg). \tag{18}$$

For equal prize sharing, $c = 0$, inequity does not change and an individual's utility function follows (11). For simplicity, we also assume linearity in inequity aversion.

We assume no disutility from advantageous inequality when group identity $z_i = 0$ and a negative effect of group identity on envy, i.e. disadvantageous inequity. This second assumptions warrants some explanation. Previous studies on in-group bias and social preferences consistently show higher tolerance towards selfishness when perpetrated by in-group members (see e.g., Baumgartner et al. 2012; Bernhard et al. 2006; Delton and Krasnow 2017; Jordan et al. 2014; McAuliffe and Dunham 2017; McLeish and Oxoby 2011; Stagnaro et al. 2018). Chen and Li (2009) find group identity to decrease aversions towards disadvantageous inequity by up to 93%. Evidence from the organizational literature suggests work-group identity to decrease victimization caused by envy (Kim and Glomb 2014) and work in psychology shows very similar processes in children (Gaviria et al. 2021). On the other hand, Jiang and Li (2019) show in a principal-agent game that in-group agents are less tolerant of low offers from the principal (disadvantageous inequity), while exerting greater reciprocity as a response to generous offers, thereby offsetting advantageous inequity. While this is in line with our prediction regarding the effect of group attachment on advantageous inequity, it also suggests that the interplay between group identity and inequity attitudes is context dependent. Prior work from e.g. Afridi et al. (2015) show that institutions, and in particular tournament competition, can mediate the effect of group identity on behavior. Overall, while the evidence for an effect of group identity on advantageous inequity aversion is more straightforward, the majority of research also points to a moderating effect on envy. We further hypothesize that the between-group competition in our model motivates individuals to strive for the prize, irrespective of identity, and that this greatly diminishes the salience of disadvantageous inequity. In contrast, it directly relates to the unavoidable consequences of changes in equality of outcomes, increasing the salience of advantageous inequity to the detriment of the group. Thus, we think that the predicted ambiguity for proportional prize schemes is primarily driven by guilt, rather than envy. However, since our experiment does not differentiate between the two modes of inequity aversion, and the relative strength of the two effects does not change our predictions since they are positively correlated, we refrain from modelling arbitrary weights conditional on the inequity term.

Proposition 1: Strong group attachment has an ambivalent effect on cooperation under proportional prize sharing.

As shown in (18), π_{ik}^{IE} equals π_{ik}^{ID} for $c = 0$. For $c = 1$, an individual experiences lower utility from π_{ik}^{BG} if $(1 - z_{ik})(\alpha_i \frac{1}{m_k - 1} \sum_{j \neq i}^{m_k} (\pi_{jk}^{BG} - \pi_{ik}^{BG})) + (1 + z_{ik})\beta_i \frac{1}{m_k - 1} \sum_{j \neq i}^{m_k} (\pi_{ik}^{BG} - \pi_{jk}^{BG}) > 0$. Subjects with proportional prize sharing therefore experience disutility when receiving a lower share of the prize than their group members, as well as when they receive a higher share. For proportional prize sharing, the utility function (18) simplifies (assuming players $j \neq i$ are symmetric) to:

$$
\begin{aligned}
\pi_{ik}^{IE} = (1 - z_{ik})(e - x_{ik}) + \frac{tX_k}{m_k} + \left(\frac{X_k}{X}\right) S \left(\frac{x_{ik}}{X_k}\right) - \\
c\left(\frac{S}{X}\right)\Big((1 - z_{ik})\alpha_i(x_{jk} - x_{ik}) + (1 + z_{ik})\beta_i(x_{ik} - x_{jk})\Big)
\end{aligned}
\tag{19}
$$

This first makes another argument why overall, cooperation under proportional prize sharing should be higher. Low contributions, x_{ik}, correlate with a higher likelihood of receiving a relatively small share of the prize and thus trigger disadvantageous inequity aversion. Second, disutility from advantageous inequity increases with group attachment z_{ik}. The positive effect of group identity on cooperation induced by disutility from not cooperating with one's group is (partially) offset by the possibility of guilt. Third, group identity decreases disutility from envy, alleviating an individual's incentives to contribute a high x_{ik} in order to secure an equal share of the prize.

Isolating the additional effect of z_{ik} on the first derivative of equation (19) $\frac{\partial \pi_{ik}^{IE}}{\partial x_{ik}}$ reveals the following necessary, but not sufficient condition for increased group identity to decrease group contributions under proportional prize sharing:[5]

$$
S > \frac{X^2}{X(\alpha + \beta) - x_{ik}(\alpha + \beta)}. \tag{20}
$$

If the contest prize S is too small, inequity aversion will not deter cooperation under proportional prize sharing. In contrast, higher prizes exacerbate the problem.

Figure 1 shows simulations for the expected utility (18) given the contribution x_{ik} and varying levels of group identity z_{ik} under the parameters of this experiment (see section 3). Here, we assume that all other players j are symmetric and contribute $x_j = 25$. We also assume $\alpha = 1$ and $\beta = 0.9$. As we show in the appendix (Figures 7–9), the general tendency holds for a variety of combinations for α, β and x_j. With moderate to low expected contributions from all other members and relatively high levels of β, increasing group identity under a proportional prize scheme shifts optimal contributions towards the minimum. Note that this effect would increase in group size m_k as well as prize S. Contrary, under egalitarian prize sharing, identity unequivocally increases the attractiveness of full contributions as selfish behavior induces additional disutility.

3 Experimental design

We modify and extend the public good game used by Gunnthorsdottir and Rapoport (2006) by introducing partners-matching and increased group identity to test the combined effect of different monetary incentive structures and different levels of group identity on cooperative behavior.

Within their groups, subjects play a standard public good game while simultaneously competing for an external and commonly-known prize (contest). The prize is distributed among members of the winning group by either the egalitarian, $c = 0$, or the proportional prize sharing rule, $c = 1$, with the probability of winning depending is determined by the contest success function (see (2)) on the contributions of all players to their public good. Payoffs are denominated in tokens, accumulated over periods and paid at the end of the experiment, with one token converts to EUR 0.01.

[5] See 5 in the appendix.

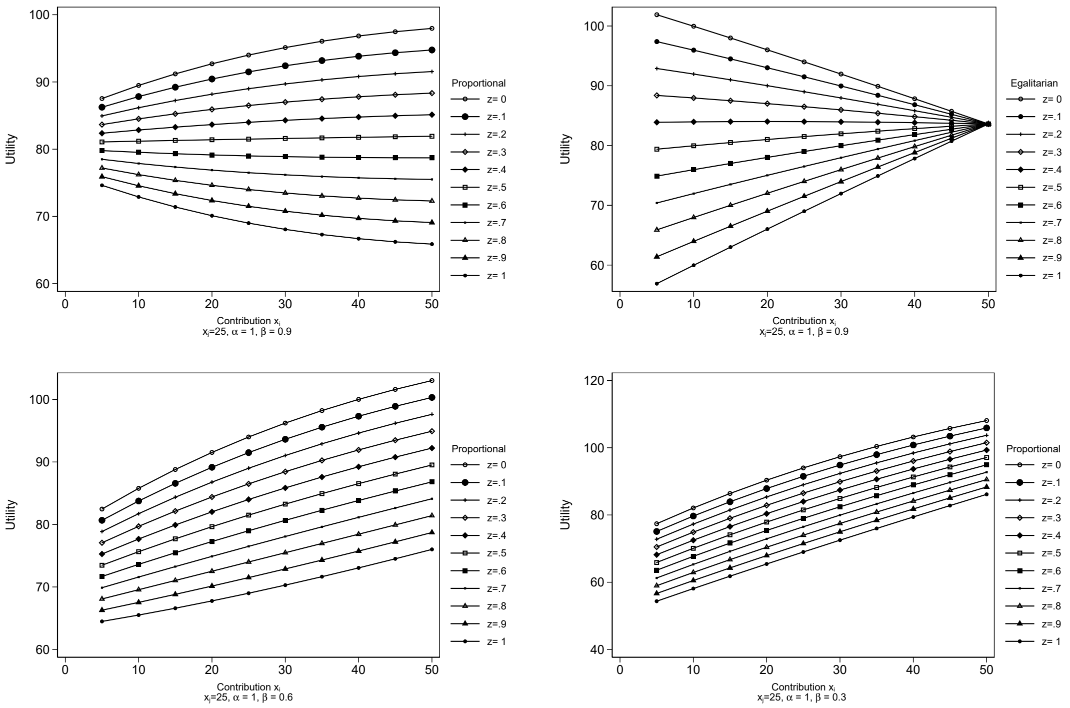


Figure 1 Expected utility for different individual contributions x_{ik} in the inequity paradigm. Since inequity considerations do not matter under egalitarian prize sharing, increased identity shifts optimal behavior towards full contributions. Contrary, under proportional prize sharing, that effect is conditional on the inequity parameters as well as the expectations regarding the behavior of other group members.

In each session, eight subjects participate and are randomly assigned to two groups of four, i.e. $n = 2$ and $m = 4$. The allocation remains constant over all ten periods, $T = 10$. At the beginning of each round, subjects decide how much of their endowment, $e = 50$, they want to invest in the public good. We set the multiplying factor $t = 2$ to generate a marginal per capita return (MPCR) of 0.5. The exogenous prize is set to $S = 152$; equal to Gunnthorsdottir and Rapoport (2006). Finally, after each round, subjects are informed about the prize-winning team, the underlying winning probabilities, their own payoff from the current round, as well as the accumulated payoff earned over all periods played so far.

To generate different values for z_{ik} we manipulate a player's in-group attachment. In *low identity* treatments, we allocate subjects randomly to either group 1 or group 2, without informing the subjects about the identity of their group members (minimal group identity).

In contrast, in *high identity* treatments, group identity is artificially increased by using the puzzle task of Eckel and Grossman (2005). In this unpaid team-building task, group members have to work together to construct a puzzle prior to the experiment. The puzzle comprises of five different pieces that add up to a square[6]. However, since no member possesses all five necessary pieces at the beginning, subjects have to engage in trading with their group members. For this purpose, they are seated within their group, and are allowed to talk and support each other while solving the task. Moreover, we also use the

[6]For detailed instructions, see the online appendix. Furthermore, to avoid agreements and strategical planning, groups were given no prior information about the following public good game they were supposed to participate in.

special labeling of "group red" and "group green" rather than group 1 and 2, and equip subjects with color tags they can pin to their clothes, indicating their group affiliation. Likewise, puzzle pieces were either red or green, according to their group color.

To enable comparison among treatments, each subject in *low identity* treatments has to solve the same -yet grey- puzzle on their own, without knowing that they will be playing in groups in the subsequent public good experiment.

After all subjects in the session finished the puzzle task and materials were collected by the experimenter, each subject was placed alone in a cubicle and instructed regarding the subsequent public good game. Cubicles were equipped with pen and paper to allow note taking. After reading the instructions, subjects had to pass four comprehension questions to ensure their understanding of the instructions.

At the end of the experiment, subjects completed a questionnaire. Besides general socioeconomic questions, we asked for individuals' in-group attachment.

Treatment conditions
According to our research question, we employed four different experimental treatment conditions that differ in terms of group identity and prize-sharing mechanism. In *low identity* treatments, subjects solve the puzzle task in separation and are subsequently randomly allocated to group 1 or group 2. Within low identity, the prize is either distributed equally (Low ID EG) or proportionally (Low ID PR) among the winning team's members. In *high identity*, subjects are allocated to group green or group red and solve the puzzle task in cooperation with their team members. Communication is allowed during the puzzle task. Again, with high identity, the prize is either distributed equally (High ID EG) or proportionally (High ID PR) among the winning team's members as well.

Hypotheses

In order to predict subjects' cooperative behavior, we insert our experimental parameters to our theoretical framework. By taking individual group identity into consideration, we first expect higher investments in public goods with increased identity, resulting from higher values of z_{ik}. Following the literature, we expect people to embody altruism and act as conditional cooperators in a group, especially when they perceive themselves as a part of it. We assume these tendencies will increase when group identity is made more salient by intergroup competition. Under the egalitarian prize sharing rule, this effect is unambiguous. Second, under the proportional prize sharing rule, we expect intergroup competition to introduce additional inequity concerns. Following our theoretical framework and the literature on group identity and social preferences, interacting inequity concerns with higher group attachment decreases investments in the public good. This is driven by decreased aversion towards disadvantageous inequity and increased aversion towards advantageous inequity. Overall, the effect of group identity on cooperation within a proportional between-group contest is ambivalent, and we therefore expect no aggregate change in contributions.

***Hypothesis 1*:** Under the egalitarian prize sharing rule, investments in public goods are higher with increased group identity.

***Hypothesis 2*:** Under the proportional prize sharing rule, investments in public goods do not change with increased group identity.

In low identity treatments, we expect investments in public goods to be higher under the proportional rather than the egalitarian prize sharing rule, given their different monetary incentives (Gunnthorsdottir and Rapoport 2006; Kugler et al. 2010). Individuals adjust their investment behavior according to the prize-sharing mechanism, with a higher investment resulting from a higher value of c. We expect subjects to invest more in public goods under the proportional sharing rule, $c = 1$, than under the egalitarian sharing rule, $c = 0$. Concurrent with the first two hypotheses, we expect that difference to largely vanish in the high identity treatments.

***Hypothesis 3*:** With low group identity, investments in public goods are higher under the proportional than under the egalitarian prize sharing rule.

***Hypothesis 4*:** With high group identity, investments in public goods do not differ between the proportional and the egalitarian prize sharing rule.

Based on our theory, if z_{ik} exceeds the value of $1 - \frac{t}{m_k}$ under the egalitarian prize sharing rule, full cooperation becomes the dominant strategy. By increasing group identity, more individuals are expected to become full cooperators, defined as individuals who invest their entire endowment, $e = 50$ tokens, in each round. Contrary, we cannot make that prediction for cooperation under proportional prize sharing, as full cooperation always increases expected disutility from advantageous inequity. Ceteris paribus, the change induced by high group identity should therefore be higher under egalitarian payout.

***Hypothesis 5*:** Under the egalitarian prize sharing rule, the number of full cooperators increases with increased group identity.

Experimental procedure and data

The experiment was conducted between May and July 2016 in the Laboratory for Behavioral Economics at the University of Goettingen. The experiment was programmed using zTree (Fischbacher 2007) and participants were recruited with ORSEE (Greiner 2015). We implemented our 2 x 2 design, crossing the dimensions Low ID/High ID and EG/PR. Before starting the main task, subjects went through the puzzle task either alone (Low ID) or as a team (High ID). The 320 subjects participated in 40 sessions lasting about 45 minutes, resulting in average earnings of EUR 11.17. Approximately 53% of the participants were female. Overall, the average participant age was 24 and roughly 41% of the participants were economics or business administration students.

4 Results and Discussion

To increase group identity above the minimum, we asked subjects to puzzle prior to the experiment. In order to check for group identity manipulation, the post-experimental questionnaire comprises three questions about in-group attachment. Participants were asked to indicate on a scale from 1 (=absolutely not) to 10 (=extremely), how strongly they feel attached to their group (Chen and Li 2009), and indicate their agreement from 1 (=not at all) to 10 (=very much) to the following statements: "I feel I am really a part of my group." (adapted from the Michigan Organizational Assessment Questionnaire Korsgaard et al. 1995) and "If I had a chance to do the same experiment for the same payout in another group, I would still stay here in this group." (Wech et al. 1998). We calculated an equally weighted additive index among all three questions (Cronbach's alpha: 0.805).

According to our theoretical model, the strength of in-group attachment, symbolized by z_{ik}, is crucial to the amount of investment. We find a significant difference in affiliation-ratings by individuals in low and high identity conditions ($\bar{x}_{\text{High ID}} = 6.19$, $\bar{x}_{\text{Low ID}} = 5.26$; Wilcoxon rank-sum test (WRS): $z = 3.273, p = 0.001$). While this could be affected by the experiment, since affiliation was surveyed post-experimentally, contributions in the high-identity treatments are significantly higher starting in round one (WRS[7]: $z = -2.072, p = 0.0382$; also see figure 2). This confirms the effectiveness of our puzzle intervention.

Descriptives

Table 1 shows mean contributions and distributive indices across all four conditions. On average, players contribute 84% (42.03 tokens) of their endowment, which is a comparably high amount (see e.g., Gunnthorsdottir and Rapoport 2006). This indicates that introducing between-group competition paired with group identity leads to highly cooperative behavior. Mean contributions are lower for low (89,5% of the endowment) than for high identity treatments (96%). Aggregate differences between prize sharing mechanisms are marginal: 83,7% of the endowment is contributed under egalitarian prize sharing, while 84,5% is contributed under proportional prize sharing. Disaggregating by identity shows that this is mainly due to the differential effect of group identity conditional on the prize scheme. Contrary to previous findings of stranger-matching public good games without contests (e.g., Andreoni 1988; Fischbacher et al. 2001; Ledyard 1995), overall contributions are fairly stable over time, despite the typical increasing effect from round one to two and an end-game effect.

Table 1 Summary statistics

		Low ID			**High ID**		
	Overall	*All*	*EG*	*PR*	*All*	*EG*	*PR*
Mean contribution	42.03	40.55	39.27	41.83	43.51	44.38	42.64
SD	10.4	10.77	10.57	10.88	9.83	9.73	9.91
25%-quantile	38.85	35	32.8	39.7	41.2	43.25	40.6
50%-quantile	45.8	44.75	40.75	46.85	48	49.85	45.5
N	320	160	80	80	160	80	80

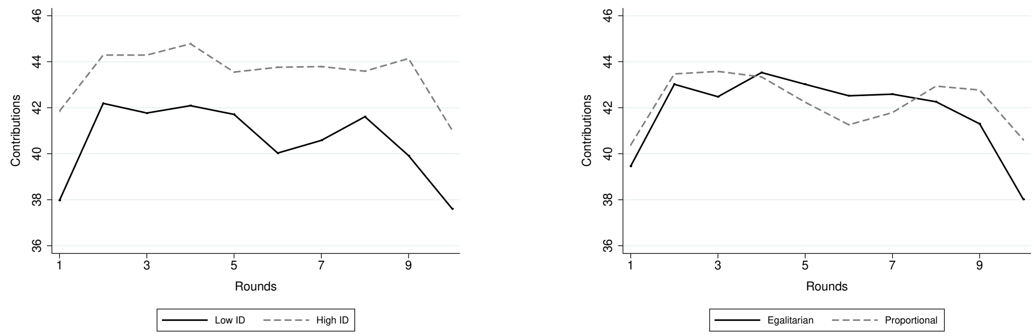
Note: Table 1 displays statistics of individuals decisions, therefore, N denotes the number of individuals in each treatment.

Group Identity and Cooperation

Average aggregated group contributions between low and high identity conditions are significantly different (WRS: $z = -2.056, p = 0.0398$). This is in line with previous research indicating that group identity positively influences the level of contribution (Chen and Li 2009; Eckel and Grossman 2005), i.e. cooperative behavior. However, distinguishing the effect according to the underlying prize sharing rule shows that the positive effect of

[7] All Wilcoxon rank-sum tests concerning cooperative behavior are conducted on session level, thus using 40 observations.

Figure 2 Average investments for low and high identity (*left*) and egalitarian and proportional prize sharing (*right*) per round.



group attachment on cooperation is solely driven by increased contributions under the egalitarian prize rule (see figure 3). Here, subjects in the high identity condition cooperate significantly more (WRS: $z = -2.343, p = 0.0191$), whereas we document no such effect for proportional prize sharing (WRS: $z = -0.454, p = 0.6501$)[8].

***Result 1*:** We confirm Hypotheses 1 and 2. Under the egalitarian prize sharing rule, investments in public goods are higher with increased group identity. Under the proportional prize sharing rule, investments in public goods do not change with increased group identity.

In line with our predictions, the unambiguously positive effect of identity on group contributions cannot be replicated once the contest potentially induces outcome differences between group members. We attribute this primarily to identity-induced concerns about advantageous inequity as well as reduced anticipation of envy. Another possibility is that an equally shared prize generally increases the fear of exploitation since free-riders proportionally gain more from shirking as compared to proportional prize sharing. In so far as the equal prize sharing mechanism aggravates the social dilemma of public goods, group identity has more room to moderate cooperation-inhibiting factors. If that were true, we would expect significantly lower contributions under egalitarian prize sharing without the identity manipulation. As we will show now, this is not the case.

Prize Sharing and Cooperation

If cooperative behavior is monetarily rewarded, cooperation becomes a rational strategy. The results of Gunnthorsdottir and Rapoport (2006) indicate that proportional rewarding can outperform equal rewarding, explained e.g. by the non-satiation axiom of choice theory. However, we find no significant difference between prize sharing rules, either over the whole sample (WRS: $z = 0.0000, p = 1.000$), between the low identity conditions (WRS: $z = 1.209, p = 0.227$), or the high identity conditions (WRS: $z = -1.436, p = 0.151$). There is a tendency for higher contributions under proportional prize sharing when identity is lower (which reverses in the high identity conditions) as predicted by our theoretical model (13), but overall, we reject Hypothesis 3 (see figure 4).

[8] Due to a cross-over between the two identity conditions (see figure 3), we checked for significant differences from round 4 to round 10, but the difference is still insignificant (WRS: $z = -0.68, p = 0.4963$)

Figure 3 Average investments for low and high identity under egalitarian and proportional prize sharing per round.

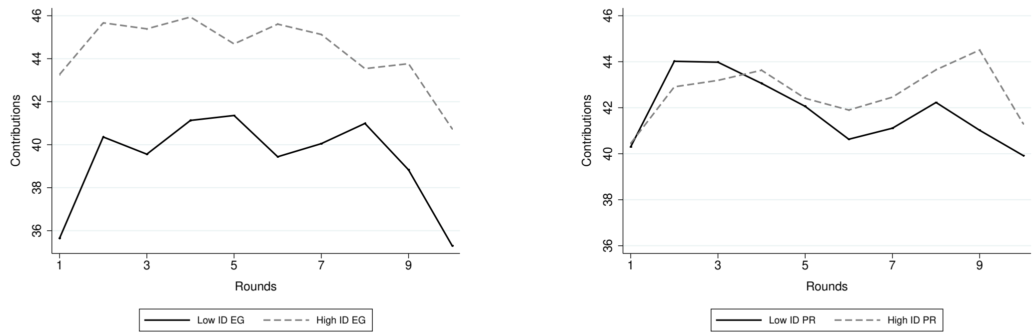
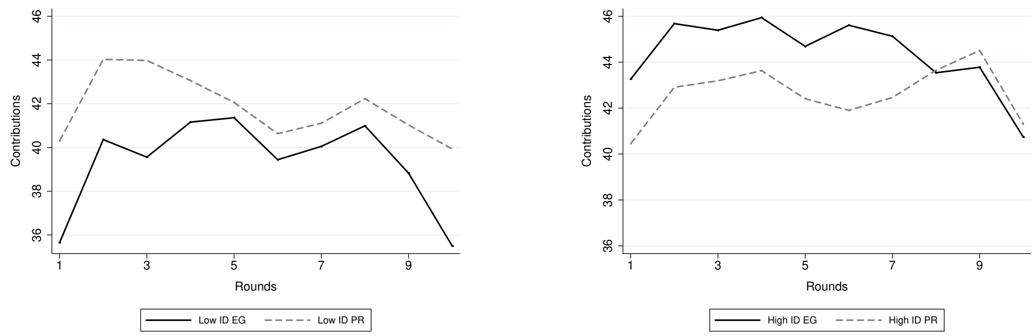


Figure 4 Average investments for egalitarian and proportional prize sharing with low and high identity per round.



Result 2: We reject Hypothesis 3 and confirm Hypothesis 4. Investments in public goods do not differ between proportional and egalitarian prize sharing.

Even though differences between prize sharing rules with high identity are insignificant over all ten rounds, figure 4 indicates a difference until round seven. Due to a cross-over, we check for significance in rounds one to seven (therefore excluding the end-game effect) and find a significant difference indicating a stronger effect of group identity under egalitarian prize sharing (WRS: $z = 1.890, p = 0.059$).

This result supports our previous findings: Higher group identity is only effective with egalitarian prize sharing and leads to higher and more persistent cooperative behavior. Moreover, group identity paired with a suitable institutional framework, i.e. egalitarian prize sharing, outperforms the monetary incentive structure of proportional prize sharing in terms of cooperative behavior.

Result 2a: Excluding the end-game effect, there is a significant reversal in contributions following induced group identity. While contributions without the intervention tended to be higher for proportional prize sharing, contributions in high identity conditions are higher with an equally shared prize.

Full-contributors

As derived by theory, if z_{ik} exceeds or equals the critical value of $1 - \frac{t}{m_k}$, full contribution becomes the dominant strategy under egalitarian prize sharing. Overall, almost a third

(31.25%) of all participants are full contributors, meaning individuals contributing 50 tokens in all 10 rounds, with significantly more full contributors in high identity than in low identity treatments ($\chi^2(1) = 11.4036, p = 0.001$, see table 2). Consistent with our theoretical model, full contributors indicate significantly higher values for z_{ik} than others (WRS: $z = -4.134, p = 0.000$), with a mean affiliation-index of 6.6 compared to 5.3.

Result 3: We find significantly more full contributors in high identity treatments.

Result 3a: Full contributors report feeling more attached to their group.

Also in line with our prediction, we only find a significant, positive effect of increased group identity on full contribution behavior under egalitarian prize sharing ($\chi^2(1) = 14.4166, p = 0.000$). The combination of egalitarian prize sharing and increased group identity results in perfect cooperative behavior for 50% of all participants.

Result 3b: We confirm Hypothesis 5. Combining high identity and egalitarian prize sharing results in a 50%-share of perfect cooperative behavior.

Individual Cooperative and Defecting Behavior

As our results show comparatively high and fairly stable contributions over time (Fischbacher et al. 2001; Ledyard 1995), we dive deeper into our data and analyze individual and single decision behavior. While we cannot account for group and session-level confounds, this analysis serves both as a robustness check as well as a pointer towards possible explanations for the documented treatment effects (see Table 2).

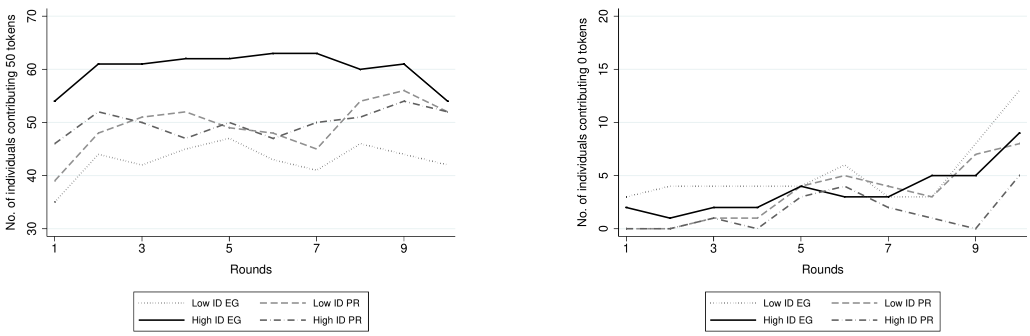
Table 2 Cooperative, defecting and conditional cooperative behavior

	Low ID		**High ID**		
	EG (T1)	*PR (T2)*	*EG (T3)*	*PR (T4)*	*Total*
No. of full contributors*	17 21.25%	19 23.75%	40 50%	24 30%	100 31.25%
No. of temporary full contributors*	70 87.5%	74 92.5%	71 88.75%	72 90%	287 89.69%
No. of single full contribution decisions**	429 53.63%	494 61.75%	601 75.13%	499 62.38%	2023 63.22%
No. of temporary full defectors***	24 30%	10 12.5%	15 18.75%	9 11.25%	58 18.13%
No. of single full defection decisions**	52 6.25%	33 4.13%	36 4.5%	16 2%	137 4.28%
Below group's average investment**	252 31.5%	235 29.38%	145 18.13%	221 27.63%	853 26.66%
Below own prior investment****	155 21.53%	111 15.42%	70 9.72%	109 15.14%	445 15.45%

* On the basis of all 320 individuals equally distributed over treatments.
** On the basis of all 3200 single decisions; 320 individuals and 10 rounds.
*** On the basis of all 320 individuals. We find no full defectors, i.e. individuals contributing zero in every round.
**** Calculations start in round two, i.e. calculation on the basis of 2800 single decisions.

For temporarily full contributors (individuals who contribute 50 token in at least one round), we find approximately similar and large shares across all four conditions

Figure 5 No. of full contributions and fully free-riding decisions.

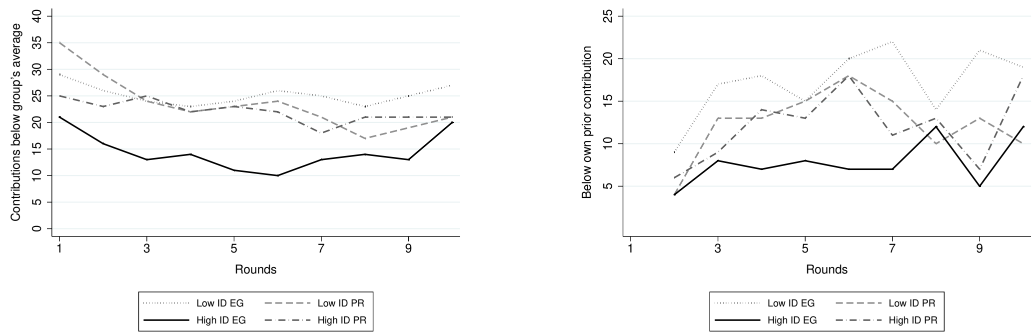


(89.69% of all decisions) and therefore, no significant differences. However, for single full contribution decisions, the lowest (53.63%) and the highest (75.13%) share occurs under egalitarian prize sharing. Increasing identity has a significant, positive effect under egalitarian prize sharing (T1 vs. T3; $\chi^2(1) = 80.62, p = 0.000$) and results in a significantly higher share of perfect cooperative behavior than under proportional prize sharing (T3 vs. T4; $\chi^2(1) = 30.27, p = 0.000$). Thus, *High ID EG* is a clear positive outlier (see figure 5 (left)).

Across all conditions, we find no individual fully defecting, i.e. no individual contributed zero tokens in all ten rounds. Indeed, only 18.13% of all 320 individuals contributed nothing at least once. Most of these defections happened with egalitarian prize sharing and low identity (T1). If group identity is low, proportional prize sharing significantly reduces defecting behavior (T1 vs. T2; $\chi^2(1) = 7.32, p = 0.007$). Increasing group identity under egalitarian prize sharing also reduces free-riding behavior (T1 vs. T3; $\chi^2(1) = 2.75, p = 0.097$). Overall, we find significant differences between prize sharing rules ($\chi^2(1) = 8.42, p = 0.004$), but not for identity. Besides, temporary full defectors rate their group affiliation significantly lower ($\bar{x}_{\text{TFD}} = 4.28$) as those who never full defected ($\bar{x}_{\text{No TFD}} = 6.04, WRS: z = 4.775, p = 0.000$).
On the single decision level, only 4.28% of all single decisions are fully free-riding decisions. From figure 5, we see very few individuals deciding to fully free-ride in the first rounds, while shares increase over time. Despite weak visual differences from figure 5, we find a significant, positive effect of identity on the reduction of single full defection decisions ($\chi^2(1) = 8.3, p = 0.004$). This holds true for egalitarian prize sharing (T1 vs. T3; ($\chi^2(1) = 3.08, p = 0.079$) and for proportional prize sharing (T2 vs. T4; $\chi^2(1) = 6.08, p = 0.014$). Controlling for identity, a proportional prize sharing mechanism reduces fully free-riding decisions (T1 vs. T2: $\chi^2(1) = 4.49, p = 0.034$ and T3 vs. T4: $\chi^2(1) = 7.95, p = 0.005$). In sum, the combination of high identity and proportional prize sharing appears most optimal for reducing single free-riding decisions.

Finally, we compare an individual's contribution compared to their group's average and their own prior contribution. Both measures indicate a self-serving bias and partially explain the declining contribution trend in standard public good games (Fischbacher et al. 2001; Ledyard 1995). As depicted in figure 6, the number of individuals contributing below their group's average is slightly U-shaped and overall relatively small. This is in line with our previous findings of positive, stable contributions over time achieved by between-group competition and some kind of group identity. Under egalitarian prize sharing, increasing group identity helps overcome the self-serving bias as represented by contributions below a

Figure 6 Below group's average and own prior contribution.



group's average (T1 vs. T3: $\chi^2(1) = 38.36, p = 0.000$). Overall, *High ID EG* again clearly outperforms all other conditions (T3 vs. T4: $\chi^2(1) = 20.46, p = 0.000$).

Figure 6 (right) displays the number of individuals whose contribution declined in subsequent rounds. Again, high identity and egalitarian prize sharing weaken self-serving behavior in form of reduced contributions when compared to one's own prior contributions (T1 vs. T3: $\chi^2(1) = 38.06, p = 0.000$). Proportional prize sharing only has a positive effect when identity is low (T1 vs. T2: $\chi^2(1) = 8.93, p = 0.003$). Again, *High ID EG* is the most effective combination (T3 vs. T4: $\chi^2(1) = 9.7, p = 0.002$).

Result 4: Over all four conditions, high group identity paired with an egalitarian prize sharing rule maximizes (full) cooperative behavior and minimizes self-serving behavior. Combining high group identity with a proportional prize minimizes full defective behavior.

5 Conclusion

We show that group identity has a significant, positive effect on cooperative behavior in a repeated contest public good game. However, this effect is moderated by the institutionalized prize sharing rule. As theoretically predicted, group identity only increases cooperation under egalitarian prize sharing. Under the proportional prize rule, identity has no measurable effect. Overall, interacting high group identity with an egalitarian between-group contest clearly outperforms all other setups.

This contrasts a lot of the prevailing literature, where proportional incentive schemes have usually been found to outperform egalitarian ones. We argue that proportional contest incentive schemes suffer from additional inequity concerns, which are exacerbated by increased group identity. Once they feel attached to their group, individuals experience more disutility from advantageous inequity while caring less about disadvantageous inequity. This reduces group contributions and harms performance. On the other side, group identity unambiguously increases contributions when the prize is shared equally amongst all group members, as people experience disutility from not cooperating with one another (e.g., Charness et al. 2007; Chen and Li 2009; Dawes and Thaler 1988).

Individual-level analysis suggests that the positive effect of identity with an equal prize treatment is primarily driven by increases in (full) cooperation as well as less self-serving behavior. The monetary incentives of proportional prize sharing appear to be most effective in reducing full defecting behavior. Further research is required to better understand these mechanisms.

We conclude that the effect of increased group identity is sensitive to the institutional setting. In our scenario, institutionalizing an equal treatment of group members seems to be important for the positive effect of group identity to unfold. This also highlights the need for practitioners to carefully consider social incentives and multi-level conflicts when designing effective incentive schemes. In contrast to what is commonly shown and assumed, being mindful of inequity within groups and installing group-level payout structures can increase cooperation and output.

Appendix A: Derivations

Contest

For simplification, we use $\sum_{i=1}^{m_k} x_{ik}^c = \sum x_{ik}^c$.

$$\begin{aligned}
\frac{\delta\pi_{ik}^e}{\delta x_{ik}} &= \frac{t}{m_k} - 1 + S\left(\frac{x_{ik}^c(X - X_k)}{X^2 \sum x_{ik}^c} + \frac{cX_k(x_{ik}^{c-1}(\sum x_{ik}^c) - x_{ik}^c(\sum x_{ik}^{c-1}))}{X \sum x_{ik}^{2c}}\right) \\
&= \frac{t}{m_k} - 1 + S\left(\frac{x_{ik}^c(X - X_k)}{X^2 \sum x_{ik}^c} + \frac{cX_k(x_{ik}^{c-1}(\sum x_{ik}^c) - x_{ik}^c(\sum x_{ik}^{c-1}))}{X \sum x_{ik}^{2c}}\right) \\
&= \frac{t}{m_k} - 1 + S\left(\frac{x_{ik}^c(\sum x_{ik}^c)(X - X_k) + cXX_k(x_{ik}^{c-1}(\sum x_{ik}^c) - x_{ik}^c(\sum x_{ik}^{c-1}))}{X^2 \sum x_{ik}^{2c}}\right).
\end{aligned}$$

Assuming symmetric players gives $x_{ik} = x_i$, $X_k = m_k x_i$, and $X = nm_k x_i$ and changes the former equation to

$$\begin{aligned}
&= \frac{t}{m_k} - 1 + Sx_i^c\left(\frac{m_k x_i^c[(n-1)m_k x_i] + cn(m_k x_i)^2(x_i^{-1} m_k x_i^c - x_i^{c-1})}{(nm_k x_i)^2[m_k]^2 x_i^{2c}}\right) \\
&= \frac{t}{m_k} - 1 + Sx_i^c\left(\frac{[m_k]^2 x_i^{c+1}(n-1) + cn(m_k x_i)^2 x_i^{c-1}(m_k - 1)}{[nm_k]^2[m_k]^2 x_i^{2+2c}}\right) \\
&= \frac{t}{m_k} - 1 + S\left(\frac{n - 1 + cn(m_k - 1)}{[nm_k]^2 x_i}\right).
\end{aligned}$$

Finally, solving for x_i gives

$$x_{ik}^* = S\left(\frac{n - 1 + cn(m_k - 1)}{[nm_k]^2(1 - \frac{t}{m_k})}\right).$$

Inequity

Isolating the effect of group identity on $\frac{\partial\pi_{ik}^{IE}}{\partial x_{ik}}$ and formulating the condition under which z_{ik} decreases cooperation is represented by:

$$\begin{aligned}
z &< S\left(\frac{X(z_{ik}\alpha + z_{ik}\beta) - x_{ik}(z_{ik}\alpha + z_{ik}\beta)}{X^2}\right) \\
z &< S\left(\frac{Xz_{ik}(\alpha + \beta)}{X^2} - \frac{x_{ik}z_{ik}(\alpha + \beta)}{X^2}\right) \\
zX^2 &< S(x_{ik}z_{ik}(\alpha + \beta) - x_{ik}z_{ik}(\alpha + \beta)) \\
X^2 &< S(X(\alpha + \beta) - x_{ik}(\alpha + \beta)) \\
\frac{X^2}{X(\alpha + \beta) - x_{ik}(\alpha + \beta)} &< S
\end{aligned}$$

Appendix B: Simulations Inequity Aversion and Group Identity

For more information and additional figures, please refer to the online repository of this paper.

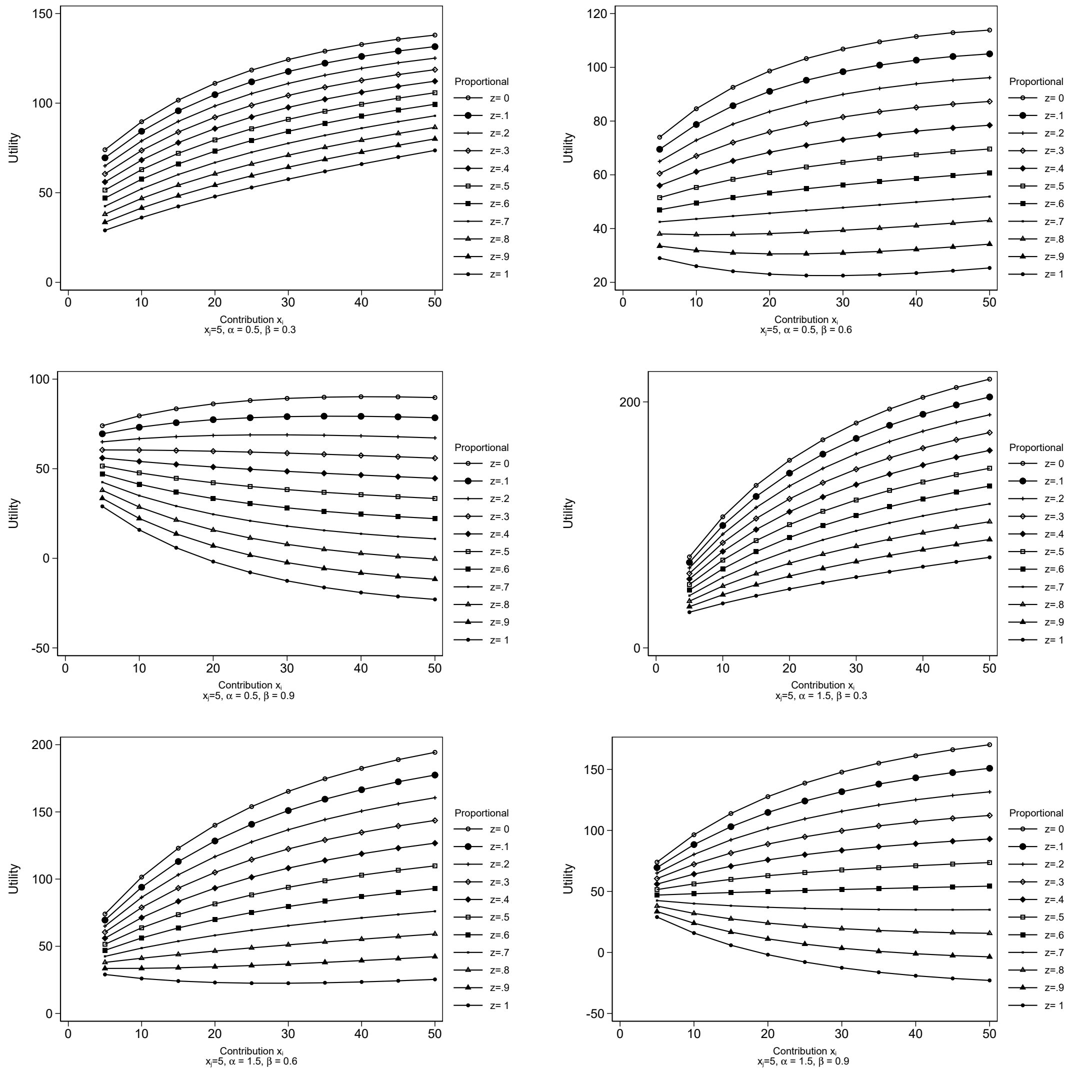


Figure 7 Proportional Prize Sharing. Expected utility for different individual contributions x_{ik} in the inequity paradigm.

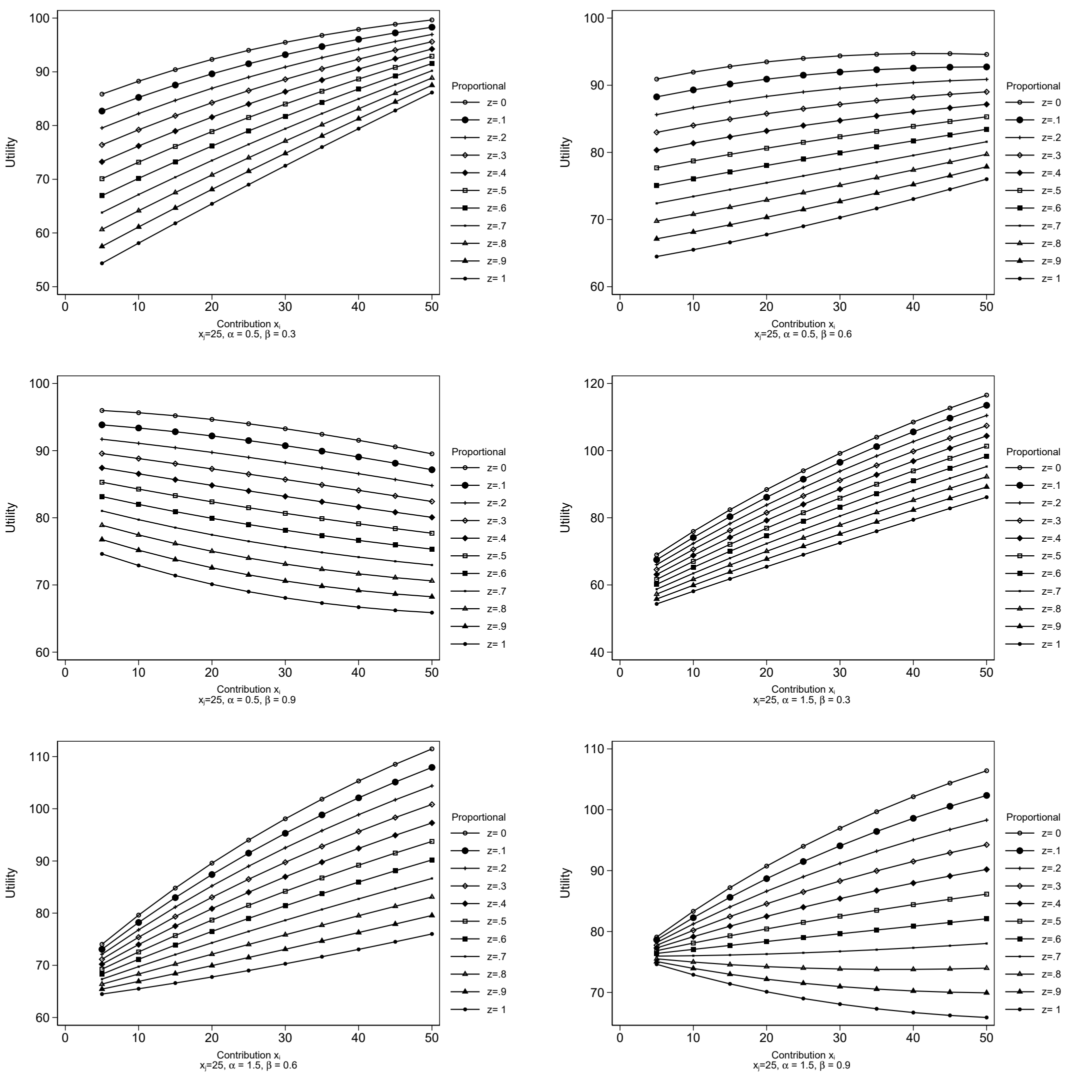


Figure 8 Proportional Prize Sharing. Expected utility for different individual contributions x_{ik} in the inequity paradigm.

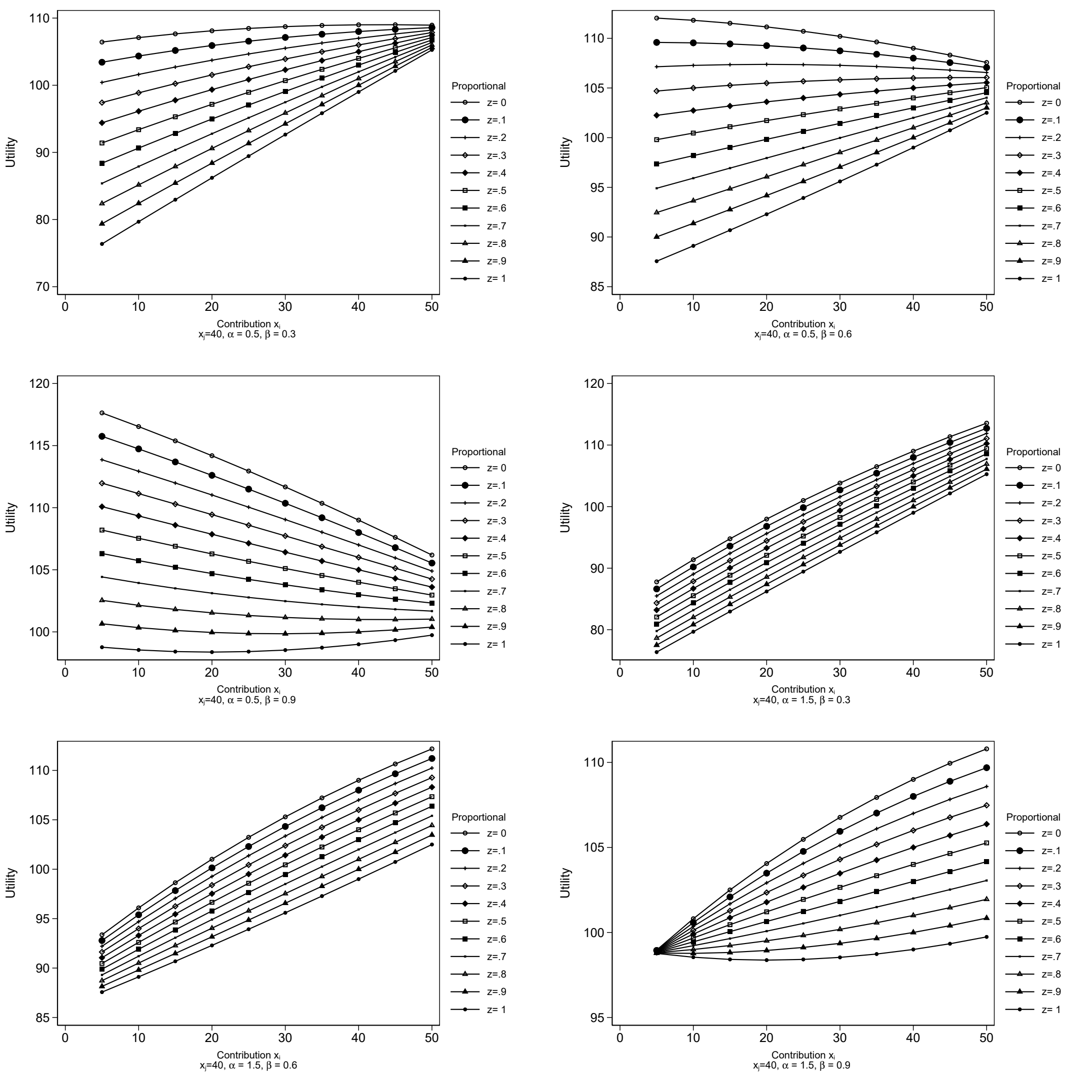


Figure 9 Proportional Prize Sharing. Expected utility for different individual contributions x_{ik} in the inequity paradigm.

References

Abbink, Klaus, Jordi Brandts, Benedikt Herrmann and Henrik Orzen (2010). 'Intergroup Conflict and Intra-Group Punishment in an Experimental Contest Game'. In: *American Economic Review* 100 (1), 420–447.

— (2012). 'Parochial altruism in inter-group conflicts'. In: *Economic Letters* 17 (1), 45–48.

Afridi, Farzana, Sherry Xin Li and Yufei Ren (2015). 'Social identity and inequality: The impact of China's hukou system'. In: *Journal of Public Economics* 123, 17–29.

Ahn, Tho-Keyong, R. Mark Isaac and Timothy C. Salmon (2011). 'Rent seeking in groups'. In: *International Journal of Industrial Organization* 29 (1), 116–125.

Akerlof, George A. (2002). 'Behavioral macroeconomics and macroeconomic behavior'. In: *The American Economic Review* 92 (3), 411–433.

Akerlof, George A. and Rachel E. Kranton (2000). 'Economics and identity'. In: *The Quarterly Journal of Economics* 115 (3), 715–735.

— (2005). 'Identity and the Economics of Organizations'. In: *The Journal of Economic Perspectives* 19 (1), 9–32.

— (2008). 'Identity, Supervision, and Work Groups'. In: *The American Economic Review* 98 (2), 212–217.

— (2010). *Identity Economics*. Princeton University Press, Princeton/New Jersey.

Andreoni, James (1988). 'Why free ride?: Strategies and learning in public goods experiments'. In: *Journal of Public Economics* 37 (3), 291–304.

Andreoni, James and Rachel Croson (2008). 'Partners versus strangers: Random rematching in public goods experiments'. In: *Handbook of Experimental Economic Results*. Ed. by C. R. Plott and V. L. Smith. Vol. 1. Amsterdam: Elsevier, 776–783.

Ashforth, Blake E. and Fred Mael (1989). 'Social identity theory and the organization'. In: *Academy of Management Review* 14 (1), 20–39.

Ashforth, Blake E., Kristie M. Rogers and Kevin G. Corley (2011). 'Identity in organizations: Exploring cross-level dynamics'. In: *Organization Science* 22 (5), 1144–1156.

Babcock, Philip, Kelly Bedard, Gary Charness, John Hartman and Heather Royer (2015). 'Letting down the team? Social effects of team incentives'. In: *Journal of the European Economic Association* 13 (5), 841–870.

Bartel, Ann P., Brianna Cardiff-Hicks and Kathryn Shaw (2017). 'Incentives for Lawyers: Moving Away from "Eat What You Kill"'. In: *ILR Review* 70 (2), 336–358.

Baumgartner, Thomas, Lorenz Götte, Rahel Gügler and Ernst Fehr (2012). 'The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement'. In: *Human Brain Mapping* 33 (6), 1452–1469.

Bernhard, Helen, Ernst Fehr and Urs Fischbacher (2006). 'Group affiliation and altruistic norm enforcement'. In: *American Economic Review* 96 (2), 217–221.

Bicskei, Marianna, Matthias Lankau and Kilian Bizer (2016). 'Negative reciprocity and its relation to anger-like emotions in identity-homogeneous and-heterogeneous groups'. In: *Journal of Economic Psychology* 54, 17–34.

Bornstein, Gary and Ido Erev (1994). 'The enhancing effect of intergroup competition on group performance'. In: *International Journal of Conflict Management* 5 (3), 271–283.

Bornstein, Gary and Uri Gneezy (2002). 'Price competition between teams'. In: *Experimental Economics* 5 (1), 29–38.

Brewer, Marilynn B. and Roderick M. Kramer (1986). 'Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing'. In: *Journal of Personality and Social Psychology* 50 (3), 543–549.

Burton-Chellew, Maxwell N and Stuart A West (2012). 'Pseudocompetition among groups increases human cooperation in a public-goods game'. In: *Animal Behaviour* 84 (4), 947–952.

Burton-Chellew, Maxwell N., Adin Ross-Gillespie and Stuart A. West (2010). 'Cooperation in humans: competition between groups and proximate emotions'. In: *Evolution and Human Behavior* 31 (2), 104–108.

Cason, Timothy N., Roman M. Sheremeta and Jingjing Zhang (2012). 'Communication and efficiency in competitive coordination games'. In: *Games and Economic Behavior* 76 (1), 26–43.

Charness, Gary and Patrick Holder (2019). 'Charity in the laboratory: Matching, competition, and group identity'. In: *Management Science* 65 (3), 1398–1407.

Charness, Gary, Luca Rigotti and Aldo Rustichini (2007). 'Individual behavior and group membership'. In: *The American Economic Review* 97 (4), 1340–1352.

Chaudhuri, Ananish (2011). 'Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature'. In: *Experimental Economics* 14 (1), 47–83.

Chen, Yan and Sherry Xin Li (2009). 'Group identity and social preferences'. In: *The American Economic Review* 99 (1), 431–457.

Chowdhury, Subhasish M., Joo Young Jeon and Abhijit Ramalingam (2016). 'Identity and group conflict'. In: *European Economic Review* 90, 107–121.

Christensen, P. Niels, Hank Rothgerber, Wendy Wood and David C. Matz (2004). 'Social norms and identity relevance: A motivational approach to normative behavior'. In: *Personality and Social Psychology Bulletin* 30 (10), 1295–1309.

Dawes, Robyn M. and Richard H. Thaler (1988). 'Anomalies: cooperation'. In: *The Journal of Economic Perspectives* 2 (3), 187–197.

Delton, Andrew W. and Max M. Krasnow (2017). 'The psychology of deterrence explains why group membership matters for third-party punishment'. In: *Evolution and Human Behavior* 38 (6), 734–743.

Drago, Robert and Geoffrey K. Turnbull (1991). 'Competition and cooperation in the workplace'. In: *Journal of Economic Behavior & Organization* 15 (3), 347–364.

Eckel, Catherine C. and Philip J. Grossman (2005). 'Managing diversity by creating team identity'. In: *Journal of Economic Behavior & Organization* 58 (3), 371–392.

Everett, Jim A.C., Nadira S. Faber and Molly Crockett (2015). 'Preferences and beliefs in ingroup favoritism'. In: *Frontiers in Behavioral Neuroscience* 9, 15.

Fehr, Ernst and Klaus M. Schmidt (1999). 'A theory of fairness, competition, and cooperation'. In: *The Quarterly Journal of Economics* 114 (3), 817–868.

Fershtman, Chaim and Uri Gneezy (2001). 'Discrimination in a segmented society: An experimental approach'. In: *Quarterly Journal of Economics* 116 (1), 351–377.

Fischbacher, Urs (2007). 'z-Tree: Zurich toolbox for ready-made economic experiments'. In: *Experimental Economics* 10 (2), 171–178.

Fischbacher, Urs, Simon Gaechter and Ernst Fehr (2001). 'Are people conditionally cooperative? Evidence from a public goods experiment'. In: *Economics Letters* 71 (3), 397–404.

Fletcher, Thomas D., Debra A. Major and Donald D. Davis (2008). 'The interactive relationship of competitive climate and trait competitiveness with workplace attitudes, stress, and performance'. In: *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 29 (7), 899–922.

Gaviria, Elena, Laura Quintanilla and Marıa José Navas (2021). 'Influence of Group Identification on Malicious and Benign Envy: A Cross-Sectional Developmental Study'. In: *Frontiers in Psychology* 12, 2353.

Goette, Lorenz, David Huffman and Stephan Meier (2006). 'The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups'. In: *American Economic Review* 96 (2), 212–216.

Goette, Lorenz, David Huffman, Stephan Meier and Matthias Sutter (2012). 'Competition Between Organizational Groups: Its Impact on Altruistic and Antisocial Motivations'. In: *Management Science* 58 (5), 948–960.

Greiner, Ben (2015). 'Subject pool recruitment procedures: organizing experiments with ORSEE'. In: *Journal of the Economic Science Association* 1 (1), 114–125.

Gunnthorsdottir, Anna and Amnon Rapoport (2006). 'Embedding social dilemmas in intergroup competition reduces free-riding'. In: *Organizational Behavior and Human Decision Processes* 101 (2), 184–199.

Güth, Werner, Matteo Ploner and Tobias Regner (2009). 'Determinants of in-group bias: Is group affiliation mediated by guilt-aversion?' In: *Journal of Economic Psychology* 30 (5), 814–827.

Halac, Marina, Navin Kartik and Qingmin Liu (2017). 'Contests for experimentation'. In: *Journal of Political Economy* 125 (5), 1523–1569.

He, Hongwei, Yehuda Baruch and Chieh-Peng Lin (2014). 'Modeling team knowledge sharing and team flexibility: The role of within-team competition'. In: *Human Relations* 67 (8), 947–978.

Heidemeier, Heike and Jenny V. Bittner (2012). 'Competition and achievement goals in work teams'. In: *Human Performance* 25 (2), 138–158.

Hoffmann, Christin and Kirsten Thommes (2020). 'Combining Egalitarian and Proportional Sharing Rules in Team Tournaments to Incentivize Energy-Efficient Behavior in a Principal-Agent Context'. In: *Organization & Environment*, 1–25.

Huettel, Scott A. and Rachel E. Kranton (2012). 'Identity economics and the brain: uncovering the mechanisms of social conflict'. In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367 (1589), 680–691.

Jensen, Martin Kaae and Maria Kozlovskaya (2016). 'A representation theorem for guilt aversion'. In: *Journal of Economic Behavior & Organization* 125, 148–161.

Jiang, Jiang and Sherry Xin Li (2019). 'Group identity and partnership'. In: *Journal of Economic Behavior & Organization* 160, 202–213.

Jones, J. Logan, Walter D. Davis and Christopher H. Thomas (2017). 'Is competition engaging? Examining the interactive effects of goal orientation and competitive work environment on engagement'. In: *Human Resource Management* 56 (3), 389–405.

Jordan, Jillian J, Katherine McAuliffe and Felix Warneken (2014). 'Development of in-group favoritism in children's third-party punishment of selfishness'. In: *Proceedings of the National Academy of Sciences* 111 (35), 12710–12715.

Kato, Takao and Pian Shu (2016). 'Competition and Social Identity in the Workplace: Evidence from a Chinese Textile Firm'. In: *Journal of Economic Behavior & Organization* 131 (Part A), 37–50.

Katz, Eliakim, Shmuel Nitzan and Jacob Rosenberg (1990). 'Rent-seeking for pure public goods'. In: *Public Choice* 65 (1), 49–60.

Ke, Changxia, Kai A. Konrad and Florian Morath (2013). 'Brothers in arms – An experiment on the alliance puzzle'. In: *Games and Economic Behavior* 77 (1), 61–76.

— (2015). 'Alliances in the shadow of conflict'. In: *Economic Inquiry* 53 (2), 854–871.

Keller, Anita C., Daniel Spurk, Franziska Baumeler and Andreas Hirschi (2016). 'Competitive climate and workaholism: Negative sides of future orientation and calling'. In: *Personality and Individual Differences* 96, 122–126.

Keser, Claudia and Frans Van Winden (2000). 'Conditional cooperation and voluntary contributions to public goods'. In: *The Scandinavian Journal of Economics* 102 (1), 23–39.

Kim, Eugene and Theresa M. Glomb (2014). 'Victimization of high performers: The roles of envy and work group identification.' In: *Journal of applied psychology* 99 (4), 619.

Korsgaard, M. Audrey, David M. Schweiger and Harry J. Sapienza (1995). 'Building Commitment, Attachment, and Trust in Strategic Decision-Making Teams: The Role of Procedural Justice'. In: *The Academy of Management Journal* 30 (1), 60–84.

Kugler, Tamar, Amnon Rapoport and Asya Pazy (2010). 'Public good provision in inter-team conflicts: Effects of asymmetry and profit-sharing rule'. In: *Journal of Behavioral Decision Making* 23 (4), 421–438.

Ledyard, John (1995). 'Handbook of Experimental Economics'. In: ed. by J. Kagel & A. Roth. Princeton University Press, Princeton, NJ. Chap. Public Goods: Some Experimental Results, 111–168.

McAuliffe, Katherine and Yarrow Dunham (2017). 'Fairness overrides group bias in children's second-party punishment.' In: *Journal of Experimental Psychology: General* 146 (4), 485.

McLeish, Kendra N. and Robert J. Oxoby (2011). 'Social interactions and the salience of social identity'. In: *Journal of Economic Psychology* 32 (1), 172–178.

Morell, Alexander (2019). 'The short arm of guilt–An experiment on group identity and guilt aversion'. In: *Journal of Economic Behavior & Organization* 166, 332–345.

Morita, Hodaka and Maroš Servátka (2013). 'Group identity and relation-specific investment: An experimental investigation'. In: *European Economic Review* 58, 95–109.

Müller, Daniel (2019). 'The anatomy of distributional preferences with group identity'. In: *Journal of Economic Behavior & Organization* 166, 785–807.

Ockenfels, Axel and Peter Werner (2014). 'Beliefs and ingroup favoritism'. In: *Journal of Economic Behavior & Organization* 108, 453–462.

Ploner, Matteo and Tobias Regner (2013). 'Self-image and moral balancing: An experimental analysis'. In: *Journal of Economic Behavior & Organization* 93, 374–383.

Puurtinen, Mikael and Tapio Mappes (2009). 'Between-group competition and human cooperation'. In: *Proceedings of the Royal Society B: Biological Sciences* 276 (1655), 355–360.

Rand, David G. and Martin A. Nowak (2013). 'Human Cooperation'. In: *Trends in Cognitive Science* 17 (8), 413–425.

Rey-Biel, Pedro (2008). 'Inequity aversion and team incentives'. In: *Scandinavian Journal of Economics* 110 (2), 297–320.

Sheremeta, Roman M. (2018). 'Behavior in group contests: A review of experimental research'. In: *Journal of Economic Surveys* 32 (3), 683–704.

Sheremeta, Roman M. and Jingjing Zhang (2010). 'Can groups solve the problem of over-bidding in contests?' In: *Social Choice and Welfare* 35 (2), 175–197.

Sherif, M.Muzafer, O. J. Harvey, B. Jack White, William R. Hood and Carolyn W. Sherif (1961). *Intergroup conflict and cooperation: The Robbers Cave experiment*. Vol. 10. University Book Exchange Norman, OK.

Solow, John L. and Nicole Kirkwood (2002). 'Group identity and gender in public goods experiments'. In: *Journal of Economic Behavior & Organization* 48 (4), 403–412.

Stagnaro, Michael N., Yarrow Dunham and David G. Rand (2018). 'Profit versus prejudice: harnessing self-interest to reduce in-group bias'. In: *Social Psychological and Personality Science* 9 (1), 50–58.

Stets, Jan E. and Peter J. Burke (2000). 'Identity theory and social identity theory'. In: *Social Psychology Quarterly* 63 (3), 224–237.

Tajfel, Henri (1974). 'Social identity and intergroup behaviour.' In: *Social Science Information/Sur les Sciences Sociales* 13 (2), 65–93.

— (1978). 'Differentiation between social groups: Studies in the social psychology of intergroup relations'. In: ed. by H. Tajfel. Academic Press. Chap. Social Categorization, Social Identity, and Social Comparison, 61–76.

Tajfel, Henri, Michael G. Billig, Robert P. Bundy and Claude Flament (1971). 'Social categorization and intergroup behaviour'. In: *European Journal of Social Psychology* 1 (2), 149–178.

Tajfel, Henri and John C. Turner (1979). 'An integrative theory of intergroup conflict'. In: *The Social Psychology of Intergroup Relations* 33 (47), 74.

Tan, Jonathan H.W. and Friedel Bolle (2007). 'Team competition and the public goods game'. In: *Economics Letters* 96 (1), 133–139.

Van Knippenberg, Daan (2000). 'Work motivation and performance: A social identity perspective'. In: *Applied Psychology* 49 (3), 357–371.

Wech, Barbara A., Kevin W. Mossholder, Robert P. Steel and Nathan Bennett (1998). 'Does work group cohesiveness affect individuals' performance and organizational commitment? A cross-level examination'. In: *Small Group Research* 29 (4), 472–494.

Weng, Qian and Fredrik Carlsson (2015). 'Cooperation in teams: The role of identity, punishment, and endowment distribution'. In: *Journal of Public Economics* 126, 25–38.

Zelmer, Jennifer (2003). 'Linear public goods experiments: A meta-analysis'. In: *Experimental Economics* 6 (3), 299–310.

Chapter 7

Conclusion

This thesis aims to highlight non-R&D innovation and contribute to a broader and more profound appreciation of learning and knowledge creation processes crucial for employee and individual-driven innovation attainment. Across multiple articles, it shows that low-threshold changes in the organizational design can help compensate a lack of explicit, formalized R&D resources by encouraging employees to unfold their innovative potential.

In particular, setting a learning goal motivates people to search for patterns and structures to organize their input resources more efficiently by increasing the visibility of innovative capabilities. However, organizations should consider tangible and clearly defined targets that allow people to measure their progress, as they create higher potential to achieve innovative solutions. Besides, in DUI-like routine tasks, people accumulate experiential knowledge, making them more susceptible for choosing low-risk, low-reward options instead of high-risk, high-reward options (often a necessary pre-requisite for innovation). This is particularly important when experienced rewards are rare and occur relatively late in the search process. Thus, setting a specific, challenging learning goal can help people to overcome comparably bad experiences and endure the difficult and rocky search for innovations.

Other tools, such as monetary incentives and delegating the compensation decision process, should be carefully embedded into given social structures. First, depending on the level of group identity and cohesion, proportional or performance-based compensations that conflicts a group target or interfere with the cooperative structure of a task can have a hampering effect. Thus, effective incentives schemes should consider social structures and multi-level conflicts that emerge from team work. Second, delegating the compensation decision should fit the task environment. When the subsequent task is ambiguous, as it is the case for innovation, people prefer a low-risk, performance independent payment. This preference for sure and fixed payments should be considered as a basis for (innovative) performance, but could be complemented with pay-for-performance elements for certain sub-tasks, as monetary incentives do not crowd-out intrinsic motivation, and subsequently innovation, per se.

Finally, the learning and knowledge creation processes examined in this thesis are especially crucial for non-R&D output, but not exclusively so. Learning from experience, by conducting a task or using tools, or learning by searching and interacting, are universal behavioral patterns. Thus, tools discussed and tested in the specific framework of learning by "doing-using-interacting" can be transferred, implemented and tested within R&D-structures, particularly as they illustrate low-threshold interventions.

Declaration of contribution to each essay of this cumulative dissertation

based on the CRediT taxonomy by Brand et al. 2015*

A New Measurement Conception for the 'Doing-Using-Interacting' Mode of Innovation

With: Harm Alhusen, Tatjana Bennat, Kilian Bizer, Uwe Cantner, Martin Kalthaus, Till Proeger, Rolf Sternberg, Stefan Töpfer
Own contribution: 5%; conceptualization, writing - review & editing, investigation
Published: Harm Alhusen, Tatjana Bennat, Kilian Bizer, Uwe Cantner, Elaine Horstmann, Martin Kalthaus, Till Proeger, Rolf Sternberg and Stefan Töpfer (2021). A New Measurement Conception for the 'Doing-Using-Interacting' Mode of Innovation. In: *Research Policy*, 50(4), 104214.

Managing Innovations: Using Learning and Performance Goals to Promote Non-R&D Innovation

With: -/-
Own contribution: 100%
Submitted to: Research Policy

The Effect of Monetary Incentives and Self-Chosen Payment Schemes on Non-R&D Innovation

With: Kilian Bizer
Own contribution: 90%; conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing – original draft, visualization.
Submitted to: Journal of Business Economics

Experience Dominates Choice Bracketing in Risky Decisions under Incomplete Information

With: Alexander Erlei
Own contribution: 50%; conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing – original draft and review & editing, visualization.
R & R: Judgement and Decision Making

Contests for Social Dilemmas: Unequal Treatment Erodes the Positive Effect of Group Identity on Cooperation

With: Ann-Kathrin Blankenberg, Alexander Erlei, Tim Schneider
Own contribution: 50%; conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing – original draft and review & editing, visualization.
R & R: Journal of Economic Behavior & Organization

Date, Signature

*Brand, Amy, Liz Allen, Micah Altman, Marjorie Hlava and Jo Scott (2015). Beyond authorship: attribution, contribution, collaboration, and credit. In: *Learned Publishing*, 28(2), 151-155.

Ph.D. program in Economics
Declaration for admission to the doctoral examination

I confirm

1. that the dissertation that I submitted *"The Making-Of: Innovation – Understanding and Designing the Environment for Non-R&D Innovation"* was produced independently without assistance from external parties, and not contrary to high scientific standards and integrity,

2. that I have adhered to the examination regulations, including upholding a high degree of scientific integrity, which includes the strict and proper use of citations so that the inclusion of other ideas in the dissertation are clearly distinguished,

3. that in the process of completing this doctoral thesis, no intermediaries were compensated to assist me neither with the admissions or preparation processes, and in this process,

   - No remuneration or equivalent compensation were provided
   - No services were engaged that may contradict the purpose of producing a doctoral thesis

4. that I have not submitted this dissertation or parts of this dissertation elsewhere.

I am aware that false claims (and the discovery of those false claims now, and in the future) with regards to the declaration for admission to the doctoral examination can lead to the invalidation or revoking of the doctoral degree.

____

Date, Signature