HIGHER-ORDER RISK PREFERENCE AND TEAMS

–

FIVE ESSAYS IN EXPERIMENTAL ECONOMICS

Dissertation

zur Erlangung des Doktorgrades

Dr. rer. pol.

der Fakultät für Wirtschaftswissenschaften der

Universität Duisburg-Essen

Campus Essen

vorgelegt

von

Alexander Haering

aus

Ratingen, Deutschland

Betreuerin:

Prof. Dr. Jeannette Brosig-Koch

Lehrstuhl für Volkswirtschaftslehre, insbesondere Quantitative Wirtschaftspolitik

Essen, Oktober 2020

Gutachterinnen und Gutachter:

Prof. Dr. Jeannette Brosig-Koch (Erstgutachterin)

Lehrstuhl für Volkswirtschaftslehre, insbesondere Quantitative Wirtschaftspolitik – Universität Duisburg-Essen

&

Prof. Dr. Timo Heinrich (Zweitgutachter)

Institute for Digital Economics (W-5) − Technische Universität Hamburg

Tag der mündlichen Prüfung:

23. März 2021

DOI: 10.17185/duepublico/74469
URN: urn:nbn:de:hbz:464-20210729-102459-6

Introduction

This dissertation investigates (higher-order) risk preferences and team settings to deepen the understanding of the impact of information on individual decision making using economic laboratory experiments. In this context, it explores the effect of framing, wage (in-)equality and the behavior of (other) team members. All these factors are information available to the subject: subjects are informed in a specific way about the riskiness of lotteries and the wage or behavior of other team members.

Information play a critical role for individual decision making as they help to reduce uncertainty and information asymmetry (see Arrow, 1973 for a basic overview how limitations of information affects subjects' behavior). Consumers, for instance, try to acquire information about products they want to buy, and firms gather information about potential employees they want to recruit. Sticking to the examples, Milgrom and Roberts (1986) investigate the role of information in form of advertisement on consumption decisions and Spence (1973) on signals as proxy for information in labor market decisions. All these observable decisions are influenced by other, more abstract factors like the individual preference to take risk or individual reactions to the decisions of others.

Risk preferences are a fundamental concept when analyzing individual decision making. Investigating the effect of information, several questions arise: How do individuals behave when they have information on the riskiness of e.g. different lotteries? And how does the visualization of these information affect their behavior? Furthermore, decisions are seldomly made by individuals in isolation. For example, research and development decisions are typically conducted by subjects in interaction with other individuals or organizations: individual decision making is potentially driven by peer effects that can be observed, e.g., in team decisions. Here, information can influence individual behavior, too: How does information on wage inequality affect behavior? And does it matter if subjects know about contributions by others to a joint research project?

Therefore, it is essential to study these factors in isolation and how they are influenced by information, as they can affect these decisions as well. To investigate these abstract concepts, independent of other (un-)observable factors, this dissertation uses economic laboratory experiments. The advantage of experiments is that one can observe behavior in a controlled environment and only modify variables of interest (Guala 2005). In addition, it is possible to control certain conditions that might affect the results like anonymity, communication, and incentives.

Subjects' preferences to take risk play a crucial role in many economic models, explaining how the degree of risk-aversion influences human behavior in different domains, i.e., educational choice (e.g. Bonin et al., 2007), health behavior (e.g. Anderson & Mellor, 2008 and Felder & Mayrhofer, 2011) or migration decisions (Jaeger et al., 2010). Beside these studies focusing on one country only, there is evidence of national differences in subjects' responses to risky decisions. For example, using hypothetical payoffs in a questionnaire setting, Hsee and Weber (1999) observed that Chinese people are more risk-loving than people from the United States.

Individual reactions to the decisions of other team members are critical when analyzing behavior as well. A factor influencing these interactions is the wage inequality between subjects (see, e.g., Pfeffer & Davis-Blake, 1990). In case of interactions between organizations, different budgets for research and development activities affect decision making as well. In contrast to wage inequality, decisions by organizations interacting with others, like, for example, in joint research and development project, can be modeled as a public goods game (Baumol 1952 and see Ledyard 1995, for an overview on public goods games) with different endowments, e.g., different budgets. Here, results from economic lab experiments are mixed. On the one hand, Cherry et al. (2005) and Aquino et al. (1992) find that different endowments of participants lead to lower contribution to a public good. On the other hand, Warr (1982, 1983) and Chan et al. (1999) show that the endowment distribution has no effect on the overall contribution

As those abstract factors are hard to measure in the field independent of other (un-)observable factors, it is convenient to use economic laboratory experiments. This dissertation does exactly this and studies individual preferences to take risk and behavior in teams in five chapters[1].

The 1st chapter "*Risk Preferences in China – Results from Experimental Economics*" describes methods that are used to elicit risk preferences in economic laboratory experiments and analyzes studies comparing individual risk-taking behavior in China and in other countries. It focuses on studies that elicit preferences for risk-taking using (i) non-incentivized questionnaires and/or (ii) methods of experimental economics. The results show that in non-incentivized surveys Chinese subjects are more willing to take risk than Germans and US-Americans. The existing studies using incentivized experiments, however, suggest that this relationship is less

[1] See https://www.dropbox.com/sh/r5vjk6hnofhgnxt/AADYboRt5btmi-qovYxdjGefa?dl=0 for a web appendix, including the raw data and introductions for all experiments.

clear. Some studies report Chinese to be less risk averse than Germans or US-Americans while others do not find significant differences.

Yet, higher-order risk preferences – beyond risk-aversion – influence individual decision making: prudence, indicated by a positive third-order derivative of the utility function (3rd order risk-aversion, Kimball, 1990) and temperance, specified by a negative fourth-order derivative (4th order risk-aversion; Kimball, 1992) impact decisions made by individuals when dealing with uncertainty. Prudence and temperance can help to explain individuals' savings and investments decisions, taking the riskiness of future income into account (Kimball 1990, 1992). Several studies in various areas confirm these findings. Research on auctions (Esö and White 2004), bargaining games (White 2008), prevention (Eeckhoudt and Gollier 2005, Courbage and Rey 2006, 2016, and Peter 2017) and medical decision making (Eeckhoudt 2002, Felder and Mayrhofer 2014, 2017) shows that higher-order risk preference can explain individual behavior. So far, little is known whether these preferences occur in different subject pools, i.e., in different countries, to the same degree.

Chapter 2 "*Exploring the consistency of higher-order risk preferences*" measures the consistency of higher-order risk preferences by investigating potential country differences between China, Germany and the United States, the effect of stake size, and how the presentation (framing) of lotteries influences individual behavior. In summary, the majority of choices in the three different countries are consistent. That means that all subjects exhibit a so-called mixed risk-loving, i.e. risk loving in odd and risk averse in even orders, or mixed risk averse, i.e. risk averse in all orders, behavior. The same holds after a tenfold increase in the stake size. However, investigating the effects with respect to the lottery framing reveals that mixed risk-loving or mixed risk averse behavior is strengthened if the lotteries are displayed in compound rather than reduced form.

To gather more evidence on the effect with regard to lottery framing observed in chapter 2, chapter 3 "*Framing decisions in experiments on higher-order risk preferences*" analyzes how the way of presentation of these lotteries (framing and lottery display type) affect the degree of higher-order risk preferences. This chapter explores differences in risk preferences resulting from displaying lotteries in reduced rather than in compound form or displaying lotteries in an urn-style rather than in a spinner-style. Overall, the findings show that individual behavior is influenced by lottery framing but not by display format.

Chapter 4 "*Costly information acquisition: The effects of wage inequality*" analyzes the effects of information on wage inequality on individual performance. Wage inequality has been studied in several non-experimental studies, which typically investigate the correlation of different outcomes rather than a causal effect (see Pfeffer, 2007; Shaw, 2014 and Downes & Choi, 2014, for overviews). In research settings, a greater wage inequality is correlated with less individual productivity (Pfeffer & Langton, 1993), with a lower number of successful patent applications (Yanadori & Cui, 2013) and with fewer introduction of new products (Wang et al., 2015). These are all clear indicators for barriers to innovation processes due to wage inequality. The results of chapter 4 reveal that disadvantageous wage inequality does not have a negative effect on agents' performances. In some cases, an increase of wage inequality lead to a steeper rise in agents' performance than a general increase in wages. This effect can be explained by agents' individual differences in loss aversion.

And Chapter 5, "*Experimental evidence on innovation barriers and collective innovation processes in automotive industry value chains*", analyzes under what conditions innovation barriers between four firms, each of which occupying a unique position along the automotive industry value chain, can be overcome. It narrows the gap between economic laboratory experiments and decision making outside the lab by (i) applying a framed laboratory experiment and (ii) investigating how firms embedded in a specific industry context may overcome factors hampering innovation processes. The results reveal that certain constellations can support cooperation: when subjects have information on the distribution of others (in sequential decision-making process), the overall welfare increases even in case of unequally distributed R&D budgets.

References

Anderson, L. R., & Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of health economics*, *27*(5), 1260-1274.

Aquino, K., Steisel, V., & Kay, A. (1992). The effects of resource distribution, voice, and decision framing on the provision of public goods. *Journal of Conflict Resolution*, *36*(4), 665-687.

Arrow, K. J. (1973). Information and economic behavior. HARVARD UNIV CAMBRIDGE MA.

Baumol, W. J. (1952). Welfare Economics and the Theory of the State. Harvard: Harvard University.

Bonin, H., Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2007). Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics*, *14*(6), 926-937.

Chan, K. S., Mestelman, S., Moir, R., & Muller, R. A. (1999). Heterogeneity and the voluntary provision of public goods. *Experimental Economics*, *2*(1), 5-30.

Cherry, T. L., Kroll, S., & Shogren, J. F. (2005). The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab. *Journal of Economic Behavior & Organization*, *57*(3), 357-365.

Courbage, C., & Rey, B. (2006). Prudence and Optimal Prevention for Health Risks. *Health Economics*, *15*(12), 1323-27.

Courbage, C., & Rey, B. (2016). Decision Thresholds and Changes in Risk for Preventive Treatment. *Health Economics*, *25*(1), 111-24.

Downes, P. E. & Choi, D. (2014). Employee reactions to pay dispersion: A typology of existing research, *Human Resource Management Review*, *24*(1), 53-66.

Eeckhoudt, L. (2002). *Risk and medical decision making* (Vol. 14). Springer Science & Business Media.

Eeckhoudt, L., & Gollier, C. (2005). The impact of prudence on optimal prevention. *Economic Theory*, *26*(4), 989-994.

Esö, P., & White, L. (2004). Precautionary bidding in auctions. *Econometrica*, *72*(1), 77-92.

Felder, S., & Mayrhofer, T. (2011). Medical decision making: a health economic primer. Springer Science & Business Media.

Felder, S., & Mayrhofer, T. (2014). Risk preferences: consequences for test and treatment thresholds and optimal cutoffs. *Medical Decision Making*, *34*(1), 33-41.

Felder, S., & Mayrhofer, T. (2017). Medical Decision Making: A Health Economic Primer. 2nd ed. Berlin Heidelberg: Springer.

Guala, F. (2005). *The methodology of experimental economics*. Cambridge University Press.

Hsee, C. K., & Weber, E. U. (1999). Cross-national differences in risk preference and lay predictions. *Journal of Behavioral Decision Making*, *12*(2), 165-179.

Jaeger, D. A., Dohmen, T., Falk, A., Huffman, D., Sunde, U., & Bonin, H. (2010). Direct evidence on risk attitudes and migration. *The Review of Economics and Statistics*, *92*(3), 684-689.

Kimball, M. S. (1990). Precautionary saving in the small and in the large, *Econometrica*, *58*(1), 53-73.

Kimball, M. S. (1992). Precautionary motives for holding assets. In: Newman, P., Milgate, M. & Falwell, J. (Eds.) *The New Palgrave Dictionary of Money and Finance*, MacMillan, London, 158-161.

Ledyard, J. (1995). Public Goods: A Survey of Experimental Research. *The Handbook of Experimental Economics,* edited by Kagel, J. & Roth, A.E., Princeton: Princeton University Press.

Milgrom, P., & Roberts, J. (1986). Price and advertising signals of product quality. *Journal of political economy*, *94*(4), 796-821.

Peter, R. (2017). Optimal self-protection in two periods: On the role of endogenous saving. *Journal of Economic Behavior & Organization*, *137*, 19-36.

Pfeffer, J. (2007). Human resources from an organizational behavior perspective: Some paradoxes explained. *Journal of Economic Perspectives*, *21*(4), 115-134.

Pfeffer, J. & Davis-Blake, A. (1990). Determinants of salary dispersion in organizations. *Industrial Relations*, *29*(1), 38-57.

Pfeffer, J. & Langton, N. (1993). The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty. *Administrative Science Quarterly*, 382-407.

Shaw, J. D. (2014). Pay dispersion, *Annual Review of Organizational Psychology and Organizational Behavior*, *1*(1), 521-544.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355-374.

Wang, T., Zhao, B. & Thornhill, S. (2015). Pay dispersion and organizational innovation: The mediation effects of employee participation and voluntary turnover. *Human Relations*, *68*(7), 1155-1181.

Warr, P. G. (1982). Pareto optimal redistribution and private charity. *Journal of Public Economics*, *19*(1), 131-138.

Warr, P. G. (1983). The private provision of a public good is independent of the distribution of income. *Economics letters*, *13*(2-3), 207-211.

White, L. (2008). Prudence in bargaining: The effect of uncertainty on bargaining outcomes. *Games and Economic Behavior*, *62*(1), 211-231.

Yanadori, Y., & Cui, V. (2013). Creating incentives for innovation? The relationship between pay dispersion in R&D groups and firm innovation performance. *Strategic Management Journal*, *34*(12), 1502-1511.

CHAPTER 1

Risk Preferences in China – Results from Experimental Economics

Published Article:

Haering, A. & Heinrich, T. (2017). Risk preferences in China – Results from Experimental Economics. ASIEN, 142, 66-88.

ASIEN 142 (Januar 2017), S. 68–88

Refereed article

Risk Preferences in China — Results from Experimental Economics*

Alexander Haering and Timo Heinrich

Summary
The propensity to take risks is a fundamental trait that determines the nature of decision making. For example, risk taking is regarded as an important driver of entrepreneurial and innovative behavior in an economy. In this paper, we survey the empirical evidence on individual risk-taking behavior in China. We focus on those studies that elicit preferences for risk taking involving real monetary stakes under controlled conditions, using the methods of Experimental Economics. The studies that we summarize compare Chinese subjects to those in other countries. While non-incentivized surveys find that Chinese subjects are more willing to take risks than Germans and Americans are, the existing experimental studies suggest that this relationship is less clear cut.

Keywords: Experimental Economics, risk preferences, China, survey, cross-cultural experiments

Alexander Haering is a Doctoral Researcher with the Research Group Behavioral Economics and East Asia, IN-EAST School of Advanced Studies, University of Duisburg-Essen, Duisburg; Research Fellow, RWI Leibniz-Institute For Economic Research, Essen.

Timo Heinrich is a Postdoctoral Researcher with the Research Group Behavioral Economics and East Asia, IN-EAST School of Advanced Studies, University of Duisburg-Essen, Duisburg.

* We thank the guest editor, Marcus Conlé, and the anonymous referees for their very valuable comments on this paper. We are grateful to Werner Pascha and Markus Taube for their guidance, and to Heike Hennig-Schmidt, Sven Horak, Toshio Yamagishi, and Chaoliang Yang for sharing their knowledge about cross-cultural experiments. We also thank our fellow members of the IN-EAST School of Advanced Studies as well as the participants of the 5th ACCER International Workshop 2014 in Duisburg for contributing to many fruitful discussions. Franziska Then and Stephanie Zimmer provided excellent research assistance. The financial support from the BMBF to the IN-EAST School of Advanced Studies (Grant No. 01UC1304) is also gratefully acknowledged.

Introduction

For economists, the taste for risk is a fundamental human trait that characterizes individual decisions taken under uncertainty. The attitude toward risk is decisive in many economic models, explaining, for example, educational choice, household savings, or health-related behavior (e. g. Bonin et al. 2007; Noussair et al. 2014; Felder and Mayrhofer 2011). When explaining regional differences in economic outcomes based on microeconomic models, it is therefore important to know whether local preferences differ with respect to risk. In China, risk taking is necessary for entrepreneurs and innovators to cope with the country's transition — as pointed out by Tan (2001). Risk taking is also considered a driver of innovation and entrepreneurship in general (see Khilstrom and Laffont 1979 for a classical model, and Åstebro, et al. 2014 for a summary of empirical evidence from Behavioral Economics).

Chinese people have been found to behave differently from those in Western countries in their strategic interactions with others (e. g. Hennig-Schmidt et al. 2008; Hennig-Schmidt and Walkowitz 2016). In this paper we ask whether Chinese people also differ systematically from those in other countries with respect to individual decision making. We focus specifically on their risk attitudes, and review evidence collected under controlled conditions using the tools of Experimental Economics. Nowadays, experimental methods are used by many micro- and macroeconomists. As Guala (2012) summarizes, the key idea of experimentation is the observation of events under controlled conditions. Control not only concerns variables that are changed by the experimenter but also the background conditions. More specifically, in Experimental Economics the background conditions are partly controlled by running the experiment in a laboratory — which allows decisions to be observed while controlling communication, anonymity, and incentives. Following Roth (1995), the aims of conducting experiments can be loosely classified into testing economic theories, observing regularities in human behavior, and generating policy advice by testing economic institutions.

Testing theories was the most common aim of early experimental economics research — and it still is today. The laboratory allows the creation of decision situations that closely follow theoretical models; observed decisions can then be contrasted with theoretical predictions. Experiments that aim to uncover patterns in human behavior are closely linked to work on testing theories: Experiments can guide the development of theories in situations for which no theories exist yet, or they can stimulate the development of new theories that are better at explaining observed behavior. The work by Kahneman and Tversky (1979, cf. Section 2) on decision making under uncertainty that led to the development of Prospect Theory may be the most prominent example hereof. Experiments that aim to inform policymakers, for example by comparing different market institutions, were pioneered by Smith (1991). These experiments are commonly applied in market

design, and have been used to study several institutions implemented in real-world markets — such as online auctions (e.g. Ockenfels et al. 2006; Brosig-Koch and Heinrich 2014), spectrum auctions (e.g. Grimm, et al. 2003; Abbink et al. 2005), and entry-level labor markets (e.g. Kagel and Roth 2000; Roth, 2002).

Recently, economic experiments have also been used to compare the behavior of different subject pools in different locations. The main contribution of our own paper is a systematic review of those experimental studies that compare the risk attitudes of Chinese people to those of inhabitants of other countries. As an additional contribution, we review different approaches to conducting cross-regional experiments.

Our paper proceeds as follows: In the next section, we describe some of the methods that are used to elicit risk preferences in experimental economics research. In the third section, we then explain the challenges of collecting comparable data in multiple locations and summarize the results of existing studies and their attempts to create comparability. The fourth section concludes the paper with a discussion of limitations and future research.

Measuring risk preferences

An iconic example of decision making under risk is the Saint Petersburg paradox. Consider a gamble that is based on a series of coin throws. If the coin comes up heads on the first throw, you earn one euro and the game ends. If it comes up tails on the first throw, the stakes are doubled and you earn two euros should it come up heads on the second throw. Should it come up tails, the stakes are doubled again and you earn four euros if it comes up heads, and so on. This gamble has an expected value of infinity. Therefore, if you maximize expected payoffs you should be willing to pay a lot of money for being allowed to play the game. Yet few people would actually do so.

Bernoulli (1738) proposed a solution for this paradox and suggested that people maximize "moral expectation" and not expected payoffs. He suggested that the marginal value of money is decreasing, meaning that a wealthy person values an additional income of one euro much less than a poor person values the same amount. If people derive utility from money, this can be expressed by maximizing a utility function that is increasing in monetary value but has a decreasing slope. This kind of concave utility function implies risk aversion. For example, consider a lottery that either pays zero or ten euros with equal probability. Risk-averse people who own a ticket for this lottery will be willing to sell it for any price above five euros (the expected value). But because they are risk averse and the slope of their utility is diminishing then they will also accept a price below five euros for the ticket. How far below five euros, however, depends on the degree of risk aversion and the curvature of their utility function. In their seminal work, von Neumann and Morgenstern (1944) showed that preferences obeying a set of simple axioms could

be expressed by maximizing a utility function of this kind (and of many other kinds too).

These authors' expected utility framework is still dominant and used in many economic models today, even though it cannot explain some behavioral patterns that have been observed when people choose between different lotteries.[1] Allais (1953) was one of the first who pointed out systematic violations of the independence axiom of expected utility. Consider three lotteries A, B, and C. The independence axiom states that lottery A is preferred to lottery B if and only if pA+(1-p)C is preferred to pB+(1-p)C, where p is a probability between 0 and 1. In other words, making lotteries A and B each part of a new compound lottery by adding the same uncertainty should not alter their relative value to the decision maker. Yet as Allais (1953) demonstrated, it often does. Another prominent violation of the theory is the observation that experimental subjects exhibit preference reversals over identical lotteries depending on whether they can sell or buy these lotteries (Lichtenstein and Slovic 1971; Lindman 1971).

Kahneman and Tversky (1979) proposed Prospect Theory, which is consistent with many of these behavioral patterns. It models people as valuing outcomes relative to a given reference point. Gains relative to this reference point are valued with an increasing concave value function, while losses are valued with an increasing convex one — in other words people are viewed as risk-averse with respect to gains, but risk-seeking with respect to losses. In addition, probabilities are weighed non-objectively — meaning small probabilities are overweighed, and large ones underweighed. Prospect Theory is able to capture many deviations from Expected Utility Theory but it also has more degrees of freedom. For this reason, many economists still prefer the more parsimonious Expected Utility Theory.

Despite different theoretical approaches in modeling behavior under uncertainty, experimentally elicited risk preferences are widely used to explain behavior in other decision-making situations. There is some evidence that they are predictive of field behavior. Anderson and Mellor (2008) observe that subjects who are more risk seeking in an experiment are also more likely to smoke cigarettes, drink heavily, to be overweight, and to not use seatbelts. Noussair et al. (2014) find risk preferences to be predictive of decision making with respect to the savings and portfolio choices of households. In addition, the answers to experimentally validated survey questions

1 It is important to note that decision makers (outside of casinos) seldom know the exact probabilities of outcomes, as assumed in Expected Utility Theory (e.g. Knight, 1921). An early extension of Expected Utility Theory is Subjective Expected Utility Theory by Savage (1954), which does not rely on objectively known probabilities. It assumes people evaluate outcomes objectively, but models probabilities as being based on subjective evaluation. This theory is more widely applicable, but suffers from similar shortcomings (see the classical study by Ellsberg, 1961). To our knowledge, there are only two studies that compare risk preferences over lotteries with unknown probabilities between citizens of China and of other countries (see Vieider et al. 2015a; Vieider et al. 2015b).

about self-assessed risk attitudes have been found to be associated with field behavior. Jaeger et al. (2010) observe those who are more risk seeking to be more likely to migrate. Bonin et al. (2007) find those who are more risk seeking to be more likely to work in occupations with a high income-related risk. But the evidence is not clear cut. Sutter et al. (2013), for example, only find a negative correlation of risk aversion with body mass index but no significant correlation with savings behavior, smoking, or alcohol consumption in adolescents.

Risk preferences can be measured with a variety of experimental procedures. In the following, we will briefly describe three popular methods that have been used in the papers that we survey. Our description is based on the overviews provided by Harrison and Rutström (2008) and Charness et al. (2013).

Multiple price list

The multiple price list (MPL) is one of the most commonly used such methods. Most prominent herein is the version by Holt and Laury (2002), but, according to Harrison and Rutström (2008), the very first to use this mechanism were Miller et al. (1969). Table 1 below shows the original price list by Holt and Laury (2002). Subjects typically face a list of two binary lotteries. In each row of the list they choose the lottery that they prefer, and one of the rows is randomly chosen and played to determine their payoff.

The payoffs from the outcomes of the lotteries remain the same between rows, but their probabilities change. The payoffs on the left (Option A) have a lower spread than those on the right (Option B). Moving down from row to row, the probability of the larger outcome within each lottery increases while the probability of the smaller outcome decreases. This makes the righthand side option more attractive in terms of expected payoff when moving down the table (see the rightmost column). From the fifth row on, it is more attractive for someone who is indifferent with respect to risk to choose Option B. Because the spreads differ between the lotteries of both options, however, some people might switch earlier and some later — depending on their taste for risk. In fact, those who switch before the fifth row can be considered “risk seeking” and those who switch later as “risk averse.” The resulting switching point gives the experimenter an estimate of an individual’s attitude towards risk. Note, however, that a subject may behave inconsistently and switch multiple times. Another problem is that the price list induces subjects to switch in the middle of the table, as Harrison and Rutström (2008) point out (see Ebert and Wiesen 2014, and Heinrich and Mayrhofer 2014 for examples).

Table 1. MPL by Holt and Laury (2002)

Row	Option A	Option B	Expected payoff difference
1	1/10 of $2.00, 9/10 of $1.60	1/10 of $3.85, 9/10 of $0.10	$1.17
2	2/10 of $2.00, 8/10 of $1.60	2/10 of $3.85, 8/10 of $0.10	$0.83
3	3/10 of $2.00, 7/10 of $1.60	3/10 of $3.85, 7/10 of $0.10	$0.50
4	4/10 of $2.00, 6/10 of $1.60	4/10 of $3.85, 6/10 of $0.10	$0.16
5	5/10 of $2.00, 5/10 of $1.60	5/10 of $3.85, 5/10 of $0.10	-$0.18
6	6/10 of $2.00, 4/10 of $1.60	6/10 of $3.85, 4/10 of $0.10	-$0.51
7	7/10 of $2.00, 3/10 of $1.60	7/10 of $3.85, 3/10 of $0.10	-$0.85
8	8/10 of $2.00, 2/10 of $1.60	8/10 of $3.85, 2/10 of $0.10	-$1.18
9	9/10 of $2.00, 1/10 of $1.60	9/10 of $3.85, 1/10 of $0.10	-$1.52
10	10/10 of $2.00, 0/10 of $1.60	10/10 of $3.85, 0/10 of $0.10	-$1.85

In some versions of the MPL, one of the two options is a degenerate lottery with certain payoffs (see Schubert et al. 1999 for an early example). In the following, we denote this elicitation method as "MPL-1L" (because it only contains one nondegenerate lottery in each choice) and the standard Holt and Laury (2002) version as "MPL-2L" (because it contains two nondegenerate lotteries in each choice).

Random lottery pairs

The random lottery pairs (RLP) procedure presents subjects with a series of choices over two lotteries. In each choice they express their preference for one of the two lotteries, or, in some procedures, indifference. One of these choices is then selected randomly to determine the payoff. A prominent example is the study by Hey and Orme (1994). They confront subjects with a pair of two-outcome lotteries in each choice. The potential outcomes are taken from the set £0, £10, £20, and £30. The probabilities vary across lotteries and are displayed visually in a pie chart. This approach is very easy to understand, but it does not yield a clear-cut measure for attitude towards risk — as the switching point in the MPL does. As Harrison and Rutström (2008) point out, some form of statistical estimation is thus needed.

Figure 1. Lottery pair by Deck and Schlesinger (2014)

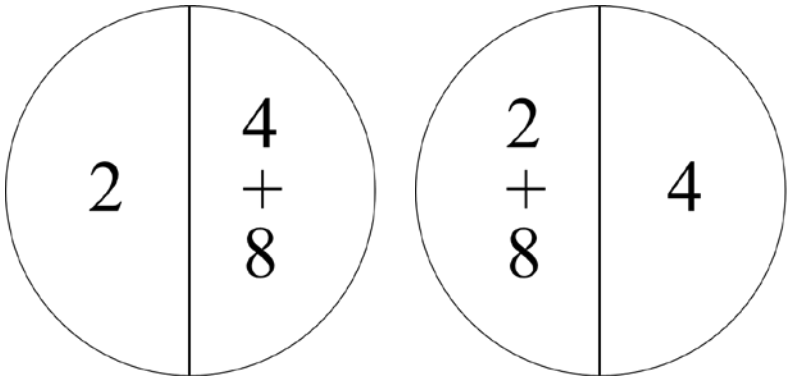


Figure 1 gives an example from the random lottery pairs developed by Deck and Schlesinger (2014), which are also used in the cross-regional comparison by Haering et al. (2017) described in the following section. Similar to Hey and Orne (1994), the probabilities are displayed in a pie chart. Different from their work, however, the random lottery pairs by Deck and Schlesinger (2014) used to elicit risk-averse or risk-seeking choices are all 50–50 lotteries — that is, all outcomes are equally likely.[2] Both lotteries have an expected value of 7 but the outcomes of the lottery on the left (2 and 12) have a larger spread than those on the right (10 and 4). This means that while a risk-neutral individual would be indifferent between both lotteries, every risk-averse person should in principle select the lottery on the right.

Becker-DeGroot-Marschak mechanism

The mechanism suggested by Becker et al. (1964) (BDM) can be used to elicit subjects' certainty equivalents for lotteries, meaning the amount that a subject has to receive with a probability of 100 percent to be willing to sell a lottery ticket that they own. This is a measure for risk attitude. A risk-averse subject will accept a price below the expected value. The more risk averse that they are, the lower this price will be. The mechanism works as follows: The subject owns a lottery and is informed about its characteristics. They also learn that a price for the lottery is picked at random, and that they can state the threshold at which they are willing to

2 The lottery pairs by Deck and Schlesinger (2014) also include lotteries to measure higher-order risk preferences like prudence and temperance, done by combining several 50–50 lotteries. Prudent individuals save more when their future income becomes more risky, while temperate individuals invest less in risky assets when their future income becomes more risky (Kimball, 1990, 1993). Because in other decision pairs the outcomes of different lotteries are added up, for consistency the addition of payoffs (4+8 and 2+8) is also used in the simple lottery that is displayed in Figure 1.

sell. If the price is above that threshold, the lottery will be sold at that price and they will be paid accordingly. If the price is below or equal to the threshold, the subject keeps the lottery and plays it. This mechanism is theoretically incentive compatible, meaning a subject will state their true threshold because they cannot gain from misstating it. Yet this logic is not always apparent. In order to pick the true threshold, one has to realize that the final selling price does not actually depend on the originally stated threshold (for recent criticism, see Cason and Plott, 2014).

Risk preferences in China

Control in cross-regional experiments

At first glance, conducting the same experiment in different regions, countries, or cultures appears a simple way to learn about behavioral differences. However, great care has to be taken to conduct experiments in a comparable way. Roth et al. (1991) ran bargaining and market experiments in Jerusalem, Ljubljana, Pittsburg, and Tokyo. They were among the first to systematically address the following confounding effects that can render observations incomparable:

Experimenter effects: If experiments in different locations are conducted by different experimenters, these experimenters or differences in their procedures might influence decisions differently. Roth et al. (1991) defined the detailed operational procedures that were followed by all experimenters. In addition, all experimenters conducted experiments in Pittsburg in order to detect pure experimenter effects.

Language effects: If languages differ across locations, it becomes necessary to translate the experiment instructions. Literal translations are usually impossible because certain words may not exist in all languages or they may differ in their connotations, which might influence behavior. Roth et al. (1991) aimed to write the original English-language instructions in terms that could be faithfully translated into other languages, "avoiding terms with heavy or ambiguous connotations" (p. 1072). Because they ran two treatments in each country, they had some additional control because — if present — a translation effect would have been observed in both. More commonly used is the back-translation procedure (Brislin 1970). In the first step, the instructions in the original language are translated into another one by a translator. In the second step, this translation is then independently translated back into the original language by a different translator. In the third step, the original and back-translated instructions are compared in order to identify and resolve any discrepancies therein.

Currency effects: Subjects in economic experiments are paid real money in order to provide salient incentives (Smith 1976). If subjects are paid in their local currency, country differences might be due to differences in the incentives that these payments provide. Or, they might be due to the different scales — for example if subjects prefer round numbers. To address the first problem, Roth et al. (1991) adjusted

payment amounts based on purchasing power in the respective countries. To address the second problem, they used "experimental currency units" — in other words, subjects in every country decided on the basis of the same number of tokens. These tokens are converted back to local currency only at the very end, when subjects are paid.

Based on the study by Roth et al. (1991), Herrmann et al. (2008a, 2008b) provide a detailed discussion of these effects and of additional measures to counteract them. For example, with respect to experimenter effects, they also highlight the importance of ensuring subjects' anonymity and of limiting their interaction with the experimenter by conducting a computerized experiment. All the effects mentioned are concerned with the procedures of the experiment itself. Even after these problems have been addressed, however, subjects are not randomly assigned to locations — as that would mean being assigned to different treatments of the same regular experiment. Differences in behavior can be due to all sorts of differences between the subject pools in different locations. Of course, it is impossible to find two subject pools that differ only with respect to their cultural background or country of origin. Therefore, Roth et al. (1991) are careful enough to suggest only that "different behavior in the different subject pools can *cautiously* be used as the basis for *preliminary conjectures* about cultural differences" (p. 1068, emphasis added).

An additional control that has been recently applied is to conduct experiments in (at least) two locations within each region. This way, differences within regions can be compared to those between regions (see Ehmke et al. 2010; and Vieider et al. 2015a). However, this approach does not help if a confounding factor is present in each and every location of a region (e.g. if recruitment procedures for subjects differ between countries for legal reasons).

Nevertheless, carefully designed experiments with hypotheses based on regional differences have discovered interesting behavioral differences across regions. Ockenfels and Weimann (1999), for example, capitalized on German reunification to conduct experiments in the east and west of Germany. Using the same language and currency, they still found pronounced behavioral differences in two regions that had been governed by opposing political systems. They found eastern subjects to behave more selfishly in anonymous laboratory settings.[3] To explain their findings, the authors argue that growing up in a socialist system may have led to solidarity and cooperative behavior in small non-anonymous groups and to egoism in large anonymous ones.

3 See also the follow-up study by Brosig-Koch et al. (2011), which was conducted 20 years after German reunification.

Henrich et al. (2005), meanwhile, conducted ultimatum, public good, and dictator game experiments in 15 small-scale societies around the world. They observed considerable heterogeneity in behavior across these societies. They reported evidence that regional differences in behavior within the societies under study are associated with differences in market integration and the payoffs from cooperation in everyday life.

Comparison of studies

To shed light on risk preferences in the People's Republic of China in comparison to in other countries, we conduct a systematic literature survey. We searched for studies that elicit risk preferences in China and in at least one other region. We only considered experimental studies, meaning those that comply with the standards of Experimental Economics. The main features hereof (in comparison to experimental research in Psychology) are the mandatory use of monetary incentives (Smith, 1976) and the ban on deliberately deceiving subjects (Ortmann and Hertwig, 2002). In addition, we only include those studies that perform statistical tests on the differences between countries.[4]

Table 2 below summarizes, at the top, the six experimental papers that fit our criteria. In addition, at the bottom we list four prominent papers employing a survey methodology (QUE) — in other words, these studies do not elicit decisions over lotteries with real monetary outcomes but ask subjects to make hypothetical decisions in a questionnaire. The first two studies (Weber and Hsee, 1998, and Hsee and Weber, 1999) were the first to focus on Chinese risk preferences. The latter two are, to the best of our knowledge, the most comprehensive survey studies on global differences with respect to risk attitudes. Regarding our research question, their findings exemplify the evidence from other surveys. The remaining surveys we found (Brumagim and Xianhua, 2005, Fan and Xiao, 2006, Lau and Ranyard, 2005, and Statman, 2008) all report the Chinese to be less risk averse than people from other countries are.

4 We used Google Scholar (http://scholar.google.com) and Ideas (http://ideas.repec.org) for a keyword search in order to identify relevant studies in the first round of filtering. The following keywords were used: risk China; risk Chinese; risk preferences China; risk preferences Chinese; risk behavior China; risk behavior Chinese; risk tolerance China; risk tolerance Chinese; Risikopräferenzen China; risk assessment China; risk cross-cultural; cross-cultural risk China; risk cross-country; risk preference cross-country; risk preference cross-country China; risk perception; cross-cultural risk; risk cross-cultural China; and, cross-cultural risk preferences. This resulted in a huge number of studies being discovered. In a second round of filtering, we focused on those studies that compare China with at least one other country. In a third step, we focused on those studies using the methods of Experimental Economics. In the final step, we excluded two studies that did not statistically compare results between countries (Bohnet et al. 2008; Bruhin et al. 2010). We nevertheless discuss these two in due course.

The second column of Table 2 below lists the countries compared in the respective studies. It makes clear that the United States is the most common reference point, followed by Germany. We therefore focus on Germany and the US in the following. The next three columns list the measures taken to ensure the comparability of data collection across locations with respect to the effects pointed out by Roth et al. (1991). The comparisons reveal near consensus with respect to language effects: of the nine studies conducted in different languages, eight use the back-translation method (Brislin 1970). Despite the drawbacks mentioned by Roth et al. (1991), eight out of ten studies opt to display varying payoffs in local currency instead of in experimental currency. This saves subjects from calculating actual payoffs and might make payoffs more salient, but it also potentially creates confounding scale effects. Nine out of ten studies also report how they converted payoffs between countries. All studies use measures that reflect the income differences between the respective subjective pools. However, there appears to be no consensus on the reference measure: some studies use the country-based purchasing power parity (PPP) measure while others rely on more local measures, such as the wages of student research assistants.

With respect to potential experimenter effects, there appears to be even more heterogeneity. Only four out of nine relevant studies actually mention the approach taken. All of these studies relied on the support of local researchers or interpreters. Those by Haering et al. (2017), Rieger et al. (2014), and Falk et al. (2015) also relied on standardized protocols. Haering et al. (2017) are the only ones to control for experimenter differences, by additionally having all experimenters conduct one session in the same location — as advocated by Roth et al. (1991).

In addition, Table 2 lists the general parameters of the studies that we survey. These use different elicitation methods, different sample sizes, and different control variables to capture subject pool differences. There is considerable heterogeneity with respect to the control variables. Ideally, researchers would include many demographic controls to exclude confounding subject pool differences when looking for cross-regional differences in behavior. Yet this also requires larger samples, creating additional costs.

The last two columns of Table 2 summarize the results. The “Risk aversion” one lists significant differences between regions, while the “Other” column lists additional findings. Let us consider the survey papers first, as this methodology has been the standard approach used by economists and other social scientists to assess risk attitudes for many years now. The three survey studies that compare China directly to other countries find Chinese people to be less risk averse than Germans and Americans are. In this respect, they are similar to other surveys not included in Table 2 (Brumagim and Xianhua 2005; Fan and Xiao 2006; Lau and Ranyard, 2005, and Statman 2008).

The fourth mentioned survey study was recently conducted by Falk, Becker, et al. (2015). It is the first study to assess risk preferences (as well as other characteristics of human decision making) in representative samples using an experimentally validated survey measure. This means that the authors also conducted another study (as described in Falk et al. 2016) in which survey answers were compared to the choices made with real monetary stakes by the same subjects. This allows for the selection of survey questions that are highly correlated with the choices made when actual money is involved.

The drawback of their approach is that the cross-regional comparison of risk preferences is only valid if the correlation between imagined and real choices is similar across regions. Vieder et al. (2015b) find that the correlation between survey questions and incentivized measures in fact varies across countries. As they point out, the correlation is significantly positive in 19 to 29 of the 30 countries that they cover (depending on the question and on the domain of payoffs). This might explain why survey questions have been found to correlate with experimentally elicited measures of risk aversion by some (Dohmen et al. 2011 and Falk et al. 2016) but not all authors (Anderson and Mellor, 2009, Lönnqvist et al. 2011).

Falk et al. (2015) do not directly compare risk preferences across countries. Instead, they correlate the average risk attitude in 76 countries with other characteristics of these countries. They find the degree of risk aversion to be significantly and positively correlated with life expectancy, less inequality (as measured by the Gini coefficient), and the higher rigidity of employment laws. It is weakly significantly correlated with a larger level of redistribution (measured as the share of government transfers of national income) and a lower number of homicides. There is no significant correlation with gross domestic product (GDP) per capita or the degree of institutionalized democracy.[5]

Let us now consider the experimental studies that collect decisions made over real monetary stakes. Even though we list six experimental papers, the results are only drawn from five datasets — because Vieider et al. (2015a) consider a subset of the data presented in Vieider et al. (2015b). In three of the five datasets, the respective authors find differences in line with the results of the survey studies: Ehmke et al. (2010) find Chinese participants to be less risk averse than those in the French and American subject pools are. Vieider et al. (2015b) find Chinese participants to be

5 Falk et al. (2015) do not provide a direct comparison of risk preferences in China and in other countries. However, based on the correlations that they provide one can derive an ordering of risk preferences: When comparing China to the US and Germany, for example, we would expect Chinese people to be the least risk averse based on life expectancy, Gini coefficient and redistribution of GDP. Based on labor regulations and the number of homicides per capita, we would expect Americans to be the least risk averse.

less risk averse than German ones.[6] Haering et al. (2017) find Chinese participants to be less risk averse than American and German ones. However, in the remaining two experimental datasets (Kachelmeier and Shehata 1992, and Liu, Meng, et al. 2014), the authors find no significant differences between locations. These results make clear that Chinese participants *cannot* be unequivocally regarded as less risk averse than German and American ones, as the previous survey evidence suggests. These findings also highlight that more research is needed to analyze why hypothetical decisions differ from real ones in a variety of ways across countries.[7]

Two further studies have experimentally elicited risk preferences in China and in other countries but are not listed in Table 2. Bohnet et al. (2008), on the one hand, compare the attitude toward risk in situations where nature resolves uncertainty to the attitude toward risk in situations in which another person resolves it. They find people in Brazil, China, Oman, Switzerland, Turkey and the US to be "betrayal averse", meaning that they prefer risks in which uncertainty is resolved by nature. They also elicit risk preferences in each country, but do not compare them directly. They compare each country to the sample mean, finding only subjects in Oman to be more risk averse than the average. Bruhin et al. (2010), on the other hand, conduct experiments in two locations in Switzerland and two in China. They are interested in identifying behavioral types, so they do not directly compare risk attitudes between countries or between locations. In both countries they find that roughly 80 percent of subjects can be classified as behaving consistent with Prospect Theory, while the remaining subjects maximize expected values. However, they point out that some of the Prospect Theory-type subjects in China strongly overweigh gain and underweigh loss probabilities — which could explain a general tendency to be less risk averse.

6 In the case of lotteries with unknown probabilities (cf. Footnote 1), Vieider et al. (2015b, online appendix) find Chinese people to be less risk averse than both Americans *and* Germans.

7 Also note that Vieider et al. (2015a) report only very small within-country differences in China and in Ethiopia, while Ehmke, Lusk and Tyner (2010) make a similar observation in the US.

Table 2. Comparison of studies

Study	Comparison countries[1]	Cross-regional controls				Elicitation method[2]
		Experimenter	Language	Currency display	Currency conversion[3]	
Kachelmeier & Shehata (1992)	CA, US	Assistance by local interpreter	Back translation	Local currency	n/a	BDM
Ehmke et al. (2010)	FR, NE, US	n/a	Back translation	Local currency	PPP	MPL-2L
Liu et al. (2014)	TW	Not required	Not required	Experimental currency	Wage of a student research assistant	MPL-2L
Vieider et al. (2015b)	DE, US, & 42 others	n/a	Back translation	Local currency	Wage of a student research assistant, PPP adjusted	MPL-1L
Vieider et al. (2015a)	ET	n/a	Back translation	Local currency	Wage of a student research assistant, PPP adjusted	MPL-1L
Haering et al. (2017)	DE, US	Detailed protocol, local experimenters, supervising experimenter	Back translation	Experimental currency	PPP, UBS, adjusted by local guidelines for subject payment	RLP
Weber & Hsee (1998)	DE, PL, US	n/a	Back translation	Local currency	Expenses of students	QUE
Hsee & Weber (1999)	US	n/a	Back translation	Local currency	Expenses of students	QUE
Rieger et al. (2014)	DE, US, & 50 others	Standardized oral introductions read aloud by the local lecturer	Translated by professional translators or translators with an economics background	Local currency	Income and expenses of students, PPP adjusted	QUE
Falk et al. (2015)	DE, US, & 73 others	Professional interviewers using a standardized procedure across Countries	Back translation	Local currency	Median household income	QUE

1: CA: Canada; CN: People's Republic of China; DE: Germany; ET: Ethiopia; FR: France; NE: Niger; PL: Poland; TW: Taiwan; US: United States of America.

2: BDM: Becker-DeGroot-Marschak; MPL-1L: Multiple Price List one lottery; MPL-2L: Multiple Price List two lotteries; QUE: Questionnaires, RLP: Random Lottery Pairs.

3: BNT: Berlin Numeracy Test; CRT: Cognitive Reflection Test; Econ: Economics; IRB form: In the US, subjects need to be presented with a form by the Institutional Review Board for experiments with human subjects beforehand; Math: Mathematics; Major: Major field of study; Stats: Statistics; UBS Prices & Earnings 2014, available online at: www.ubs.com/pricesandearnings.

Table 2. (continued)

	Subject Pool		Results	
Study	Country: Location or University (*N*)[1]	Control variables[3]	Risk aversion[1]	Other
Kachelmeier & Shehata (1992)	CN: Beijing Univ. (40) CA: "Medium sized university" (32) US: "Large university" (28)	None	No significant differences.	Subjects in China are more risk averse when monetary payoffs are increased tenfold
Ehmke et al. (2010)	CN: Hangzhou (96), FR: Grenoble (70), US: West Lafayette (63), Manhattan (57), NE: Niamey (60)	Gender	CN & NE < FR & US	Within-country differences in risk preferences are small
Liu et al. (2014)	CN: Beijing Univ. (185), TW: National Taiwan Univ. (195)	Gender, age, graduate student, major, conservative upbringing, father's education, mother's education	No significant differences.	Beijing University students become significantly more risk loving after being primed with Confucianism
Vieider et al. (2015b)	31 universities in 30 countries (2,939)	Gender, age, major, GDP/capita, Gini coefficient	ET < CN < DE	Incentivized measures correlate with survey questions in a majority of countries
Vieider et al. (2015a)	CN: Jiao Tong Univ. (124), Beijing Normal Univ. (80), ET: Two campuses of Addis Ababa Univ. (83 & 62)	Gender, age, major	ET < CN	Within-country differences in risk preferences are small
Haering et al. (2017)	CN: Nankai Univ. (140), DE: Univ. of Duisburg-Essen (145), US: Harvard Business School (129)	Experimenter, gender, age, CRT score, BNT score, sum of math, stats & econ courses, IRB form	CN < US & DE	Subjects in China are more risk averse when monetary payoffs are increased tenfold
Weber & Hsee (1998)	CN: (85), DEU: (31), PO: (81), USA: (86) "Major urban universities"	Major	CN < PO < DE & US	Chinese are closer to risk neutral in pricing options
Hsee & Weber (1999)	US: Univ. of Chicago (99), Ohio State Univ. (66), CN: Chengjian Univ. (110), Jiao Tong Univ. (65)	n/a	CN < US	Chinese are more risk seeking in investments but not in medical or academic decisions
Rieger et al. (2014)	>60 universities in 53 countries (6,912)	Gender, age, GDP/capita, individualism, uncertainty avoidance index	CN < US < DE	People in richer countries are more risk averse in gains
Falk et al. (2015)	Representative samples in 76 countries (>80,000)	None	n/a	Risk aversion correlates with life expectancy, Gini coefficient, redistribution of GDP, labor regulation, and number of homicides on a country level

Conclusion

We started out with the aim of assessing the risk attitude of Chinese people in comparison to the inhabitants of other countries. Most commonly, survey studies have been used to compare risk attitudes across countries. These studies are based on choices over hypothetical stakes. Virtually all of them find a higher propensity of Chinese participants to take risks relative to American or German participants. However, in Experimental Economics we are interested in preferences over actual monetary outcomes. If we want to draw conclusions about these types of preference based on survey studies using hypothetical outcomes, we have to assume that choices over hypothetical outcomes correlate with choices over monetary ones too — but this is not always the case, as observed by Vieider et al. (2015b). When comparing answers across countries, we also have to assume that this correlation is similar.

However, with respect to China, it is not always clear that instruments for empirical data collection that have been developed in Western countries can be readily transferred to that national context (see Roy et al. 2001 and Stening and Zhang 2007 for overviews). For example, there appears to be evidence for a tendency of Chinese respondents to choose midpoints on Likert scales in questionnaires (Shenkar 1994). The experimental studies that are based on choices over real monetary stakes suggest that differences in preferences are less clear: three studies find Chinese people to be less risk averse than Germans or Americans are, while two studies find no significant differences between them.

However, not all of the studies that we cover can be readily compared because of their varying designs. For example, several reported studies display the varying payoffs in local currency which might lead to confounding scale effects. For a more extensive discussion of how differences in experimental design may account for differences in behavior, see Goerg et al. (2016). It is also possible that our comparison of studies is confounded by regional differences within countries, or by changes in risk attitudes over time. We have not focused on the last point in this paper. Yet macroeconomic conditions have been found to influence decision making under risk (e. g. Browne, Jaeger, et al. 2015, Cohn, Engelmann, et al. 2015), and these conditions have changed quite dramatically in China in recent decades. Also note that the number of experimental studies comparing the risk preferences of the Chinese to those of other peoples is relatively small. If more data becomes available, a quantitative meta-analysis would be the next step.[8]

8 In our review, we only considered individual decision making. Yet cross-regional experimental studies comparing behavior in strategic interactions in Western and Eastern countries generally observe a high degree of dissimilarity (see, for example, Oosterbeek, et al. 2004). By exploring the negotiation behavior of teams from China and Germany, for example, Hennig-Schmidt and Walkowitz (2016) observe that the latter put great weight on fairness issues and try to reach an

In general, our overview summarizes the popular design approaches taken in cross-regional experiments. It highlights the importance of general standards, such as the back-translation procedure, for the comparability of results. One method that is not widely used yet is to run experiments in different locations within the same region, as a control. This is a promising approach because it allows research to compare within-region differences to between-region ones. Due to their cost, experiments are usually restricted to small samples from student subject pools. One alternative to experimental studies are experimentally validated survey measures. These can be applied to representative samples more efficiently. However, their contribution with respect to risk preferences over real monetary stakes is based on a rather strong assumption: the validation that took place in one country is assumed to hold true within all countries under study.

Even if risk preferences are found to differ systematically between individuals or regions, little is known at present about the underlying drivers thereof. Often variations in risk preferences are attributed to cultural differences between countries. For example, Hsee and Weber (1999) found that Chinese people are more likely to take risks than Americans are when deciding over hypothetical payoffs. They explain their finding by the much lower individualism in China relative to in the US, which was also observed by Hofstede (1980). Based on the "cushion hypothesis" people from China are therefore less likely than those from the US to deal with the consequences of risky decisions on their own.

Hofstede (1980) originally identified four dimensions that characterize a culture: power distance, individualism, masculinity, and uncertainty avoidance. Uncertainty avoidance has also been reported to be associated with risk taking. A higher degree of uncertainty avoidance means that members of a society try harder to avoid situations with high uncertainties. This is not synonymous with risk aversion, however. Instead, people might also take additional risks to avoid ambiguity — that is, situations in which the probabilities of outcomes are not known. Nevertheless, in their own survey Rieger et al. (2015) observe that more uncertainty avoidance is associated with less risk taking. However, with respect to uncertainty avoidance China differs less from Western countries — for example the US ranks 57th and China 63rd out of the 69 countries that were surveyed (Hofstede et al. 2010).

It has also been observed that risk preferences are transmitted from one generation to the next (Dohmen et al. 2012), and that they are at least partly genetically determined (Cesarini et al. 2009). Quite recent observations by Becker et al. (2015) suggest that differences in risk preferences between countries (elicited through representative surveys) can be explained by genetic and migratory distance. Their

acceptable payoff within a reasonable timeframe. In contrast, teams from China try to collect as much information on their negotiation partners as possible so as to anticipate their behavior.

results also highlight the importance of environmental factors, including the prevailing institutions for the shaping of risk preferences (see also, Callen et al. 2014 and Browne et al. 2016).

Given the rapid change in living conditions as well as in the institutional environment in China, it thus remains to be seen how risk preferences develop along economic and cultural parameters there. In future, longitudinal studies that combine experiments (or experimentally validated survey measures) with representative samples will help to disentangle the different drivers of decision making under uncertainty. The observation of behavior over time could inform new theories to help explain individual decision making. As an important application, such data may also be able to help explain regional differences in innovativeness and its development.

References

Abbink, Klaus; Irlenbusch, Bernd; Pezanis-Christou, Paul; Rockenbach, Bettina; Sadrieh, Abdolkarim; Selten, Reinhard (2005): "An experimental test of design alternatives for the British 3G/UMTS auction", in: *European Economic Review*, 49: 505–530

Allais, Maurice (1953): "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine", in : *Econometrica*, 21: 503–546

Anderson, Lisa R.; Mellor, Jennifer M. (2008): "Predicting health behaviors with an experimental measure of risk preference", in: *Journal of Health Economics*, 27.5: 1260–1274

— (2009): "Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure", in: *Journal of Risk and Uncertainty*, 39.2: 137–160

Åstebro, Thomas; Herz, Holger; Nanda, Ramana; Weber, Roberto A. (2014): "Seeking the roots of entrepreneurship: insights from behavioral economics", in: *Journal of Economic Perspectives*, 28.3: 49–69

Becker, Anke; Enke, Benjamin; Falk, Armin (2015): "The ancient origins of the cross-country heterogeneity in risk preferences". https://www.cens.uni-bonn.de/team/board/armin-falk/falk-risk-sent.pdf (accessed 2016-03-30)

Becker, Gordon M.; DeGroot, Morris H.; Marschak, Jacob (1964): "Measuring utility by a single-response sequential method", in: *Systems Research and Behavioral Science*, 9: 226–232

Bernoulli, Daniel (1738): "Specimen theoriae novae de mensura sortis", in: *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 5: 175–192

Bohnet, Iris; Greig, Fiona; Herrmann, Benedikt; Zeckhauser, Richard (2008): "Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States", in: *The American Economic Review*, 98.1: 294–310

Bonin, Holger; Dohmen, Thomas; Falk, Armin; Huffman, David; Sunde, Uwe (2007): "Cross-sectional earnings risk and occupational sorting: The role of risk attitudes", in: *Labour Economics*, 14.6: 926–937

Brislin, Richard W. (1970): "Back-translation for cross-cultural research", in: *Journal of Cross-Cultural Psychology*, 1.3: 185–216

Brosig-Koch, Jeannette; Helbach, Christoph; Ockenfels, Axel; Weimann, Joachim (2011): "Still different after all these years: Solidarity behavior in East and West Germany", in: *Journal of Public Economics*, 95.11: 1373–1376

Brosig-Koch, Jeannette; Heinrich, Timo (2014): "Reputation and mechanism choice in procurement auctions: An experiment", in: *Production and Operations Management*, 23.2: 210–220

Browne, Mark J.; Jaeger, Verena; Steinorth, Petra (2015): "Impact of economic conditions on individual risk attitude", SSRN 2631066

Browne, Mark J.; Jaeger, Verena; Richter, Andreas; Steinorth, Petra (2016): "Familiy transitions and risk attitude", in: *MRIC Working Paper No. 32*

Bruhin, Adrian; Fehr-Duda, Helga; Epper, Thomas (2010): "Risk and rationality: Uncovering heterogeneity in probability distortion", in: *Econometrica*, 78.4: 1375–1412

Brumagim, Alan L.; Xianhua, Wu (2005): "An examination of cross-cultural differences in attitudes towards risk: Testing prospect theory in the People's Republic of China", in: *Multinational Business Review*, 13.3: 67–86

Callen, Michael; Isaqzadeh, Mohammad; Long, James D.; Sprenger, Charles (2014): "Violence and risk preference: Experimental evidence from Afghanistan", in: *American Economic Review*, 104.1: 123–148

Cason, Timothy N.; Plott, Charles R. (2014): "Misconceptions and game form recognition: challenges to theories of revealed preference and framing." *Journal of Political Economy*, 122.6: 1235–1270

Cesarini, David; Dawes, Christopher T.; Johannesson, Magnus; Lichtenstein, Paul; Wallace, Björn (2009): "Genetic variation in preferences for giving and risk taking", in: *Quarterly Journal of Economics*, May 2009: 809–842

Charness, Gary; Gneezy, Uri; Imas, Alex (2013): "Experimental methods: Eliciting risk preferences", in: *Journal of Economic Behavior & Organization*, 87: 43–51

Cohn, Alain; Engelmann, Jan; Fehr, Ernst; Maréchal, Michel A. (2015): "Evidence for countercyclical risk aversion: an experiment with financial professionals", in: *American Economic Review*, 105.2: 860–885

Deck, Cary; Schlesinger, Harris (2014): "Consistency of higher order risk preferences", in: *Econometrica*, 82.5: 1913–1943

Dohmen, Thomas; Falk, Armin; Huffman, David; Sunde, Uwe; Schupp, Jürgen; Wagner, Gert G. (2011): "Individual risk attitudes: Measurement, determinants, and behavioral consequences", in: *Journal of the European Economic Association*, 9.3: 522–550

Dohmen, Thomas; Falk, Armin; Sunde, Uwe (2012): "The intergenerational transmission of risk and trust attitudes", in: *Review of Economic Studies*, 72: 645–677

Ebert, Sebastian; Wiesen, Daniel (2014): "Joint measurement of risk aversion, prudence, and temperance", in: *Journal of Risk and Uncertainty*, 48.3: 231–252

Ehmke, Mariah; Lusk, Jayson; Tyner, Wallace (2010): "Multidimensional tests for economic behavior differences across cultures", in: *Journal of Socio-Economics*, 39.1: 37–45

Ellsberg, Daniel (1961): "Risk, ambiguity, and the Savage axioms", in: *Quarterly Journal of Economics*, 75: 643–669

Falk, Armin, Becker, Anke, Dohmen, Thomas; Huffman, David; Sunde, Uwe (2016): "The preference survey module: A validated instrument for measuring risk, time, and social preferences.", in: *IZA Discussion Paper (9674)*

Falk, Armin; Becker, Anke; Dohmen, Thomas; Enke, Benjamin; Huffman, David; Sunde, Uwe (2015): "The nature and predictive power of preferences: Global Evidence", in: *IZA Discussion Paper (9504)*

Fan, Jessie X.; Xiao, Jing Jian (2006): "A cross-cultural study in risk tolerance: Comparing Chinese and Americans", in: *Journal of Personal Finance*, 5.3: 54–75

Felder, Stefan; Mayrhofer, Thomas (2011): *Medical decision making: A health economic primer*. Heidelberg: Springer Science & Business Media

Goerg, Sebastian J.; Hennig-Schmidt, Heike; Walkowitz, Gari; Winter, Eyal (2016): "In wrong anticipation-miscalibrated beliefs between Germans, Israelis, and Palestinians", in: *PLoS One*, 11.6, https://doi.org/10.1371/journal.pone.0156998

Grimm, Veronika; Riedel, Frank; Wolfstetter, Elmar (2003): "Low price equilibrium in multiunit auctions: The GSM spectrum auction in Germany", in: *International Journal of Industrial Organization*, 21: 1557–1569

Guala, Francesco (2012): "Experimentation in economics", in: Mäki, U. (ed.): *Philosophy of Economics* (Handbook of the Philosophy of Science, 13). Oxford, UK and Amsterdam: Elsevier, 597–640

Haering, Alexander; Heinrich, Timo; Mayrhofer, Thomas (2017): "Exploring the consistency of higher-order risk preferences", in: *Ruhr Economics Papers* #688

Harrison, Glenn W.; Rutström, Elisabet E. (2008): "Risk aversion in the laboratory", in: Cox, James C.; Harrison, Glenn W. (eds.): *Risk Aversion in Experiments* (Research in Experimental Economics, 12). Bingley: Emerald, 41–196

Heinrich, Timo; Mayrhofer, Thomas (2014): "Higher-order risk preferences in social settings", in: *Ruhr Economic Paper* #508

Hennig-Schmidt, Heike; Li, Zhu-Yu; Yang, Chaoliang (2008): "Why people reject advantageous offers – Non-monotonic strategies in ultimatum bargaining: Evaluating a video experiment run in PR China", in: *Journal of Economic Behavior & Organization*, 65.2: 373–384

Hennig-Schmidt, Heike; Walkowitz, Gari (2016): "Negotiations among Chinese and Germans – An experimental case study", in: *Homo Oeconomicus*, 32.3/4: 451–488

Henrich, Joseph; Boyd, Robert; Bowles, Samuel; Camerer, Colin; Fehr, Ernst; Gintis, Herbert; McElreath, Richard; Alvard, Michael; Barr, Abigail; Ensminger, Jean; Henrich, Natalie S.; Hill,

Kim; Gil-White, Francisco; Gurven, Michael; Marlowe, Frank W.; Patton, John Q.; Tracer, David (2005): "'Economic man' in cross-cultural perspective: Behavioral experiments in 15 small-scale societies", in: *Behavioral and Brain Sciences*, 28.6: 795–815

Herrmann, Benedikt; Thöni, Christian; Gächter, Simon (2008a): "Antisocial punishment across societies", in: *Science*, 319.5868: 1362–1367

— (2008b): "Antisocial punishment across societies", in: *Science*, 319.5868: Supplementary Material

Hey, John D.; Orme, Chris (1994): "Investigating generalizations of expected utility theory using experimental data", in: *Econometrica*, 62.6: 1291–1326

Holt, Charles A.; Laury, Susan K. (2002): "Risk aversion and incentive effects", in: *American Economic Review*, 92.5: 1644–1655

Hsee, Christopher K.; Weber, Elke U. (1999): "Cross-national differences in risk preference and lay predictions", in: *Journal of Behavioral Decision Making*, 12: 165–179

Hofstede, Geert (1980): *Culture's consequences: International differences in work-related values*. Beverly Hills: Sage Publications

Hofstede, Geert; Hofstede, Gert J.; Minkov, Michael (2010): *Cultures and Organizations - Software of the Mind: Intercultural Cooperation and Its Importance for Survival (Business Skills and Development)*. London: McGraw-Hill

Jaeger, David A.; Dohmen, Thomas; Falk, Armin; Huffman, David; Sunde, Uwe.; Bonin, Holger (2010): "Direct evidence on risk attitudes and migration", in: *Review of Economics and Statistics*, 92.3: 684–689

Kagel, John H.; Roth, Alvin E. (2000): "The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment", in: *Quarterly Journal of Economics*, 115: 201–235

Kachelmeier, Steven J.; Shehata, Mohamed (1992): "Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China", in: *American Economic Review*, 82.5: 1120–1141

Kahneman, Daniel; Tversky, Amos (1979): "Prospect theory: An analysis of decision under risk", in: *Econometrica*, 47: 263–291

Khilstrom, Richard E.; Laffont, Jean-Jacques (1979): "A general equilibrium entrepreneurial theory of firm formation based on risk aversion", in: *Journal of Political Economy*, 87.4: 719–748

Kimball, Miles S. (1990): "Precautionary saving in the small and in the large", in: *Econometrica*, 58.1: 53–73

— (1992): "Precautionary motives for holding assets", in: Newman, P.; Milgate, M.; Falwell, J. (eds.): *The New Palgrave Dictionary of Money and Finance*. London: MacMillan, 158–161

Knight, Frank H. (1921): *Risk, uncertainty and profit*. New York: Hart, Schaffner and Marx

Lau, Lai-Yin; Ranyard, Rob (2005): "Chinese and English probabilistic thinking and risk taking in gambling", in: *Journal of Cross-Cultural Psychology*, 36.5: 621–627

Lichtenstein, Sarah; Slovic, Paul (1971): "Reversals of preference between bids and choices in gambling decisions", in: *Journal of Experimental Psychology*, 89.1: 46–55

Lindman, Harold R. (1971): "Inconsistent preferences among gambles", in: *Journal of Experimental Psychology*, 89.2: 390–397

Liu, Elaine M.; Meng, Juanjuan; Wang, Joseph T. Y. (2014): "Confucianism and preferences: Evidence from lab experiments in Taiwan and China", in: *Journal of Economic Behavior & Organization*, 104: 106–122

Lönnqvist, Jan-Erik; Verkasalo, Markku; Walkowitz, Gari (2011): "It pays to pay – Big Five personality influences on co-operative behavior in an incentivized and hypothetical prisoner's dilemma game", in: *Personality and Individual Differences*, 50.2: 300–304

Miller, Louis; Meyer, David E.; Lanzetta, John T. (1969): "Choice among equal expected value alternatives: Sequential effects of winning probability level on risk preferences", in: *Journal of Experimental Psychology*, 79.3: 419–423

Noussair, Charles N.; Trautmann, Stefan T.; Van de Kuilen, Gijs (2014): "Higher order risk attitudes, demographics, and financial decisions", in: *Review of Economic Studies*, 81.1: 325–355

Ockenfels, Axel; Reiley, David; Sadrieh, Abdolkarim (2006): "Online auctions", in: Hendershott, T. (ed.): *Handbooks in Information Systems, vol. 1*. Amsterdam: Elsevier, 571–628

Ockenfels, Axel; Weimann, Joachim (1999): "Types and patterns: An experimental East-West-German comparison of cooperation and solidarity", in: *Journal of Public Economics*, 71: 275–287

Oosterbeek, Hessel; Sloof, Randolph; Van De Kuilen, Gijs (2004): "Cultural differences in ultimatum game experiments: Evidence from a meta-analysis", in: *Experimental Economics*, 7.2: 171–188

Ortmann, Andreas; Hertwig, Ralph (2002): "The costs of deception: Evidence from psychology", in: *Experimental Economics*, 5.2: 111–131

Rieger, Marc O.; Wang, Mei; Hens, Thorsten (2014): "Risk preferences around the world", in: *Management Science*, 61.3: 637–648

Roth, Alvin E. (1995): "Bargaining experiments", in: Kagel, John H.; Roth, Alvin E. (eds.): *Handbook of Experimental Economics*. Princeton: Princeton University Press, 254–348

— (2002): "The economist as engineer: Game theory, experimentation, and computation as tools for design economics", in: *Econometrica*, 70: 1341–1378

Roth, Alvin E.; Prasnikar, Vesna; Okuno-Fujiwara, Masahiro; Zamir, Shmuel (1991): "Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study", in: *American Economic Review*, 81.5: 1068–1095

Roy, Abhik; Walters, Peter G.; Luk, Sherriff T. (2001): "Chinese puzzles and paradoxes: Conducting business research in China", in: *Journal of Business Research*, 52.2: 203–210

Savage, Leonard J. (1954): *The Foundations of Statistics*. New York: Wiley Publications in Statistics

Schubert, Renate; Brown, Martin; Gysler, Matthias; Brachinger, Hans W. (1999): "Financial decision-making: Are women really more risk-averse?", in: *American Economic Review*, 89.2: 381–385

Shenkar, Oded (1994): "The People's Republic of China: Raising the bamboo screen through international management research", in: *International Studies of Management & Organization*, 24.1–2: 9–34

Smith, Vernon L. (1976): "Experimental economics: Induced value theory", in: *American Economic Review*, 66.2: 274–279

— (1991): *Papers in Experimental Economics*. Cambridge: Cambridge University Press

Statman, Meir (2008): "Countries and culture in behavioral finance", in: *CFA Institute Conference Proceedings Quarterly*, 25.3: 38–44

Stening, Bruce W.; Zhang, Marina Y. (2007): "Methodological challenges confronted when conducting management research in China", in: *International Journal of Cross Cultural Management*, 7.1: 121–142

Sutter, Matthias; Kocher, Martin G.; Glätzle-Rützler, Daniela; Trautmann, Stefan T. (2013): "Impatience and uncertainty: Experimental decisions predict adolescents' field behavior", in: *American Economic Review* 103.1: 510–531

Tan, Justin (2001): "Innovation and risk-taking in a transitional economy: A comparative study of Chinese managers and entrepreneurs", in: *Journal of Business Venturing*, 16.4: 359–376

Vieider, Ferdinand M.; Chmura, Thorsten; Fisher, Tyler; Kusakawa, Takao; Martinsson, Peter; Thompson, Frauke M.; Sunday, Adewara (2015a): "Within-versus between-country differences in risk attitudes: implications for cultural comparisons", in: *Theory and Decision*, 78.2: 209–218

Vieider, Ferdinand M.; Lefebvre, Mathieu; Bouchouicha, Ranoua; Chmura, Thorsten; Hakimov, Rustamdjan; Krawczyk, Michal; Martinsson, Peter (2015b): "Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries", in: *Journal of the European Economic Association*, 13.3: 421–452

Von Neumann, John; Morgenstern, Oskar (1944): *Theory of Games and Economic Behavior*. Princeton: Princeton University Press

Weber, Elke U.; Hsee, Christopher (1998): "Cross-cultural differences in risk perception but cross-cultural similarities in attitudes towards perceived risk", in: *Management Science*, 44.9: 1205–1217

CHAPTER 2

Exploring the Consistency of Higher Order Risk Preferences

Published Article:

Haering, A., Heinrich, T. & Mayrhofer, T. (2020). Exploring the Consistency of Higher Order Risk Preferences. International Economic Review, 61(1), 283-320.

INTERNATIONAL ECONOMIC REVIEW
Vol. 61, No. 1, February 2020 DOI: 10.1111/iere.1

EXPLORING THE CONSISTENCY OF HIGHER ORDER RISK PREFERENCES*

BY ALEXANDER HAERING, TIMO HEINRICH, AND THOMAS MAYRHOFER [1]

RWI – Leibniz-Institute for Economic Research, Germany, Fresenius University of Applied Sciences, Germany; Durham University Business School, U.K.; Harvard University, U.S.A., Stralsund University of Applied Sciences, Germany

This study measures higher order risk preferences and their consistency. We explore the role of country differences, the variation of stakes, and the framing of lotteries. We observe a robust dichotomous pattern of choice behavior in China, the United States, and Germany. The majority of choices are consistent with mixed risk aversion or mixed risk-loving behavior. We also find this pattern after a 10-fold increase in the stakes. Finally, our results reveal that this pattern is strengthened if the lotteries are displayed in compound instead of reduced form. In a follow-up study, we explore potential explanations for this framing effect.

1. INTRODUCTION

Within the expected utility framework, most of the commonly used utility functions (e.g., $\ln(x)$ and $x^{0.5}$) imply "mixed risk aversion," which means that the derivatives of the utility functions exhibit alternating signs (see Brockett and Golden, 1987; Caballé and Pomansky, 1996). Therefore, these utility functions assume second-order risk aversion ($U^{\mathrm{II}} < 0$), as well as higher order risk preferences, such as prudence ($U^{\mathrm{III}} > 0$)—also called third-order risk aversion—and temperance ($U^{\mathrm{IV}} < 0$)—also called fourth-order risk aversion. More recently, higher order risk preferences have also been defined as preferences over binary lotteries by Eeckhoudt and Schlesinger (2006). These model-free definitions do not require assumptions as far reaching as expected utility theory and lend themselves to experimental investigation. Based on the binary lotteries by Eeckhoudt and Schlesinger (2006), Eeckhoudt et al. (2009) define mixed risk aversion as a preference for combining "good" outcomes with "bad" ones. Crainich et al. (2013) then introduced the concept of "mixed risk-loving" behavior, which they define as a preference for combining "good" outcomes with "good" ones. In an expected utility framework, this would imply a utility function for which all the derivatives are strictly positive.[2] Although mixed risk averters are second-order risk-averse, prudent, and temperate, mixed risk-loving

*Manuscript received October 2018; revised April 2019.

[1] We thank the co-editor Masaki Aoyagi and two anonymous referees for very helpful comments. We are also grateful to Han Bleichrodt, Jeannette Brosig-Koch, Louis Eeckhoudt, Glenn Harrison, Heike Hennig-Schmidt, Zwetelina Iliewa, Johanna Kokot, Gijs van de Kuilen, Isabel Marcin, Peter Martinsson, Sonja Opper, Richard Peter, Hendrik Schmitz, Jan Siebert, Stefan Traub, Stefan Trautmann, and Toshio Yamagishi for their valuable feedback. We have also benefited from the comments of seminar participants at the Ruhr University Bochum, the Helmut Schmidt University Hamburg, IE Business School Madrid, and the University of Paderborn, as well as those of conference participants at the GfeW Annual Meeting 2015 in Hamburg, the CEAR/MRIC Behavioral Insurance Workshop 2015 in Atlanta, the Experimental Finance conference 2016 in Mannheim, the FUR conference 2016 at the University of Warwick, the INFORMS Annual Meeting 2016 in Nashville, the 6th ACCER International Workshop 2017 in Duisburg, the Tagung des Sozialwissenschaftlichen Ausschusses 2018 in Essen, and the International ESA Meeting 2018 in Berlin. Furthermore, we thank Christine E. Looser, Haiying Qin, Weijing Le, Chaoliang Yang, and Guanzhong Yang for their support in organizing the experiments. Constantin Carstens, Johanna Rosebrock, Franziska Then, and Stephanie Zimmer provided excellent research assistance. Financial support from the BMBF through the IN-EAST School of Advanced Studies (Grant No. 01UC1304) is gratefully acknowledged. Please address correspondence to: Timo Heinrich, Durham University Business School, Mill Hill Lane, Durham, DH1 3LB, UK. E-mail: *timo.heinrich@uni-due.de*.

[2] As Ebert (2013) points out, neither property follows from risk-loving preferences per se.

individuals are second-order risk-loving, prudent, and intemperate. To put it more generally, mixed risk averters coincide in their choices with mixed risk lovers in the odd orders (e.g., in prudence) but differ in the even orders (e.g., in temperance).

Recently, Deck and Schlesinger (2014) used an economic laboratory experiment to study mixed risk-averse and mixed risk-loving behavior. They made two major observations: First, in their data, a nonnegligible minority of individuals make consistently second-order risk-loving choices. Second, in line with the theoretical prediction, they observed a consistent pattern of mixed risk-averse and mixed risk-loving behavior. In this article, we study whether this dichotomy can be regarded as a widespread pattern explaining the heterogeneity of choices under risk. We conduct a large-scale experiment with a total of 605 participants and add to the literature by measuring higher order risk preferences (up to order 6) across different countries by employing distinct subject pools in China, in the United States, and in Germany. Furthermore, we contribute to previous findings by studying the effect of a 10-fold increase in the stakes and the effect of a straightforward change in the framing of the lotteries. In a follow-up study with an additional 224 participants, we explore the influence of framing further.[3]

Previous experimental studies by Deck and Schlesinger (2010), Ebert and Wiesen (2011, 2014), Maier and Rüger (2012), and Noussair et al. (2014) suggest that a majority of aggregate choices are in line with prudence and—with the exception of the studies by Deck and Schlesinger (2010) and Baillon et al. (2017)—in line with temperance (see Appendix A.1 for a more detailed comparison).[4] In addition, based on representative data from the Netherlands, Noussair et al. (2014) find that lottery choices are correlated with behavior in the field. The effects they observe are in line with theoretical predictions. Prudent lottery choices are associated with greater wealth, a greater likelihood of having a savings account, and a lower likelihood of credit card debt. Temperate lottery choices are associated with less risky investment portfolios.

Ideally, considering individual differences in higher order risk preferences will help to build more realistic economic models. In lifecycle savings models, for example, prudence and temperance determine how current savings are influenced by the riskiness of future income (Kimball, 1990, 1992). Other areas in which higher order risk preferences have been theoretically shown to impact behavior include auctions (Esö and White, 2004), bargaining games (White, 2008), research and development expenditures (Nocetti, 2015), prevention (Eeckhoudt and Gollier, 2005; Courbage and Rey, 2006, 2016; Peter, 2017) and medical decision making (Eeckhoudt, 2002; Felder and Mayrhofer, 2014, 2017). However, what still needs to be established is the degree to which the previous findings on individual differences in higher order risk preferences are robust and sufficiently general in different contexts.

[3] Deck and Schlesinger (2014) were not only the first to study mixed risk-averse and mixed risk-loving behavior experimentally, but were also the first to assess risk preference of orders greater than 4. Risk aversion of order 5 (called "edginess" by Lajeri-Chaherli, 2004) or even 6 (named "bentness" by Miles S. Kimball at a conference to honor Louis Eeckhoudt in 2012) have so far rarely been studied. However, utility functions typically imply assumptions across *all* orders of risk aversion, and there is no compelling reason why these assumptions should not be subject to empirical scrutiny. In addition, in an intertemporal consumption problem, an increase in the nth order risk of future income yields an increase in savings if and only if $(n + 1)$-th order risk aversion is present (as shown by Eeckhoudt and Schlesinger, 2008, in an expected utility theory framework). In other words, anyone who thinks that nth order risk matters to decision makers will care about their nth and $(n + 1)$-th order risk attitudes in an intertemporal setting. Also note that in Deck and Schlesinger's (2014) design, the elicitation of higher order risk attitudes requires rather complex lotteries. We believe that, because of this complexity, assessing behavior in the respective lotteries with fifth- and sixth-order variations of risk is quite useful because it provides an even tougher test for the theoretical predictions. Very recently, Ebert et al. (2017) proposed an alternative method to elicit higher order risk preferences. Their theory is based on greater mutual aggravation and does not require complex doubly-compounded lotteries.

[4] Baillon et al. (2017) also measure ambiguity prudence and ambiguity temperance based on the preference conditions by Baillon (2017). Other experimental studies have observed that higher order risks influence precautionary savings (Bostian and Heinzel, 2012), as well as behavior in auctions (Kocher et al., 2015) and medical treatment and prevention decisions (Krieger and Mayrhofer, 2012, 2017). Moreover, higher order risk preferences have also been studied experimentally in social settings (Heinrich and Mayrhofer, 2018), across multiple domains (Ebert and van de Kuilen, 2015; Deck and Schlesinger, 2017) and with children (Heinrich and Shachat, 2018). For a detailed review of the recent experimental literature on higher order risk preferences, please see Trautmann and van de Kuilen (2018).

During recent decades, it has become evident that many behavioral patterns identified in Western subject populations are by no means universal human traits. For example, economists discovered that human behavior in strategic interaction varies widely across societies, with aggregate behavior covering virtually the complete strategy space (see, e.g., Roth et al., 1991; Oosterbeek et al., 2004; Herrmann et al., 2008). With regard to second-order risk aversion, research in economics and psychology has provided evidence of differences in risk attitudes across countries (see, e.g., Weber and Hsee, 1998; Vieider, Chmura et al., 2015; Vieider, Lefebvre et al., 2015) and across stake sizes (see, e.g., Binswanger, 1980, 1981; Kachelmeier and Shehata, 1992b). Measuring risk preferences in different international subject pools provides a tougher test of the generalizability of a theory. Furthermore, experimental research in economics has often been criticized for using small samples and small stakes (e.g., Levitt and List, 2007). Conducting the experiment in three countries provides us with a larger aggregate sample. It also allows us to exploit differences in purchasing power to conduct high stakes experiments for (relatively) low cost.

Additionally, it has been observed that displaying lotteries in a reduced instead of compound form may impact the degree of second-order risk aversion (see, e.g., Abdellaoui et al., 2015; Harrison et al., 2015). With respect to mixed risk-averse and mixed risk-loving behavior, Deck and Schlesinger (2014) conjecture the compound presentation "admittedly also facilitates viewing the problem as 'combining good with bad' or 'combining good with good,' instead of presenting the lotteries in a reduced form, which might obfuscate this interpretation" (Deck and Schlesinger, 2014, 1921ff).

Although subject pool differences, stake size, and the framing of lotteries have been studied with respect to second-order risk aversion, there has been very little work on higher order risk preferences. Since second-order risk aversion and higher order risk preferences are related theoretically, as well as empirically, we expect that these factors also influence higher order risk preferences.

The results we report in this article suggest that a majority of people across national contexts are second-order risk-averse. Moreover, we confirm the main observations by Deck and Schlesinger (2014): A considerable proportion of people are second-order risk-loving and choices can be explained rather well by a dichotomy of preference types. In total, up to 62% of the participants can be classified as mixed risk-averse and up to 14% as mixed risk-loving. We present the first comparison of higher order risk preferences across countries (i.e., China, the United States, and Germany) and under high stakes. Our study reveals that mixed risk aversion is somewhat more prevalent among Germans than among Chinese and that the dichotomy of preference types persists when stakes increase. We also discover that the dichotomy can be strengthened through the framing of the lottery.[5] When we display the lotteries in compound instead of reduced form, we observe significantly more prudent and temperate behavior *within* the same subjects. A follow-up study reveals that the justifications for specific lottery choices differ significantly between both types of framings.

2. THEORETICAL BACKGROUND AND HYPOTHESES

2.1. *Theoretical Background.* In this subsection, we present the theoretical background to our experiment, which follows Deck and Schlesinger (2014). In their experiment, they use a variety of lotteries, which are based on the theoretical work by Eeckhoudt and Schlesinger (2006), Eeckhoudt et al. (2009) and Crainich et al. (2013). The lotteries are binary with equal probabilities, that is, $[x, y]$ denotes a lottery with a 50–50 chance of receiving either outcome x or outcome y. However, x and y might themselves be lotteries.

[5] A very recent study by Deck and Schlesinger (2017) that was conducted in parallel with ours makes a similar observation with respect to the framing of lotteries (see Subsection 2.2.3 and Section 5). Furthermore, they replicate the observations made in Deck and Schlesinger (2014) and also consider choices when the payoffs are nonmonetary.

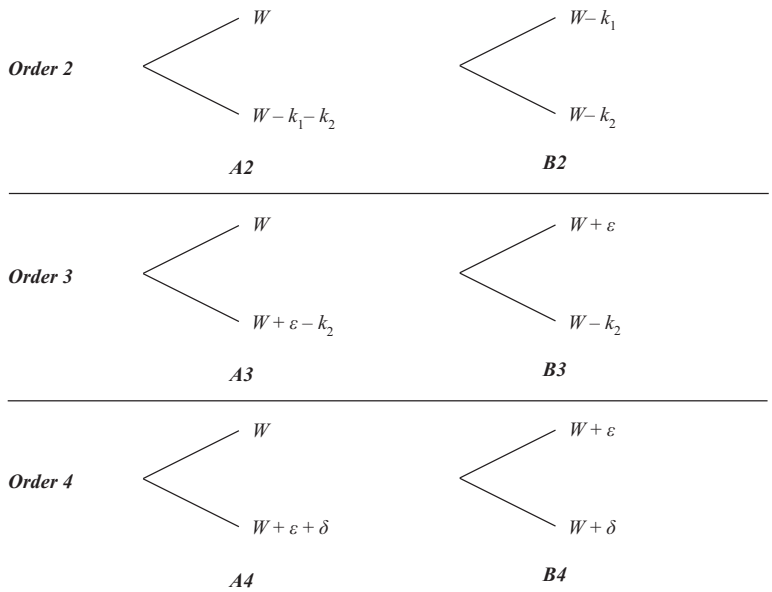


FIGURE 1

LOTTERIES FOR ELICITING RISK PREFERENCES UP TO ORDER 4

In the following, we will refer to risk aversion of nth degree as "n-RA" and to risk-loving behavior of nth degree as "n-RL." Figure 1 shows lotteries for eliciting risk preferences up to order 4. Let us assume that W is the initial wealth of an individual, with $W > 0$, and that k_1 and k_2 are sure losses, with $k_1 > 0$ and $k_2 > 0$. Furthermore, ε and δ are independent zero-mean background risks, that is, lotteries with an expected value of zero.

The first row in Figure 1 illustrates a second-order risk aversion, that is, 2-RA, task in which risk aversion is a preference for disaggregating harms, that is, the sure losses k_1 and k_2. Disaggregating these two "bad" payoffs reduces the spread between the two possible outcomes. This corresponds to a lower variance, which is a necessary assumption for lower second-order risk. Lottery $A2$ has a greater spread (and thus variance) than lottery $B2$. A risk-averse individual would choose lottery $B2$ over lottery $A2$ and, vice versa, a risk-loving individual would choose lottery $A2$ over lottery $B2$. In other words, although both regard a sure loss as "bad," second-order risk is only "bad" for the risk-averse individual but "good" for the risk-loving individual.

The second row in Figure 1 shows a 3-RA task. In this case, the sure loss k_1 is replaced by a zero-mean background risk ε. Eeckhoudt and Schlesinger (2006) define 3-RA as a preference for disaggregating a sure loss and an additional zero-mean background risk. Therefore, a 3-RA individual would prefer lottery $B3$ over lottery $A3$, whereas a 3-RL individual would prefer $A3$. Note that in this case a risk-averse and a risk-loving individual would agree that avoiding a sure loss is "good," but both differ in their judgment of the zero-mean background risk. The upper arm of $B3$ yields a combination of "good" with "bad" for risk-averse individuals and a combination of "good" with "good" for risk-loving individuals.

The third row in Figure 1 exemplifies a 4-RA task. Now the second loss k_2 is also replaced by a second zero-mean background risk δ (which is independent of ε). Eeckhoudt and Schlesinger (2006) define 4-RA as a preference for disaggregating two independent zero-mean background risks. Thus, a 4-RA individual would prefer lottery $B4$ over lottery $A4$, whereas a 4-RL individual would prefer lottery $A4$ over lottery $B4$. The lower arm of $A4$ yields a combination of "bad" with "bad" for risk-averse individuals. However, for risk-loving individuals, this represents a combination of "good" with "good."

Deck and Schlesinger (2014) now define mixed risk aversion as a preference for combining "good" with "bad," and mixed risk-loving behavior as a preference for combining "good" with

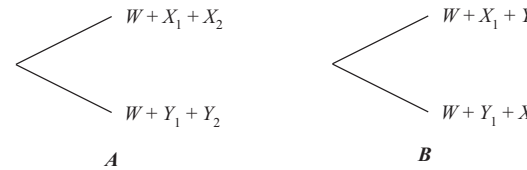


FIGURE 2

LOTTERIES FOR ELICITING RISK PREFERENCES IN A GENERAL FRAMEWORK

"good." This yields a pattern in which both types coincide in their lottery choices of odd orders but differ in even orders.

For orders higher than 4, Deck and Schlesinger (2014) use a more general approach that is based on the theoretical work by Eeckhoudt et al. (2009) and is illustrated in Figure 2. Following Deck and Schlesinger (2014), we consider a pair of random variables $[X_1, Y_1]$, where Y_1 has more nth order risk than X_1. According to Ekern (1980), Y_1 has more nth order risk than X_1 if X_1 is nth order stochastic dominant compared to Y_1 and the two random variables have the same $n-1$ moments (for $n > 1$). Moreover, let us consider a second pair of random variables $[X_2, Y_2]$ where Y_2 has more mth order risk than X_2. All random variables are statistically independent of each other. Eeckhoudt et al. (2009) show that, for this setting, the 50–50 lottery $[W + X_1 + X_2, W + Y_1 + Y_2]$ has more $(m + n)$-th order risk than the 50–50 lottery $[W + X_1 + Y_2, W + Y_1 + X_2]$. An individual who prefers lotteries with lower $(m + n)$-th order risk is "$(m + n)$-th order risk-averse." An individual who is $(m + n)$-th order risk-averse would choose lottery B over lottery A in Figure 2. This approach is more general, and can be used for all orders.

Moreover, from the viewpoint of a mixed risk-averse individual, both random variables X_i can be considered as relatively "good," and both random variables Y_i as relatively "bad." Lottery A in Figure 2 shows a 50–50 chance of receiving either "good" with "good" (upper lottery arm) or "bad" with "bad" (lower lottery arm), whereas lottery B shows a combination of "good" with "bad" in both lottery arms. Lottery B therefore apportions the good and bad outcomes. A mixed risk-averse individual dislikes risk of any order and therefore always prefers lottery B. However, a mixed risk-loving individual only dislikes risk of odd orders and therefore only prefers lottery B if $m + n$ is odd (and lottery A otherwise).

2.2. *Hypotheses.*

2.2.1. *Cross-country differences.* In selecting China, the United States, and Germany, we aim to strike a balance between the economic relevance of the subject pools and their heterogeneity. On the one hand, these countries are those with the highest population on their respective continents, as well as the largest economies in terms of total GDP. On the other hand, these countries differ culturally. In one of the first studies analyzing second-order risk preferences in an international comparison, Hsee and Weber (1999) found that Chinese people are more likely to take risks than Americans with respect to hypothetical payoffs (see also Weber and Hsee, 1998; Statman, 2008). They explain their findings based on the cultural trait of individualism as introduced by Hofstede (1980). According to the cushion hypothesis, people from China are less individualistic and thus less likely than Americans to deal on their own with the consequences of risky decisions. In fact, the most recent data on the cultural dimensions of 69 countries by Hofstede et al. (2010) also reveal that Germany, the United States, and China differ widely with respect to individualism: the United States ranks 1st, Germany 17th, and China 52nd out of these 69 countries.

Hofstede (1980) originally identified power distance, masculinity, and uncertainty avoidance next to individualism as dimensions that characterize a culture. Of these four dimensions, uncertainty avoidance has been found to be associated with risk preferences: In a comprehensive survey covering 53 countries, Rieger et al. (2015) observe that uncertainty avoidance is

associated with higher second-order risk aversion. The data on uncertainty avoidance suggest smaller disparities between the countries: Germany ranks 40th, the United States 57th, and China 63rd among the 69 countries. Thus, based on individualism and uncertainty avoidance, we would expect Chinese people to be the least risk-averse.[6]

Although there are many international comparison studies on second-order risk aversion (see Haering and Heinrich 2017 for an overview), nothing is known about differences in higher order risk preferences. We follow Deck and Schlesinger's (2014) argument and assume that human behavior is driven by a basic tendency to combine either "good" with "bad" or "good" with "good." Under this assumption, one may assume that the observed differences in second-order risk aversion indicate differences in the distribution of the two types between subject pools. Accordingly, based on the evidence on second-order risk aversion, we expect less mixed risk-averse and more mixed risk-loving behavior in China:

> **Hypothesis 1**: *Chinese people make fewer mixed risk-averse and more mixed risk-loving choices than Americans and Germans.*

2.2.2. *Differences in stake sizes.* Markowitz (1952) was among the first who argued that second-order risk preferences could change with increasing wealth. He suggested that the utility function, for levels of wealth above present wealth, is first convex and then concave. Therefore, Markowitz assumed that individuals are second-order risk-loving when the stakes are small and second-order risk-averse when the stakes are high. Pratt (1964) and Arrow (1965), who introduced—independently of each other—the now famous Arrow–Pratt coefficients as measurements for absolute and relative risk aversion, also assumed increasing risk aversion with increasing wealth. Similar assumptions were made by Eeckhoudt and Kimball (1992) and Kimball (1992) regarding 3-RA and by Eeckhoudt et al. (1996) and Gollier and Pratt (1996) regarding 4-RA.

Empirically testing these theoretical assumptions has been a challenge, since it requires a considerable variation of the stake size. The most common approach to this is to conduct experiments in developing countries, where large monetary incentives can be provided at lower cost than in developed countries. These studies typically observe more second-order risk-averse choices with higher stakes when eliciting risk preferences using binary gambles (Binswanger, 1980, 1981; Grisley and Kellog, 1987; Wik et al., 2004) or tasks based on eliciting certainty equivalents (Kachelmeier and Shehata, 1992a; Fehr-Duda et al., 2010). However, similar observations have been made in developed countries. Holt and Laury (2002, 2005) elicit second-order risk aversion (using a price list format) in the United States. They find that (relative) risk aversion increases with real stakes but not with hypothetical stakes.

There are only two experimental papers that consider the relationship between stake size and higher order risk preferences. Deck and Schlesinger (2010) confront subjects with 10 choices between lottery pairs. These lotteries have an overall expected payoff of $25.80. The comparison of two choices allows them to study the influence of a fivefold increase in the stake on 3-RA;

[6] Of course, China, the United States, and Germany also differ in economic, social, and political measures that may correlate with risk preferences. The existing evidence is broadly consistent, with Chinese people being the least second-order risk-averse. Falk et al. (2015) conducted the first representative survey comparing economic preferences around the globe. The authors correlate the average risk attitude in 76 countries with other country characteristics. Based on the five (weakly) significant correlations they observe, second-order risk aversion should be greatest in Germany. With regard to three measures (degree of redistribution, life expectancy, and degree of inequality), we would expect Chinese people to be the least risk-averse. With regard to two other measures (rigidity of employment laws and number of homicides per capita), we would expect Americans to be the least risk-averse. Rieger et al. (2015) find a positive correlation between (log) GDP per capita and second-order risk aversion in the gain domain. A similar observation is made in the large experimental study by Vieider et al. (2015), using monetary incentives. These authors elicit the risk preferences of students in 30 countries and observe a positive correlation between (log) GDP per capita and second-order risk aversion. Based on these correlations, we would expect Chinese people to be the least risk-averse compared to people from Germany and the United States. Further experimental comparisons between China and Western countries have been conducted by Kachelmeier and Shehata (1992a) and Ehmke et al. (2010).

two more comparisons of lottery choices allow them to study the influence on 4-RA. Deck and Schlesinger (2010) find weak support for the hypothesis that 3-RA preferences are more pronounced when stake sizes are higher (approximately one-third of their subjects changed their behavior when the stake size increased, and 70% of them made more 3-RA choices). Although they find mostly 4-RL behavior in their subject population, 4-RL behavior is less common when the stakes are higher.

Noussair et al. (2014) study the prevalence of 3-RA and 4-RA in a laboratory experiment, as well as in a large representative sample of the Dutch population. In expectation, participants in their real payoff treatments earn €7 (because the lotteries have an expected value of €70 but only one in 10 participants is paid). Noussair et al. (2014) find that 2-RA and 4-RA increase when the hypothetical stakes are increased (from €70 to €10,500). They find no significant difference between the real monetary payoff treatments and a treatment with hypothetical payoffs (in which lotteries have an expected value of €70, but no one is paid). They also do not find any stake size effect for 3-RA. However, in a direct test of higher order risk preferences and their relationship to an endowment to risk ratio, Noussair et al. (2014) find decreasing absolute 3-RA and decreasing absolute 4-RA. Moreover, their estimated parameters regarding their expo-power utility functions show increasing relative 3-RA and increasing relative 4-RA.

Although the theory suggesting two simple types of preferences (for either (i) combining "good" with "bad" or (ii) combining "good" with "good") does not predict a change of type if the stake size changes, the empirical evidence indicates that (relative) 2-RA increases with higher stakes. In addition, there is limited evidence indicating an increase in (relative) 3-RA and (relative) 4-RA. We formulate our second hypothesis accordingly:

> **Hypothesis 2**: *The number of mixed risk-averse choices increases and the number of mixed risk-loving choices decreases when the stake size increases.*

2.2.3. *Differences through displaying reduced instead of compound lotteries.* According to most theories of decision making, displaying actuarially equivalent lotteries as compound or reduced lotteries does not influence choices.

With respect to second-order risk aversion, it has been known for a while that reduced lotteries are often valued differently than compound lotteries. Early experiments in psychology report, for example, that people overestimate the joint probabilities in compound lotteries (see, e.g., Slovic, 1969; Bar-Hillel, 1973 and the overview in Budescu and Fischer, 2001). In economics, the observation that reduced lotteries are valued differently has been used to explain the pattern of preference reversals as observed by Lichtenstein and Slovic (1971), Lindman (1971), and Grether and Plott (1979), as well as ambiguity aversion as identified by Ellsberg (1961). For example, Segal (1988) shows that a violation of the reduction of compound lotteries axiom (ROCL) can generate preference reversals even if the independence axiom holds.[7] Segal (1987, 1990) also shows that ambiguity aversion can be explained by relaxing the ROCL and applying Quiggin's (1982) rank-dependent utility model.

Higher order risk preferences are typically elicited using compound lotteries. To our knowledge, only Maier and Rüger (2012), Deck and Schlesinger (2017) and Baillon et al. (2017) have

[7] Consider a two-stage lottery and a one-stage lottery yielding the same prizes as the two-stage lottery with the probabilities multiplied out. The ROCL then states that a decision maker is indifferent between these two lotteries (see Samuelson, 1952). Note that these theoretical results are directly related to the so-called "random lottery incentive mechanism," that is, the random selection of one of several lotteries for paying subjects in experiments, while treating choices within lotteries as if made in isolation. This is done to elicit preferences across multiple lotteries in an incentive-compatible way while keeping wealth constant. This procedure has become the norm in experimental economics (Baltussen et al., 2012). It is typically justified with reports of small or unsystematic differences between behaviors under different payment protocols (see, e.g., Starmer and Sugden, 1991; Beattie and Loomes, 1997, Cubitt et al., 1998). Recently, however, the random lottery incentive mechanism has been criticized by Harrison and Swarthout (2014), Harrison et al. (2015), and Cox et al. (2015). They mainly point out the logical inconsistency in assuming the independence axiom holds when paying based on the random lottery mechanism, while taking violations of the independence axiom across lottery choices at face value.

TABLE 1
TREATMENTS AND DESIGN

	N	Lotteries by Deck and Schlesinger (2014) (Order, C = Compound, R = Reduced)	ECU to Local Currency[1]	Average Payoff Local Currency[2]
CHN	140	1; 2; 3C; 4C; 5C; 6C	2.90	47.06
USA	129	1; 2; 3C; 4C; 5C; 6C	0.93	28.14
GER	145	1; 2; 3C; 4C; 5C; 6C	0.61	18.10
CHN 10×	48	1; 2; 3C; 4C; 5C; 6C	29.00	571.92
Compound & Reduced	143	1; 2; iC; iR; jC; jR $i, j \in \{3, 4, 5, 6\}, i \neq j$	0.61	17.40
Follow-up Experiment	224	iC or iR $i \in \{3, 4\}$	0.61	12.05

[1]Equals \$0.47 (CHN), \$0.68 (GER), and \$4.67 (CHN 10×) at the time of the experiment.
[2]Equals \$7.59 (CHN), \$20.18 (GER), \$92.24 (CHN 10×), and \$19.49 (Compound & Reduced) and \$13.43 (Follow-up) at the time of the experiment. All payments except for the Follow-up Experiment included a show-up fee of \$8.50, which was adjusted for China and Germany using the respective exchange rates.

used reduced lotteries to elicit higher order risk preferences. In the gain domain, Maier and Rüger (2012) observe 55% of choices to be 2-RA, 60% to be 3-RA, and 58% to be 4-RA.[8] These percentages are at the lower end of the range of observed frequencies in other studies (see also Table A.1 in the Appendix) and thus support the conjecture by Deck and Schlesinger (2014) that the reduced form may obfuscate the "good" with "bad" or "good" with "good" interpretation. In a study conducted in parallel to ours, Deck and Schlesinger (2017) find a significant framing effect between the compound and the reduced presentation of the lotteries. They observe less 4-RA and 5-RA in reduced lotteries but no difference in the frequency of 3-RA choices when lotteries are displayed in a reduced instead of a compound form. Furthermore, in a recent study, Baillon et al. (2017) also elicit preferences using reduced lotteries and find little 4-RA: In their study, 84% of choices are 2-RA, 71% are 3-RA, and 43% are 4-RA.

In summary, the presentation of lotteries in a reduced instead of a compound form might influence choices, if decision makers violate the independence axiom or the ROCL. Based on the prior empirical results and the conjecture by Deck and Schlesinger (2014), we expect less mixed risk-averse and less mixed risk-loving choices when using a reduced framing:

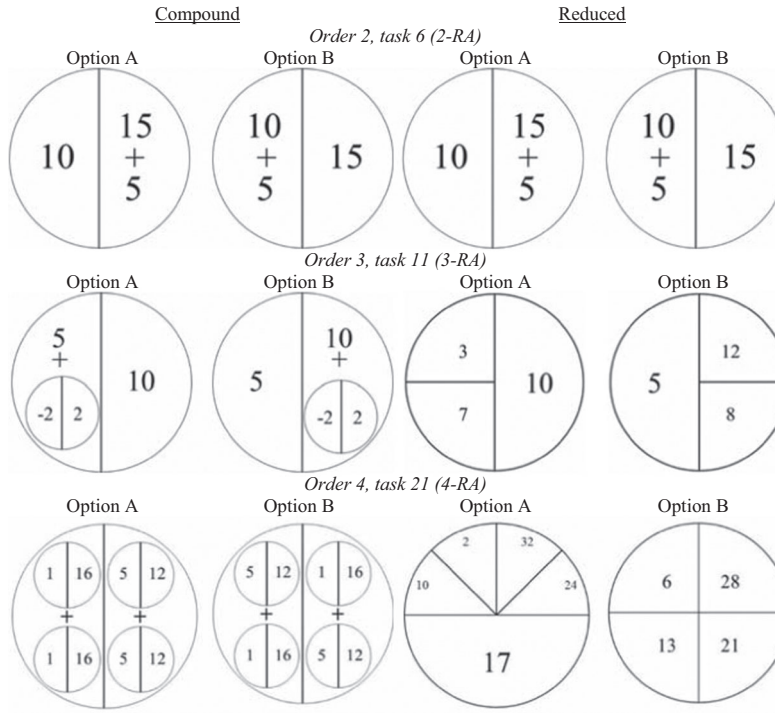
Hypothesis 3: *The number of mixed risk-averse choices and the number of mixed risk-loving choices decrease when lotteries are displayed in a reduced instead of a compound framing.*

3. EXPERIMENTAL DESIGN

3.1. *Elicitation Method.* Our elicitation method follows Deck and Schlesinger (2014) that comprises 38 tasks (see Deck and Schlesinger 2014, 1922ff; Online Appendix O1). Each of these tasks involves choosing between Option A and Option B. Examples of the 2-RA, 3-RA, and 4-RA lotteries (i.e., lotteries of orders 2, 3 and 4) as presented to the participants in compound form are shown on the left-hand side of Figure 3. Each option involves different amounts of money, and each 50–50 lottery is represented as a circle with a line through the middle.

For example, Option A in the 2-RA task (order 2, task 6) involves a 50–50 chance of winning either 10 ECU or 20 ECU (where ECU stands for experimental-currency-units; see Subsection 3.3 for details and Table 1 for the exchange rate of ECU to the local currency). Following Deck and Schlesinger (2014), all outcomes are shown in the domain of gains. Let us assume that $W = 20$ and $k_1 = k_2 = 5$, where W denotes wealth and k_1 and k_2 (certain) losses that are subtracted from wealth (see Figure 1). In Option A, 10 ECU represents $W - k_1 - k_2 = 10$, whereas $(15 + 5)$ ECU represents the initial wealth $W = 20$. Option B represents the lottery where the harms are disaggregated, that is, $[W - k_1, W - k_2]$. In this example, this corresponds to a sure outcome of 15 ECU. Both lotteries have the same expected value of 15 ECU. However, Option A is risky,

[8] Loss and gain framings have been compared previously when eliciting higher order risk preferences, with little or no difference being reported (Deck and Schlesinger, 2010; Maier and Rüger, 2012).



NOTES: First- and second-order tasks (here: task 6) do not differ in the compound and reduced presentation.

FIGURE 3

EXPERIMENTAL EXAMPLES OF LOTTERIES OF ORDERS 2, 3, AND 4 AS PRESENTED TO PARTICIPANTS

whereas Option B is not. Thus, a 2-RA individual should choose the certain option over the risky one when the expected values are the same.

In the 3-RA task (order 3, task 11), the outcomes of a 50–50 lottery may contain another lottery. For example, Option A involves a second lottery with a 50–50 chance of winning either –2 or 2 ECU. Thus, the participant has a 25% probability of winning $5 - 2 = 3$ ECU, a 25% probability of winning $5 + 2 = 7$ ECU, and a 50% probability of winning 10 ECU. Because $[X, Y]$ denotes a lottery where there is a 50–50 chance of receiving X and a 50–50 chance of receiving Y, then Option A can also be written as $[5 + [-2, 2], 10]$. Let us assume that $W = 10$ and $k_2 = 5$. Moreover, the sure loss k_1 is replaced by a zero-mean background risk ε, which itself is a lottery (here $[-2, 2]$). Then Option A corresponds to $[W - k_2 + \varepsilon, W]$ and Option B to $[W - k_2, W + \varepsilon]$. Eeckhoudt and Schlesinger (2006) define 3-RA as a preference for disaggregating a sure loss and an additional zero-mean background risk. Therefore, a 3-RA individual should prefer Option B over Option A.

In the 4-RA task lottery (order 4, task 21), the outcomes in a 50–50 lottery may contain not just one but two other lotteries. The example shown is a composition of (2+2)-th-order risk, because Option A can be written as $[[1, 16] + [1, 16], [5, 12] + [5, 12]]$ and Option B as $[[5, 12] + [1, 16], [1, 16] + [5, 12]]$. As the lottery $[1, 16]$ has more second-order risk than the lottery $[5, 12]$, the different compositions of both lotteries differ in their fourth-order risk. An individual who prefers lotteries with lower fourth-order risk is 4-RA and would choose Option B over Option A. Because the outcomes of option B are composed of the more risky and the less risky lottery, it generates less fourth-order risk than option A (cf. Subsection 2.1).

The right-hand side of Figure 3 also shows the corresponding reduced lottery pair for each compound lottery pair. The reduced lotteries can be derived by multiplying out the probabilities of the potential outcomes. The resulting lottery is actuarially equivalent to the compound lottery—that is, it yields the same probability distribution over outcomes.

3.2. *Experimental Treatments.* We initially conducted an economic laboratory experiment with sessions in China, the United States, and Germany. Subjects faced 38 tasks in randomized order, and one of the tasks was randomly selected for payment. In each task, the position (left or right) of the two lotteries was determined randomly. Subjects had to choose between them, revealing their risk preference (the instructions are shown in Appendix A.2).[9]

The treatments, the orders of the lotteries, and the number of subjects are shown in Table 1. Lottery pairs that were displayed in the original compound framing are identified by the suffix "C." Reduced lottery pairs are identified by the suffix "R." In addition, Table 1 includes the country-specific exchange rate and average payoffs in the local currency.

To investigate the effects of the stake size, we increased the payoff 10-fold for 48 additional Chinese subjects. The participants in the CHN 10× treatment participated in the same sessions as the Chinese subjects with regular payment. This allowed us to randomize the assignment of Chinese subjects to treatments.

To investigate the effect on choices of compound and reduced lotteries, we ran additional sessions in Germany (Compound & Reduced). All of the 143 participants faced the original choices in order 1 and order 2 (and with the exception of one lottery in order 1, none of these were displayed in compound framing). Each subject faced lotteries of two additional orders in the original compound framing, as well as in the reduced framing. This allows us to compare the differences in behavior toward compound and reduced lotteries of orders 3–6 *within* subjects. All six combinations were run in each session and subjects were randomly assigned to orders.

To shed some more light on the effects of framing, we conducted a Follow-up Experiment in Germany with 224 subjects divided randomly into four different conditions in a 2 × 2 between-subjects design. We confronted participants with an incentivized 3-RA lottery choice or an incentivized 4-RA lottery choice. In both conditions, approximately half of the participants saw the respective lottery pair in the compound framing (58 in 3-RA and 54 in 4-RA), whereas the others saw it in the reduced framing (56 in 3-RA and 56 in 4-RA). Participants in all four conditions were matched into groups of two and had to make a choice between two lotteries they were facing.[10]

Adapting an experimental design by Burchardi and Penczynski (2014) to lottery choices, our Follow-up Experiment incentivizes decision makers to reveal the reasoning behind their choice. More specifically, participants were asked to send one written free-form message, together with their preferred choice, to the other participant in their group. Subjects knew that their choices could be revised after both pieces of information were exchanged. The instructions stated "*Before you enter your final decision, you have the opportunity to influence the final choice of your partner: Before the decisions are entered, you will send a preferred choice together with a text message to your partner.*" Subjects also knew that the final choice of one member of the

[9] Paying one randomly determined task adds another layer of compounding to the lotteries. Following Deck and Schlesinger (2014), we nevertheless used the random payment technique, because the subjects' wealth is not influenced during the course of preference elicitation and because this allows for straightforward comparison with previous studies on higher order risk preferences, all of which use this method (see Appendix A.1 for more details). Furthermore, collecting multiple lottery choices from each subject is essential for answering our research questions. Eliciting only one decision per subject would not allow us to identify mixed risk-averse or mixed risk-loving types. Finally, as Azrieli et al. (2018) point out, the random payment technique "is essentially the only incentive compatible mechanism." In all sessions, the elicitation of lottery preferences was preceded by four control questions. These control questions were also used by Deck and Schlesinger (2014). The subjects were asked to state the potential payoffs in two lotteries, as well as the maximum and minimum payoffs of a compound lottery, as in Deck and Schlesinger (2014) and as shown in Online Appendix O2. All subjects were able to answer these four questions correctly (see Online Appendices O4, O6, and O8 for more details). The elicitation of lottery preferences was followed by the administration of a questionnaire containing basic demographic questions and questions to determine whether the participant had migrated to the current country (these questions were similar to those used by Sutter et al., 2013).

[10] We selected the 3-RA lottery, in which we observed the largest share of preference reversals within-subjects in the experiments described above (subject to having three different potential outcomes). In addition, we selected the 4-RA lottery, in which we observed the largest share of preferences reversals. These are tasks 11 and 21 shown in Figure 3 in Subsection 3.1. For both lottery pairs, we randomly varied the position of the more 3-RA or more 4-RA option (left or right).

group would be randomly selected after both members had entered their final decision. The final lottery choice of the selected member would determine the payoffs of both. Thus, the message was the only way to influence the other group member, who decided on the payoff-relevant lottery with a probability of one half.

We developed a classification scheme to analyze the content of these messages. This classification scheme was based on prior considerations and a nonincentivized survey that we reported in our working paper (Haering et al. 2017). Two research assistants who were unaware of the research questions and not involved in any other experimental studies first coded the old survey data, based on the existing classification scheme. Then any discrepancies in their classification were discussed with one of the authors to clarify misunderstandings. In addition, examples for each content category were selected. These examples served the coders as a reference during the classification of the 224 messages from the Follow-up Experiment, which they coded independently.

In all experiments, we also implemented the cognitive reflection test (CRT), as well as the Berlin Numeracy Test (BNT). The CRT consists of three questions and was developed by Frederick (2005) to assess the ability to resist reporting the response that first comes to mind. He finds that this measure is correlated with different measures of cognitive ability and varies widely between American universities. In his study, those who answer more questions correctly are also less second-order risk-averse in the gain domain. It has been reported that higher cognitive ability is associated with lower second-order risk aversion (e.g., by Burks et al. 2009; Dohmen et al. 2010). Therefore, we concluded that this test may capture differences in risk taking that are due to differences in cognitive ability between our subject pools. However, note that Noussair et al. (2014) find no such relationship, although in their student sample those who score more highly on the CRT are significantly more 3-RA. The BNT was developed by Cokely et al. (2012) and consists of four questions that aim to assess statistical numeracy and risk literacy. Cokely et al. report that the BNT successfully discriminates between participants on the basis of their numeracy in 15 countries, including China, the United States, and Germany. Furthermore, they find that the BNT is highly predictive of the ability to make a correct assessment of the everyday risks associated with consumption, health, or medical choices.

The sessions of our experiments were conducted at the experimental lab at Nankai University in Tianjin (China), at CLER at Harvard Business School in Boston (USA), and at the elfe laboratory at the University of Duisburg-Essen in Essen (Germany). No subject participated in more than one session. The experiment was computerized and programmed using zTree (Fischbacher, 2007). Screenshots are provided in Online Appendix O2.

3.3. *Experimental Conditions across Countries.* To create similar conditions in the CHN, USA, and GER treatments, we followed best practice as described below. To minimize currency effects, the payoffs in ECU were the same in all sessions, but the exchange rate for one ECU was different in every location (see Bohnet et al., 2008; Herrmann et al., 2008; Özer et al., 2014 for similar approaches). We selected exchange rates by putting equal weight on the UBS Prices & Earnings survey (UBS, 2014) data (this measure is also used by Özer et al., 2014), and the country-level purchasing power parity provided by the OECD (2015) (this measure is also used by Roth et al., 1991; Buchan and Croson, 2004; and Ehmke et al., 2010). This procedure led to payments that were inside the feasible bandwidth for subject payments in Tianjin and Boston but were somewhat higher than the usual average payoff in Tianjin and somewhat lower than the usual average payoff in Boston. Therefore, we adjusted the payments by 5% in the direction of the usual average payoff.[11]

[11] There was no reliable data on purchasing power available for Tianjin, Boston, and Essen. Thus, we based our calculations on the UBS data for Beijing, New York City, and Berlin. This includes that country-level data adjust for the fact that some students commute into the metropolitan areas and many spend a significant amount of time in more rural areas. The rules of the laboratory in Essen, Germany require experimenters to base expected payments on an hourly student wage of €12.50. Using this anchor, we calculated payments in China and the United States. Note that Vieider (2012) finds no influence of small variations in payoffs (±20%) on second-order risk aversion.

To minimize potential experimenter effects, all experimenters followed the same detailed protocol in all countries (see, e.g., Roth et al., 1991; Buchan and Croson, 2004; and Herrmann et al., 2008 for similar approaches). The experiments in China and in the United States were conducted by local experimenters who also spoke German. The two local experimenters also conducted one session each in Germany, which allowed us to control for idiosyncratic experimenter effects (Bohnet et al., 2008; Özer et al., 2014). These measures have also been advocated by Roth et al. (1991). As an additional measure of control, one lead experimenter from Germany was present (but not visible to subjects) to oversee the procedures in China and in the United States (see Buchan and Croson, 2004; Herrmann et al., 2008, for a similar approach). To ensure that the instructions were similar, we only used written instructions. These, along with all the computer pages, were translated using the back translation procedure (Brislin, 1970). This procedure is now commonly applied in cross-cultural research in economics (see, e.g., Buchan and Croson, 2004; Bohnet et al., 2008; Herrmann et al., 2008; Ehmke et al., 2010; Özer et al., 2014).

We attempted to conduct our study with subject pools that were as similar as possible, despite their different locations. Therefore, we only used student subjects, because they have a similar educational level and are of a similar age. In all three countries, the subjects were recruited from a subject database.[12] We were able to recruit samples that were similar in their gender composition in all three countries. However, the databases were either not large enough or did not contain enough information to allow us to recruit samples that were similar for additional demographic characteristics. Therefore, we used the additional information on the participants we collected using a postexperimental questionnaire to control for differences between the subject pools in our analysis. Also note that Vieider et al. (2015) and Ehmke et al. (2010) find little difference in experimentally elicited risk preferences between student subject pools at different locations within the same country.[13]

4. RESULTS

4.1. *Summary Statistics.* Table 2 summarizes the characteristics of all participants. In the sessions for country comparison, slightly more women than men participated in all three countries (CHN, the USA, and GER). Because we were not able to recruit subjects based on their age and gender in China and the United States, we conducted the sessions in Germany last. In Germany, we aimed to stratify our sample on the basis of the composition of the subjects recruited in the other two countries. However, we were not able to match the previous samples fully because the age structure of student populations differs across countries. Thus, the age distribution of the German participants differs significantly from the joint distribution of the Chinese and American subjects ($p = 0.010$, two-sided Mann–Whitney U test). The proportion of female subjects does not differ significantly between the German and the joint subject pools ($p = 0.883$, Fisher's exact test).

[12] In the United States and Germany, we relied on existing databases. In both countries, this procedure was handled via ORSEE (Greiner, 2015). In China, however, we had to build a database from scratch. Recruitment for this database was comparable to the procedures employed in the United States and in Germany. Two student assistants advertised participation by distributing flyers on campus and giving presentations in lectures. The advertisement promised the opportunity to earn a monetary reward for participation in an economic experiment. Potential participants could register via e-mail or text message.

[13] We had to make two adjustments because of American regulations, for which we control in regression analyses. First, in the United States, it was necessary to inform subjects about the expected payoff and the nature of our experiment in the recruitment e-mail. Second, it was necessary to present subjects with an IRB consent form in the laboratory prior to the experiments. The IRB form contained additional information regarding the experimental procedure, a short description of the task, and the expected payoffs. Neither of these two measures was required in Germany and China. Therefore, the Chinese participants received neither prior information in the recruitment e-mail nor an IRB consent form. To control for this difference, we used the American procedures in half of the sessions conducted in Germany. In other words, in these sessions German subjects were recruited via a German version of the American e-mail invitation and received a translation of the IRB consent form prior to the experiment.

TABLE 2
SUMMARY STATISTICS

	Demographics		Tests	
	Female	Age (SD)	CRT (SD)	BNT (SD)
CHN ($N = 140$)	57.9%	22.186 (2.337)	1.628 (0.840)	2.879 (1.254)
USA ($N = 129$)	62.0%	23.054 (5.039)	1.667 (1.106)	2.047 (1.262)
GER ($N = 145$)	61.4%	22.993 (2.835)	1.290 (1.154)	1.393 (1.144)
CHN 10× ($N = 48$)	50.0%	22.604 (2.574)	1.688 (0.879)	3.000 (1.187)
Compound & Reduced ($N = 143$)	68.5%	23.818 (3.320)	1.280 (0.982)	1.329 (1.099)
Follow-up Experiment ($N = 224$)	67.9%	24.470 (6.852)	1.201 (1.092)	1.277 (1.001)

Notes: *N*, number of participants; SD, standard deviation; CRT, number of correct answers out of three in the Cognitive Reflection Test; BNT, number of correct answers out of four in the Berlin Numeracy Test.

With respect to the BNT, on average the Chinese participants were able to solve 2.879 of the four questions correctly, which is higher than the 2.047 correct answers in the United States ($p < 0.001$, two-sided Mann–Whitney U test). With 1.393 correct answers on average, the German subjects provided even fewer correct answers than the Americans ($p < 0.001$). With respect to the CRT, participants in China and the United States did not differ significantly ($p = 0.568$). They were able to solve a little more than half of the three questions correctly. In Germany, the rate was lower, with 1.290 correct answers ($p \leq 0.007$).[14]

Table 2 also summarizes the characteristics of the Chinese subjects, who participated in the high stakes treatment (CHN 10×). There are no significant differences between these participants and the subjects facing regular stakes (CHN) ($p \geq 0.318$, two-sided Mann–Whitney U tests, for age and test scores; $p = 0.401$, Fisher's exact test, for gender composition).

Moreover, Table 2 shows the summary statistics of the German participants, who were confronted with different lottery formats (Compound & Reduced). For this analysis, we did not stratify the selection of participants because we are interested in the within-subject comparison. In this experiment, the proportion of women does not differ from that for the remaining German (GER) data ($p = 0.219$, Fisher's exact test). The subjects are older ($p = 0.048$, two-sided Mann–Whitney U tests), but neither the CRT nor the BNT scores differ between the two groups ($p \geq 0.627$).

The sample recruited for the Follow-up Experiment does not differ significantly from the sample that participated in Compound & Reduced treatment ($p \geq 0.365$, two-sided Mann–Whitney U tests, for age and test scores; $p = 0.909$, Fisher's exact test, for proportion of female subjects).

4.2. *Higher Order Risk Preferences across Countries.*

4.2.1. *Aggregate risk preferences.* There were seven choices to be made for each order except the first. Following Deck and Schlesinger (2014), we use the number of n-RL choices as a measure of nth order risk aversion—that is, the more the n-RL choices are, the lower the nth order risk aversion. We assume that all participants prefer more money to less money. This assumption is supported by the data in treatments CHN, the USA, and GER: 90% of the subjects in China, 99% of the subjects in the United States, and 97% of the subjects in Germany never choose a dominated payoff in order 1. The number is slightly smaller in China than in the United States and in Germany ($p = 0.001$, Fisher's exact test). This is similar to the figure of more than 92% observed by Deck and Schlesinger (2014). We expect participants to differ in their preferences in orders 2–6, but consider the aggregate data first before analyzing individual patterns.

[14] Frederick (2005) observed that students at Princeton University answered 1.63 questions correctly on average ($N = 121$), whereas students at the University of Michigan (Ann Arbor) answered 1.18 questions correctly ($N = 1{,}267$). Brañas-Garza et al. (2015) provide a meta-study.

TABLE 3
*N*TH ORDER RISK-LOVING CHOICES ACROSS COUNTRIES

Order:	1	2	3	4	5	6
H_0:	1.5	3.5	3.5	3.5	3.5	3.5
CHN						
Mean	0.100***	1.971***	1.721***	2.514***	2.343***	3.007***
Std. Dev.	(0.301)	(1.952)	(1.763)	(1.868)	(1.691)	(1.879)
Median	0	1	1	2	2	3
USA						
Mean	0.008***	1.628***	1.612***	2.558***	2.791***	3.291
Std. Dev.	(0.088)	(1.957)	(2.063)	(1.849)	(1.560)	(1.622)
Median	0	1	1	2	3	3
GER						
Mean	0.034***	1.628***	1.676***	2.500***	2.676***	3.021***
Std. Dev.	(0.183)	(1.900)	(1.700)	(1.680)	(1.615)	(1.516)
Median	0	1	1	2	3	3

*$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$, two-sided, Wilcoxon signed-rank tests.

Table 3 shows the number of *n*-RL choices in each country. In all countries, we observe a general tendency of subjects to avoid the more risky lotteries. The number of *n*-RL choices is always significantly lower than 3.5 (which is the expected average count with random behavior) ($p < 0.001$, two-sided one-sample Wilcoxon signed-rank tests). Only in the United States does the choice frequency for order 6 not differ significantly from 3.5 ($p = 0.242$).[15] The correlations of the individuals' share of *n*-RL choices between orders 2 and 6 are shown in Online Appendix O10 for all our treatments.

Comparing the frequencies of *n*-RL choices between countries only indicates a difference for order 2. To control for subject pool differences, we run an ordinary least squares (OLS) regression for each order with the number of *n*-RL choices as the dependent variable, dummies for China and Germany (so the United States acts as the baseline category) and various controls as independent variables (see Appendix A.3 for an overview of the variables, Appendix A.4 for the regression results, and Online Appendix O3 for details on our estimation strategy).[16] For order 2, the regression suggests that the Chinese subjects make more risk-loving choices than the German ones. The Chinese country dummy indicates no significant difference between Chinese and American subjects ($\beta = 0.981$, robust SE $= 0.672$, $p = 0.145$, two-sided). The same regression also yields no difference between the United States and Germany, as measured by the German country dummy ($\beta = 0.066$, robust SE $= 0.590$, $p = 0.910$). However, the country dummies of China and Germany differ significantly ($p = 0.009$, two-sided Wald test).

[15] The distributions of choice frequencies within each order across the three countries, across stakes, and across lottery formats are shown in Online Appendix O9.1.

[16] The regression analyses include the demographic and test results listed in Table 2, as well as controls for the experimenter and the IRB (cf. footnote 13). The latter two are influential. First, we observe a significant influence of our Chinese experimenter. In the German session conducted by him, participants behaved in a more second-order risk-averse manner. Second, we find that in the German sessions in which the subjects were provided with IRB information, the subjects behaved in a less risk-averse manner. Because of a computer error, we could not collect the CRT and BNT scores for eight subjects in China. Therefore, we also report additional regressions without controlling for CRT and BNT. In Online Appendix O4, we present further robustness checks of this model. We asked subjects about their migration background. In an additional analysis, we use a control variable for those who were not born in the respective country or who did not answer the relevant question. This was the case for 26 subjects in the United States and seven subjects in Germany. However, none of these further robustness checks suggests a different interpretation of our data. Furthermore, making dominated choices in order 1 is an obvious error and can be viewed as a check for data quality. In the regressions presented in Online Appendices O4, O6, and O8, we provide robustness checks controlling for dominated choices.

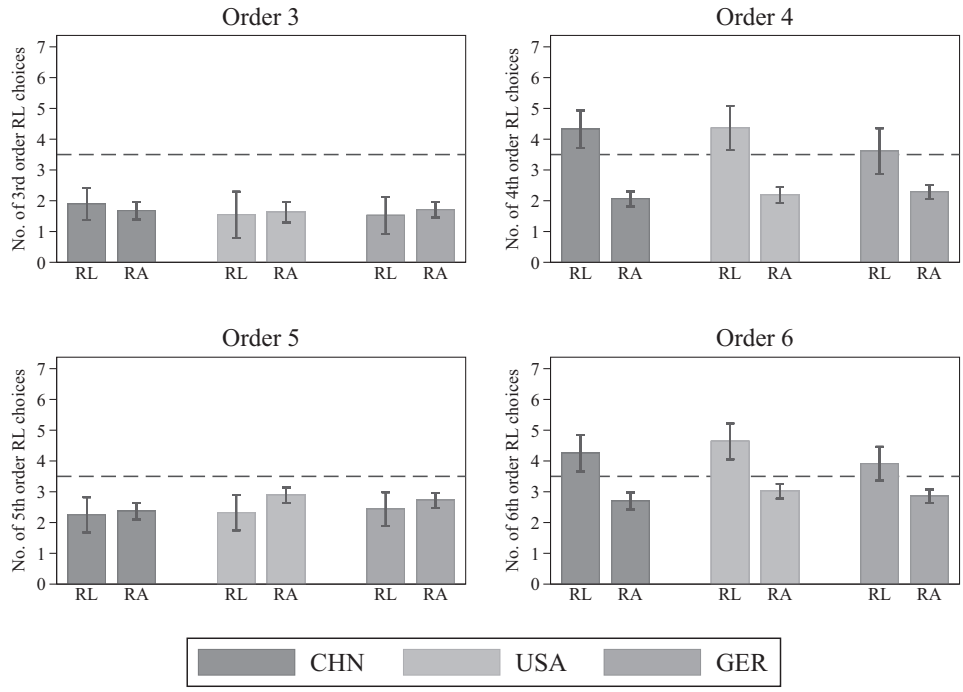


FIGURE 4

AVERAGE NUMBER OF *N*TH ORDER RISK-LOVING (N-RL) CHOICES BY RISK-AVERSE (RA) AND RISK-LOVING (RL) SUBJECTS

4.2.2. *Consistency of risk preferences.* Based on a preference for combining "good" with "bad" as described by Deck and Schlesinger (2014), individuals should be 2-, 4- and 6-RA: that is, they should exhibit mixed risk-averse behavior. In contrast, individuals who have a preference for combining "good" with "good" should be 2-, 4- and 6-RL: that is, they should exhibit mixed risk-loving behavior. Both mixed risk averters and mixed risk lovers should be 3-RA and 5-RA (see Subsection 3.1). In a first step, we follow Deck and Schlesinger (2014) and study consistency in the higher orders relative to order 2. In a second step, we classify the subjects based on all orders.

For the first step, we classify subjects as second-order risk-averse or risk-loving according to whether they make choices in line with this preference in the majority of their seven decisions in order 2. With this classification scheme, 80% of the Chinese participants, 83% of the Americans, and 84% of the Germans are classified as second-order risk-averse, and the remaining subjects as second-order risk-loving. To identify mixed risk-averse and mixed risk-loving behavior, we consider the behavior of the two groups for the higher orders.

Figure 4 displays the average number of *n*-RL choices for second-order risk averters ("RA") and second-order risk lovers ("RL") across the three countries. In addition, it includes a dashed line at 3.5 indicating the number of *n*-RL choices expected under random behavior, as well as 90% confidence intervals.[17] Note that the confidence intervals are larger for the risk lovers because of the smaller number of observations. For the odd orders 3 and 5, the graph reveals a preference for the more risk-averse option for second-order risk averters and risk lovers. For the even orders 4 and 6, the two types differ, and only risk averters tend to prefer the less risky options in the higher orders. This is exactly the pattern that would be expected when decisions are mainly made by mixed risk averters and mixed risk lovers.

[17] The distributions of choice frequencies within each order across the three countries, across stakes, and across lottery formats are shown Online Appendix O9.2 separately for risk-loving and risk-averse subjects.

TABLE 4
PERCENTAGE OF SUBJECTS WHO ARE CLASSIFIED AS MIXED RISK-AVERSE OR MIXED RISK-LOVING

	Mixed Risk-Averse				Mixed Risk-Loving				Mixed Risk-Averse or Risk-Loving+			
Threshold	CHN (%)	USA (%)	GER (%)	All (%)	CHN (%)	USA (%)	GER (%)	All (%)	CHN (%)	USA (%)	GER (%)	All (%)
$p < 0.01$	42	45	42	43	9	9	6	8	51	54	48	51
$p < 0.05$	54	53	57	55	13	12	8	11	67	65	65	66
$p < 0.10$	64	60	63	62	15	15	11	14	79	75	74	76

Notes: Classification based on binomial tests with different significance thresholds: $p < 0.01$, $p < 0.05$, or $p < 0.10$, which represent 27, 26, or 25 consistent choices out of 38 possible choices.
+In all countries, the share of classified subjects is significantly different from the share that would be expected under random behavior ($p < 0.001$, Fisher's exact tests).

This impression is confirmed by nonparametric tests: second-order risk averters and risk lovers in all countries are 3-RA and 5-RA when comparing the number of n-RL choices to the benchmark of 3.5 ($p \leq 0.005$, two-sided one-sample Wilcoxon signed-rank tests), but only Second-order risk averters are also 4-RA and 6-RA ($p \leq 0.005$). Second-order risk lovers are instead 4-RL and 6-RL in China and in the United States ($p \leq 0.072$) but not in Germany ($p \geq 0.259$).

For the second step of the analysis, we consider choices of all orders at once, because—strictly speaking—the theory does not differentiate between any of the even orders or between any of the odd orders. All of a subject's individual choices can be classified as being consistent or inconsistent with mixed risk-averse behavior, for example. The classification yields a binary variable with 38 observations for each subject. Based on this, we can classify the subjects into types. Running a binomial test for each subject allows us to test the null hypothesis that half of his or her 38 choices adhere to the mixed risk-averse type, for example. If we can reject this hypothesis and most choices adhere to the pattern, we classify the subject as mixed risk-averse. The same procedure is applied for mixed risk-loving behavior.

Table 4 summarizes the proportion of subjects that can be classified into the two types. It lists the distributions for three significance thresholds.[18] Under the strictest criterion, between 42% and 45% are mixed risk-averse and between 6% and 9% are mixed risk-loving across the countries. If the criterion is relaxed, these percentages go up to between 60% and 64%, or 11% and 15%, respectively. The proportion of all subjects consistent with one type or the other ranges from 51% to 76%.

Next, we use individual behavioral patterns to compare the consistency across countries, by counting for each subject (i) the number of choices consistent with mixed risk-averse behavior, (ii) the number of choices consistent with mixed risk-loving behavior, and (iii) the maximum of both. As already suggested in Figure 4, there appears to be no difference in the behavioral patterns across countries. Running separate OLS regressions (see Appendix A.4) with these three dependent variables provides additional evidence with respect to Hypothesis 1. First, in the regression of the number of mixed risk-averse choices on country dummies, the dummies for China ($\beta = -1.191$, robust standard error (SE) $= 2.041$, $p = 0.560$, two-sided) and Germany are insignificant ($\beta = 0.689$, robust SE $= 1.716$, $p = 0.689$). But—reflecting the difference in 2-RA—the dummies differ weakly[19] from each other, suggesting that the Chinese subjects make somewhat fewer mixed risk-averse choices than the Germans ($p = 0.099$, two-sided Wald test). Second, in the regression of the number of mixed risk-loving choices on country dummies, the

[18] Note that only the 1% threshold guarantees a mutually exclusive classification when the subjects make 38 decisions across orders 1–6. An individual may be classified as being consistent with respect to both types when applying the 5% or the 10% threshold. However, when applying the 5% threshold, this is only the case for one subject in China and one subject in the United States. When applying the 10% threshold, this is the case for four subjects in China, one subject in Germany and five subjects in the United States.

[19] By weakly we mean significant at the 10% level.

TABLE 5
*N*TH ORDER RISK-LOVING (N-RL) CHOICES ACROSS STAKES

Order:	1	2	3	4	5	6
H_0:	1.5	3.5	3.5	3.5	3.5	3.5
CHN						
Mean	0.100***	1.971***	1.721***	2.514***	2.343***	3.007***
Std. Dev.	(0.301)	(1.952)	(1.763)	(1.868)	(1.691)	(1.879)
Median	0	1	1	2	2	3
CHN 10×						
Mean	0.083***	1.417***	1.646***	2.021***	2.167***	2.646***
Std. Dev.	(0.279)	(1.820)	(2.005)	(1.780)	(1.521)	(1.521)
Median	0	1	1	2	2	3

*$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$, two-sided Wilcoxon signed-rank tests.

dummies for China ($\beta = 1.858$, robust SE $= 1.630$, $p = 0.255$) and Germany ($\beta = 0.388$, robust SE $= 1.473$, $p = 0.792$) are insignificant. Again, the two dummies differ weakly ($p = 0.094$), meaning that the Chinese subjects make somewhat more mixed risk-loving choices. Third, when using the maximum of the two variables for each subject as the dependent variable in an OLS regression, we find no significant country differences ($p \geq 0.436$).

Observation 1: *Between* 51% *and* 76% *of all subjects can be classified as adhering to either mixed risk-averse or mixed risk-loving behavior across countries. After controlling for procedural differences and the subjects' characteristics, the Chinese participants are found to make weakly fewer mixed risk-averse choices and weakly more mixed risk-loving choices than the Germans.*

4.3. *Higher Order Risk Preferences across Stakes.*

4.3.1. *Aggregate risk preference.* As in the previous analyses, we interpret the number of *n*-RL choices as a measure of *n*th order risk aversion, and assume that all participants prefer more money to less. In CHN, 90% never choose a dominated payoff in order 1. In CHN 10×, this share is 92% and therefore not significantly larger ($p = 0.494$, Fisher's exact test).

Table 5 shows the number of *n*-RL choices under both incentive structures in all orders. The number of *n*-RL choices is significantly lower than would be expected under random behavior in all orders ($p \leq 0.010$, two-sided one-sample Wilcoxon signed-rank tests).

The previous evidence indicates that more second-order risk-averse choices are made when the stakes increase. The data presented in Table 5 suggest a similar effect: With regular stakes, participants make, on average, 1.971 decisions in a risky way, but under high stakes, the average is 1.417 decisions. To control for subject pool differences within China, we run OLS regressions separately for each order, with the number of *n*-RL choices as the dependent variable on a dummy for high stakes (and regular stakes as the baseline), as well as various controls (see Appendix A.5 for the regression results and Online Appendix O5 for details on our estimation strategy). For order 2, there is a weakly negative effect of increased stakes on the number of risky choices ($\beta = -0.616$, robust SE $= 0.319$, $p = 0.055$, two-sided), and there are no significant differences for the other orders ($p \geq 0.309$).

4.3.2. *Consistency of risk preferences.* In this subsection, we analyze whether the patterns of mixed risk-averse and mixed risk-loving behavior prevail under high stakes. Under the high stakes in CHN 10×, we also expect second-order risk averters to coincide in their choices with second-order risk lovers in the odd orders and to differ from them in the even orders.

Again, we analyze the consistency in two steps. In the first step, we classify the subjects into second-order risk-averse or risk-loving and analyze their behavior in the higher orders.

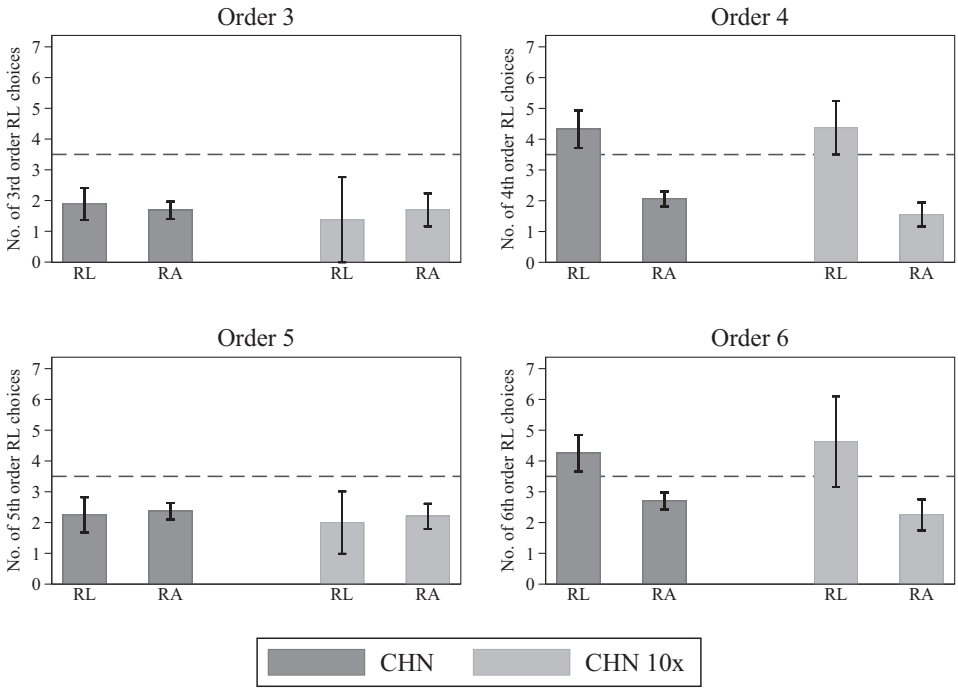


FIGURE 5

AVERAGE NUMBER OF *N*TH ORDER RISK-LOVING (N-RL) CHOICES BY RISK-AVERSE (RA) AND RISK-LOVING (RL) SUBJECTS

Classifying everyone with a majority of risk-averse choices as risk-averse yields 83% second-order risk-averse subjects in the CHN 10×, compared to 80% in CHN (the remaining subjects being classified as risk-loving).

Figure 5 displays the average number of *n*-RL choices for second-order risk averters ("RA") and second-order risk lovers ("RL") in the two treatments. For the odd orders 3 and 5 of CHN 10×, both types appear to favor the risk-averse option more frequently, whereas for orders 4 and 6 the two types appear to differ. This is the pattern suggested by the theory of mixed risk-averse and mixed risk-loving behavior.

Nonparametric tests also confirm this interpretation for CHN 10×, when comparing the number of choices to the 3.5 *n*-RL choices that would be expected under random behavior: Second-order risk averters and risk lovers significantly favor the more 3-RA or 5-RA options ($p \leq 0.048$, two-sided one-sample Wilcoxon signed-rank tests), whereas only second-order risk averters favor the more 4-RA and 6-RA options ($p \leq 0.001$). Second-order risk lovers weakly prefer the more risky options ($p = 0.084$) for order 4, whereas there is no significant tendency for order 6 ($p = 0.222$).

In the second step, we consider the decisions for all orders jointly. Using binomial tests, we check whether each subject makes decisions that are in line with either of the patterns. Table 6 presents the resulting proportions of the subjects in the CHN 10× treatment. In this treatment, 58% of subjects can be classified as either mixed risk-averse or mixed risk-loving under the strictest threshold of 1% significance. This goes up to a total of 88% with the 10% significance threshold.[20] Table 6 also lists the shares from the CHN treatment for comparison, in which between 51% and 79% could be classified in this way.

In addition, we again run separate OLS regressions with the number of choices consistent with (i) mixed risk-averse behavior, (ii) mixed risk-loving behavior, and (iii) either type as

[20] Note that when applying the 10% threshold, one subject in CHN 10× is classified as being both mixed risk-averse and mixed risk-loving. In the other two cases, the resulting classifications are mutually exclusive.

TABLE 6
SHARE OF SUBJECTS WHO ARE CLASSIFIED AS MIXED RISK-AVERSE OR MIXED RISK-LOVING

	Mixed Risk-Averse			Mixed Risk-Loving			Mixed Risk-Averse or -Loving+		
Threshold	CHN (%)	CHN 10× (%)	All (%)	CHN (%)	CHN 10× (%)	All (%)	CHN (%)	CHN 10× (%)	All (%)
$p < 0.01$	42	54	45	9	4	8	51	58	53
$p < 0.05$	54	65	57	13	13	13	67	78	70
$p < 0.10$	64	73	66	15	15	15	79	88	81

Notes: Classification based on binomial tests with different significance thresholds: $p < 0.01$, $p < 0.05$, or $p < 0.10$, which represent 27, 26, or 25 consistent choices out of 38 possible choices.
+In both treatments, the share of classified subjects is significantly different from the share that would be expected under random behavior ($p < 0.001$, Fisher's exact tests).

TABLE 7
*N*TH ORDER RISK-LOVING (N-RL) CHOICES ACROSS LOTTERY FORMATS

Order:	1	2	3	4	5	6
H_0:	1.5	3.5	3.5	3.5	3.5	3.5
Compound						
Mean	0.028***	1.266***	2.254***	2.466***	2.806***	3.243
Std. Dev.	(0.165)	(1.404)	(1.810)	(1.708)	(1.526)	(1.268)
Median	0	1	2	2	3	3
Reduced						
Mean			3.099*	3.479	2.722***	3.285
Std. Dev.			(1.790)	(1.872)	(1.730)	(1.342)
Median			3	3	3	3

*$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$, two-sided Wilcoxon signed-rank tests.

dependent variables (see Appendix A.5). When regressing the number of choices consistent with mixed risk-aversion on a treatment dummy and the control variables, we do not observe a significant effect of the treatment dummy of CHN 10× ($\beta = 1.160$, robust SE $= 1.083$, $p = 0.286$, two-sided). Also, when analyzing the number of choices consistent with mixed risk-loving behavior, we do not observe a significant influence of the treatment dummy ($\beta = -0.921$, robust SE $= 0.843$, $p = 0.276$). Also, there appears to be no significant influence on the maximum of either variable ($\beta = 0.592$, robust SE $= 0.909$, $p = 0.516$).

Observation 2: *Between* 58% *and* 88% *of all subjects can be classified as adhering to either mixed risk-averse or mixed risk-loving behavior when the stakes are increased* 10*-fold. After controlling for the subjects' characteristics, we do not find a significant difference in the number of mixed risk-averse or mixed risk-loving choices when the stakes increase.*

4.4. *Higher Order Risk Preferences across Lottery Formats.*

4.4.1. *Aggregate risk preferences.* Again, we consider the number of *n*-RL choices first and, with respect to order 1, we assume that all participants prefer more money to less. In GER, 97% never choose a dominated payoff in order 1. In the Compound & Reduced treatment, this share is also 97% ($p = 1.000$, Fisher's exact test).[21]

Table 7 presents the number of *n*-RL choices in this treatment. Please note that only the lotteries of orders 3–6 were displayed in compound and reduced form. Although the data for

[21] Comparing the number of *n*-RL choices between the compound lotteries of the Compound & Reduced treatment and the GER treatment reveals that participants in Compound & Reduced are significantly more 3-RA than participants in GER ($p = 0.019$, two-sided Mann–Whitney U test). In the remaining orders, we do not observe any significant differences ($p \geq 0.189$).

orders 1 and 2 are based on the choices of all participants, the data for the higher orders 3–6 are based on approximately half the sample (71 participants for order 3, 73 for order 4, 72 for order 5, and 70 for order 6). Each participant made choices for two of the higher orders and in both framings.

As before, there is a tendency of participants to prefer the less risky alternative for orders 3–5. Comparing choice frequencies to the 3.5 n-RL choices that would be expected under random behavior, the behavior for orders 3–5 differs significantly from the benchmark ($p < 0.001$, two-sided one-sample Wilcoxon tests), but the behavior for order 6 does not ($p = 0.163$). With respect to the reduced lotteries, however, the difference from the benchmark is only significant for order 5 ($p < 0.001$). It is weakly significant for order 3 ($p = 0.075$) and insignificant for orders 4 and 6 ($p \geq 0.113$).

Our results suggest that the lottery format influences choices for orders 3 and 4. We run linear panel regressions with individual random effects separately for each order, with the number of n-RL choices as the dependent variable. The regressions include a dummy for choices in reduced lotteries, as well as various controls (see Appendix A.6 for the regression results and Online Appendix O7 for details about our estimation strategy). The regressions indicate 0.845 less 3-RA choices in reduced lotteries ($\beta = 0.845$, robust SE $= 0.223$, $p < 0.001$, two-sided) and 1.014 less 4-RA choices ($\beta = 1.014$, robust SE $= 0.279$, $p < 0.001$).

4.4.2. *Consistency of risk preferences.* Above we reported a robust pattern of mixed risk-averse and mixed risk-loving behavior in three subject pools and under varying stakes based on the use of compound lotteries. However, the predictions we test are independent of the lottery format. They always suggest that second-order risk averters coincide in their choices with second-order risk lovers in the odd orders, whereas they differ in the even orders. Yet the compound format might facilitate viewing a lottery as a combination of "good" and "bad" outcomes.

Although the treatment presented in this subsection possesses a slightly different data structure, we proceed in the same way as before. For the Compound & Reduced treatment, 93% of the subjects are classified as second-order risk-averse ("RA") and 7% as second-order risk-loving ("RL"). Figure 6 displays the average frequency of n-RL choices made by both types across orders 3–6. The pattern of choices is less clear cut than in the previous analyses. With respect to second-order risk averters, we replicate the previous findings using compound lotteries: for orders 3–5, second-order risk averters favor the less risky lotteries, if we compare their choices to the 3.5 benchmark ($p \leq 0.001$, two-sided one-sample Wilcoxon signed-rank tests). In the case of order 6, we do not observe this tendency ($p = 0.136$). When using reduced lotteries, we still find at least a weak tendency of second-order risk averters to favor less risky lotteries for orders 3, 5, and 6 ($p \leq 0.080$), but for order 4 their choices no longer differ significantly from the 3.5 benchmark ($p = 0.738$). Independent of the order, second-order risk lovers do not systematically favor one of the options ($p \geq 0.262$), in case of compound lotteries. In case of reduced lotteries, they do favor more risky lotteries for order 4 ($p = 0.083$), but not for any other order ($p \geq 0.480$). However, there are relatively few second-order risk lovers (only between three and seven for each order) compared to the treatments discussed above. This might be driven by differences in the subject pool composition (cf. Table 2).[22]

This pattern suggests that some individuals exhibit preference reversals. On the individual level, 27% of the subjects make more n-RL choices in the reduced than in the compound lotteries of order 3, whereas 10% make fewer n-RL choices. These percentages are 32% and 15% for order 4, 20% and 19% for order 5, and 18% and 17% for order 6. To shed some light on the drivers of this change in preferences, we also run a logit regression. The dependent variable in this regression is a dummy indicating whether a subject's number of n-RL choices

[22] Although subjects in the compound lotteries of Compound & Reduced and GER do not differ with respect to the number of 2-RA choices (cf. Footnote 21), the share of subjects classified as second-order risk-averse ("RA") is significantly higher in the compound lotteries of Compound & Reduced than in GER ($p = 0.025$, Fisher's exact test).

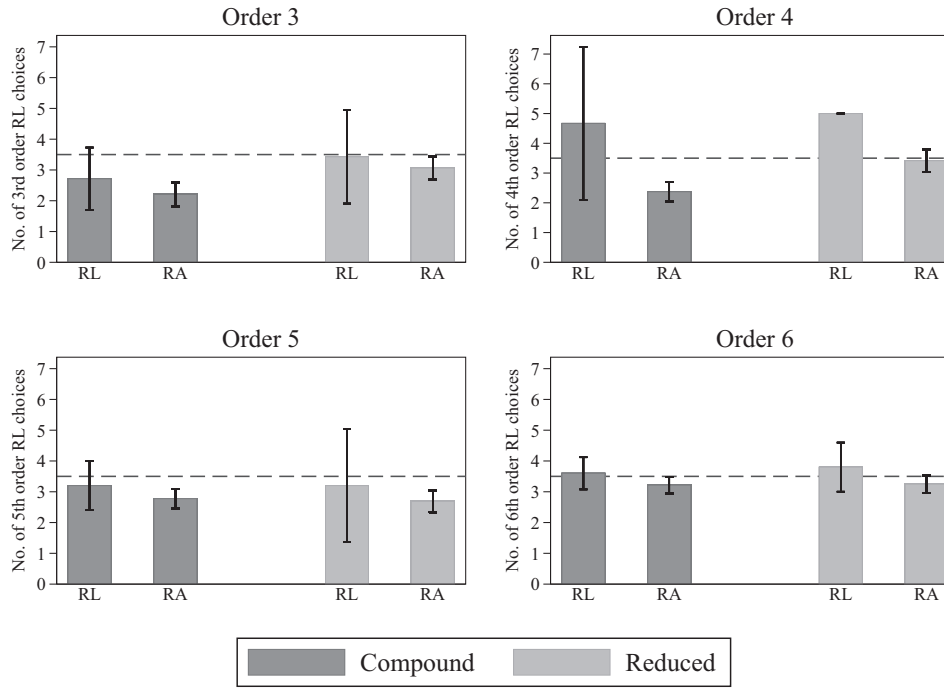


FIGURE 6

AVERAGE NUMBER OF *N*TH ORDER RISK-LOVING (N-RL) CHOICES BY RISK-AVERSE (RA) AND RISK-LOVING (RL) SUBJECTS

TABLE 8

PERCENTAGE OF SUBJECTS WHO ARE CLASSIFIED AS MIXED RISK-AVERSE OR MIXED RISK-LOVING

	Mixed Risk-Averse			Mixed Risk-Loving			Mixed Risk-Averse or Risk-Loving^{+}		
Threshold	Compound (%)	Reduced (%)	Mean (%)	Compound (%)	Reduced (%)	Mean (%)	Compound (%)	Reduced (%)	Mean (%)
$p < 0.01$	41	27	34	1	1	1	42	28	35
$p < 0.05$	50	42	46	1	1	1	51	43	47
$p < 0.10$	57	52	56	2	3	3	59	55	59

Notes: Classification based on binomial tests with different significance thresholds: $p < 0.01$, $p < 0.05$, or $p < 0.10$, which represent 19, 18, or 17 consistent choices out of 24 possible choices.
$^{+}$In both formats, the share of classified subjects is significantly different from the share that would be expected under random behavior ($p < 0.001$, Fisher's exact tests).

differed between the two treatments (see Appendix A.6 for the regression results and the Online Appendix O7 for details on our estimation strategy). As the explanatory variable, we use the characteristics of the subjects displayed in Table 2. This regression indicates that numeracy is somewhat associated with preference reversal: Those with a higher score in the BNT are weakly less likely to switch (average marginal effect = 0.047, standard error = 0.027, and $p = 0.081$). We do not observe any significant influences of the other control variables.

In a second step, we classify subjects as mixed risk-averse and mixed risk-loving. Participants made choices for two of the higher orders in both framings. Although in the previous classification we could use all 38 decisions at once, we now rely on 24 choices to classify each participant: 10 choices from order 1 and 2, and 14 of 28 choices from two of the higher orders (either 14 from the compound or 14 from the reduced framing).

Table 8 presents the percentage of subjects who are classified as either mixed risk-averse or mixed risk-loving based on the binomial tests. With the 1% significance threshold, 42% of

the subjects can be classified as belonging to one of the two types, when using the compound lotteries. When using the reduced lotteries, only 28% of the subjects can be classified in this way. These shares go up to 59% and 55%, respectively, when applying the 10% significance threshold.[23]

In addition, we compare the number of choices that are consistent with the two types between the two formats. We run separate linear random effects panel regressions with the number of choices consistent with (i) mixed risk-averse behavior, (ii) mixed risk-loving behavior, or (iii) either type as dependent variables. The regressions include the usual control variables (see Appendix A.6). Considering the number of mixed risk-averse choices, we find a negative effect of the dummy indicating choices from reduced lotteries ($\beta = -0.916$, robust SE $= 0.235$, $p < 0.001$, two-sided). Conducting the same analysis for the number of mixed risk-loving choices, we do not find a significant influence of the format ($\beta = 0.161$, robust SE $= 0.218$, $p = 0.460$). On aggregate, the regression results also indicate that consistency with either type is smaller in the reduced lotteries ($\beta = -0.811$, robust SE $= 0.227$, $p < 0.001$).

Observation 3: *Between* 28% *and* 55% *of all subjects can be classified as adhering to either mixed risk-averse or mixed risk-loving behavior when the lotteries are displayed in the reduced format. The number of mixed risk-averse choices increases significantly when the lotteries are displayed in the compound format, whereas the number of mixed risk-loving choices does not change.*

4.4.3. *Explaining the framing effect.* In Subsection 2.2.3, we outlined previous results on the differences in choices observed between reduced and compound lotteries. To our knowledge, no previous study offers an explanation for why we observe more 3-RA and more 4-RA in compound than in reduced lotteries. To gather further evidence, we conducted a follow-up study varying the type of lottery choice (3-RA and 4-RA) as well as the framing (reduced and compound) between-subjects (cf. the experimental design described in Subsection 3.2).

First, with respect to choices in the 3-RA lottery, we find an even more pronounced framing effect in this Follow-up Experiment than in the Compound & Reduced treatment. In the compound framing of the Follow-up Experiment, 76% of the 58 participants choose the more 3-RA lottery. Only 45% of the 56 participants do so in the reduced framing ($p = 0.001$, Fisher's exact test). In the respective task of the Compound & Reduced treatment, 70% of subjects choose the more 3-RA lottery in the compound framing, whereas only 49% do so in the reduced framing ($p = 0.006$, McNemar's test). With respect to the 4-RA lottery, we do observe less 4-RA choices in the Follow-up Experiment, and we do not find a significant framing effect. In the Follow-up Experiment, 57% of the 54 participants choose the more 4-RA lottery and 43% of the 56 participants do so in the reduced framing ($p = 0.182$). In the Compound & Reduced treatment, 78% of subjects choose the more 4-RA lottery in the compound framing, whereas 51% do so in the reduced lottery ($p < 0.001$).

Second, to study the reasoning behind participants' choices, they were asked to send one written free-form message, together with their preferred choice, to the other participant in their group.

For three arguments from our classification scheme, the frequency differs significantly between framings in the 3-RA or the 4-RA lottery. These are:

(i) Maximization of the largest potential payoff.
(ii) Maximization of the smallest potential payoff.
(iii) Maximization of the payoff for the most likely outcome.

[23] When the subjects make 24 decisions across orders 1 and 2, as well as two more orders $i, j \in \{3, 4, 5, 6\}$ with $i \neq j$, it depends on the combination of odd and even orders whether the classification of the two types is theoretically mutually exclusive. However, it is only when applying the 10% threshold that one subject in Compound and one in Reduced are classified as being mixed risk-averse *and* mixed risk-loving.

These arguments were reliably identified by two independent coders as indicated by values of Krippendorff's alpha above the commonly used threshold of 0.67 (see Krippendorff, 2004 and Online Appendix O11 for the complete list of arguments).

With respect to the 3-RA lottery, the frequency of all of the three arguments differs between both framings ($p \leq 0.004$, Fisher's exact tests). The first two arguments suggest that one should choose the more 3-RA lottery (and all except one of the participants using these arguments do so). They are, respectively, used by 22% and 40% of the participants in the compound framing and only by 3% and 7% in the reduced framing. The third argument suggests that one should choose the 3-RL lottery (and all except one of the participants do so). It is used by 17% of the participants in the reduced framing and by only 3% in the compound framing. With respect to the 4-RA lottery, only the frequency of the second argument differs between both framings ($p = 0.023$). It suggests that one should choose the more 4-RA lottery (and all except one of the participants do so). It is used by 26% of the participants in the compound framing and only by 6% in the reduced framing.

Overall, it appears that the compound display of lotteries leads subjects to focus more on the smallest potential payoff in the 3-RA, as well as in the 4-RA lottery. This could drive the differences in choices we observe between compound and reduced lotteries for 3-RA and 4-RA.

Observation 4: *The most commonly used argument to justify* 3-*RA (prudent) and* 4-*RA (temperate) choices is the maximization of the smallest potential payoff. It is used significantly more often in compound than in reduced lotteries.*

5. CONCLUSION

In this study, we analyze the consistency of higher order risk preferences. We contribute to this topic by exploring the role of country differences, the variation of stakes, and the framing of lotteries. In our American subject pool, we replicate the findings of Deck and Schlesinger (2014) and we identify a similar pattern in subject pools in Germany and in China. Across all three countries, a majority of participants can be classified as mixed risk averters or as mixed risk lovers (between 51% and 76% of all subjects depending on the significance level).

Existing evidence from nonincentivized and incentivized studies suggests that Chinese are more second-order risk-averse than Americans and Germans. We can only confirm this finding with respect to Chinese and Germans. We do not observe a significant difference in second-order risk aversion between Chinese and Americans. We have formulated our first hypothesis based on the assumption that differences in second-order risk aversion indicate differences in the underlying distribution of mixed risk averters and mixed risk lovers. In line with our first hypothesis, mixed risk averters are somewhat more common in Germany than in China, whereas mixed risk lovers are less common. Contrary to our first hypothesis, we do not observe differences in the prevalence of both types between our Chinese and American samples. However, it is important to note that these findings reflect the differences in second-order risk aversion in our sample. In fact, the evidence from incentivized studies on differences between Chinese and Americans is less clear cut than the evidence from nonincentivized studies (cf. Haering and Heinrich 2017). For example, in the first experimental comparison, Kachelmeier and Shehata (1992a) also do not find significant differences between Chinese and Americans.

Moreover, we provide the first analysis of higher order risk preferences with large monetary payoffs. We also observe a majority of choices to be in line with mixed risk-averse and mixed risk-loving behavior under high stakes. In line with prior evidence, we observe an increase in second-order risk aversion when the stakes are increased 10-fold. However, contrary to our second hypothesis, we find no significant change in the number of mixed risk-averse and mixed risk-loving choices when the stakes increase.

A dichotomous population with respect to higher order risk preferences may have important real-world implications. Although mixed risk averters and mixed risk lovers coincide in their choices in the odd orders, they differ in the even orders. This means that a measurement of second-order risk aversion is not sufficient for estimating the prevalence of higher order risk preferences per se. A measurement of second-order risk aversion will be indicative for temperance (4-RA) and sixth-order risk aversion (6-RA) but not for prudence (3-RA) or fifth-order risk aversion (5-RA). This is relevant for prevention decisions, for instance. From an economic perspective, prevention can be classified as self-protection or self-insurance. Self-protection lowers the *probability* of the occurrence of a loss, while the *size* of the loss is exogenous. In contrast, self-insurance aims at reducing the *size* of a loss while the *probability* of occurrence is exogenous (see Ehrlich and Becker, 1972). In medicine, self-protection is known as primary prevention and self-insurance as secondary prevention.

Although higher second-order risk aversion unambiguously leads individuals to choose higher levels of self-insurance, risk aversion is not sufficient to determine an individual's level of self-protection (see Dionne and Eeckhoudt, 1985; Briys and Schlesinger, 1990). Eeckhoudt and Gollier (2005) and Courbage and Rey (2006) show that more prudent individuals will expand less effort in self-protection.

An example for self-protection is the influenza vaccination. In general, the flu shot decreases the probability of getting the flu, but the harm of the flu itself is not affected. Prudent individuals try to avoid the worst outcome, which is facing the disutility that comes with the flu shot and still getting the flu. Therefore, more prudent individuals should be less likely to undergo an influenza vaccination. Indeed, Mayrhofer and Schmitz (2019) find that for high-risk individuals, such as individuals over 60 years of age, prudence has a significant negative impact on the likelihood of undergoing influenza vaccination. Because both mixed risk averters and mixed risk lovers are prudent, they will expend the same effort for self-protection measures like flu shots. However, this is different with regard to self-insurance. For example, cancer screenings do not decrease the likelihood of getting cancer, but an early detection can lead to early treatment and thus less harm. In this case, mixed risk-averse individuals opt for screening more often or earlier than mixed risk lovers (see Felder and Mayrhofer, 2014).

We also observe that subjects choose the prudent and temperate options less often, when the options are displayed in a reduced instead of a compound form. In the reduced lotteries, there is weak evidence that subjects generally behave prudently, and no evidence that they are generally temperate. In other words, the proportion of subjects who can be classified as mixed risk averters or mixed risk lovers decreases considerably, when reduced lotteries are used. This is in line with our third hypothesis and the conjecture by Deck and Schlesinger (2014) who point out that compound lotteries may facilitate the interpretation of lotteries as combinations of "good" with "bad" or "good" outcomes.

To our knowledge, no previous study offers an empirical explanation for why we observe more prudent and more temperate behavior in compound lotteries than we do in reduced lotteries. To shed light on our findings, we conducted a follow-up study that was aimed at revealing the reasoning behind the subjects' choices. Overall, it appears that the compound display of lotteries leads subjects to focus more on the smallest potential payoff in prudence as well as in temperance lotteries. This could drive the differences in choices we observe between compound and reduced lotteries for prudence and temperance.

As Abdellaoui et al. (2015) point out, different attitudes toward compound versus reduced risks might have big implications for marketing, policy, and economics. For example, if people are less temperate with respect to reduced risks, they would invest more in risky assets if the associated risks are presented in reduced instead of compound form.

APPENDIX

A.1. *Comparison of related papers.*

TABLE A.1

COMPARISON OF RELATED PAPERS

Study	Location(s)	Average Payoff	Payment	Elicitation Method	Lottery Type	Share of Risk-Averse/Prudent/Temperate Choices
Deck and Schlesinger (2010)	USA	$25.56	1 of 10 choices	Binary choice	Compound	–/61%/38%
Ebert and Wiesen (2011)	Germany	€18.50	1 of 34 choices	Binary choice	Compound	–/65%/–[1]
Maier and Rüger (2012)	Germany	–	1 of 84 choices	Binary choice	Reduced	56%/56%/56%[2]
Deck and Schlesinger (2014)	USA	$20.92	1 of 38 choices	Binary choice	Compound	74%/77%/58%
Ebert and Wiesen (2014)	Germany	€17.50[3]	1 of 120 choices	Risk premia	Compound	66%/88%/75%[1]
Heinrich and Mayrhofer (2018)	Germany	€18.09	1 of 240 choices	Risk premia	Compound	70%/90%/76%[1]
Noussair et al. (2014)	Netherlands	Real: 1/10 chance of €70.00[3] Hypothetical: €10,500.00[3]	1 of 17 choices	Binary choice	Compound	72%/89%/62%
Deck and Schlesinger (2017)	USA	$16.66	1 of 52 choices	Binary choice	Compound Reduced	–/73% /64%[1] –/77%/47%[1]
Baillon et al. (2018)	Netherlands	€18.50[3]	1 of 30 choices	Binary choice	Reduced	84%/71%/43%
This study (2018)	China, Germany, and the USA	¥20.06/¥544.92[4] €12.50/€11.80[4] $19.64	1 of 38 choices	Binary choice	Compound Reduced	75%/76%/64%[5] –/56%/50%

Notes: Table following Noussair et al. (2014). The dash (–) indicates that the values are not reported. Average payoff does not include show-up fee.

[1]The amounts represent the share of subjects.

[2]The shares represent the choices across domains. In case of gains, these values are 55%/60%/58% and in case of losses 57%/55%/54%.

[3]This value represents the expected payoff.

[4]The average payoffs in China represent the CHN/CHN 10× values and in case of Germany the GER/Compound & Reduced values.

[5]The shares represent the pooled choices for CHN, USA, and GER treatments. In case of CHN 10×, these values are 80%/76%/71% and in case of Compound 82%/68%/65%.

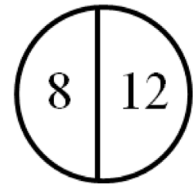
A.2. Instructions (English version). You are participating in a research study on decision making under uncertainty. At the end of the study, you will be paid your earnings in cash and it is important that you understand how your decisions affect your payoff. If you have questions at any point during the study, please raise your hand and someone will assist you. Otherwise, please do not talk during this study and turn off your cell phone.

[**CHN, USA, GER, CHN 10× and Compound & Reduced**: In this study, there is a series of 38 tasks. Each of these tasks involves choosing between Option A and Option B. Once you have completed these tasks, one of the 38 will be randomly selected to determine your payoff. All values are given in experimental currency unit (ECU).]

[**Follow-up Experiment**: There is one task. This task involves choosing between Option A and Option B. This decision can influence your payoff. All values are given in experimental currency unit (ECU).]

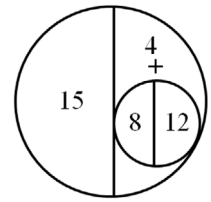
For ECU 1 you will receive $ 0.93.

Each option will involve amounts of money and possibly one or more 50–50 lotteries represented as a circle with a line through the middle. A 50–50 lottery means there is a 50% chance of receiving the item to the left of the line and a 50% chance of receiving the item to the right of the line. For example,

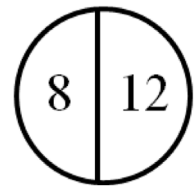


is a 50–50 lottery in which you would receive either ECU 8 or ECU 12, each with an equal chance. To determine the outcome of any 50–50 lottery, we will use a computerized random-number generator.

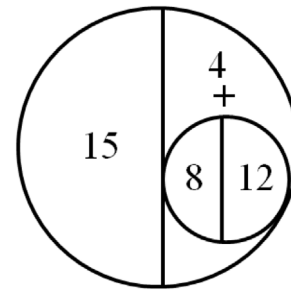
In some cases, one of the lottery outcomes in a 50–50 lottery may contain another lottery. For example,



is a 50–50 lottery where you receive either ECU 15 or you receive ECU 4 plus the 50–50 lottery

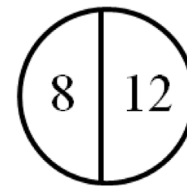


Continuing with the example,



,

there is a 50% chance that you would receive ECU 15 in the first 50–50 lottery and that would be it. There is also a 50% chance that you would receive ECU 4 +



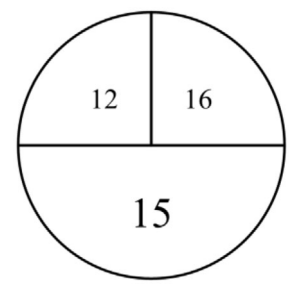
in the first 50–50 lottery.

Conditional on this outcome for the first 50–50 lottery, you would then have a 50% chance of receiving an extra ECU 8 and a 50% chance of receiving an extra ECU 12 in addition to the ECU 4. Therefore, the chance that you would end up with $4 + 8 =$ ECU 12 is $0.5 \times 0.5 = 0.25 = 25\%$. The chance that you would end up with $4 + 12 =$ ECU 16 is $0.5 \times 0.5 = 0.25 = 25\%$.

[Compound & Reduced and Follow-up Experiment:

The illustration of this option can also take place with the aid of a circle with different probabilities of the lottery results.

Like in the sample above

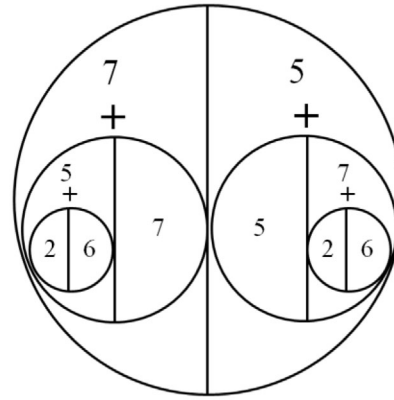


is a lottery in which you can either receive 12 ECU, 15 ECU, or 16 ECU.

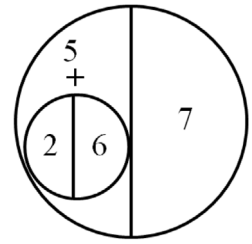
Again, there is a 50% probability that you will receive 15 ECU. In addition, the probability that you get 12 ECU or 16 ECU is 25% each.**]**

[CHN, USA, GER, CHN 10× and Compound & Reduced:

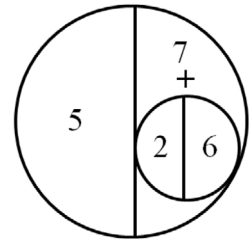
Let us look at a more complicated example.



is a 50–50 lottery where you receive either ECU 7 plus the 50–50 lottery



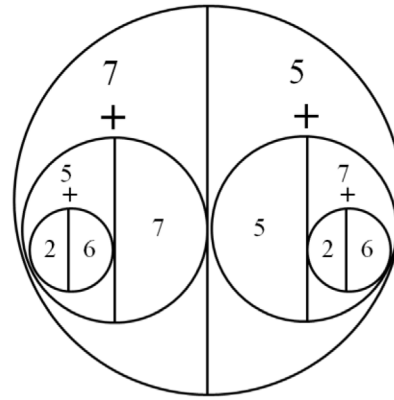
or you receive ECU 5 plus the 50–50 lottery



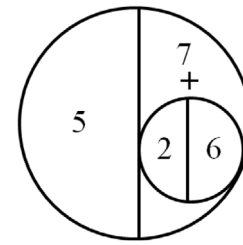
,

both of which include an additional 50–50 lottery.

In



you could earn ECU 10 if you get ECU 5 +



in the first lottery and then earn ECU 5 in the second lottery. This occurs with a $0.5 \times 0.5 = 0.25 = 25\%$ chance. Alternatively, you could earn ECU 14 with a 50% chance. Note that you could earn ECU 14 in three ways:

by (i) earning ECU 7 (in the first lottery) + ECU 5 (in the second lottery) + ECU 2 (third lottery) that happens with a $0.5 \times 0.5 \times 0.5 = 0.125 = 12.5\%$ chance,

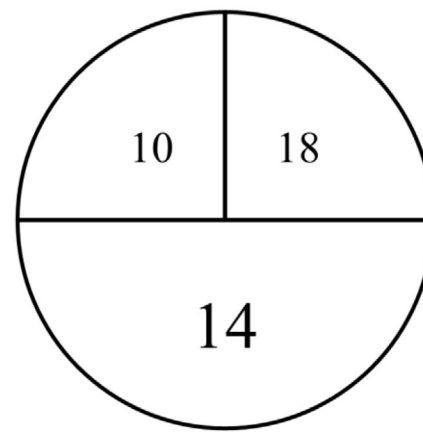
or (ii) earning ECU 7 (in the first lottery) + ECU 7 (in the second lottery) that happens with a $0.5 \times 0.5 = 0.25 = 25\%$ chance,

or (iii) earning ECU 5 (in the first lottery) + ECU 7 (in the second lottery) + ECU 2 (third lottery) that happens with a $0.5 \times 0.5 \times 0.5 = 0.125 = 12.5\%$ chance.

Finally, there are two ways that you could earn ECU 18 that occurs with a $0.5 \times 0.5 \times 0.5 + 0.5 \times 0.5 \times 0.5 = 0.25 = 25\%$ chance.**]**

[Compound & Reduced:

This option can also be illustrated with the aid of a circle with different probabilities of the lottery results (see next page).



Just like in the example on the previous two pages, you can either receive 10 ECU, 14 ECU, or 18 ECU. Again, there is a 50% chance that you will earn 14 ECU. In addition, the probability of receiving 10 ECU or 18 ECU is 25%.]

[Follow-up Experiment:

Choice of an option

You will be randomly assigned to another participant in the experiment as a partner with whom you will form a team. Your payoffs will be determined by the decisions of your team.

How does your team decision come about? Both team members will enter a final decision regarding the choice of Option A or Option B. However, only *one* of the decisions is chosen randomly and with equal probability as the team decision (with the help of a computerized random-number generator).

The chosen final decision counts for *both* team members. (Note that both team members receive the respective payoff of the task. If the payout of a lottery is 15 ECU, for example, both team members receive 15 ECU *each*).

Before you enter your final decision, you have the opportunity to influence the final choice of your partner: Before the decisions are entered, you will send a preferred choice together with a text message to your partner. Likewise, your partner sends a text message with his proposed option to you.

Only after both team members have received the other's text messages, they are allowed to enter their final decision. After both made their decision, one of the two decisions will be randomly chosen for the team. A questionnaire follows and finally your payoff is determined based on the chosen option.

Note: All participants of the experiment receive the same instructions.

Note for the text messages

The content of the text message is up to you. But be aware, that your text message is the only chance to persuade your partner of your proposed option. Thus, use the text message to explain your proposal.

It is forbidden, to provide any personal details such as name, age, address, and field of study! If you violate the rules of communication, you can be excluded from the experiment without receiving your payoff. Every text message may include 420 characters at most (approximately 3 lines). Note: To send a typed text message you have to click "send."

]

A.3. Summary of variables.

TABLE A.2
SUMMARY OF VARIABLES

Variable	Description
Dependent variables:	
Order n	Subject's number of *n*-RL choices in order *n*
No. of MRA/MRL choices	Subject's number of mixed risk-averse/risk-loving choices in all orders
MRA or MRL	The maximum of *No. of MRA choices* and *No. of MRL choices*
Comp >/</ = Redu	Dummy variable indicating that a subject's risk-loving choices in orders 3–6 are greater/smaller/equal in Compound compared to Reduced
Independent variables:	
Exp.USA	Dummy variable indicating experimenter from the United States of America
Exp.CHN	Dummy variable indicating experimenter from China
IRB	Dummy variable indicating the use of an IRB form
Female	Dummy variable indicating female subjects
Age 18–20	Dummy variable indicating subjects age 18–20
Age > 23	Dummy variable indicating subjects age 24 and above
CRT	Number of correct answers CRT (0–3)
BNT	Number of correct answers BNT (0–4)

A.4. Regression results on higher order risk preferences across countries.

TABLE A.3
OLS REGRESSION

	Order 1		Order 2		Order 3		Order 4		Order 5		Order 6	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
CHN	0.073	0.078	0.981	1.100	0.136	0.244	0.615	0.719	−0.542	−0.430	−0.072	0.115
	(0.085)	(0.084)	(0.672)	(0.675)	(0.646)	(0.644)	(0.647)	(0.647)	(0.596)	(0.582)	(0.604)	(0.595)
GER	0.049	0.051	0.066	−0.016	−0.180	−0.188	0.102	0.098	−0.407	−0.425	−0.319	−0.344
	(0.058)	(0.059)	(0.590)	(0.595)	(0.584)	(0.573)	(0.572)	(0.563)	(0.491)	(0.489)	(0.515)	(0.506)
Exp.USA	0.029	0.028	−0.590	−0.561	−0.112	−0.107	−0.384	−0.385	−0.398	−0.393	−0.244	−0.235
	(0.048)	(0.048)	(0.460)	(0.463)	(0.481)	(0.484)	(0.427)	(0.424)	(0.339)	(0.339)	(0.420)	(0.409)
Exp.CHN	0.074	0.072	−0.892**	−0.899**	−0.458	−0.478	−0.662**	−0.686**	−0.420	−0.436	−0.568*	−0.603*
	(0.057)	(0.058)	(0.377)	(0.371)	(0.337)	(0.336)	(0.330)	(0.337)	(0.350)	(0.351)	(0.324)	(0.327)
IRB	0.029	0.028	0.448	0.433	−0.258	−0.267	0.480	0.468	−0.031	−0.043	0.064	0.045
	(0.036)	(0.036)	(0.405)	(0.416)	(0.329)	(0.327)	(0.385)	(0.383)	(0.381)	(0.380)	(0.323)	(0.324)
Female	−0.007	−0.007	−0.204	−0.304	0.593***	0.622***	0.086	0.090	0.318*	0.297*	−0.019	−0.001
	(0.021)	(0.021)	(0.208)	(0.203)	(0.185)	(0.177)	(0.188)	(0.184)	(0.173)	(0.167)	(0.176)	(0.172)
Age 18–20	0.008	0.002	−0.283	−0.250	−0.163	−0.122	−0.219	−0.256	0.086	0.063	−0.416*	−0.401*
	(0.022)	(0.022)	(0.240)	(0.233)	(0.228)	(0.224)	(0.231)	(0.223)	(0.215)	(0.211)	(0.219)	(0.216)
Age > 23	0.069**	0.069**	0.082	0.056	0.213	0.240	−0.215	−0.212	0.215	0.196	−0.201	−0.179
	(0.029)	(0.030)	(0.229)	(0.229)	(0.220)	(0.219)	(0.207)	(0.206)	(0.182)	(0.181)	(0.194)	(0.193)
CRT	−0.014*		0.086		−0.106		−0.124		−0.056		−0.164*	
	(0.008)		(0.098)		(0.097)		(0.091)		(0.085)		(0.088)	
BNT	0.002		0.104		0.073		0.064		0.055		0.128*	
	(0.007)		(0.083)		(0.083)		(0.079)		(0.076)		(0.074)	
p-value CHN = GER	0.684	0.656	0.009	0.000	0.328	0.151	0.126	0.054	0.692	0.987	0.445	0.142
N	406	414	406	414	406	414	406	414	406	414	406	414
AIC	−114.688	−106.587	1696.322	1724.476	1652.363	1680.044	1640.922	1666.755	1552.701	1585.141	1580.230	1612.268
BIC	−70.619	−70.355	1740.391	1760.709	1696.433	1716.276	1684.992	1702.988	1596.770	1621.373	1624.300	1648.501

Notes: Constant not reported and robust standard errors in parentheses.
$^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

TABLE A.4
OLS REGRESSION ALL ORDERS

	No. of MRA choices		No. of MRL choices		MRA or MRL	
	(1)	(2)	(1)	(2)	(1)	(2)
CHN	−1.191	−1.826	1.858	2.041	0.492	0.053
	(2.041)	(2.028)	(1.630)	(1.647)	(1.802)	(1.785)
GER	0.689	0.823	0.388	0.300	1.180	1.107
	(1.716)	(1.712)	(1.473)	(1.440)	(1.513)	(1.482)
Exp.USA	1.697	1.653	−0.738	−0.709	1.882	1.904*
	(1.366)	(1.369)	(1.042)	(1.035)	(1.169)	(1.150)
Exp.CHN	2.927**	3.031***	−1.318	−1.345	2.017*	2.106*
	(1.156)	(1.165)	(0.900)	(0.891)	(1.049)	(1.087)
IRB	−0.731	−0.663	1.252	1.228	0.014	0.044
	(1.099)	(1.099)	(1.056)	(1.058)	(1.013)	(1.017)
Female	−0.768	−0.698	−1.042*	−1.127**	−1.485***	−1.632***
	(0.566)	(0.565)	(0.550)	(0.517)	(0.496)	(0.487)
Age 18–20	0.986	0.965	−0.849	−0.850	0.706	0.628
	(0.704)	(0.694)	(0.655)	(0.627)	(0.624)	(0.620)
Age > 23	−0.162	−0.170	−0.831	−0.841	−0.564	−0.648
	(0.637)	(0.638)	(0.559)	(0.555)	(0.554)	(0.550)
CRT	0.378		−0.026		0.585**	
	(0.286)		(0.241)		(0.257)	
BNT	−0.426*		0.166		−0.180	
	(0.244)		(0.215)		(0.215)	
p-value CHN = GER	0.099	0.015	0.094	0.033	0.500	0.290
N	406	414	406	414	406	414
AIC	2545.650	2595.508	2462.897	2501.033	2428.318	2477.862
BIC	2589.720	2631.741	2506.967	2537.266	2472.388	2514.095

Notes: Constant not reported and robust standard errors in parentheses.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

A.5. Regression results on higher order risk preferences across stakes.

TABLE A.5
OLS REGRESSION

	Order 1		Order 2		Order 3		Order 4		Order 5		Order 6	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
CHN10×	−0.024	−0.035	−0.616*	−0.604*	0.011	−0.073	−0.314	−0.380	−0.106	−0.137	−0.111	−0.197
	(0.054)	(0.051)	(0.319)	(0.313)	(0.352)	(0.328)	(0.307)	(0.304)	(0.272)	(0.260)	(0.351)	(0.345)
Female	−0.055	−0.051	−0.236	−0.347	0.292	0.355	0.261	0.213	0.508*	0.450*	0.475	0.412
	(0.047)	(0.044)	(0.300)	(0.284)	(0.289)	(0.280)	(0.279)	(0.272)	(0.266)	(0.256)	(0.293)	(0.284)
Age 18–20	−0.026	−0.045	−0.370	−0.240	−0.075	−0.081	−0.256	−0.366	−0.051	−0.068	−0.604*	−0.583*
	(0.041)	(0.037)	(0.330)	(0.315)	(0.368)	(0.338)	(0.363)	(0.339)	(0.330)	(0.315)	(0.337)	(0.330)
Age > 23	0.129**	0.130*	0.120	0.242	0.188	0.222	−0.774**	−0.673**	0.086	−0.010	−1.023***	−0.887**
	(0.064)	(0.066)	(0.380)	(0.368)	(0.336)	(0.330)	(0.307)	(0.303)	(0.282)	(0.273)	(0.360)	(0.355)
CRT	−0.025		−0.018		−0.315*		−0.114		0.006		−0.248	
	(0.022)		(0.159)		(0.188)		(0.155)		(0.152)		(0.161)	
BNT	−0.010		0.055		0.131		−0.029		−0.007		0.160	
	(0.017)		(0.114)		(0.108)		(0.101)		(0.114)		(0.111)	
N	177	188	177	188	177	188	177	188	177	188	177	188
AIC	70.421	71.174	740.298	783.117	724.952	765.952	726.697	767.163	685.291	726.357	732.864	782.145
BIC	92.654	87.356	762.531	799.299	747.185	782.134	748.930	783.345	707.524	742.540	755.097	798.327

Notes: Constant not reported and robust standard errors in parentheses.
$^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

TABLE A.6
OLS REGRESSION ALL ORDERS

	No. of MRA choices		No. of MRL choices		MRA or MRL	
	(1)	(2)	(1)	(2)	(1)	(2)
CHN10×	1.160	1.427	−0.921	−0.936	0.592	1.114
	(1.083)	(1.045)	(0.843)	(0.860)	(0.909)	(0.869)
Female	−1.245	−1.032	−0.245	−0.477	−1.205	−1.278*
	(0.894)	(0.895)	(0.783)	(0.759)	(0.768)	(0.765)
Age 18–20	1.382	1.382	−1.078	−0.995	0.455	0.509
	(1.100)	(1.070)	(0.956)	(0.899)	(0.951)	(0.923)
Age > 23	1.274	0.975	−2.080**	−1.661*	−0.253	−0.316
	(1.074)	(1.052)	(0.891)	(0.877)	(0.916)	(0.885)
CRT	0.715		−0.046		0.606	
	(0.528)		(0.416)		(0.481)	
BNT	−0.301		0.072		−0.013	
	(0.369)		(0.266)		(0.307)	
N	177	188	177	188	177	188
AIC	1134.686	1204.294	1088.558	1155.179	1076.959	1140.411
BIC	1156.919	1220.476	1110.791	1171.362	1099.192	1156.593

Notes: Constant not reported and robust standard errors in parentheses.
$^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

A.6. Regression results on higher order risk preferences across lottery formats.

TABLE A.7
RANDOM-EFFECTS GLS

	Order 3	Order 4	Order 5	Order 6
Reduced	0.845***	1.014***	−0.083	0.043
	(0.223)	(0.279)	(0.231)	(0.204)
Female	0.111	0.166	−0.442	−0.245
	(0.451)	(0.333)	(0.353)	(0.275)
Age 18–20	−0.682	−0.575	−1.108***	−0.450
	(0.563)	(0.397)	(0.376)	(0.377)
Age > 23	−0.560	−0.590*	−0.544	−0.146
	(0.406)	(0.329)	(0.362)	(0.248)
CRT	−0.160	−0.439**	−0.111	−0.067
	(0.197)	(0.199)	(0.130)	(0.127)
BNT	−0.027	−0.157	−0.103	−0.044
	(0.183)	(0.155)	(0.175)	(0.129)
N	142	146	144	140
N in group	71	73	72	70
χ^2	18.890	34.462	13.285	5.520

Notes: Constant not reported and robust standard errors in parentheses.
$^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

TABLE A.8
RANDOM-EFFECTS GLS ALL ORDERS

	No. of MRA choices	No. of MRL choices	MRA or MRL
Reduced	−0.916***	0.161	−0.811***
	(0.235)	(0.218)	(0.227)
Female	0.645	−0.260	0.337
	(0.525)	(0.416)	(0.444)
Age 18–20	1.622**	0.504	1.384**
	(0.712)	(0.564)	(0.602)
Age > 23	0.915*	0.233	0.894*
	(0.546)	(0.432)	(0.461)
CRT	0.247	−0.087	0.291
	(0.270)	(0.214)	(0.228)
BNT	0.017	0.216	0.062
	(0.241)	(0.191)	(0.204)
N	286	286	286
N in group	143	143	143
χ^2	22.806	3.198	21.639

Notes: Constant not reported and robust standard errors in parentheses.
*$p < 0.10$, **$p < 0.05$, and***$p < 0.01$.

TABLE A.9
LOGIT REGRESSION

	Comp > Redu	*Comp < Redu*	*Comp = Redu*
Female	−0.051	0.067	−0.013
	(0.081)	(0.089)	(0.062)
Age 18–20	−0.095	0.007	0.089
	(0.112)	(0.122)	(0.086)
Age > 23	−0.066	0.001	0.072
	(0.083)	(0.093)	(0.072)
CRT	−0.035	−0.004	0.038
	(0.042)	(0.046)	(0.032)
BNT	−0.015	−0.040	0.047*
	(0.038)	(0.041)	(0.027)
N	143	143	143
AIC	180.918	206.374	125.872
BIC	198.695	224.151	143.649

Notes: Calculation of marginal effects: Delta method, constant not reported, and robust standard errors in parentheses.
*$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure O2.1: Test of understanding
Figure O2.2: Lottery choice
Table O3.1: Summary of variables
Table O4.1: OLS regression
Table O4.2: OLS regression all orders
Table O6.1: OLS regression
Table O6.2: OLS regression all orders
Table O8.1: Random-effects GLS
Table O8.2: Random-effects GLS

Table O8.3: Logit regression
Table O10.1: Correlation of individual nth order risk-loving (n-RL) choices between orders 2 and 6 across countries
Table O10.2: Correlation of individual nth order risk-loving (n-RL) choices between orders 2 and 6 for CHN 10×
Table O10.3: Correlation of individual nth order risk-loving (n-RL) choices between orders 2 and 6 across lottery formats
Table O10: Message classification 3-RA
Table O11: Message classification 4-RA

REFERENCES

ABDELLAOUI, M., P. KLIBANOFF, AND L. PLACIDO, "Experiments on Compound Risk in Relation to Simple Risk and to Ambiguity," *Management Science* 61(6) (2015), 1306–22.

ARROW, K., *Aspects of the Theory of Risk-bearing*. (Helsinki: Yrjö Jahnsson Foundation, 1965).

AZRIELI, Y, C. P. CHAMBERS, AND P. J. HEALY, "Incentives in Experiments: A Theoretical Analysis," *Journal of Political Economy* 126(4) (2018), 1472–1503.

BAILLON, A., "Prudence with Respect to Ambiguity," *Economic Journal* 127(604) (2017), 1731–55.

BAILLON, A., H. SCHLESINGER, AND G. VAN DE KUILEN, "Measuring Higher Order Ambiguity Preferences," *Experimental Economics* 21 (2017), 233–56.

BALTUSSEN, G., G. T. POST, M. J. VAN DEN ASSEM, AND P. P. WAKKER, "Random Incentive Systems in a Dynamic Choice Experiment," *Experimental Economics* 15(3) (2012), 418–43.

BAR-HILLEL, M., "On the Subjective Probability of Compound Events," *Organizational Behavior and Human Performance* 9(3) (1973), 396–406.

BEATTIE, J., AND G. LOOMES, "The Impact of Incentives upon Risky Choice Experiments," *Journal of Risk and Uncertainty* 14(2) (1997), 155–68.

BINSWANGER, H. P., "Attitudes Toward Risk: Experimental Measurement in Rural India," *American Journal of Agricultural Economics* 62(3) (1980), 395–407.

———, "Attitudes Toward Risk: Theoretical Implications of an Experiment in Rural India," *Economic Journal* 91 (1981), 867–90.

BOHNET, I., F. GREIG, B. HERRMANN, AND R. ZECKHAUSER, "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States," *American Economic Review* 98(1) (2008), 294–310.

BOSTIAN, A. J. AND C. HEINZEL, "Prudential Saving: Evidence from a Laboratory Experiment," Working Paper, University of Virginia, Charlottesville, VA, 2012.

BRAÑAS-GARZA, P., P. KUJAL, AND B. LENKEI, "Cognitive Reflection Test: Whom, How, When," MPRA Working Paper No. 68049, 2015.

BRISLIN, R. W., "Back-Translation for Cross-cultural Research," *Journal of Cross-Cultural Psychology* 1(3) (1970), 185–216.

BRIYS, E., AND H. SCHLESINGER, "Risk Aversion and the Propensities for Self-insurance and Self-protection," *Southern Economic Journal* 57(2) (1990), 458–67.

BROCKETT, P. L., AND L. L. GOLDEN, "A Class of Utility Functions containing all the Common Utility Functions," *Management Science* 33(8) (1987), 955–64.

BUCHAN, N., AND R. CROSON, "The Boundaries of Trust: Own and Others' Actions in the US and China," *Journal of Economic Behavior & Organization* 55(4) (2004), 485–504.

BUDESCU, D. V., AND I. FISCHER, "The Same but Different: An Empirical Investigation of the Reducibility Principle," *Journal of Behavioral Decision Making* 14(3) (2001), 187–206.

BURCHARDI, K. B., AND S. P. PENCZYNSKI, "Out of Your Mind: Eliciting Individual Reasoning in One Shot Games," *Games and Economic Behavior* 84 (2014), 39–57.

BURKS, S. V., J. P. CARPENTER, L. GOETTE, AND A. RUSTICHINI, "Cognitive Skills Affect Economic Preferences, Strategic Behavior, and Job Attachment," *Proceedings of the National Academy of Sciences* 106(19) (2009), 7745–50.

CABALLÉ, J., AND A. POMANSKY, "Mixed Risk Aversion," *Journal of Economic Theory* 71 (1996), 485–513.

COKELY, E. T., M. GALESIC, E. SCHULZ, S. GHAZAL, AND R. GARCIA-RETAMERO,"Measuring Risk Literacy: The Berlin Numeracy Test," *Judgement and Decision Making* 7(1) (2012), 25–47.

COURBAGE, C., AND B. REY, "Prudence and Optimal Prevention for Health Risks," *Health Economics* 15 (2006), 1323–27.

———, AND ———, "Decision Thresholds and Changes in Risk for Preventive Treatment," *Health Economics* 25 (2016), 111–24.

COX, J. C., V. SADIRAJ, AND U. SCHMIDT, "Paradoxes and Mechanisms for Choice under Risk," *Experimental Economics* 18(2) (2015), 215–50.

CRAINICH, D., L. EECKHOUDT, AND A. TRANNOY, "Even (Mixed) Risk Seekers are Prudent," *American Economic Review* 103(4) (2013), 1529–35.

CUBITT, R., R. STARMER, AND R. SUGDEN, "On the Validity of the Random Lottery Incentive System," *Experimental Economics* 1(2) (1998), 115–31.

DECK, C., AND H. SCHLESINGER, "Exploring Higher Order Risk Effects," *Review of Economic Studies* 77(4) (2010), 1403–20.

———, AND ———, "Consistency of Higher Order Risk Preferences," *Econometrica* 82(5) (2014), 1913–43.

———, AND ———, "On the Robustness of Higher Order Risk Preferences," *Journal of Risk and Insurance* 85(2) (2017), 313–33.

DIONNE, G., AND L. EECKHOUDT, "Self-insurance, Self-protection and Increased Risk Aversion," *Economics Letters* 17 (1985), 39–42.

DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE, "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100(3) (2010), 1238–60.

EBERT, S., "Even (Mixed) Risk-Seekers are Prudent: Comment," *American Economic Review* 103(4) (2013), 1536–37.

EBERT, S., D. NOCETTI, AND H. SCHLESINGER, "Greater Mutual Aggravation," *Management Science* 64(6) (2017), 2809–11.

EBERT, S., AND G. VAN DE KUILEN, "Measuring Multivariate Risk Preferences," Working Paper, 2015.

EBERT, S., AND D. WIESEN, "Testing for Prudence and Skewness Seeking," *Management Science* 57(7) (2011), 1334–49.

———, AND ———, "Joint Measurement of Risk Aversion, Prudence, and Temperance." *Journal of Risk and Uncertainty* 48(3) (2014), 231–52.

EECKHOUDT, L., *Risk and Medical Decision Making* (Boston: Kluwer, 2002).

EECKHOUDT, L., AND C. GOLLIER, "The Impact of Prudence on Optimal Prevention," *Economic Theory* 26(4) (2005), 989–94.

EECKHOUDT, L., C. GOLLIER, AND H. SCHLESINGER, "Changes in Background Risk and Risk Taking Behavior," *Econometrica* 64(3) (1996), 683–89.

EECKHOUDT, L., AND M. KIMBALL, "Background Risk, Prudence, and the Demand for Insurance." in G. Dionne, ed., *Contributions to Insurance Economics*, (Netherlands: Springer, 1992), 239–54.

EECKHOUDT, L., AND H. SCHLESINGER, "Putting Risk in Its Proper Place," *American Economic Review* 96 (2006), 280–89.

———, AND ———, "Changes in Risk and the Demand for Saving," *Journal of Monetary Economics* 55(7) (2008), 1329–36.

EECKHOUDT, L., H. SCHLESINGER, AND I. TSETLIN, "Apportioning of Risks via Stochastic Dominance," *Journal of Economic Theory* 144 (2009), 994–1003.

EHMKE, M., J. LUSK, AND W. TYNER, "Multidimensional Tests for Economic Behavior Differences across Cultures," *Journal of Socio-Economics* 39(1) (2010), 37–45.

EHRLICH, I., AND G. BECKER, "Market Insurance, Self-insurance, and Self-protection," *Journal of Political Economy* 80(4) (1972), 623–48.

EKERN, S., "Increasing Nth Degree Risk," *Economics Letters* 6(4) (1980), 329–33.

ELLSBERG, D., "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics* 75(4) (1961), 643–69.

ESÖ, P., AND L. WHITE, "Precautionary Bidding in Auctions," *Econometrica* 72(1) (2004), 77–92.

FALK, A., A. BECKER, T. J. DOHMEN, B. ENKE, D. HUFFMAN, AND U. SUNDE, "The Nature and Predictive Power of Preferences: Global Evidence," IZA Discussion Paper 9504, 2015.

FEHR-DUDA, H., A. BRUHIN, T. EPPER, AND R. SCHUBERT, "Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size," *Journal of Risk and Uncertainty* 40 (2010), 147–80.

FELDER, S., AND T. MAYRHOFER, "Risk Preferences: Consequences for Test and Treatment Thresholds and Optimal Cutoffs," *Medical Decision Making* 34(1) (2014), 33–41.

———, AND ———, *Medical Decision Making: A Health Economic Primer*, 2nd edition (Berlin Heidelberg: Springer, 2017).

FISCHBACHER, U., "Z-tree: Zurich Toolbox for Readymade Economic Experiments," *Experimental Economics* 10(2) (2007), 171–78.

FREDERICK, S., "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives* 19(4) (2005), 25–42.

GOLLIER, C., AND J. W. PRATT, "Risk Vulnerability and the Tempering Effect of Background Risk," *Econometrica* 64(5) (1996), 1109–23.

GREINER, B., "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE," *Journal of the Economic Science Association* 1(1) (2015), 114–25.

GRETHER, D. M., AND C. R. PLOTT, "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review* 69(4) (1979), 623–38.

GRISLEY, W., AND E. KELLOG, "Risk-taking Preferences of Farmers in Northern Thailand: Measurements and Implications," *Agricultural Economics* 1 (1987), 127–42.

HAERING, A., AND T. HEINRICH, "Risk preferences in China: Results from experimental economics." ASIEN 142 (2017), 66–88.

———, ———, AND T. MAYRHOFER, "Exploring the consistency of higher-order risk preferences," Ruhr Economic Papers No. 688, 2017.

HARRISON, G. W., J. MARTÍNEZ-CORREA, AND J. T. SWARTHOUT, "Reduction of Compound Lotteries with Objective Probabilities: Theory and Evidence," *Journal of Economic Behavior and Organization* 119 (2015), 32–55.

HARRISON, G. W., AND J. T. SWARTHOUT, "Experimental Payment Protocols and the Bipolar Behaviorist," *Theory and Decision* 77(3) (2014), 423–38.

HEINRICH, T., AND T. MAYRHOFER, "Higher-order Risk Preferences in Social Settings," *Experimental Economics* 21(2) (2018), 434–56.

HEINRICH, T., AND J. SHACHAT, "The Development of Risk Aversion and Prudence in Chinese Children and Adolescents," Mimeo, 2018.

HERRMANN, B., C. THÖNI, AND S. GÄCHTER, "Antisocial Punishment across Societies," *Science* 319(5868) (2008), 1362–67.

HOFSTEDE, G., *Culture's Consequences: International Differences in Work-related Values* (Beverly Hills, CA: Sage Publications, 1980).

HOFSTEDE, G., G. J. HOFSTEDE, AND M. MINKOV, *Cultures and Organizations: Software of the Mind* Volume 3 (London: McGraw-Hill, 2010).

HOLT, C. A., AND S. K. LAURY, "Risk Aversion and Incentive Effects," *American Economic Review* 92(5) (2002), 1644–55.

———, AND ———, "Risk Aversion and Incentive Effects: New Data without Order Effects," *American Economic Review* 95(3) (2005), 902–04.

HSEE, C., AND E. U. WEBER, "Cross-national Differences in Risk Preference and Lay Predictions," *Journal of Behavioral Decision Making* 12 (1999), 165–79.

KACHELMEIER, S. J., AND M. SHEHATA, "Examining Risk Preferences under High Monetary Incentives: Experimental Evidence from the People's Republic of China," *American Economic Review* 82(5) (1992a), 1120–41.

———, "Culture and Competition: A Laboratory Market Comparison between China and the West," *Journal of Economic Behavior and Organization* 19(2) (1992b), 145–68.

KIMBALL, M. S., "Precautionary Saving in the Small and in the Large," *Econometrica* 58(1) (1990), 53–73.

———, "Precautionary Motives for Holding Assets," in P. Newman, M. Milgate, and J. Falwell, eds., *The New Palgrave Dictionary of Money and Finance* (London: Macmillan, 1992), 158–161.

KOCHER, M. G., J. PAHLKE, AND S. T. TRAUTMANN, "An Experimental Study of Precautionary Bidding," *European Economic Review* 78 (2015), 27–38.

KRIEGER, M., AND T. MAYRHOFER, "Patient Preferences and Treatment Thresholds under Diagnostic Risk: An Economic Laboratory Experiment," Ruhr Economic Papers No. 321, 2012.

———, AND ———, "Prudence and Prevention: An Economic Laboratory Experiment," *Applied Economics Letters* 24(1) (2017), 19–24.

KRIPPENDORFF, K., *Content Analysis: An Introduction to its Methodology* (Thousand Oaks, CA: Sage Publications, 2004).

LAJERI-CHAHERLI, F., "Proper Prudence, Standard Prudence and Precautionary Vulnerability," *Economics Letters* 82(1) (2004), 29–34.

LEVITT, S. D., AND J. A. LIST, "What do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *Journal of Economic Perspectives* 21(2) (2007), 153–74.

LICHTENSTEIN, S., AND P. SLOVIC, "Reversals of Preference between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology* 89(1) (1971), 46–55.

LINDMAN, H. R., "Inconsistent Preferences among Gambles," *Journal of Experimental Psychology* 89(2) (1971), 390–97.

MAIER, J., AND M. RÜGER, "Experimental Evidence on Higher-Order Risk Preferences with Real Monetary Losses," Working Paper, University of Munich, 2012.

MARKOWITZ, H., "The Utility of Wealth," *Journal of Political Economy* 60(2) (1952), 151–58.

MAYRHOFER, T., AND H. SCHMITZ, "Prudence and Prevention: An Empirical Investigation," Mimeo, 2019.

NOCETTI, D. C., "Robust Comparative Statics of Risk Changes," *Management Science* 62(5) (2015), 1381–92.

NOUSSAIR, C. N., S. T. TRAUTMANN, AND G. VAN DE KUILEN, "Higher Order Risk Attitudes, Demographics, and Financial Decisions," *Review of Economic Studies* 81(1) (2014), 325–55.

OECD, "Purchasing Power Parities (PPP)," https://doi.org/10.1787/1290ee5a-en (Accessed on 02 March 2015), 2015.

OOSTERBEEK, H., R. SLOOF, AND G. VAN DE KUILEN, "Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-analysis," *Experimental Economics* 7(2) (2004), 171–88.

Özer, Ö., Y. Zheng, and Y. Ren, "Trust, Trustworthiness, and Information Sharing in Supply Chains Bridging China and the United States," *Management Science* 60(10) (2014), 2435–60.

Peter, R., "Optimal Self-protection in Two Periods: On the Role of Endogenous Saving," *Journal of Economic Behavior & Organization* 137 (2017), 19–36.

Pratt, J., "Risk Aversion in the Small and in the Large," *Econometrica* 32(1-2) (1964), 122–36.

Quiggin, J., "A Theory of Anticipated Utility," *Journal of Economic Behavior & Organization* 3(4) (1982), 323–43.

Rieger, M. O., M. Wang, and T. Hens, "Risk Preferences around the World," *Management Science* 61(3) (2015), 637–48.

Roth, A. E., V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir, "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *American Economic Review* 81(5) (1991), 1068–95.

Samuelson, P. A., "Probability, Utility, and the Independence Axiom," *Econometrica* 20(4) (1952), 670–78.

Segal, U., "The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach," *International Economic Review* 28(1) (1987), 175–202.

———, "Does the Preference Reversal Phenomenon Necessarily Contradict the Independence Axiom?" *American Economic Review* 78(1) (1988), 233–36.

———, "Two-stage Lotteries without the Reduction Axiom," *Econometrica* 58(2) (1990), 349–77.

Slovic, P., "Manipulating the Attractiveness of a Gamble Without Changing its Expected Value," *Journal of Experimental Psychology* 79(1 Pt1) (1969), 139–45.

Starmer, C., and R. Sugden, "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation," *American Economic Review* 81(4) (1991), 971–78.

Statman, M., "Countries and Culture in Behavioral Finance," *CFA Institute Conference Proceedings Quarterly* 25(3) (2008), 38–44.

Sutter, M., M. G. Kocher, D. Glätzle-Rützler, and S. T. Trautmann, "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior," *American Economic Review* 103(1) (2013), 510–31.

Trautmann, S. T., and G. van de Kuilen, "Higher Order Risk Attitudes: A Review of Experimental Evidence," *European Economic Review* 103 (2018), 108–124.

UBS, "Prices and Earnings," UBS Report, 2014. https://www.ubs.com/microsites/prices-earnings/open-data.html

Vieider, F. M., "Moderate Stake Variations for Risk and Uncertainty, Gains and Losses: Methodological Implications for Comparative Studies," *Economics Letters* 117(3) (2012), 718–21.

Vieider, F. M., T. Chmura, T. Fisher, T. Kusakawa, P. Martinsson, F. M. Thompson, and A. Sunday, "Within- Versus Between-Country Differences in Risk Attitudes: Implications for Cultural Comparisons," *Theory and Decision* 78(2) (2015), 209–18.

Vieider, F. M., M. Lefebvre, R. Bouchouicha, T. Chmura, R. Hakimov, M. Krawczyk, and P. Martinsson, "Common Components of Risk and Uncertainty Attitudes across Contexts and Domains: Evidence from 30 Countries," *Journal of the European Economic Association* 13(3) (2015), 421–52.

Weber, E. U., and C. Hsee, "Cross-cultural Differences in Risk Perception but Cross-cultural Similarities in Attitudes towards Perceived Risk," *Management Science* 44(9) (1998), 1205–17.

White, L., "Prudence in Bargaining: The Effect of Uncertainty on Bargaining Outcomes," *Games and Economic Behavior* 62(1) (2008), 211–31.

Wik, M., T. Kebede, O. Bergland, and S. Holden, "On the Measurement of Risk Aversion from Experimental Data," *Applied Economics* 36 (2004), 2443–51.

CHAPTER 3

Framing Decisions in Experiments on Higher-Order Risk Preferences

Unpublished Manuscript:

Haering, A. (2020). Framing Decisions in Experiments on Higher-Order Risk Preferences.

Framing decisions in experiments on higher-order risk preferences

ALEXANDER HAERING*

May 2020

Abstract − In this study I analyze how lottery framing and lottery display type affects the degree of higher order risk-preferences. I explore differences through displaying reduced rather than compound lotteries and differences through displaying lotteries in an urn-style rather than in a spinner-style. Overall, my findings show that individual behavior is influenced by lottery framing but not by display format.

JEL Classification: C91, D81

Keywords: Risk aversion, prudence, temperance, higher-order risk preferences, lottery framing

* RWI – Leibniz Institute for Economic Research and Fresenius University of Applied Sciences.
E-mail: alexander.haering@rwi-essen.de
Financial support from the BMBF (Grant No. 01UC1304) is gratefully acknowledged.
I thank Jeannette Brosig-Koch, Timo Heinrich and Matthias Kaeding for very helpful comments. Special thanks to Anna Werbeck for outstanding comments and support. Franziska Then provided excellent research assistance.

1. Introduction

Risk aversion is the established concept when analyzing decision-making under uncertainty. Most economic models assume that the majority of people is risk-averse, meaning that they do not like risk involved in decisions. In general, this is captured by a negative second-order derivative of the utility function in an expected utility framework. However, it turned out that risk preferences of individuals are only captured partially by this concept. Higher-order risk preferences like prudence (positive third-order derivative of the utility function; Kimball, 1990) and temperance (negative fourth-order derivative; Kimball, 1992) also impact decisions made by individuals when facing uncertainty (see e.g. Esö & White, 2004, Eeckhoudt & Gollier, 2005 and White, 2008). These theoretical studies typically assume an unobservable utility function of a decision-maker who maximizes her expected utility. Contrary to this, Eeckhoudt & Schlesinger (2006) derivate a definition of higher-order risk preferences outside the expected utility framework. They defined the concepts of prudence and temperance linked to individual preferences over certain lottery pairs. In general, they showed that prudence, temperance and even that risk preferences of any order can be associated with the preference to disaggregate harms. This work paved the way for measuring higher-order risk preferences in economic lab experiments.

In my experiment, I rely on the lotteries introduced by Deck & Schlesinger (2010). I answer the question how lottery framing and lottery display type influence

[1] the degree of individual risk preferences from order 1 to order 6, using a binary choice approach. My results show that lottery framing influences individual behavior, whereas lottery display type does not. Hence, my study contributes to the understanding of higher-order risk preferences measured in the lab by being the first to analyze both, lottery framing and display type in a between subject design.

In general, previous work (see Noussair et al., 2014 and Trautmann & van de Kuilen, 2018 for overviews), using the theoretical framework of Eeckhoudt & Schlesinger (2006), differs in the way prudence, temperance and even higher-order risk preferences are measured: Some authors elicit individuals' risk premia, others use a binary choice approach.

In case of risk premia elicitation, the subjects are asked how much their valuations of a, for instance, prudent lottery differs compared to an imprudent lottery. This method provides information of the exact degree of prudence and temperance because subjects can be ordered with regards to their premia. Furthermore, the risk premia can be used to estimate an individual's utility function. Ebert & Wiesen (2014) first used the risk premia approach, classifying the huge majority of subjects as prudent and temperate.

[1] In the proper meaning of words, reducing a compound lottery also leads to a different "display type" of a lottery. But I choose these descriptions to separate the effect caused by pure visualization from the effect that might arise because reduced lotteries are often valued differently by subjects (see e.g. Budescu & Fischer, 2001 for an overview).

Using the Ebert & Wiesen (2014) method in a social interaction experiment, Heinrich & Mayrhofer (2018) confirmed their findings.

Contrary to the above described risk premia approach, the binary choice approach is the common way to measure higher-order risk preferences in the lab. It presents several pairs of lotteries to the subjects, e.g. one prudent and one imprudent. Subsequently, subjects are asked which lottery they prefer. By counting the amount of prudent, temperate etc. choices, subjects are classified accordingly. The binary choice method is used by the majority of studies. Deck & Schlesinger (2010) first used the binary choice approach classifying a majority of subjects as prudent, but only a minority as temperate. Following Deck & Schlesinger (2010), several studies confirm their findings in case of prudence (Deck & Schlesinger, 2014, 2018; Noussair et al., 2014; Baillon et al. 2018 and Haering et al., 2020). In contrast, the findings concerning temperance are less clear. On the one hand, Deck & Schlesinger (2014, 2018), Noussair et al. (2014) and Haering et al. (2020) observe an above average share of temperate choices in their subject pool. On the other hand, Baillon et al. (2018) observed only around 43% of temperate choices, confirming the first findings by Deck & Schlesinger (2010).

It is important to note that these experiments partially differ in the presentation of the lotteries to the subjects. They differ in the lottery framing (compound or reduced) and how the lotteries are displayed (spinner or urn). The majority of studies uses compound lotteries (e.g. Deck & Schlesinger, 2010, 2014; Noussair et al., 2014; Ebert & Wiesen, 2014 and Heinrich & Mayrhofer, 2018), less use reduced lotteries (e.g. Baillon et al. 2018) and few studies use both framings (Deck & Schlesinger, 2018; Haering et al. 2020). Exploring differences with regards to the lottery display type reveals that a spinner design is the common way (e.g. Deck & Schlesinger, 2010, 2014, 2018 and Haering et al. 2020) and less studies use an urn design (e.g. Ebert & Wiesen, 2014 and Heinrich & Mayrhofer, 2018). In addition, some studies only measure higher-order risk preferences up to order 4 (temperance).

I, therefore, focus on risk preferences up to order 6 and rely on the most commonly used binary choice approach. I investigate how lottery framing and lottery display type influence the degree of individual risk preferences from order 1 to order 6. My two explicit research questions are: Is individual behavior influenced by the lottery framing (compound or reduced)? And how does the lottery display type (spinner or urn) affect behavior?

The remaining paper is structured as follows. In section two, I summarize the theoretical background. In section three, I present the experimental design. I summarize my findings in section four and discuss them in section five.

2. Theoretical background

This section briefly reviews the theoretical background based on Deck & Schlesinger (2014). To measure higher-order risk preferences, they use different sets of compound lotteries, which are based on the theoretical background established by Eeckhoudt & Schlesinger (2006), Eeckhoudt et al. (2009) and Crainich et al. (2013). These lottery sets consist of binary lotteries with equal probabilities, [x, y]. Here, the lottery contains two potential outcomes x and y, with a 50% chance of receiving x and a 50%t chance of receiving y. In case of lotteries with an order higher than 2, x and y might themselves be lotteries.

Based on the theoretical background of Deck & Schlesinger (2014), Figure 1 shows risk apportionment up to order 4 and for any order. Here, W is an individual's initial wealth ($W > 0$), and k_1 and k_2 are fixed values ($k_1 > 0$ and $k_2 > 0$). Due to the negative sign, they represent two sure losses. Two independent zero-mean background risks are represented by δ and ε.

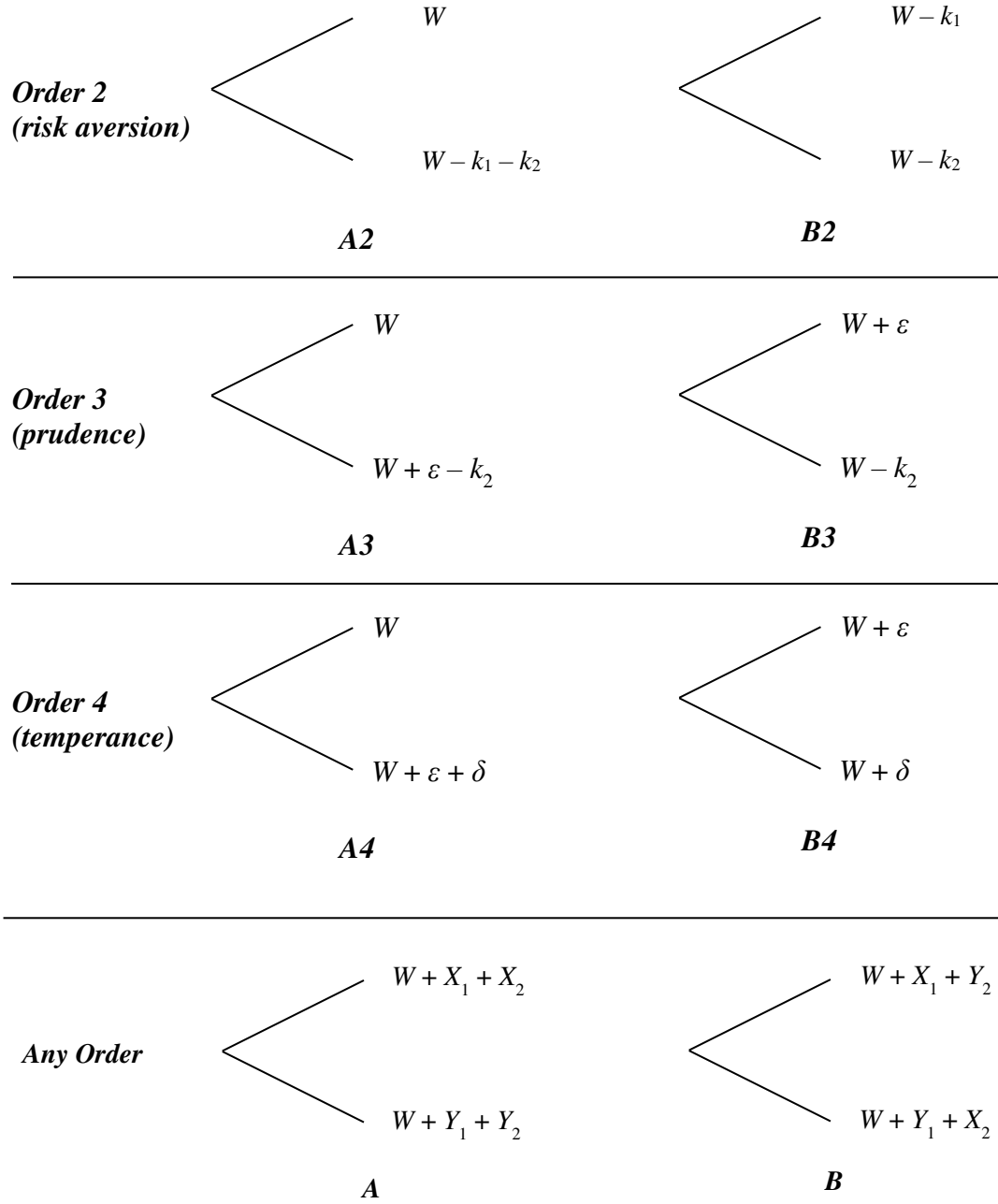
In the first row, risk apportionment of order 2 (risk aversion) is illustrated. A risk averse individual prefers Lottery B2 over A2 because of the lower variance resulting in a disaggregation of the two sure losses k_1 and k_2 in Lottery B2. In other words, a lower variance is associated with a lower second-order risk. Vice versa, a risk-loving individual prefers Lottery A2.

The second row of Figure 1 shows risk apportionment of order 3 (prudence). Here k_1, a sure loss, is replaced by the zero-mean background risk ε. Following Eeckhoudt & Schlesinger (2006) prudence is defined as a preference for disaggregating a sure loss and an additional zero-mean risk. Hence, a prudent individual prefers Lottery B3 and an imprudent individual prefers Lottery A3.

The third row displays risk apportionment of order 4 (temperance). Here, the second sure loss k_2 is replaced by δ, the additional zero-mean risk that is independent of ε. In this case, temperance can be defined as a preference for disaggregating the two risks ε and δ (Eeckhoudt & Schlesinger, 2006). Therefore, a temperate individual prefers Lottery B4 over A4 and, vice versa, an intemperate individual prefers A4 over B4.

A general approach for orders higher than order 4 by Deck & Schlesinger (2014), based on the theoretical approach by Eeckhoudt et al. (2009), is shown in the last row of Figure 1. The approach assumes that the tasks consists of two random variables [X_1,Y_1]. Here, Y_1 has more n-th degree risk than X_1. Y_1 has more n-th degree risk than X_1 if the following two conditions are fulfilled: (1) X_1 and Y_1 have the same $n - 1$ moments ($n > 0$) and (2) if X_1 is n-th order stochastic dominant to Y_1. Beside that, [X_2,Y_2] is a second pair of random variables, where Y_2 has more m-th degree risk than X_2. Here, all random variables are statistically independent of each other. Subjects which prefer lotteries with a lower (m + n)-th degree risk are "risk apportioning of order $m + n$". And subjects which are risk apportioning of order $m + n$ prefer lottery B over lottery A in Figure 1.

Figure 1:Risk apportionment up to order 4 and for any order as lottery preferences



3. Experimental design

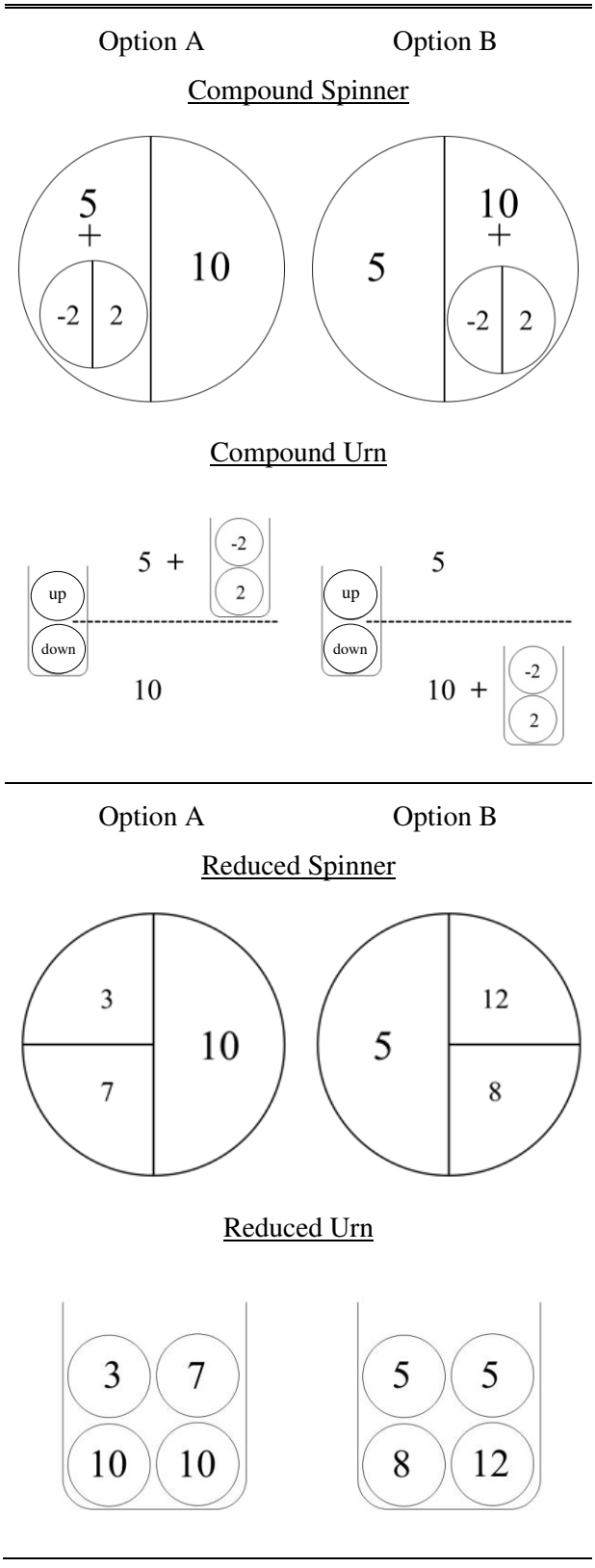
Elicitation method

The general experimental design and the lottery parameters follow the elicitation method by Deck & Schlesinger (2014). It contains 38 tasks (3 in case of order 1 and 7 each in orders 2 to 6) and subjects choose between an Option A and an Option B. In contrast to Deck & Schlesinger (2014), I introduce a reduced framing and an additional way to display the lotteries to the subjects, the urn format. Figure 2 shows examples of a prudent task (order 3) the way it is presented to the subjects, depending of the framings (compound and reduced) and display types (spinner and urn) I use in my experiment. In case of spinners, the probability of winning is represented by the share of the circle (like a wheel of fortune) and in the urns display format by the number of balls in them.

In Figure 2 "Compound Spinner" and "Compound Urn" − Option A − involves a 50% chance of winning 10 ECU (experimental-currency-units) or a 50% chance of winning 5 ECU. But when a subject receives the 5 ECU, it involves a second lottery which either results in a 2 ECU win or 2 ECU lose, which is added or subtracted from the 5 ECU. In "Reduced Spinner" and "Reduced Urn" − Option A−, participants can win 5 – 2 = 3 ECU with a probability of 25%, 5 + 2 = 7 ECU with a probability of 25%, and 10 ECU with a probability of 50%. So, the reduced framing can be derived by multiplying out the probabilities of the potential outcomes.

Using the notation from Section 2, where W is an individual's initial wealth, k_2 is a fixed value, ε is a zero-mean background risk and $[x, y]$ denotes a lottery with a 50% chance of receiving x and a 50% chance of receiving y. In this example, the parameters of the task are $W = 10$ and $k_2 = 5$. The zero-mean background risk ε is represented by an additional lottery [-2,2]. Therefore, the one lottery "Option A" corresponds to $[W - k_2 + \varepsilon, W]$ and the other lottery "Option B" to $[W - k_2, W + \varepsilon]$. Following Eeckhoudt & Schlesinger (2006) a prudent individual should prefer "Option B" over "Option A", independent of framing and lottery display type.

Figure 2: Examples of lotteries of order 3 (Task 11) as presented to participants



Experimental treatments and protocol

The economic laboratory experiment was conducted at the Laboratory for Experimental Economics (elfe) in Essen, Germany. All participants faced 34 tasks[2] in randomized order, and one of the tasks was randomly selected for payment.[3] In each task, the subjects had to choose between two lotteries, Option A or Option B. The arrangement of Option A and B was displayed randomly to the participants, meaning that Option A was randomly displayed on the left or right side of the screen and Option B vice versa. Table 1 shows my treatments as well as the lotteries` display type (spinner or urn) and framing (compound or reduced).

Table 1:Treatments

Treatment	Lotteries by Deck & Schlesinger (2014) (Order, S = Spinner or U = Urn, C = Compound or R = Reduced)
Spinner_C	1SC; 2SC; 3SC; 4SC; 5SC; 6SC
Urn_C	1UC; 2UC; 3UC; 4UC; 5UC; 6UC
Spinner_R	1SR; 2SR; 3SR; 4SR; 5SR; 6SR
Urn_R	1UR; 2UR; 3UR; 4UR; 5UR; 6UR

I use a between-subjects design and each subject only participated in one treatment. For recruiting the participants, I use ORSEE (Greiner, 2015); for programming my experiment z-Tree (Fischbacher, 2007). Overall, 143 student subjects took part. 36.4% were students of economics, 23.8% students of engineering and 34.4% students of other disciplines. They earned on average Euro 25.41 and a session lasted roughly two hours[4]. Participants were randomly assigned to one of the four treatments in each session by drawing a card from a stack of cards. After entering their respective cubical, subjects read the instructions and all remaining questions were answered in private. Before the actual tasks started, participants had to pass a test of understanding (containing four questions, see Appendix A1). In case participants had problems with a question, they were helped privately by one of the experimenters. Intermediate questions during the experiment were allowed and introduced by hand signals, question and answer were handled privately. At the end of the experiment payments were given out in private.

[2] Unfortunately, duo to a computer error I was not able to use one task in order 4 and 5 each and two tasks in order 6.

[3] This might add an additional compound layer to the lotteries. But I followed Deck and Schlesinger (2014) and used a random payment technique, which allows to compare my results with previous studies. In addition, as Azrieli et al. (2018) point out, the random payment technique "is essentially the only incentive compatible mechanism".

[4] The sessions were conducted as the first part of another experiment. Here only this first part is analyzed.

Before and after the task, the participants answered four questions. The following first two questions were asked before the lottery task, but after the subjects read the introductions: "*What do you think, how confident will you be with your choices?*" and "*What do you think, how well did you understand the instructions?*". After finishing the lottery tasks, the subjects had to answer the two final questions: "*How understandable were the lotteries?*" and "*How confident are you with your decisions?*". Participants answered the questions by rating them with school grades (A to F).

4. Results

Subject pool and summary statistics

Table 2 summarizes the demographics of the subject pool I used in my experiment. The share of female participants is slightly above 50% in **Spinner_C** and **Urn_C,** but the differences between the four treatments are not significant (p =1, Fisher's exact test).

Table 2: Summary statistics

				Demographics		Course of studies	
Treatment	Display Type	Framing	N	Female	Age (SD)	Econ.	Eng.
Spinner_C	Spinner	Compound	35	51.4%	24.514 (5.198)	22.9%	31.4%
Urn_C	Urn	Compound	36	52.8%	23.333 (2.449)	41.7%	19.4%
Spinner_R	Spinner	Reduced	36	50.0%	23.917 (2.116)	38.9%	22.2%
Urn_R	Urn	Reduced	36	50.0%	24.333 (3.538)	41.7%	22.2%

Note: N is the number of participants and SD is the standard deviation.
In course of studies "Econ." represents economics and "Eng." Engineering,
the remaining participants are enrolled in other disciplines.

The observed differences regarding the average subjects' age are statistically insignificant ($p \geq 0.185$, two-sided Mann-Whitney U test). The same holds true for the share of participants studying economics ($p \geq 0.129$, Fisher's exact test) or engineering ($p \geq 0.285$). To provide a robustness check of my results, I add the subjects age and gender to OLS regressions (see Appendix A2 for details).

Analyses of individual behaviour

In this subsection the individual behavior is analyzed by comparing the four treatments (**Spinner_C, Spinner_R, Urn_C** and **Urn_R**) with each other. I measure if subjects exhibit a tendency for *n*-th order risk-loving behavior in each treatment and test whether participants' behavior differs between the treatments. I, then, pool my four treatments by lottery framing (**Compound** vs **Reduced**) and display type (**Spinner** vs **Urn**). My goal of the latter step is to measure the sole influence of the lottery framing and display type respectively.

Finally, I investigate how the subjects rate the lotteries by school grades before and after the tasks and the time they need to make their decisions.

Higher-order risk preferences across treatments − In each task, subjects were able to choose between a risk-loving and a risk-averse choice. The number of choices differs between the orders. There are three choices in order 1, seven in orders 2 & 3, six in orders 4 & 5 and five in order 6. Like Deck and Schlesinger (2014), I use the number of risk-loving choices in each order as a measure of *n*-th order risk aversion: the more *n*-th order risk-loving choices a subject selects, the lower is her degree of *n*-th order risk aversion. In general, I assume that all subjects prefer more money over less, as measured by the tasks in order 1. That is the case for the vast majority of participants (98.6%). Only two subjects[5] choose an option with a lower payoff in the **Spinner_C** treatment once. Yet, this share is not statistically different when comparing **Spinner_C** with any other treatment ($p = 0.239$, Fisher's exact test).

Firstly, I investigate whether the participants exhibit a tendency for *n*-th order risk-loving behavior in each treatment separately. Table 3 summarizes the average number and median of *n*-th order risk-loving choices in each treatment separated by order 1 to 6. In addition, it shows p-values of two-sided Wilcoxon signed-rank tests against the median that I would expect due to random behavior by the subjects (H_0 WSR test). I, therefore, use this test strategy to measure a tendency for risk-loving or risk-averse behavior in each order.

Testing against random behavior reveals that in all treatments up to order 5, except in the **Spinner_R** treatment, I can (weakly) significantly reject random behavior by the participants ($p \leq 0.088$, two-sided Wilcoxon signed-rank test). The participants display a tendency for *n*-th order risk-aversion up to order 5, apart from order 3 in **Urn_R**. Here, the subjects exhibit 3-rd order risk-loving (imprudent) behavior due to the higher number of 3-rd order risk-loving choices.

I assume that subjects stated their *n*-th order risk preferences in a nonrandom way in all four treatments up to order 5 (except order 4 in **Spinner_R**). They dislike risk in any order with order 3 in **Urn_R** treatment being the only exception. In this instance, participants prefer the 3-rd order risk loving option,

[5] All my results reported in this paper are robust when I drop these two subjects from my analysis.

an indicator of imprudent behavior. In order 6, I can only reject random behavior in **Spinner_R**. Actually, such random behavior comes as no surprise, as the lotteries in order 6 are quite complex, even in case of reduced framing.

Table 3: n-th order risk-loving choices across treatments

Order	**1**	**2**	**3**	**4**	**5**	**6**
# of choices	3	7	7	6	6	5
H_0 WSR test	1.5	3.5	3.5	3	3	2.5
Spinner_C						
Mean	0.057	1.229	1.429	1.914	2.429	2.143
Std. Dev.	0.236	1.682	1.596	1.704	1.290	1.438
Median	0	1	1	2	2	2
p-value	**0.000**	**0.000**	**0.000**	**0.001**	**0.021**	0.145
Urn_C						
Mean	0.000	1.111	1	2.278	2.167	2.361
Std. Dev.	0.000	1.785	1.219	1.632	1.444	1.355
Median	0	0.5	1	2	2	2
p-value	**0.000**	**0.000**	**0.000**	**0.018**	**0.002**	0.590
Spinner_R						
Mean	0.000	1.472	2.889	2.917	1.833	1.889
Std. Dev.	0.000	1.765	2.095	1.381	1.384	1.282
Median	0	1	2	3	2	2
p-value	**0.000**	**0.000**	**0.088**	0.859	**0.000**	**0.012**
Urn_R						
Mean	0.000	1.222	4.111	2.444	1.583	2.194
Std. Dev.	0.000	1.476	1.894	1.698	1.273	1.369
Median	0	1	4.5	2	1	2
p-value	**0.000**	**0.000**	**0.063**	**0.072**	**0.000**	0.220

Note: Std. Dev represents the standard deviation. p-values calculated by two-sided Wilcoxon signed-rank test (WSR) against H_0.

Secondly, I investigate differences in the share of *n*-th order risk-loving choices between each treatment separately. I test whether lottery framing, display type or both jointly influence individual risk-loving behavior in each order, respectively.

I start with lottery framings. Comparing the compound and reduced form of Spinner display types (**Spinner_C** and **Spinner_R)** reveals that participants choose highly significantly more often the 3-rd and 4-th order risk-loving option in **Spinner_R** ($p \leq 0.004$, two-sided Mann-Whitney *U* test). In case of order 5, they choose weakly significantly less often the risk-loving choice ($p = 0.054$). All other observed differences between **Spinner_C** and **Spinner_R** are insignificant ($p \geq 0.508$).

Investigating differences between **Urn_C** and **Urn_R** indicates that subjects choose highly significantly more often the 3-rd order risk-loving option and weakly significantly less often the 5-th order

risk-loving option (p = 0.071) in **Urn_R** (p =0.000). All other differences between **Urn_C** and **Urn_R** are insignificant (p ≥ 0.511).

Considering differences between display types, a comparison of the reduced form of the two display types (**Spinner_R** and **Urn_R)** shows that subjects choose highly significantly more 3-rd order risk-loving options in **Urn_R** (p = 0.013). All other observed differences between these two treatments are insignificant ($p \geq 0.224$). Finally, in case of the compound form of the display types (**Spinner_C** and **Urn_C),** I do not observe any significant differences ($p \geq 0.284$).

All findings above are robust with regards to the subjects' gender and age (see Appendix A2 and A3 for the regression specification and additional OLS regressions).

In summary, my results of n-th order risk-loving choices between treatments reveal two things. (1) The majority of differences occur between the lottery framings, i.e. between **Spinner_C** and **Spinner_R** and between **Urn_C** and **Urn_R**. (2) When I consider the display types, comparing **Spinner_R** and **Urn_R** or **Spinner_C** and **Urn_C**, subjects only behave differently between **Spinner_R** and **Urn_R** in order 3. Stated differently, the display format does not influence a subject's n-th order risk-loving behavior much. This finding gives ample reason that lottery framing is the driving force of differences in n-th order risk-loving behavior.

Higher-order risk preferences across Framings − To verify the finding that lottery framing influences subjects' behavior and display type does not, I pool my observations to compare both framings and both lottery display types solely. Therefore, I pool the two display type treatments depending on the lottery framing and the two framing treatments with regards to display type.

I start by comparing the aggregated **Compound** and **Reduced** treatments against each other. Again, I interpret the number of n-th order risk-loving choices as a measure of n-th order risk aversion (more n-th order risk-loving choices the lower is her degree of n-th order risk aversion).

Firstly, I explore whether the participants exhibit a tendency for n-th order risk-loving behavior in the pooled **Compound** and the pooled **Reduced** treatments. Table 4 displays the average number of n-th order risk-loving choices in the **Compound** and the **Reduced** treatments separated by orders. It shows the p-values of two-sided Wilcoxon signed-rank tests against the median, I would expect under random behavior by the subjects (H_0 WSR test).

In case of **Compound** framing, I observe a clear non-random behavior of the subjects up to order 5 (p = 0.000, two-sided Wilcoxon signed-rank test). They prefer highly significantly less often the n-th order risk-loving choices and therefore exhibit a clear tendency for n-th order risk-aversion.

Table 4: n-th order risk-loving choices across Framings

Order	**1**	**2**	**3**	**4**	**5**	**6**
# of choices	3	7	7	6	6	5
H_0 WSR test	1.5	3.5	3.5	3	3	2.5
Compound						
Mean	0.028	1.169	1.211	2.099	2.296	2.254
Std. Dev.	0.167	1.724	1.423	1.666	1.367	1.391
Median	0	1	1	2	2	2
p-value	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.148
Reduced						
Mean	0	1.347	3.5	2.681	1.708	2.042
Std. Dev.	0	1.62	2.076	1.555	1.326	1.326
Median	0	1	4	3	1.5	2
p-value	**0.000**	**0.000**	0.986	0.146	**0.000**	**0.008**

Note: Std. Dev represents the standard deviation. P-values calculated by two-sided Wilcoxon signed-rank test against H_0.

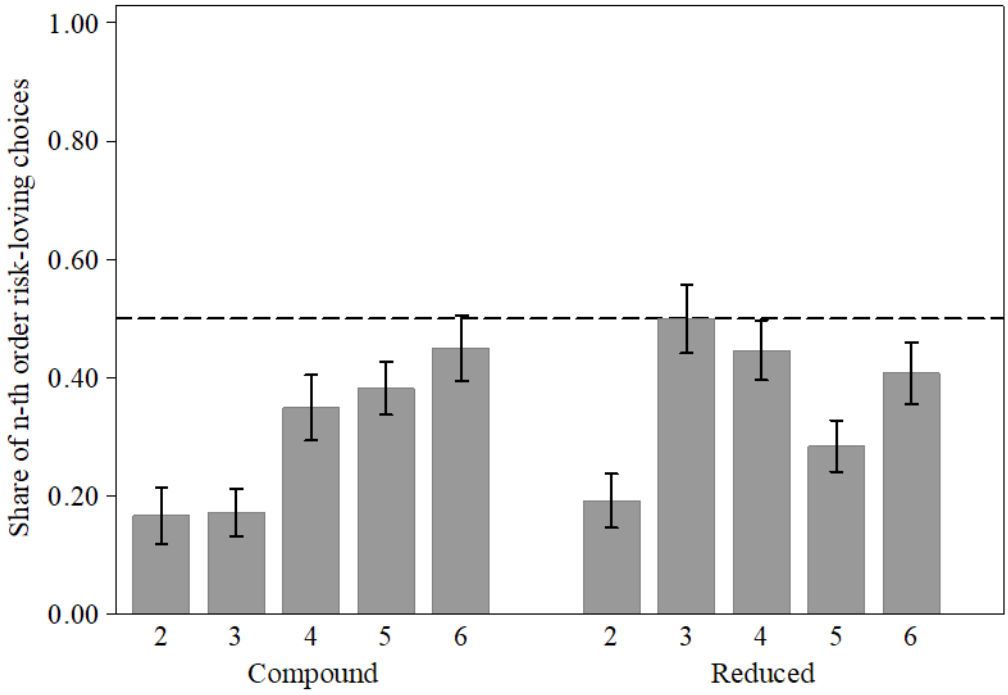
When facing the **Reduced** framing, the pattern is less clear. Participants highly significantly less often choose the *n*-th order risk-loving option in order 1, 2, 5 and 6 ($p \leq 0.008$). Yet, I cannot reject random behavior by the participants in order 3 and 4 ($p \geq 0.146$).

In summary, I observe that in pooled **Compound** framing, subjects stated their *n*-th order risk preferences in a nonrandom way. They dislike risk in every order up to order 5. But in **Reduced** framing, individuals' behavior is not that clear. They are neither risk-loving nor risk-averse in order 3 and 4.

Secondly, I explore differences in the share of *n*-th order risk-loving choices between the pooled **Compound** and pooled **Reduced** framings. I test whether lottery framing influences individual risk-loving behavior in each order. Figure 3 shows the average share of *n*-th order risk-loving choices as well as 90 percent confidence intervals. It compares both Framing treatments against each other, separated by order.

In order 3 and 4, the subjects choose (highly) significantly more often the *n*-th order risk-loving choice in **Reduced** framing ($p \leq 0.025$, two-sided Mann-Whitney *U* test) and in order 5 highly significantly less often ($p = 0.009$), instead. The behavior in order 2 and 6 is not significantly different ($p \geq 0.359$). All observations are robust when I add a female dummy and a subjects age as control variables in OLS regressions (see Appendix A3 for the regression results).

Figure 3: Share of n-th order risk-loving choices across Framings



In summary, I observe a clear non-random pattern of risk-averse, prudent, temperate and edgy behavior in **Compound** lotteries. Yet, in **Reduced** lotteries, the pattern is less clear. I also observe evidence for a different behavior by the subjects due to the lottery framing. Subjects choose the *n*-th order risk-loving option more often in orders 3 and 4 in the **Reduced** framing and less often in order 5. I, therefore, find that the framing of the lotteries influences individual behavior. But the results should be interpreted with caution. I cannot reject random behavior by the participants in the **Reduced** framing.

Higher-order risk preferences across display types − In the second step, I compare the aggregated **Spinner** and **Urn** treatments against each other. My goal is to explore the effect of display type. I interpret the number of *n*-th order risk-loving choices as a measure of *n*-th order risk aversion.

First, I examine whether the subjects show a tendency for *n*-th order risk-loving behavior in the pooled **Spinner** and the pooled **Urn** treatments. Table 5 summarizes the results. Again, the table displays the average number of *n*-th order risk-loving choices in the pooled **Spinner** as well as the pooled **Urn** treatments, separated by order. It reports p-values of two-sided Wilcoxon signed-rank tests against the median, which would occur due to random behavior by the subjects (H_0 WSR test).

In both treatment pools, **Spinner** and **Urn**, participants highly significantly more often ($p \leq 0.006$, two-sided Wilcoxon signed-rank test) prefer the *n*-th order risk-averse choice. The only exception is order 6 in case of **Urn** ($p = 0.209$).

Table 5: n-th order risk-loving choices across display types

Order	**1**	**2**	**3**	**4**	**5**	**6**
# of choices	3	7	7	6	6	5
H_0 WSR test	1.5	3.5	3.5	3	3	2.5
Spinner						
Mean	0.028	1.352	2.169	2.423	2.127	2.014
Std. Dev.	0.167	1.716	1.993	1.618	1.362	1.357
Median	0	1	2	2	2	2
p-value	**0.000**	**0.000**	**0.000**	**0.006**	**0.000**	**0.005**
Urn						
Mean	0.000	1.167	2.556	2.361	1.875	2.278
Std. Dev.	0.000	1.627	2.226	1.656	1.383	1.355
Median	0	1	2	2	2	2
p-value	**0.000**	**0.000**	**0.001**	**0.003**	**0.000**	0.209

Note: Std. Dev represents the standard deviation. P-values calculated by two-sided Wilcoxon signed-rank test against H_0.

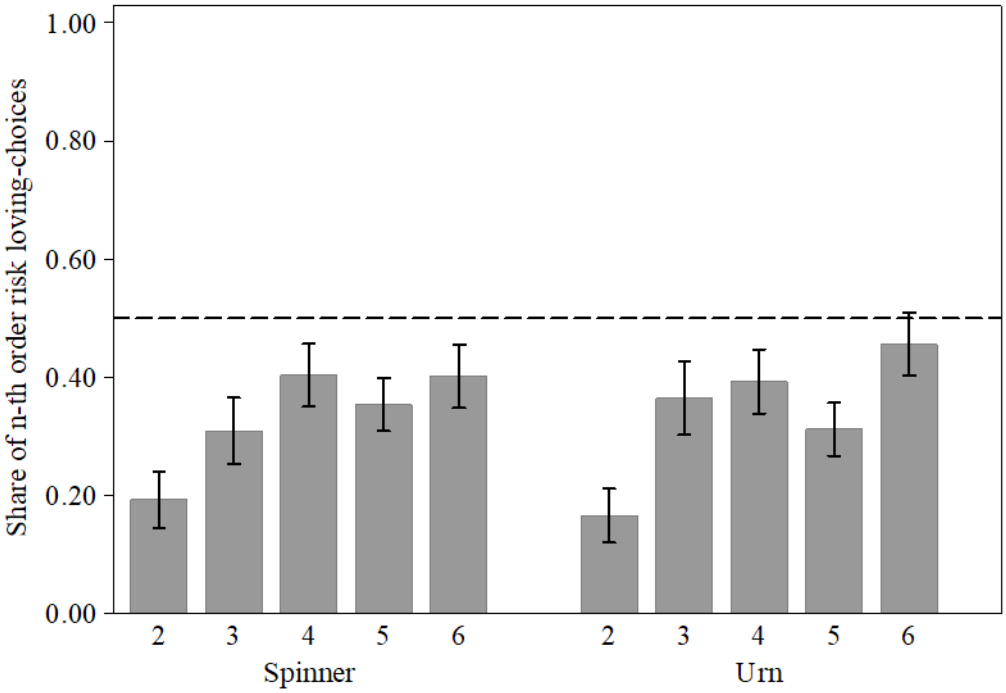
Consequently, I observe a non-random and a clear *n*-th order risk-averse behavior up to order 5 in both treatments. In case of my pooled **Spinner** treatments even up to order 6. Regardless of lottery display type, the subjects stated their *n*-th order risk preferences in a nonrandom way and dislike risk of any order.

Secondly, I examine differences in the share of *n*-th order risk-loving choices between the pooled **Spinner** and pooled **Urn** display treatments. Figure 4 displays the average share of *n*-th order risk-loving choices as well as 90 percent confidence intervals for both display type treatments separated by order.

Comparing the differences between the two display types reveals that all minor differences are not statistically significant ($p \geq 0.192$, two-sided Mann-Whitney *U* test). This observation is also confirmed by the OLS regressions (controlling for the subjects' age and gender) in Appendix A3.

In summary, I observe a non-random behavior and a clear *n*-th order risk-averse behavior up to order 5 in both treatments. I do not find evidence for a different behavior by the subjects due to the lottery display type. I, therefore, find that the display type of the lotteries does not influence individual *n*-th order risk-loving behavior.

Figure 4:Share of n-th order risk-loving choices across display types



Potential differences in lottery evaluation and time needed by the subjects – In a final step, I gather more exploratory evidence about the subjects' behavior due to the four treatments. Therefore, I examine how the subjects rated the lotteries by school grades before and after the tasks. And I consider the time needed by the participants to make their decisions. I compare all my four treatments (**Spinner_C**, **Urn_C**, **Spinner_R** and **Urn_R**) separately.

To measure the confidence of the subjects before and after the lottery task as well as to investigate how understandable the introductions and the lotteries were, I used four questions. I asked the following first two questions before the lottery task, but after the subjects read the introductions. These are #1 "*What do you think, how confident will you be with your choices?*" and #2 "*What do you think, how well did you understand the instructions?*". After completion of the tasks, I asked the final two questions: #3 "*How understandable were the lotteries?*" and #4 "*How confident are you with your decisions?*". The questions as well as the median grade given by the subjects are displayed in Table 6.

Overall, I observe a similar pattern with minor differences. Comparing the grades between **Spinner_C** and **Spinner_R** treatments reveals that subjects rated question #1 weakly significantly better ($p = 0.095$, two-sided Mann-Whitney *U* test) and question # 3 highly significantly better ($p = 0.011$) in the reduced framing. And comparing **Urn_C** and **Urn_R** treatments reveals that subjects rated question #3 highly significantly better ($p = 0.006$) in **Urn_R**. All other minor differences are insignificant ($p \geq 0.192$).

Table 6: Median grades across Treatments

#	Before	*Spinner_C*	*Urn_C*	*Spinner_R*	*Urn_R*
1	What do you think, how confident will you be with your choices?	B	B	B	B
2	What do you think, how well did you understand the instructions?	A	A	A	A
	After				
3	How understandable were the lotteries?	B	B	A	A
4	How confident are you with your decisions?	C	C	B-	C

Note: Questions are rated using school grades, A for "very good" to F for "unsatisfactory".

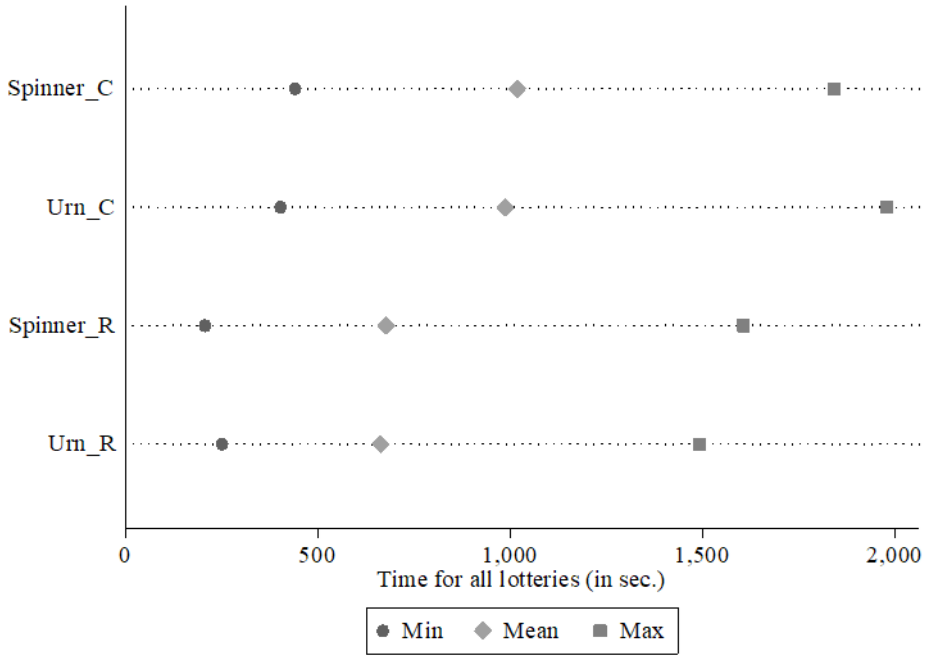
In a nutshell, subjects thought that they will be less confident when they are confronted with the compound framing with a spinner display format. They also think that the compound lotteries are less understandable in both display type formats.

In a next step, I analyse the time needed by the subjects to make their decisions. I investigate whether the subjects need more time in the more complex compound treatments compared to the reduced treatments. Figure 5 summarizes the average time (mean) as well as the minimum (min) and maximum (max) separated by treatment.

On average, the participants needed highly significantly ($p = 0.000$, two-sided Mann-Whitney *U* test) more time in **Spinner_C** than in **Spinner_R**. I observe the same pattern in the Urn treatments when I compare **Urn_C** and **Urn_R** ($p = 0.001$). Comparing **Spinner_C** and **Urn_C** as well as **Spinner_R** and **Urn_R** revels no significant differences ($p \geq 0.550$).

In summary, subjects needed more time for the compound than for the reduced framed lotteries, independent of display type. This does not come as a surprise. The compound framed lotteries (especially in higher orders) are more complex and harder to understand. But the extra time they needed is a good indicator that they understand the task and try to solve the decision problem they are facing.

Figure 5: Time needed for all lotteries across Treatments



5. Conclusion

In this study, I analyze how lottery framing and lottery display type affect the degree of higher order risk-preferences. Beside risk-aversion (2-nd order), I focus on prudence (3-rd order), temperance (4-th order), edginess (5-th order) and risk apportioning of order 6. Based on the elicitation method introduced by Deck & Schlesinger (2014), I explore differences through displaying reduced rather than compound lotteries and differences through displaying lotteries in an urn display format than in a spinner display format.

Overall, my findings show that the lottery framing influences individual behavior: Confronted with a spinner display format, subjects choose less prudent and temperate (3-rd and 4-th order) but more edgy (5-th order) options in a reduced lottery framing. This observation holds true for orders 3 and 5 when subjects are confronted with an urn display format. Comparing the differences emerging from lottery display format reveals that only in order 3 subjects behave differently. They choose less prudent options due to an urn display format in a reduced framing. Put differently, the display format does not influence a subject's *n*-th order risk-loving behavior, but the lottery framing is the driving force of differences. These findings are confirmed when I pool my observations to compare both framings and both lottery display types solely.

My finding regarding the effect of lottery framing is confirmed by recent studies. Deck & Schlesinger (2018) compare subjects' behavior due to compound and reduced framing. They apply a spinner display format and use a within subject design. They observe a significant framing effect: subjects are less temperate (4-th order) and more edgy (5-th order) but − in contrast to my study − slightly more prudent (3-rd order) due to a reduced framing. Using a within subject design, Haering et al. (2020) observe that the framing influences individuals' *n*-th order risk loving behavior, too. They compare compound with reduced lotteries using a spinner display format. Their findings confirm that subjects choose the 3-rd and 4-th order risk-loving option less often in compound lotteries. But they do not observe significant differences in order 5. Maier & Rüger (2012) study higher-order risk preferences using reduced lotteries only. In their study the share of prudent and temperate subjects is on average lower than in most studies using compound lotteries. But the authors do not directly compare lottery framings.

In addition, Haering et al. (2020) shed some light into the driving factors behind these differences in *n*-th order risk-loving behavior duo to lottery framing. They show that subjects' reasoning for a prudent and temperate choice is the maximization of the smallest potential payoff when facing compound lotteries. This finding gives ample reason that subjects do not value compound and reduced lotteries in the same way (Trautmann & van de Kuilen, 2018) and might fail to reduce compound lotteries by themselves in a proper way (see e.g. Starmer & Sugden, 1991). This might lead to a different behavior by subjects depending on the lottery framing in the context of higher-order risk preferences.

As, to my knowledge, I am the first to use different display types in one study, a verification of my finding is not straightforward. Yet, a comparison of two recent studies using a spinner display type (Deck & Schlesinger, 2018 and Haering et al., 2020) with two studies using an urn-style display type (Heinrich & Mayrhofer, 2018 and Bleichrodt & van Bruggen, 2018) reveals an ambiguous picture. The papers using a spinner design both observe that the majority of subjects exhibit prudent and temperate behavior. The papers using an urn-style display type observe inconsistent results. Heinrich & Mayrhofer (2018) find that the majority of subjects can be classified as prudent and temperate, whereas Bleichrodt & van Bruggen (2018) classified only slightly above half of the subjects as prudent and only less than half as temperate. This gives ample reason that my observation of a non-existing display type effect should be interpreted with caution. Though, it cannot be ruled out that there are other differences between these studies that influence individual behavior.

Overall, my results contribute to the understanding of higher-order risk preferences measured in the lab. I enrich the growing literature by being the first to analyze both, lottery framing and display type in a between subject design up to order 6. My findings can help researchers when designing new experiments on individual behavior with a focus on higher-order risk preferences.

References

Azrieli, Y., Chambers, C. P., & Healy, P. J. (2018). Incentives in experiments: A theoretical analysis. *Journal of Political Economy*, 126(4), 1472-1503.

Baillon, A., Schlesinger, H., & van de Kuilen, G. (2018). Measuring higher order ambiguity preferences. *Experimental Economics*, *21*(2), 233-256.

Bleichrodt, H., & van Bruggen, P. (2018). Higher order risk preferences for gains and losses. *Working Paper.*

Budescu, D. V., & Fischer, I. (2001). The same but different: an empirical investigation of the reducibility principle. *Journal of Behavioral Decision Making*, 14(3), 187-206.

Crainich, D., Eeckhoudt, L., & Trannoy, A. (2013). Even (mixed) risk lovers are prudent. *American Economic Review*, *103*(4), 1529-35.

Deck, C., & Schlesinger, H. (2010). Exploring higher order risk effects. *The Review of Economic Studies*, *77*(4), 1403-1420.

Deck, C. & Schlesinger, H. (2014). Consistency of higher order risk preferences, *Econometrica*, 82(5), 1913-1943.

Deck, C., & Schlesinger, H. (2018). On the robustness of higher order risk preferences. *Journal of Risk and Insurance*, *85*(2), 313-333.

Eeckhoudt, L., & Gollier, C. (2005). The impact of prudence on optimal prevention. *Economic Theory*, 26(4), 989-994.

Eeckhoudt, L., & Schlesinger, H. (2006). Putting risk in its proper place. *American Economic Review*, *96*(1), 280-289.

Eeckhoudt, L., Schlesinger, H., & Tsetlin, I. (2009). Apportioning of risks via stochastic dominance. *Journal of Economic Theory*, *144*(3), 994-1003.

Ebert, S., & Wiesen, D. (2014). Joint measurement of risk aversion, prudence, and temperance. *Journal of Risk and Uncertainty*, *48*(3), 231-252.

Esö, P., & White, L. (2004). Precautionary bidding in auctions. *Econometrica*, *72*(1), 77-92.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2), 171-178.

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

Haering, A., Heinrich, T. & Mayrhofer, T. (2020). Exploring the Consistency of Higher-order Risk Preferences. *International Economic Review*, 61(1),*283-320*.

Heinrich, T., & Mayrhofer, T. (2018). Higher-order risk preferences in social settings. *Experimental Economics*, *21*(2), 434-456.

Kimball, M. S. (1990). Precautionary saving in the small and in the large, *Econometrica*, 58(1), 53-73.

Kimball, M. S. (1992). Precautionary motives for holding assets. In: Newman, P., Milgate, M. & Falwell, J. (Eds.) *The New Palgrave Dictionary of Money and Finance*, MacMillan, London, 158-161.

Maier, J., & Rüger, M. (2012). "Experimental Evidence on Higher-order Risk Preferences with Real Monetary Losses." *Working Paper*, University of Munich

Noussair, C. N., Trautmann, S. T. & van de Kuilen, G. (2014). Higher order risk attitudes, demographics, and financial decisions, *Review of Economic Studies*, 81(1), 325-355.

Starmer, C., & Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation. *The American Economic Review*, 81(4), 971-978.

Trautmann, S. T., & van de Kuilen, G. (2018). Higher order risk attitudes: A review of experimental evidence. *European Economic Review*, *103*, 108-124.

White, L. (2008). Prudence in bargaining: The effect of uncertainty on bargaining outcomes. *Games and Economic Behavior*, *62*(1), 211-231.

Appendix

A1. Test of understanding

The test is translated from German. The questions were the same in all treatments, but the graphic presented varied depending on the treatment. Participants had to choose one of the answers. The correct answer is marked bold.

Page 1

Question 1: If you were to observe the following choices and selected Option A, you would receive...

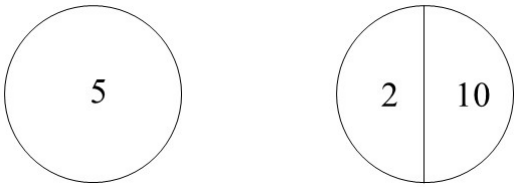
Answers 1: "2" or "**5**" or "10"

Question 2: If you were to observe the following choices and selected Option B, you would receive...

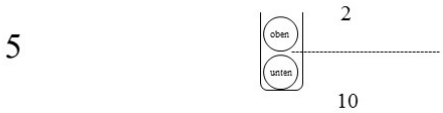
Answers 2: "5" or "**2 or 10, each with an equal chance**"

Option A **Option B**

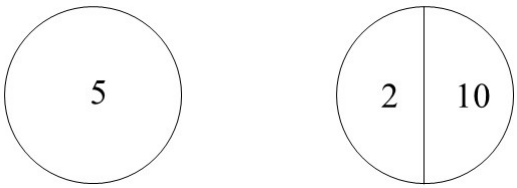
Spinner_C



Urn_C



Spinner_R



Urn_R

Page 2

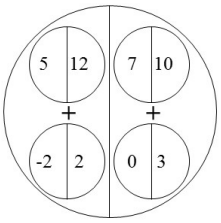
Question 3: If you were to select the following lottery, the smallest amount of money you could earn is...

Answers 3: "-2" or "0" or "**3**"

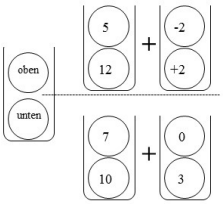
Question 4: If you were to select the following lottery, the largest amount of money you could earn is...

Answers 4:"13" or "**14**" or "17"

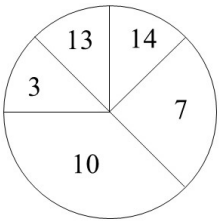
Spinner_C



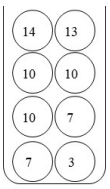
Urn_C



Spinner_R



Urn_R



A2: Summary of variables and specification regression

Table A1: Summary of variables

Variable	Description
y_i :	
Order n	Subject's number of *n*-th order risk-loving choices in order *n*
$\boldsymbol{\Gamma}'_i$:	
Urn_C	Dummy variable indicating "Urn Compound" treatment
Spinner_R	Dummy variable indicating "Spinner Reduced" treatment
Urn_R	Dummy variable indicating "Urn Reduced" treatment
Compound	Dummy variable indicating both "Compound" treatment
Spinner	Dummy variable indicating both "Spinner" treatment
$\boldsymbol{X}'_{it}$:	
Female	Dummy variable indicating female subjects
Age	The subjects age

I estimate an OLS regression to investigate differences in higher-order risk preferences across my four treatments using the following equation:

$$y_i = \beta_0 + \boldsymbol{\Gamma}'_i\tau + \boldsymbol{X}'_i\gamma + \varepsilon_i \tag{1.1}$$

In equation 1.1 y_i represents an individual's number risk-loving choices within one order *n*. Vector $\boldsymbol{\Gamma}'_i$ contains Urn_C_i, $Spinner\_R_i$ and Urn_R_i which are dummy variables indicating a subjects' treatment or the dummy indicator $Compound_i$ or $Spinner_i$ respectively in case of pooled analysis. The vector $\boldsymbol{X}'_i$ contains additional explanatory variables to consider potential effects of individual's demographics (*Female, Age*). To avoid problems due to a correlation between the error terms ε_i between subjects in a specific session (heteroscedasticity), I use robust standard errors.

A3: Regression results

Table A2: OLS regression number of n-th order risk-loving choices

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	O2	O2	O3	O3	O4	O4	O5	O5	O6	O6
Urn_C	-0.117	-0.076	-0.429	-0.472	0.363	0.340	-0.262	-0.255	0.218	0.193
	(0.411)	(0.407)	(0.338)	(0.359)	(0.396)	(0.400)	(0.325)	(0.331)	(0.332)	(0.336)
Spinner_R	0.244	0.255	1.460***	1.448***	1.002***	0.989***	-0.595*	-0.593*	-0.254	-0.263
	(0.409)	(0.403)	(0.441)	(0.445)	(0.369)	(0.371)	(0.317)	(0.322)	(0.324)	(0.326)
Urn_R	-0.006	-0.007	2.683***	2.683***	0.530	0.525	-0.845***	-0.846***	0.052	0.051
	(0.376)	(0.378)	(0.415)	(0.419)	(0.404)	(0.406)	(0.304)	(0.306)	(0.333)	(0.333)
Female		-0.452		0.438		-0.074		-0.092		0.187
		(0.284)		(0.291)		(0.273)		(0.228)		(0.229)
Age		0.030		-0.032		-0.021		0.005		-0.020
		(0.039)		(0.037)		(0.042)		(0.032)		(0.026)
p-values										
Urn_C vs Spinner_R	0.390	0.422	0.000	0.000	0.075	0.072	0.319	0.314	0.131	0.148
Urn_R vs Spinner_R	0.515	0.493	0.010	0.009	0.198	0.204	0.426	0.427	0.330	0.315
Urn_C vs Urn_R	0.774	0.859	0.000	0.000	0.672	0.636	0.071	0.072	0.605	0.660
N	143	143	143	143	143	143	143	143	143	143
AIC	558.356	558.914	567.095	567.930	545.685	549.334	495.586	499.379	498.084	500.938
BIC	570.207	576.691	578.946	585.707	557.536	567.111	507.437	517.156	509.935	518.715

Note: Constant not reported, robust standard errors in parentheses, asterisks indicate the significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: OLS regression number of n-th order risk-loving choices

	(1) O2	(2) O2	(3) O3	(4) O3	(5) O4	(6) O4	(7) O5	(8) O5	(9) O6	(10) O6
Compound	-0.178	-0.162	-2.289***	-2.302***	-0.582**	-0.586**	0.587**	0.591**	0.212	0.204
	(0.280)	(0.278)	(0.297)	(0.296)	(0.270)	(0.270)	(0.225)	(0.226)	(0.227)	(0.228)
Female		-0.452		0.441		-0.075		-0.092		0.188
		(0.283)		(0.300)		(0.273)		(0.228)		(0.228)
Age		0.030		-0.021		-0.027		0.007		-0.021
		(0.039)		(0.035)		(0.043)		(0.032)		(0.026)
N	143	143	143	143	143	143	143	143	143	143
AIC	554.853	555.413	573.061	574.458	544.192	547.649	492.902	496.666	495.478	498.289
BIC	560.779	567.265	578.987	586.310	550.117	559.500	498.828	508.518	501.404	510.141

Note: Constant not reported, robust standard errors in parentheses, asterisks indicate the significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: OLS regression number of n-th order risk-loving choices

	(1) O2	(2) O2	(3) O3	(4) O3	(5) O4	(6) O4	(7) O5	(8) O5	(9) O6	(10) O6
Compound	0.185	0.171	-0.387	-0.381	0.061	0.070	0.252	0.250	-0.264	-0.255
	(0.280)	(0.277)	(0.353)	(0.354)	(0.274)	(0.276)	(0.230)	(0.231)	(0.227)	(0.228)
Female		-0.455		0.397		-0.086		-0.081		0.191
		(0.282)		(0.355)		(0.278)		(0.232)		(0.228)
Age		0.029		-0.008		-0.025		0.003		-0.019
		(0.039)		(0.040)		(0.043)		(0.032)		(0.026)
N	143	143	143	143	143	143	143	143	143	143
AIC	554.819	555.374	621.811	624.466	548.795	552.312	498.427	502.289	494.993	497.824
BIC	560.745	567.226	627.736	636.317	554.721	564.163	504.352	514.140	500.919	509.675

Note: Constant not reported, robust standard errors in parentheses, asterisks indicate the significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

CHAPTER 4

Costly Information Acquisition: The Effects of Wage Inequality

Unpublished Manuscript:

Heinrich, T., Haering, A. & Horak, S. (2018). Costly Information Acquisition: The Effects of Wage Inequality.

Costly information acquisition: The effects of wage inequality *

Timo Heinrich†, Alexander Haering‡ and Sven Horak§

June 2018

Abstract

Information is often acquired within organizations. These social settings are usually characterized by wage inequality but it is difficult to study the influence of wages on performance systematically. In this paper we analyze the causal effects of wage inequality on information acquisition performance and vary the pay of agents and their peers. Our experimental results reveal that disadvantageous inequality in wages does not have a negative effect on agents' performance. It can even be more effective in improving performance of agents than a general increase in wages. This effect can be explained by agents' individual differences in loss aversion.

Keywords: Information acquisition, wage inequality, experiment.

JEL Codes: C91, D83, M52.

---

* Financial support from the BMBF through the IN-EAST School of Advanced Studies (Grant No. 01UC1304) is gratefully acknowledged. This study was also supported through a grant to Timo Heinrich from the Mercator Research Center Ruhr.

† Timo Heinrich, Durham University Business School, Mill Hill Lane, Durham, DH1 3LB, UK; email: timo.heinrich@ibes.uni-due.de (corresponding author).

‡ RWI – Leibniz-Institute for Economic Research, Essen and Fresenius University of Applied Sciences; email: alexander.haering@rwi-essen.de.

§ Peter J. Tobin College of Business, St. John's University; email: horaks@stjohns.edu.

1 Introduction

Following Arrow (1969), technological progress can be regarded as a reduction in uncertainty. Uncertainty is reduced by acquiring information about the world through research and development. Arrow already highlights the problem of determining the optimal sequence of activities during this process. In the same vein, ten years later, Weitzman (1979) motivates his classical paper on optimal search with an example of the research and development process of a large organization. The organization's researchers consider new production technologies and can sequentially develop them to resolve uncertainty about the rewards they deliver. The risks and rewards of these technologies differ, and only one technology can be used for production. Weitzman then derives the optimal sequence of developing production technologies for an individual decision-maker under a range of conditions.

But, as in Weitzman's example, research is usually conducted within the social settings of organizations and not by individuals in isolation. Like other organizations, research organizations are often characterized by wage inequality (see, e.g., Pfeffer & Davis-Blake, 1990), even though equal wages have frequently been identified as a critical factor for the motivation of workers (see, e.g., Bewley, 1999). In this paper, we ask how wage inequality affects the quality of choices in information acquisition. This question is highly relevant, for example, to managers who need to incentivize individuals in research and development teams. Does higher performance pay for individuals in charge improve outcomes? Or should managers aim to create homogenous incentives to foster cohesion?

The causal effect of wage inequality on the quality of decision making in information acquisition is extremely difficult to observe in the field. Wage inequality may be the cause of bad decision making but it might also be its effect (for example, when managers respond by linking pay more strongly to performance). As researchers, it is difficult to vary wages exogenously and usually impossible to judge the quality of decisions objectively. However, Herbst & Mas (2015) have recently found that laboratory experiments yield estimates of peer effects on worker productivity that are very close to those observed in the field. We therefore study information acquisition under controlled laboratory conditions that allow us to systematically vary the pay structure and observe the quality of decision making. We build on a sequential search task as studied by Weitzman (1979)

in which decision-makers can explore different options with known risks and rewards. This task has been widely used to model innovation behavior (see, e.g., Poblete & Spulber, 2017) and has a straightforward optimal decision sequence. This allows us to compare the performance of all subjects to the same objective benchmark.[1]

Behavior in the task by Weitzman (1979) has already been studied experimentally by Slonim (1994) and Gabaix et al. (2006). Both papers only consider individual choices made in isolation. Slonim (1994) focuses on experience and on the method to elicit choices. He argues that the cognitive effort associated with identifying the optimal search path may decrease with experience. In fact, he finds that behavior gets closer to the optimal path when the tasks are played repeatedly. He also observes that eliciting all possible choices for a task rather than eliciting only the relevant choices sequentially leads to better performance. Gabaix et al. (2006) propose the directed cognition algorithm based on myopic cost-benefit calculations as a boundedly rational model for information acquisition behavior. They find that the directed cognition algorithm predicts aggregate information acquisition patterns quite well. Whether the path suggested by directed cognition differs from the optimal path depends on the parameterization of the tasks. When both paths diverge, they find that directed cognition does a better job of matching laboratory evidence.[2] To our knowledge, however, none of the previous studies on sequential search have studied search

[1] In a small but growing experimental literature, different tasks have been used to study different aspects of innovation (see Brüggemann & Bizer, 2016, for an overview). For example, Ederer & Manso (2013), Ederer (2013) and Herz et al. (2014) use bandit problems to analyze exploration and exploitation behavior while Eckartz et al. (2012) and Mohnen & Ostermaier (2013) use word finding tasks to assess creativity.

[2] The task by Weitzman (1979) differs from sequential or simultaneous search tasks in which agents repeatedly explore the *same* alternative by making draws from a known payoff distribution (see Chade et al., 2017, for an overview on the theory of search models). In simultaneous search with one alternative an agent has to decide ex ante on the number of draws he wants to make. In sequential search with one alternative an agent can decide whether to stop or to continue exploring after each draw. Note that the setting we study also differs from the bandit problem in which at least one of the available options has an unknown payoff distribution (see Bergemann & Välimäki, 2008, for an overview of theoretical models). We are not interested in learning behavior but focus on the identification of optimal behavior in our modeling. Even though it captures some important aspects of search behavior, the drawback of the bandit problem is that already in moderately situations the optimal path is not readily computable.

behavior in different social settings even though differences in payoffs, for example, have widely been argued to drive work effort.

As far as we know, wage inequality has not been studied experimentally with respect to search behavior. Yet, since researchers study employer-employee interaction, a number of writers have argued that fairness concerns are important for an agent's effort provision. For example, Frank (1984) shows theoretically that wages will be less dispersed than marginal products of work when employees care about the wages of their coworkers and are free to choose their coworkers. Lazear (1989) shows theoretically that it can be more efficient for a firm to pay workers equally to foster cooperative behavior. Based on a series of interviews with managers, Bewley (1999) also concludes that equal pay is a critical factor for the motivation of workers.

In laboratory experiments, the causal effects of wage inequality have usually been studied in a gift-exchange setting in which a principal interacts with two or more agents (see Charness & Kuhn, 2011, for an overview). In these settings, an agent's effort is purely a choice variable, and the experimenter determines the effort cost. The agent can reciprocate a fair wage offer by exerting high effort, thus, comparing his income not only to the other agents but also to the principal. In this study, we are mainly interested in the horizontal comparison of wages. Bartling and von Siemens (2011) also focus on horizontal wage comparisons. Like us, they study a setting without a principal but with two identical agents. However, in their setting both agents can collaborate to increase their payoffs, which are also determined based on stated efforts. Depending on the treatment, their wages are either equal or unequal. The authors do not find a significant effect of inequality on individually stated effort. Different from their study, effort is not transparent in a cognitive task like information acquisition. It is more closely related to real-effort tasks, where the experimenter cannot control for effort costs. In laboratory settings, where total payoffs of agents are not transparent, studies find a detrimental effect of pay inequality on quality of work (Greiner et al., 2011; Bracha et al., 2015). Also, in field experiments, a negative effect on job satisfaction (Card et al., 2012) and job performance has been observed (Cohn et al., 2014; Breza et al., 2015).

There are also a number of non-experimental studies that correlate wage dispersion with performance across different tasks (see Pfeffer, 2007; Shaw, 2014; and Downes & Choi, 2014, for general overviews). With respect to research tasks, Pfeffer & Langton (1993) have observed that larger pay dispersion within a department is associated with less individual productivity (based on

data from university faculty, measured as publications over a two-year span). Yanadori & Cui (2013) consider field data of research teams. They observe that larger dispersion in wages is associated with a lower number of successful patent applications. Wang et al. (2015) consider the relationship between pay dispersion and organizational innovation, measured as whether a firm introduces new products or processes in a given time frame. They report a positive correlation between a firm's employee turnover and pay dispersion as well as a negative correlation between turnover and organizational innovation. However, to our knowledge, no study has considered the causal effects of wage inequality on information acquisition performance.

In our experimental analysis, we employ the sequential search task by Weitzman (1979) with the same parameterization as Gabaix et al. (2006) but create a social setting.[3] Thus, we study a sequential search problem without discounting and with different binary outcomes with known probabilities. An agent and his peer are paid based on the decisions made by the agent. The agent enters choices which can be observed (but not influenced) by his peer. We abstract away from principal-agent interaction or free-riding within groups. Instead, we always consider a social setting in which incentives are perfectly aligned. However, we exogenously vary absolute and relative pay levels of the agent and his peer.[4]

For deriving our hypotheses, we assume that people are in principle able to choose rationally but have to exert costly cognitive effort for increasing their expected payoff. They are not myopic and know the expected payoffs associated with each decision sequence. We then model the choice of a decision sequence as a simple maximization problem in which agents choose their cognitive effort, depending on the wage level. In addition, we assume that morale effects can influence choices. This analysis allows us to derive a set of competing hypotheses based on pure self-interest,

[3] Gabaix et al. (2006) also conduct a second experiment that studies a more-complex choice problem in which the optimal path cannot be determined analytically. We only consider the parameters of their first experiment.

[4] In two somewhat related studies, Ederer (2013) and Boyce et al. (2016) analyze behavior in a bandit problem in social settings. However, they do not consider wage inequality. Ederer (2013) studies a setting in which two agents work in parallel. He compares a pay-for-performance contract to an exploration contract. In addition, he varies whether the two agents are able to learn from each other's decisions. He finds theoretically and empirically that subjects identify better strategies under the exploration contract and when social learning is possible. In addition, performance is further improved when profits are shared between the two agents. Boyce et al. (2016) focus on the strategic interaction of players, finding that subjects free ride on the information acquired by others, as theoretically expected.

inequality-averse preferences (Fehr & Schmidt, 1999) and reference-dependent preferences (Kahneman & Tversky, 1979; Kőszegi & Rabin, 2006, 2007).[5]

In our experiment, we do not consider advantageous inequality but focus on disadvantageous inequality. We find that disadvantageous inequality in wages does not have a negative effective on agents' performance. A unilateral increase in the peer's wage can be more effective in improving performance of agents than a general increase in wages (i.e. more effective than an increase in wages for both of them). We present additional analyses which reveal that this result is partly driven by agents' individual differences in loss aversion.

[5] Reference-dependent preferences have been found to influence behaviour in experiments on sequential search with one alternative (Schunk, 2009; Schunk & Winter, 2009; McGee, 2014; Soetevent & Bružikas, 2016). All of these studies, however, focus on decisions made by individuals in isolation.

2 Experimental design

We focus on the sequential search problem presented by Weitzman (1979) and adopt the notation by Gabaix et al. (2006) in the following. In a task j, an agent i has to choose between three projects k. Each project has a known probability p_{jk} of generating a known payoff V_{jk}, if the respective project is successful and zero otherwise. For a payment of c, the agent can acquire information about project k and learns whether it was successful. Once the agent has decided to stop exploring, he can pick the highest-paying project and earns the respective payoff minus total search costs. The agent can only select projects where he already learned about the success or failure. Optimally, the risk-neutral agent[6] would always explore projects based on the index Z_{jk}. For projects with unknown payoffs, the index is calculated as:

$$Z_{jk} = (p_{jk}V_{jk}-c)/p_{jk}. \tag{1}$$

The index is equal to V_{jk} for a successful project and 0 for a failed project. If the project with the highest value Z_{jk} has an unknown outcome, the agent would acquire information about it. If the project has a known outcome, he would pick this project and stop exploring. Note that the index implies that if two projects have the same expected net gain, the one with the smaller probability of success will be selected – low-probability, high-payoff projects should be investigated first, because they will end the search.

Gabaix et al. (2006) also propose an alternative way of selecting the sequence of projects as a boundedly rational model for actual choice behavior. They propose the "directed cognition algorithm", which suggests that people explore projects in the sequence of decreasing expected gains

$$G_i = p_{jk}(V_{jk}-S_{jt})-c_{jk} \tag{2}$$

[6] We present the standard solution for a risk neutral decision-maker here. In the analysis of our experimental data we will compare behavior to the risk neutral benchmark as well as to an individual benchmark. This individual benchmark is built on individual risk preferences elicited in a separate task. Based on these estimates each individual's optimal path can be calculated.

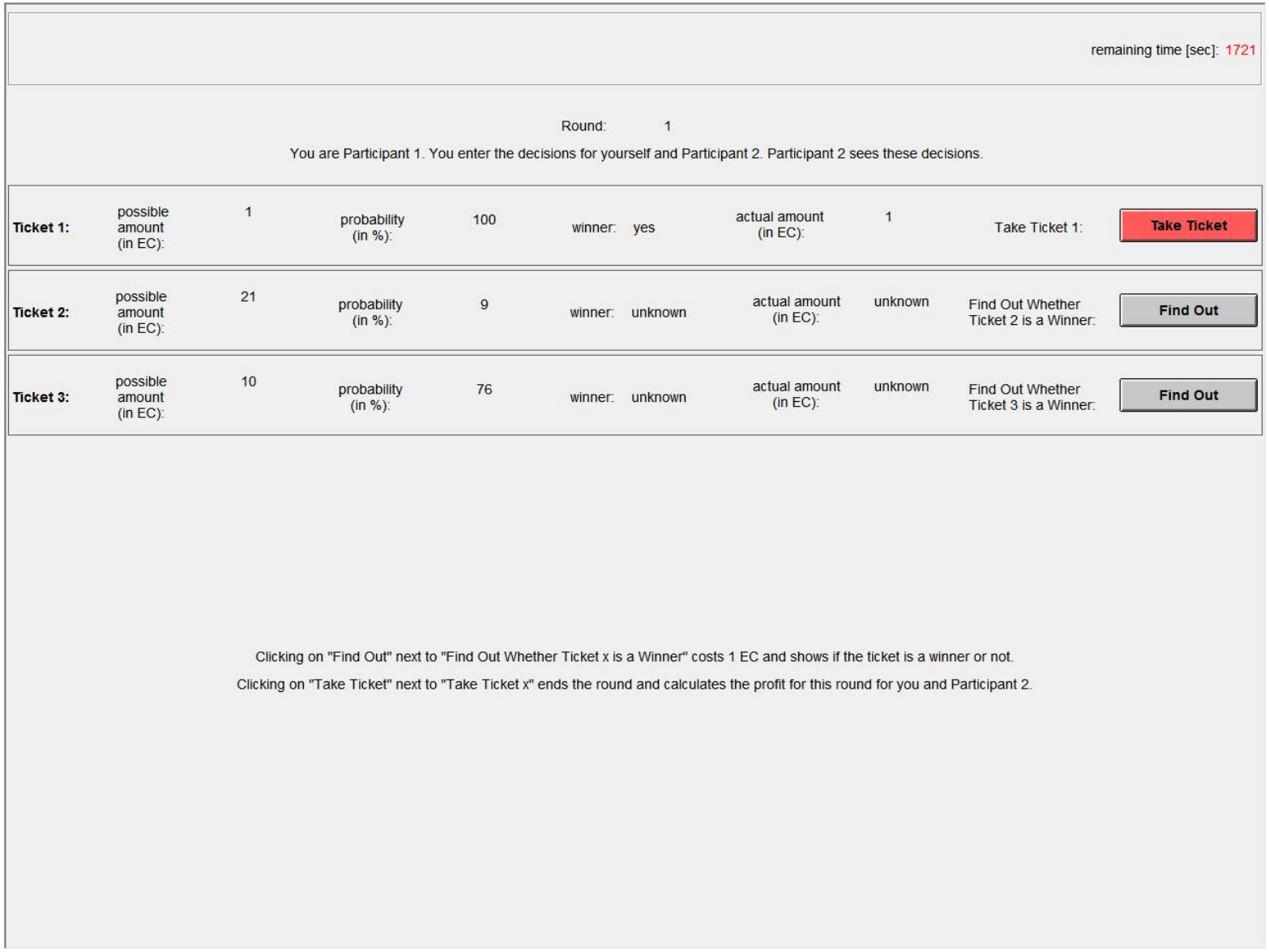
where S_{jt} is the value of the best currently known winning project. If the highest gain is provided by a project with known outcome, the algorithm stops, and this project is selected. Based on this heuristic, agents maximize expected gains myopically and choose as if each choice was the last to be made. We provide an example below.

Subjects in all treatments of the experiment acquire information about their projects in ten tasks. The tasks are ordered randomly, subjects face each task only once and one of the tasks is randomly selected for payment at the end of the experiment. We use the same parameterization as Gabaix et al. (2006), so that in five of the tasks the sequence of optimal choices diverges from the directed cognition sequence (tasks A to E). In the other five, the two coincide (tasks F to J). See Appendix A for the parameters of all 10 tasks.

Subjects face the tasks A to J in pairs. The first subject acts as the decision-making agent entering the decisions. He is matched to a second subject who witnesses project characteristics, choices and outcomes and acts as the peer in order to create a social setting. Both subjects cannot communicate but they know that the choices of the agent affect the payoffs of both (see Appendix B for the instructions including screenshots). We opted for a passive peer in order to increase the salience of inequality in wages. In addition, this approach avoids any uncertainty about inequality in effort costs because the agent's peer does not exert any effort. The experiment was implemented via z-Tree (Fischbacher, 2007).

Figure 1 contains a screenshot of the agent's screen with the three projects. It shows that probabilities and payoffs are known to subjects. Payoff amounts are given in experimental currency (EC). One unit of EC was worth Euro 1.50 to subjects. For each task we observe one of the possible decision paths. To illustrate the possible decision paths, let us consider this as an example: Project 1 is the safe option, guaranteeing a payoff of EC 1, because it is always successful. Project 2 has a payoff of EC 21 in case of success and a probability of success of 9%. Project 3 yields EC 10 in case of success. Its success rate is 76%. Projects 2 and 3 yield no payoff if they fail. Would a subject acquire information about the outcome of the two uncertain projects? Probably, because, in expectation, both yield a payoff that is larger than the payment of $c = 1$. But which one should be explored first?

Figure 1: Screenshot of agent's screen (task A)



Calculating the index Z_{Ak} for the three projects as in (1) yields the optimal sequence. The values are $Z_{A1} = 1$, $Z_{A2} = (0.09*21-1)/0.09 = 9.89$ and $Z_{A3} = (0.76*10-1)/0.76 = 8.68$. Therefore, project 2 should be explored first and taken if successful, because its outcome is certain and $Z_{A2} = 21$. If not, project 3 should be explored and selected if successful. Only if neither is successful, project 1 is selected. The decision sequence will be different for someone applying the directed cognition algorithm. This heuristic only considers the gain from the next move being made. In the beginning, the best-known successful project is project 3, therefore $S_t = 1$. Based on (2), the gains from the three projects are $G_{A1} = 1(1-1)-0 = 0$, $G_{A2} = 0.09(21-1)-1 = 0.8$ and $G_{A3} = 0.76(10-1)-1 = 5.84$, so now project 3 is considered first. If it is successful, the gains change to $G_{A1} = 1(1-10)-0 = -9$, $G_{A2} = 0.09(21-10)-1 = -0.01$ and $G_{A3} = 1(10-10)-0 = 0$, and the sequence ends. If project 3 fails, project 2 will be explored.

We ran three experimental treatments that systematically vary the wages within pairs of subjects, as shown in Table 1.[11] The treatments are designed to disentangle self-interested, inequality averse and reference-dependent motivations in search. In the **LOW** and the **HIGH** treatment, both the agent and his peer earn the same (equal) prize V_{jK} where K is the project selected in one randomly chosen round j. However, we vary the stakes both can earn. Either both receive 30% less than the nominal payoff ($\delta = 0.7$) or both receive a bonus of the same size ($\delta = 1.3$). In the **UNEQUAL** treatment, we create disadvantageous inequality from the agent's point of view: His peer receives a bonus of 30% on top of his payoff V_{jK} ($\delta = 1.3$), while he earns 30% less ($\delta = 0.7$).[12]

[11] We also ran a control treatment ($\delta = 1$) not reported here, replicating the first experiment in Gabaix et al. (2006). In this treatment, agents made choices individually and not within pairs. For example, we observe very similar frequencies of subjects choosing optimally in their first moves: 34% in tasks A to E of both experiments and 82% in tasks F to J of our experiment, compared to 74% in Gabaix et al. (2006).

[12] We chose this parameterization to create a realistic representation of wage differentials and a salient unfairness at the same time. In our experiment, those earning less receive 35 percent of the total wage. In standard ultimatum games using typical stakes, for example, offers in the range between 30 and 40 percent of the pie have been found to be rejected 30 percent of the time by List & Cherry (2000) and 42 percent of the time by Slonim & Roth (1998).

Table 1: Experimental treatments

Treatment	Wage differentials		N
	Agent	Peer	
LOW	$\delta = 0.7$	$\delta = 0.7$	72
HIGH	$\delta = 1.3$	$\delta = 1.3$	72
UNEQUAL	$\delta = 0.7$	$\delta = 1.3$	72

Upon entering the laboratory, subjects drew a ball from an urn which assigned them to a cubicle. This way subjects were also randomly assigned to a treatment. In each session we ran all three treatments in order to minimize potential session effects. The distribution of female and male agents was balanced across treatments by using two different urns. Only for one subject we had to deviate from this procedure because of several no-shows in one session. Thus, we have an equal number of male and female agents in treatments **LOW** and **UNEQUAL** and 47% female participants in **HIGH**. After taking a seat in their cubicle, the subjects read the experimental instructions. Questions about the instructions were answered in private. Before the actual series of search tasks started, subjects participated in a test of understanding. In this test, they made choices in one exemplary search task, which was not payoff relevant. After making their choices, they had to calculate possible payoffs from this task themselves (which all were able to do).

The actual series of payoff-relevant search tasks was followed by a preference elicitation task for which new instructions were handed out (also included in Appendix B). This task elicits subjects' loss aversion. It allows us to control for individual loss aversion when assessing search performance across the three treatments. The preference elicitation was adapted from Karle et al. (2015) who developed it to assess loss aversion after subjects made a consumption choice. In order to elicit subjects' preferences, they first had to make a series of six choices between a lottery paying either 0.00 or 1.00 Euro with equal probability and a sure payoff S. S took the values 0.10, 0.20, 0.30, 0.40, 0.50 and 0.60 Euro. Second, they had to make a series of six choices between a sure payoff of 0.00 Euro and lotteries creating either a gain of 1.00 Euro with probability 1/3 or a loss of R with probability 2/3. R was 0.10, 0.20, 0.30, 0.40, 0.50 and 0.60. Subjects knew that one of the 12 choices in total would be randomly selected for payoff. This payoff was added to an initial endowment of 2.00 Euro. The different attitudes towards lotteries with strictly positive payoffs and towards

lotteries with a potential loss can then be used to assess an individual's degree of loss aversion and to classify subjects accordingly.[13]

This task was incentivized and was selected because it is short and very easy for subjects to understand.[14] The experiment concluded with a brief questionnaire. Then subjects were paid in private based on the random draws and the decisions they made in the search task and the loss aversion task. Finally, they left the laboratory individually.

The experiment was conducted at the Essen Laboratory for Experimental Economics (elfe) in Germany. We implemented a between-subjects design, so each participant only participated in one treatment. In total, 216 student subjects took part in our experiment. They earned Euro 18.92 on average from a session that lasted approximately 90 minutes.

[13] More precisely: In the first series a subject should prefer the lottery in every decision up to a cutoff value S_k. For a risk neutral individual this cutoff value is 0.50 Euro, for which he is indifferent between a sure payoff of 0.50 Euro and a lottery paying the same amount in expectation. In the second series a subject should prefer the lottery in every decision up to a cutoff value $|R_k|$. For a risk neutral individual this value would be |-0.50|, for which he is indifferent between a sure payoff of 0.00 Euro and a lottery with 0.00 Euro in expected payoff. Due to the potential losses in the second series a loss averse individual would avoid risk more strongly in the second series than in the first. Following Karle et al. (2015) the elicited cutoff values are then used to solve for the parameters of the exponential utility representation by Tversky & Kahneman (1992). It characterizes preferences by a risk aversion parameter and a loss aversion parameter. For more details see Karle et al. (2015, p. 109f).

[14] The preference elicitation task is solely based on individual decision making. Any incentivized task for assessing inequality aversion necessarily involves payoffs to another person. Thus, we did not elicit individual inequality aversion because it is likely to be influenced by outcomes from the search task (or vice versa).

3 Hypotheses

We are interested in the behavioral changes that result from wage inequality in the type of task described in Section 2. In the following, we assume that, with increasing cognitive effort, people improve their choices and are more likely to follow the optimal path, i.e., to choose projects in decreasing order of Z_{ij} (cf. Slonim, 1994). For deriving the hypotheses, we assume that agents form rational expectations about the additional wage they earn from exerting additional effort and identifying a path with higher-expected payoff. For example, it might be more costly to identify the optimal path when it differs from the path described by the directed cognition algorithm, as the results by Gabaix et al. (2006) suggest. This approach follows Bolton & Faure-Grimaud (2009) in that we assume that people deal optimally with their cognitive limitations (but see also the discussion by Conlisk, 1996, and Lipman, 1991, 1999, on the infinite regress problem). It differs from their approach in that we assume people to correctly anticipate the costs and benefits of exerting additional effort. Our approach is also similar to the papers on search behavior by Caplin et al. (2011) and Reutskaja et al. (2011) in that we treat search as a real-effort task in as much as we cannot control for the subjects' effort costs. In other words, we assume that there is an unobserved search cost (in addition to c) in identifying the optimal path. We also take a reduced-form approach: We only consider the resulting payoffs of the dynamic decision task and do not model each individual decision, i.e., each step of the decision path within a task.

We follow Breza et al. (2015) in adapting the framework by DellaVigna et al. (2016) and assume that morale effects can influence the choice of effort. An agent i's utility function is given by:

$$U(e_i, w_i, y) = e_i w_i - g(e_i) - M(e_i, w_i, y) \tag{3}$$

where e_i is his cognitive effort and w_i is his wage, i.e., the amount by which his payoff increases from additional cognitive effort. The effort costs are convex and given by $g(e_i)$ with $g'(e_i) > 0$, $g''(e_i) > 0$ and $g(0) = 0$. The morale effect $M(e_i, w_i, y)$ depends on the agent's wage w_i, the effort e_i and some reference point y.

In our basic model, there is no morale effect. The optimal effort is simply derived from the first-order condition $w_i - g'(e_i) = 0$ as:

$$e_i^* = g'^{-1}(w_i) \tag{4}$$

where $g'^{-1}(\ )$ is the inverse of the effort cost function and monotonically increasing. Because $-g''(e^*) < 0$, this is a maximum. A wage increase results in a higher optimal effort choice yielding our first hypotheses.

Hypothesis 1a (self-interest): *An increase of the agent's wage leads to a* <u>*larger*</u> *propensity to acquire information optimally.*

Hypothesis 1b (self-interest): *An increase of the peer's wage* <u>*does not*</u> *influence the propensity to acquire information optimally.*

We now consider how this prediction changes once we introduce morale effects. Like Breza et al. (2015), we assume that the reference point y is formed based on the wage per unit of effort received by peers. In our case, the reference wage is equal to the wage the peer earns from the agent's actions ($y = w_{-i}e_i$). We model this based on the theory by Fehr & Schmidt (1999) of inequality aversion (IA) so that an agent suffers from unequal pay. Following them, we assume a linear relationship between inequality and utility. The morale effect that is subtracted in (3) is then given by:

$$M_{IA}(e_i, w_i, w_{-i}) = \alpha\max\{e_i w_{-i} - e_i w_i, 0\} + \beta\max\{e_i w_i - e_i w_{-i}, 0\}. \tag{5}$$

The positive parameters α and β capture two types of disutility from inequality ($\alpha, \beta \geq 0$): Agents suffer more from inequality that is disadvantageous to them than from inequality that is advantageous to them ($\alpha \geq \beta$). Furthermore, agents always like to have more money than less ($\beta < 1$).[15]

Depending on the type of inequality, solving for the optimal effort yields:

$$e_i^* = g'^{-1}(w_i - \alpha(w_{-i} - w_i)) \text{ if } w_{-i} > w_i \text{ and} \tag{6}$$

$$e_i^* = g'^{-1}(w_i - \beta(w_i - w_{-i})) \text{ if } w_i \geq w_{-i}.$$

[15] As Fehr & Schmit (1999) and Breza et al. (2015) point out, other assumptions about α and β can be made. The assumption that people do not like being paid less than their peers captured through a positive α is quite common. However, if people are status seeking, β may also be negative, which would increase the effect.

In our setting, wage inequality cannot be changed by the agents but is exogenously determined through the pay structure. Because $g'^{-1}(\ )$ is monotonically increasing, a change in its input parameters also changes optimal effort accordingly. Based on the theory of Fehr & Schmidt (1999) and their parameterization of α and β we can derive hypotheses about the reactions to wage changes: In our case, we consider disadvantageous inequality where the agent earns less than his peer ($w_i < w_{-i}$). Here, an increase of the agent's wage leads to an increase in effort. However, an increase in the peer's wage has the following effect: the larger the peer's wage the lower the agent's effort. This behavior differs from that of a self-interested agent. Thus we formulate the following hypotheses:

Hypothesis 2a (disadvantageous inequality aversion): *An increase of the agent's wage leads to a <u>larger</u> propensity to acquire information optimally.*

Hypothesis 2b (disadvantageous inequality aversion): *An increase of the peer's wage leads to a <u>lower</u> propensity to acquire information optimally.*

We now turn to the role of prior expectations as reference points. We assume that people form an expectation about the total wage they earn prior to learning about the specific nature of the task. For simplicity, we assume that they evaluate their current payoff relative to a point estimate of total wage W_i ($y = W_i$). We follow the approach by Fehr & Goette (2007) and Pokorny (2008) and assume reference-dependent preferences (RD) in the spirit of Kahneman & Tversky (1979) and Kőszegi & Rabin (2006, 2007). Following the notation above, the morale effect that is subtracted in (3) can be written as:

$$M_{RD}(e_i, w_i, W_i) = \alpha\max\{W_i - e_i w_i, 0\} + \beta\max\{e_i w_i - W_i, 0\}. \tag{7}$$

In this case, α is also positive, because people dislike payoffs below the reference point ($\alpha > 0$). However, β is negative ($\beta < 0$), because they enjoy payoffs above the reference point. Furthermore, the former effect is stronger than the latter ($|\alpha| > |\beta|$): For a given wage the decrease is smaller after the expected total wage is reached (in W_i/e_i). While agents cannot influence wage inequality, they can influence deviations from their expectations. By adjusting their effort, they can move closer to their reference point. The critical effort that yields their expected total wage W_i is W_i/w_i. The first order condition then differs for ranges above and below this critical effort and is given by:

$$U'(e_i, w_i, w_{-i}) = w_i - g'(e_i) + \alpha w_i \text{ for } e_i < W_i/w_i \text{ and} \tag{8}$$

$$U'(e_i, w_i, w_{-i}) = w_i - g'(e_i) - \beta w_i \text{ for } e_i \geq W_i/w_i.$$

These conditions show that, because of loss aversion, the marginal utility from effort is larger below the reference point than above ($w_i+\alpha w_i > w_i-\beta w_i$): The higher the wages, the higher the likelihood that the total wage an agent expected is already reached. This kink in the utility function yields three different cases of optimal effort choice (cf. Pokorny, 2008):

$$e_i^* = g'^{-1}(w_i+\alpha w_i) \text{ if } g'(W_i/w_i) > w_i+\alpha w_i,$$

$$e_i^* = W_i/w_i \text{ if } w_i-\beta w_i < g'(W_i/w_i) < w_i+\alpha w_i, \tag{9}$$

$$e_i^* = g'^{-1}(w_i-\beta w_i) \text{ if } g'(W_i/w_i) < w_i-\beta w_i.$$

In the first case, the marginal utility from income meets marginal costs at some point that yields a total wage *below* the reference point. In the third case, they meet at a point that yields a total wage *above* the reference point. In both cases, the optimal effort is increasing in the wage. These conditions are most likely to be met if the range between $w_i-\beta w_i$ and $w_i+\alpha w_i$ is small. The size of this range is characterized by the degree of an agent's loss aversion (measured as $|\alpha|-|\beta|$). If loss aversion is small enough, relative to the wage increase, the first and the third case are more likely than the second. However, in these situations we would expect the same reaction towards wage changes as under purely self-interested behavior. Thus, we focus on the second case: If loss aversion is large relative to the wage increase, agents are more likely to choose an effort that yields their expected total wage before and after a wage increase. Because the reference point W_i/w_i is lower for a higher wage, the exerted effort will decrease in this case. This is captured in our last hypothesis.

Hypothesis 3a (reference dependency with large loss aversion): *An increase of the agent's wage leads to a <u>lower</u> propensity to acquire information optimally.*

Hypothesis 3b (reference dependency with large loss aversion): *An increase of the peer's wage <u>does not</u> influence the propensity to acquire information optimally.*

These hypotheses are based on the assumption that agents only consider their own payoff when forming expectations. Thus, they are loss averse relative to this reference point only. Of course, it is also reasonable to expect that behavior is governed by reference dependency relative to expectations about their own payoff *and* loss aversion relative to other reference points like others' payoffs. Fehr & Schmidt (1999) already point out that their theory implies "that a subject is loss averse in social comparisons" (p. 824). We will come back to this point in the discussion of our results.[16]

Table 2 summarizes the prediction with regards to our three treatments (see Section 2). It shows the respective ordering of treatments based on the expected number of optimal choices. Hypotheses 1a and 1b are solely based on the agent's self-interest. Hypothesis 1a suggests that optimal decisions should be made most frequently in the treatment with the high individual incentive **HIGH**. Correspondingly, it suggests that optimal decisions should be made least frequently in the **LOW** and the **UNEQUAL** treatment, where the incentives for the agent are lower. Because incentives are the same for the agent in the latter two treatments, the performance should also be the same based on Hypothesis 1b. While the self-interest model ignores the peer, disadvantageous inequality aversion takes differences of pay between them into account. Like in the self-interest model this theoretical approach suggests that subjects decide in an optimal way most often in **HIGH** because pay is highest and there is no disutility from inequality. This prediction is based on Hypothesis 2a. Different from the self-interest model, however, disadvantageous inequality aversion suggests less optimal decisions in **UNEQUAL** than in **LOW** and **HIGH**. This prediction is based on Hypothesis 2b and the agent's disutility from being paid less than his peer. With respect to wage increases of the agent, reference-dependent preferences can make a prediction that is different from the other models: For a large level of loss aversion performance in **HIGH** should be worse than in **LOW** and **UNEQUAL** based on Hypothesis 3a. Reference-dependent preferences

[16] Also note that like loss aversion satisficing in the spirit of Simon (1955) suggests that search is inhibited if the most-preferred alternative yields a payoff above some target, while it is stimulated if this alternative is below the target as pointed out by March (1991). The higher the payoffs, the more alternatives in the choice set may become satisfactory. That means a decision-maker will stop evaluating the existing alternatives earlier and is less likely to end up with the optimal strategy.

suggest the same ordering as the self-interest model with respect to inequality in payoffs based on Hypothesis 3b.

Table 2: Summary of predictions

Model	Hypothesis	Share of optimal choices
Self-interest	1a 1b	**LOW < HIGH** and **UNEQUAL < HIGH** **LOW = UNEQUAL**
Disadvantageous inequality aversion	2a 2b	**LOW < HIGH** and **UNEQUAL < HIGH** **LOW > UNEQUAL** and **UNEQUAL < HIGH**
Reference-dependency with large loss aversion	3a 3b	**LOW > HIGH** and **UNEQUAL > HIGH** **LOW = UNEQUAL**

4 Results

In the following, we analyze the behavior in the ten tasks across our three treatments. We analyze the causal effect of changes in wages and peer wages on information acquisition performance. We judge performance by whether an agent followed the optimal decision sequence. Therefore, we calculated the optimal decision sequence for each task separately by taking into account a subject's previous choices within a task.[17] First, we compare behavior across tasks. Second, we compare behavior across treatments. In a third analysis, we consider behavior across treatments but split the sample based on the individual loss aversion parameters. Appendix C contains the raw data.

For the third analysis we can only consider participants who decided consistently in the assessment of risk preference elicitation task by Karle et al. (2015). In the price list task subjects could switch multiple times. Of the 108 agents 15 had to be excluded due to multiple switches in the task.[18] Thus we are not able to calculate their loss aversion and risk aversion parameters. For consistency the respective 15 participants are excluded for all of the remaining analyses. Note that none of the analyses which can be conducted with the whole sample yield qualitatively different results. Also note that for the analyses presented here, we consider the optimal path to be the optimal path under risk neutrality and, thus, to be the same for all participants. This way our results can be readily compared to the analyses of Gabaix et al. (2006). Also, from a firm's perspective the normative prediction under risk-neutrality is arguably the more informative benchmark. However, based on our assessment of the risk aversion parameter we are also able to determine the individually optimal path for every participant. We will point out in the following when the analysis of the individually optimal path allows for a different interpretation of our data.

[17] Looking at mistakes made by the agents reveals that e.g. in case of tasks A to E 6.25% (**LOW**), 7.14% (**HIGH**) and 6.06% (**UNEQUAL**) of the agents choose a failed project in one of the tasks. In case of tasks F to J we observe 6.25% (**LOW**), 10.71% (**HIGH**) and 0.00% (**UNEQUAL**). However, these shares differ weakly significantly between **HIGH** and **UNEQUAL** in case of tasks F to J only ($p = 0.091$, Fisher's exact test).

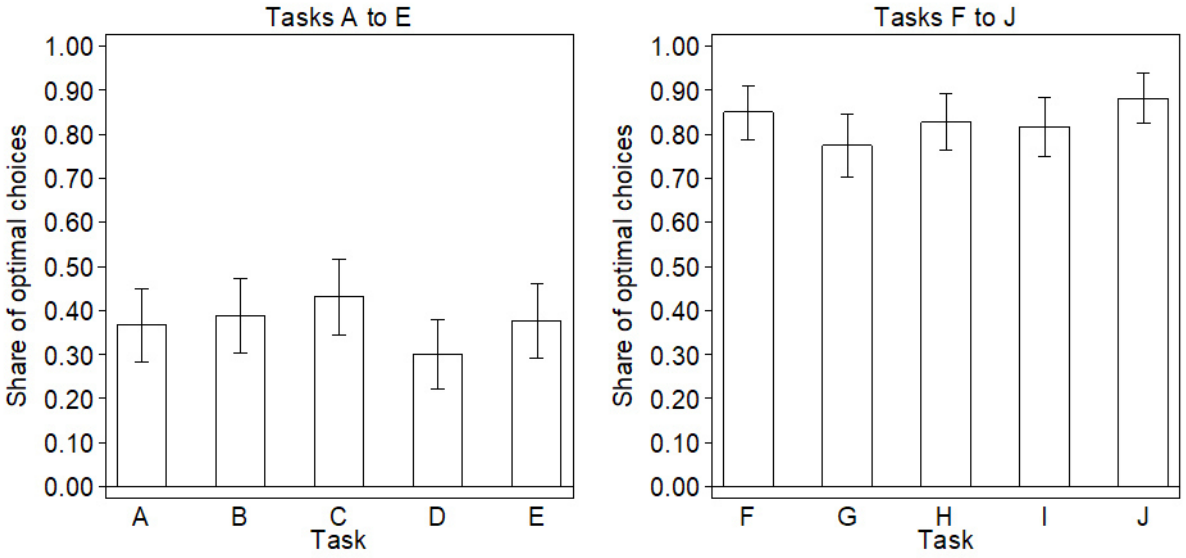
[18] With 14% the share of inconsistent decisions is somewhat larger than the 6% observed by Karle et al. (2015). Loss aversion of our agents does not differ significantly from their sample ($p = 0.174$, two-sided Mann-Whitney-U test) but agents in our experiment are significantly less risk averse ($p < 0.001$).

After excluding the inconsistent agents, we are left with 32 agents in **LOW**, 28 in **HIGH** and 33 in **UNEQUAL** (and the equal number of peers). Comparison tests based on the remaining sample suggest that allocation to treatments is still balanced between treatments with respect to gender ($p \geq 0.613$, Fisher's exact tests). Also, there are no significant differences in age ($p \geq 0.315$, two-sided Mann-Whitney-U tests), risk aversion ($p \geq 0.896$) or loss aversion ($p \geq 0.661$) between treatments.

Behavior across tasks

When deriving the competing hypotheses in Section 3, we assumed that agents optimally choose their cognitive effort when selecting a decision sequence. Depending on the parameters of the task, the optimal sequence can coincide or diverge from the sequence suggested by the directed cognition algorithm. In tasks A to E, the optimal sequence diverges. It demands agents to explore the low-probability, high-payoff project first, even though its expected payoff is lower than that of the high-probability, low-payoff project. In tasks F to J, both sequences coincide, i.e., the optimal sequence in these tasks is characterized by choosing projects in the order of decreasing expected payoffs.

Figure 2: Share of optimal choices across tasks for (all treatments pooled, 95% confidence intervals)



The findings by Gabaix et al. (2006) suggest that many people use the directed cognition algorithm as a heuristic. Aggregating the data from our three treatments, we observe a similar pattern for decisions in social settings, as shown in Figure 2: In tasks A to E, only 37% of the agents choose

the optimal path, while 55% follow the directed cognition algorithm. In tasks F to J, 83% choose the optimal path, but both strategies cannot be distinguished. The share of optimal play is significantly larger in tasks F to J than in tasks A to E ($p < 0.001$, two-sided Wilcoxon signed-rank test). This is summarized in our first observation:

Observation 1: *The share of optimal behavior is significantly larger in tasks F to J than in tasks A to E.*

That means the share of optimal information acquisition behavior is significantly larger in tasks that can be solved optimally by the directed cognition algorithm. When deriving our hypotheses, we assumed that increasing expected payoff requires additional cognitive effort, which is associated with higher costs. One way to interpret the difference in performance between both kinds of tasks is that increasing expected payoff is considerably cheaper when the directed cognition algorithm is sufficient to identify the optimal path. Thus, in the following we would expect larger treatment differences in tasks A to E than in tasks E to J. In the former set of tasks agents need more cognitive effort to identify the optimal path than in the latter set of tasks in which performance is quite close to the optimum already.

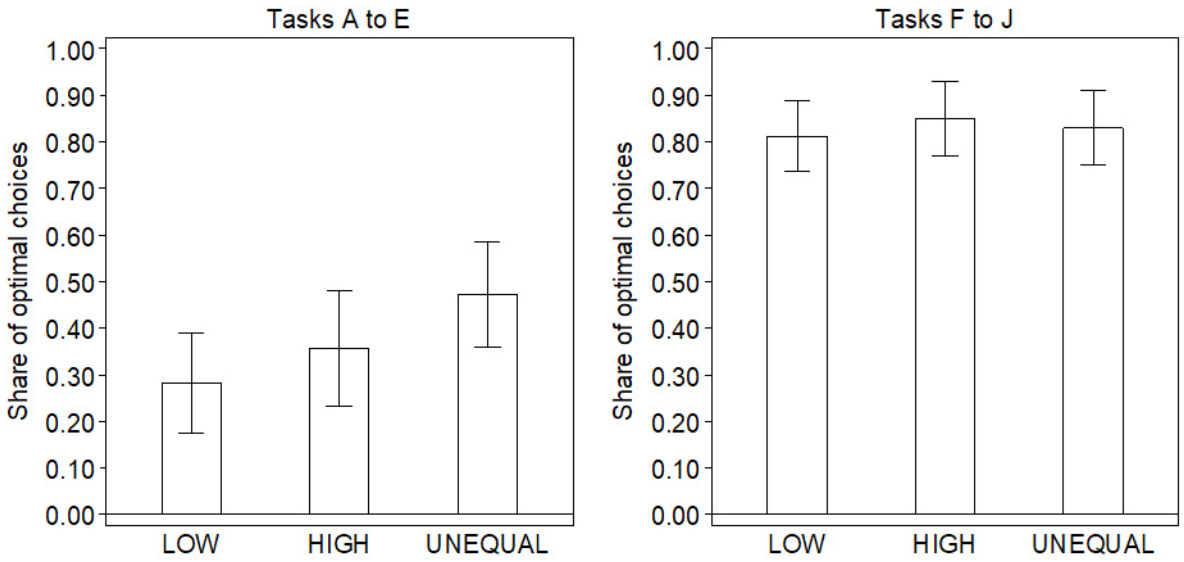
Treatment comparisons

Considering the behavior separately by treatments reveals that the differences between the two types of tasks is present within treatments as well. Figure 3 displays the share of optimal choices averaged across tasks separately by treatment. Average performance over tasks F to J ranges from 81.3% to 85% while the average performance over tasks A to E ranges from 28.1% to 47.3% only. The tasks A to E (which cannot be solved optimally by the directed cognition algorithm) are played optimally less often than tasks F to J within each of the treatments ($p < 0.001$, two-sided Wilcoxon signed-rank tests). As expected, performance also varies more strongly across treatments in tasks A to E than in tasks F to J.

Comparing performance across the three treatments allows us to assess the influence of different wage schedules on search performance. When comparing the share of optimal choices, we do not find any significant differences with respect to tasks F to J ($p \geq 0.381$, two-sided Mann-Whitney-*U* tests). When comparing performance in tasks A to E the difference between the **UNEQUAL**

treatment (in which agents perform best) and the **LOW** treatment (in which they perform worst) is significant ($p = 0.035$) while the other two treatment comparisons are insignificant ($p \geq 0.204$).[19]

Figure 3: Share of optimal choices across tasks by treatment (95% confidence intervals)



As a robustness check, we run the ordinary least squares (OLS) regressions shown in Table 3. They take the share of tasks an agent played optimally as a dependent variable and are run separately for tasks A to E and tasks F to J. Depending on the model specification, they also take the agent's gender and age as explanatory variables next to the treatment dummies **LOW** and **HIGH** (with **UNEQUAL** as the baseline category). The regression results are in line with the pattern described above: While there are significant treatment differences in all models for tasks A to E, we find no significant differences in tasks F to J across regression models (1), (2) and (3) ($p \geq 0.560$, two-sided Wald tests). With respect to tasks A to E, the OLS regressions confirm that performance in **LOW** is worse than in the **UNEQUAL** treatment ($p \leq 0.041$).[20] Based on the estimates the share of

[19] When considering optimal choices based on individual risk preferences, performance is still a lot better in the **UNEQUAL** treatment than in the **LOW** treatment (42.4% versus 27.5% of choices are optimal). The two-sided comparison test does not pass the 5% or 10% significant threshold anymore ($p = 0.125$). But we can still reject the directed hypothesis that unequal wages have a detrimental effect on performance at the 10% level ($p = 0.063$).

[20] Again, when taking individual risk preferences into account, only the directed hypothesis that inequality in wages has a detrimental effect on performance can be rejected across the three regression models ($p \leq 0.060$).

optimal choices is around 19% lower in **LOW**. There is no significant difference between **UNEQUAL** and **HIGH** or between **LOW** and **HIGH** ($p \geq 0.245$).[21]

Table 3: OLS regressions, dependent variable: Share of optimal choices

	Tasks A to E			Tasks F to J		
	(1)	(2)	(3)	(1)	(2)	(3)
LOW	-0.191**	-0.193**	-0.193**	-0.018	-0.016	-0.017
	(0.092)	(0.093)	(0.093)	(0.065)	(0.063)	(0.063)
HIGH	-0.116	-0.111	-0.111	0.020	0.012	0.010
	(0.099)	(0.098)	(0.099)	(0.067)	(0.066)	(0.066)
Female		0.089	0.090		-0.134**	-0.130**
		(0.078)	(0.082)		(0.053)	(0.053)
Age			0.001			0.003
			(0.012)			(0.008)
N	93	93	93	93	93	93
R^2	0.045	0.059	0.059	0.004	0.072	0.073

Standard errors in parentheses, * $p<0.100$, ** $p<0.050$, *** $p<0.010$.

The comparison tests above do not condition on the individual loss aversion. Thus, they are only informative with respect to our Hypotheses 1a, 1b, 2a and 2b. The first two hypotheses are standard hypothesis based on purely self-interested decision-makers. The latter two hypotheses are based on inequality aversion. According to Hypothesis 1a an increase in wages leads to an increase in performance. This prediction coincides with the prediction made under inequality aversion formulated in Hypothesis 2a. With respect to an increase in wages neither of the hypotheses can be rejected: We observe no significant differences when comparing treatments **LOW** and **UNEQUAL** to **HIGH**.

With respect to inequality in wages the predictions of the standard model and the inequality aversion model differ. According to Hypothesis 1b, a self-interested agent is not influenced by peer wages. According to Hypothesis 2b, however, an inequality averse agent will perform worse when

[21] With respect to tasks F to J there appears to be a gender effect. Female agents are slightly less likely to solve these tasks optimally than male agents in models (2) and (3) ($p \leq 0.016$). However, this does not affect our results with respect to insignificant differences between our treatments.

disadvantageous inequality in payoffs increases (ceteris paribus). But we observe better performance in **UNEQUAL** than in **LOW**: Performance increases with peer wages which leads us to reject Hypothesis 2b. Accordingly, we formulate the following observation:

Observation 2: *Unilaterally increasing the pay of an agent's peer does not have a negative effect on the agent's performance.*

Considering agents' loss aversion

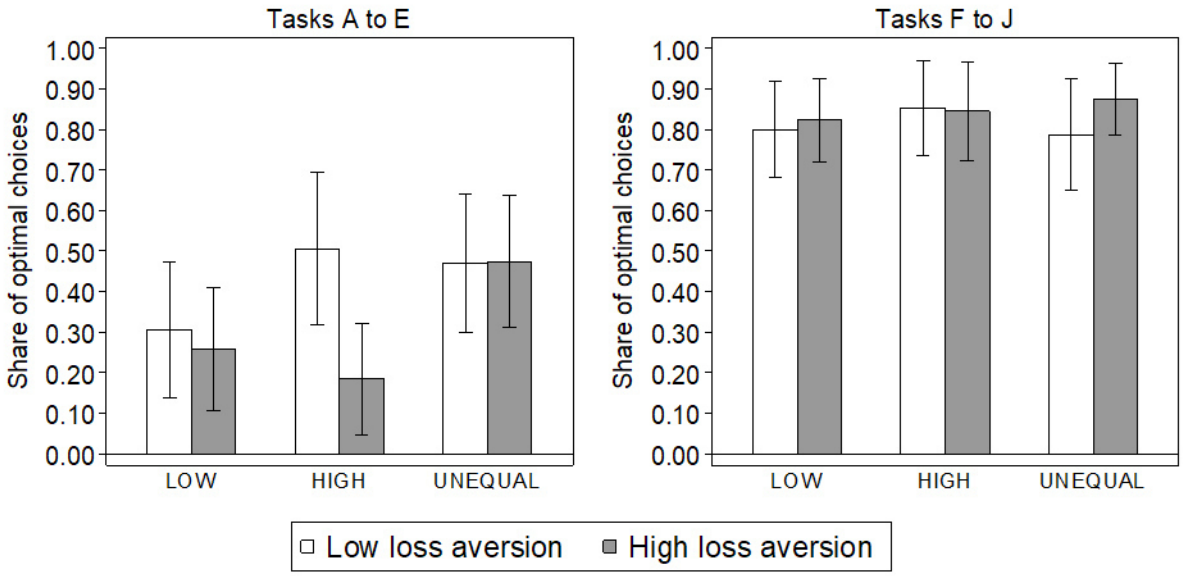
Agents' reactions to varying incentives might be driven by individual differences. As outlined in Section 3, predictions for reference-dependent preferences depend on the degree of an agent's loss aversion. Agents with low loss aversion will react to changes in wages in the same way as described by the standard model based on self-interested agents, i.e. higher wages will lead them to acquire information optimally more often. Yet, agents with high loss aversion may exhibit the opposite pattern: An increase in wages may lead them to choose optimally less often. This relationship is formulated in Hypothesis 3a. With respect to inequality in wages, we would expect no effect of a change in peer wages on performance. This is formulated in Hypothesis 3b (which is the same as Hypothesis 1b of the standard model).

In order to test the predictions of the model with reference-dependent preferences we measured loss aversion in a separate task. Of course, the agents' effort cost functions are unknown to us. Thus, we cannot pin down the individuals for which we expect a negative influence of higher wages on performance. Nevertheless, we can make a prediction about aggregate behavior: The negative relationship between an agent's wage and his performance should be observed more often for agents with relatively large loss aversion than for agents with low loss aversion. In the following we therefore use our measurements of loss aversion to divide our agents into "high" and "low" loss aversion agents by a median split.[22] Comparing both groups with respect to gender composition yields no significant difference ($p = 0.535$, Fisher's exact test) but subjects with high loss aversion are on average 1.5 years older ($p = 0.062$, two-sided Mann-Whitney-U test).

[22] Our loss aversion median equals 1.67 which is slightly lower than the median of 1.80 observed by Karle et al. (2015).

Figure 4 displays the shares of optimal choice sequences in both types of tasks. It compares performance for high and low loss aversion agents between treatments. We do not find any significant treatment differences in tasks F to J ($p \geq 0.481$, two-sided Mann-Whitney-*U* tests). In the tasks A to E (in which identifying the optimal path is more demanding) high wages appear to increase performance by agents with low loss aversion but not by agents with high loss aversion. A unilateral increase in peer wages, however, appears to increase performance for both types of agents. The comparison tests support this interpretation with respect to the high loss aversion agents: Agents solve 47.5% of the tasks optimally in **UNEQUAL** and only 25.9% in **LOW** and 18.5% in **HIGH**, which means that performance in **UNEQUAL** is significantly better than in **HIGH** ($p = 0.017$) and weakly significantly better than in **LOW** ($p = 0.058$). Performance does not differ significantly between **HIGH** and **LOW** ($p = 0.610$). With respect to low loss aversion agents, however, the improvements from 30.7% in **LOW** to 50.7% in **HIGH** or 47.1% in **UNEQUAL** are not significant ($p \geq 0.170$).

Figure 4: Share of optimal choices across tasks by treatment for agents with high and low loss aversion (based on median split, 95% confidence intervals)



In Tables 4a and 4b we present additional ordinary least squares (OLS) regressions, which allow us to control for the small difference in age between the high and the low loss aversion group. Again, we consider the influences on performance separately for tasks A to E and tasks F to J but split the sample in high and low loss aversion agents. For low loss aversion agents, we do not observe any significant treatment differences ($p \geq 0.169$, two-sided Wald tests).[23] For high loss aversion agents, we find performance to be significantly worse in **HIGH** than in **UNEQUAL** in all three models ($p \leq 0.021$). The difference between **LOW** and **UNEQUAL** is weakly significant in model (1) ($p = 0.097$).[24] A one-sided test of the hypothesis that unequal pay is detrimental to performance is rejected in any case ($p \leq 0.059$).[25]

[23] However, we find that female agents are slightly less likely to solve tasks F to J optimally than male agents ($p \leq 0.065$).

[24] Please note that this observation is not robust if we add gender in models (2) and (3) ($p \geq 0.111$).

[25] In addition, we find no gender differences for high loss aversion agents ($p \geq 0.127$). Thus, the gender effect observed in tasks F to J appears to be driven by agents with low loss aversion.

Table 4a: OLS regressions for low loss aversion (based on median split), dependent variable: Share of optimal choices

	Tasks A to E			Tasks F to J		
	(1)	(2)	(3)	(1)	(2)	(3)
LOW	-0.164	-0.154	-0.171	0.012	0.001	-0.015
	(0.137)	(0.136)	(0.136)	(0.104)	(0.101)	(0.099)
HIGH	0.036	0.035	0.019	0.065	0.066	0.049
	(0.145)	(0.143)	(0.152)	(0.103)	(0.100)	(0.097)
Female		0.151	0.165		-0.164**	-0.149*
		(0.115)	(0.120)		(0.080)	(0.078)
Age			0.011			0.012
			(0.019)			(0.011)
N	47	47	47	47	47	47
R^2	0.047	0.084	0.091	0.010	0.097	0.114

Standard errors in parentheses, * $p<0.100$, ** $p<0.050$, *** $p<0.010$.

Table 4b: OLS regressions for high loss aversion (based on median split), dependent variable: Share of optimal choices

	Tasks A to E			Tasks F to J		
	(1)	(2)	(3)	(1)	(2)	(3)
LOW	-0.216*	-0.215	-0.217	-0.051	-0.042	-0.049
	(0.128)	(0.132)	(0.136)	(0.078)	(0.077)	(0.074)
HIGH	-0.290**	-0.293**	-0.293**	-0.029	-0.042	-0.043
	(0.120)	(0.122)	(0.123)	(0.085)	(0.087)	(0.087)
Female		-0.018	-0.021		-0.102	-0.111
		(0.108)	(0.111)		(0.070)	(0.072)
Age			-0.003			-0.007
			(0.016)			(0.010)
N	46	46	46	46	46	46
R^2	0.119	0.120	0.121	0.009	0.059	0.070

Standard errors in parentheses, * $p<0.100$, ** $p<0.050$, *** $p<0.010$.

As we do not find any significant differences in performance between treatments for low loss aversion agents, we cannot reject any of our hypotheses with respect to this subsample. Performance is somewhat better in **HIGH** than in **LOW** which is in line with the predictions of our models of selfish and inequality averse behavior. It is also in line with reference-dependent preferences as long as loss aversion is sufficiently low. The high-level performance in **UNEQUAL** is not predicted by any of the models. However, as performance in **UNEQUAL** is not significantly different from **HIGH** or **LOW**, the respective hypotheses are not rejected.

With respect to the subsample of high loss aversion agents we can reject Hypotheses 1a and 2a which both predict better performance with higher pay. Yet, agents in **UNEQUAL** outperform agents in **HIGH** even though they are paid less in **UNEQUAL** than in **HIGH**. This is in line with Hypothesis 3a based on reference-dependent preferences. Note, however, that reference-dependent preferences based on expected individual payoffs also suggest the same performance in **UNEQUAL** and in **LOW** as stated in Hypothesis 3b. The same prediction is made for a rational self-interested agent in Hypothesis 1b. Both hypotheses are rejected at a weak significance level[26] in our experiment. We do not find any evidence for a detrimental effect of inequality in pay. In fact, Hypothesis 2b, which predicts worse performance in **UNEQUAL** than in **HIGH** and in **LOW,** is rejected in all tests. We summarize our results in the following observation:

Observation 3: *For a high loss aversion agent unilaterally increasing the pay of his peer leads to better performance than increasing the pay of both (the agent and his peer).*

[26] Without controlling for gender differences.

5 Conclusion

In research and development processes people have to make a series of decisions under uncertainty, for example, when considering new production technologies. As described in Weitzman's (1979) example they can sequentially develop these technologies to resolve uncertainty about the rewards they deliver. Yet, people typically make these decisions in social settings. Within an organization their peers will also profit from their effort financially. However, even if incentives are perfectly aligned between the members of an organization, pay structures are often very heterogeneous. We ask whether disadvantageous inequality in wages may be detrimental to performance in this type of setting. To answer our question, we created an experimental design to identify the causal influence of wage inequality on information acquisition performance and to compare behavior to a normative benchmark.

We focus on a setting where one agent is responsible for decision-making. The agent learns about a peer who profits from the decisions the agent makes. First, we confirm the result by Gabaix et al. (2006), who find that their directed cognition algorithm describes actual behavior quite well. We find a similar decision pattern for agents acting in a social setting. In addition, we observe that the pay structure influences the quality of decisions in tasks that cannot be easily solved optimally by the directed cognition algorithm. Our results seem to suggest that raising the peer's wage to a level above the agent's wage is more effective in improving performance than raising the peer's *and* the agent's wage to that level. More precisely, we observe that disadvantageous inequality in wages does not have a negative effect on performance. Considering individual loss aversion of agents, we find that a high loss aversion agent does not improve his performance when he is paid more. Increasing the pay of his peer, however, leads to a better performance than increasing the pay of both (the agent and his peer).

The observation that an increase in wages does not necessarily lead to more effort is in line with the results of several empirical studies that consider individual-effort provision in real-effort settings. For example, laboratory experiments by Gneezy & Rustichini (2000), Pokorny (2008) and Ariely et al. (2009) have reported that larger incentives may lead to lower effort. With respect to field data, Camerer et al. (1997) conducted probably the most prominent study. They study the behavior of cab drivers in New York City and observe that cab drivers work less on days with higher wages. Camerer et al. argue that cab drivers' effort is determined by reference-dependent

preferences, depending on some daily income target. Farber (2005, 2008) has criticized this interpretation, but results by Crawford & Meng (2011) confirm the original interpretation. In a field experiment with bike messengers, Fehr & Goette (2007) exogenously vary their individual commission rate and find a similar response.

However, none of the three models we consider fully captures the effects we observe. The reference-dependent preference model is consistent with the comparative statics of a wage increase based on an individual's loss aversion. However, it does not explain the high performance of high and low loss aversion agents in the **UNEQUAL** treatment. Why do low loss aversion agents perform on the same level in **HIGH** and **UNEQUAL**? Our results suggest that agents treat expectation-based reference points differently from social reference points from unequal pay. Loss aversion with respect to others' pay could explain an increase in performance in **UNEQUAL** relative to **LOW**. However, previous evidence for this increase is rather weak: Linde & Sonnemans (2012) observe more (and not less) risk aversion when people choose over lotteries that pay them at most as much as a peer with a fixed payoff.

References

Ariely, D., Gneezy, U., Loewenstein, G. & Mazar, N. (2009): Large stakes and big mistakes, *Review of Economic Studies*, 76(2), 451-469.

Arrow, K. J. (1969): Classificatory notes on the production and transmission of technological knowledge, *American Economic Review*, 59(2), 29-35.

Bartling, B. & von Siemens, F. A. (2011): Wage inequality and team production: An experimental analysis, *Journal of Economic Psychology*, 32(1), 1-16.

Bergemann, D. & Välimäki, J. (2008): Bandit problems, *The New Palgrave Dictionary of Economics*, 2nd ed., Macmillan Press.

Bewley, T. F. (1999): Why wages don't fall during a recession, Harvard University Press.

Bolton, P. & Faure-Grimaud, A. (2009): Thinking ahead: The decision problem, *Review of Economic Studies*, 76(4), 1205-1238.

Boyce, J. R., Bruner, D. M. & McKee, M. (2016): Strategic experimentation in the lab. *Managerial and Decision Economics*, 37(6), 375-391.

Bracha, A., Gneezy, U. & Loewenstein, G. (2015): Relative pay and labor supply, *Journal of Labor Economics*, 33(2), 297-315.

Breza, E., Kaur, S. & Shamdasani, Y. (2015): The morale effects of pay inequality, *Working paper*.

Brüggemann, J. & Bizer, K. (2016): Laboratory experiments in innovation research: A methodological overview and a review of the current literature, *Journal of Innovation and Entrepreneurship*, 5(1), 1-13.

Camerer, C., Babcock, L., Loewenstein, G. & Thaler, R. (1997): Labor supply of New York City cabdrivers: One day at a time, *Quarterly Journal of Economics*, 112(2), 407-441.

Caplin, A., Dean, M. & Martin, D. (2011): Search and satisficing, *American Economic Review*, 101(7), 2899-2922.

Card, D., Mas, A., Moretti, E. & Saez, E. (2012): Inequality at work: The effect of peer salaries on job satisfaction, *American Economic Review*, 102(6), 2981-3003.

Chade, H., Eeckhout, J. & Smith, L. (2017): Sorting through search and matching models in economics, *Journal of Economic Literature*, 55(2), 493-544.

Charness, G. & Kuhn, P. (2011): Lab labor: What can labor economists learn from the lab?. In: Card, D. & Ashenfelter, O. (eds.): *Handbook of Labor Economics*, Vol. 4, 229-330.

Cohn, A., Fehr, E., Herrmann, B. & Schneider, F. (2014): Social comparison and effort provi-sion: Evidence from a field experiment, *Journal of the European Economic Association*, 12(4), 877-898.

Conlisk, J. (1996): Why bounded rationality?, *Journal of Economic Literature*, 34(2), 669-700.

Crawford, V. P. & Meng, J. (2011): New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational expectations targets for hours and income, *American Economic Review*, 101(5), 1912-1932.

DellaVigna, S., List, J. A., Malmendier, U. & Rao, G. (2016): Estimating social preferences and gift exchange at work, *Working paper*.

Downes, P. E. & Choi, D. (2014): Employee reactions to pay dispersion: A typology of existing research, *Human Resource Management Review*, 24(1), 53-66.

Eckartz, K., Kirchkamp, O. & Schunk, D. (2012): How do incentives affect creativity?, *Working paper*.

Ederer, F. (2013): Incentives for parallel innovation, *Working paper.*

Ederer, F. & Manso, G. (2013): Is pay for performance detrimental to innovation?, *Management Science*, 59(7), 1496-1513.

Farber, H. S. (2005): Is tomorrow another day? The labor supply of New York City cabdrivers, *Journal of Political Economy*, 113(1), 46-82.

Farber, H. S. (2008): Reference-dependent preferences and labor supply: The case of New York City taxi drivers, *American Economic Review*, 98(3), 1069-1082.

Fehr, E. & Goette, L. (2007): Do workers work more if wages are high? Evidence from a ran-domized field experiment, *American Economic Review*, 97(1), 298-317.

Fehr, E. & Schmidt, K. M. (1999): A theory of fairness, competition, and cooperation, *Quarterly Journal of Economics*, 114(3), 817-868.

Fischbacher, U. (2007): z-Tree: Zurich toolbox for ready-made economic experiments, *Experimental economics*, 10(2), 171-178.

Frank, R. H. (1984): Are workers paid their marginal products?, *American Economic Review*, 74(4), 549-571.

Gabaix, X., Laibson, D., Moloche, G. & Weinberg, S. (2006): Costly information acquisition: Experimental analysis of a boundedly rational model, *American Economic Review*, 96(4), 1043-1068.

Gneezy, U. & Rustichini, A. (2000): Pay enough or don't pay at all, *Quarterly Journal of Economics*, 115(3), 791-810.

Greiner, B., Ockenfels, A. & Werner, P. (2011): Wage transparency and performance: A real-effort experiment, *Economics Letters*, 111(3), 236-238.

Herbst, D., & Mas, A. (2015): Peer effects on worker output in the laboratory generalize to the field, *Science*, 350(6260), 545-549.

Herz, H., Schunk, D. & Zehnder, C. (2014): How do judgmental overconfidence and overoptimism shape innovative activity?, *Games and Economic Behavior*, 83, 1-23.

Kahneman, D. & Tversky, A. (1979): Prospect theory: An analysis of decision under risk, *Econometrica*, 47(2), 263-291.

Karle, H., Kirchsteiger, G. & Peitz, M. (2015): Loss aversion and consumption choice: Theory and experimental evidence. *American Economic Journal: Microeconomics*, 7(2), 101-120.

Kőszegi, B. & Rabin, M. (2007): Reference-dependent risk attitudes, *American Economic Review*, 97(4), 1047-1073.

Kőszegi, B. & Rabin, M. (2006): A model of reference-dependent preferences, *Quarterly Journal of Economics*, 121(4), 1133-1165.

Lazear, E. P. (1989): Pay equality and industrial politics, *Journal of Political Economy*, 97(3), 561-580.

Linde, J. & Sonnemans, J. (2012): Social comparison and risky choices. *Journal of Risk and Uncertainty*, 44(1), 45-72.

Lipman, B. L. (1999): Decision theory without logical omniscience: Toward an axiomatic framework for bounded rationality, *Review of Economic Studies*, 66(2), 339-361.

Lipman, B. L. (1991): How to decide how to decide how to...: Modeling limited rationality, *Econometrica*, 59(4), 1105-1125.

List, J. A. & Cherry, T. L. (2000): Learning to accept in ultimatum games: Evidence from an experimental design that generates low offers, *Experimental Economics*, 3(1), 11-29.

March, J. G. (1991): Exploration and exploitation in organizational learning, *Organization Science*, 2(1), 71-87.

McGee, P. (2014). Asymmetric consumer search and reference prices, *Working paper*.

Mohnen, A. & Ostermaier, A. (2013): Incentives for creativity: Limits of objective performance evaluation, *Working paper*.

Pfeffer, J. (2007): Human resources from an organizational behavior perspective: Some paradoxes explained, *Journal of Economic Perspectives*, 21(4), 115-134.

Pfeffer, J. & Davis-Blake, A. (1990): Determinants of salary dispersion in organizations, *Industrial Relations*, 29(1), 38-57.

Pfeffer, J. & Langton, N. (1993): The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty, *Administrative Science Quarterly*, 382-407.

Poblete, J. & Spulber, D. (2017): Managing innovation: Optimal incentive contracts for delegated R&D with double moral hazard, *European Economic Review*, 95, 38-61.

Pokorny, K. (2008): Pay - but do not pay too much: An experimental study on the impact of incentives, *Journal of Economic Behavior & Organization*, 66(2), 251-264.

Reutskaja, E., Nagel, R., Camerer, C. F. & Rangel, A. (2011): Search dynamics in consumer choice under time pressure: An eye-tracking study, *American Economic Review*, 101(2), 900-926.

Schunk, D. & Winter, J. (2009): The relationship between risk attitudes and heuristics in search tasks: A laboratory experiment, *Journal of Economic Behavior & Organization*, 71(2), 347-360.

Schunk, D. (2009): Behavioral heterogeneity in dynamic search situations: Theory and experimental evidence, *Journal of Economic Dynamics and Control*, 33(9), 1719-1738.

Shaw, J. D. (2014): Pay dispersion, *Annual Review of Organizational Psychology and Organizational Behavior*, 1(1), 521-544.

Simon, H. A. (1955): A behavioral model of rational choice, *Quarterly Journal of Economics*, 69(1), 99-118.

Slonim, R. & Roth, A. E. (1998): Learning in high stakes ultimatum games: An experiment in the Slovak Republic, *Econometrica*, 66(3), 569-596.

Slonim, R. (1994): Learning in a search-for-the-best-alternative experiment. *Journal of Economic Behavior & Organization*, 25(2), 141-165.

Soetevent, A. R. & Bružikas, T. (2016): Risk and loss aversion, price uncertainty and the implications for consumer search, *Working paper.*

Tversky, A. & Kahneman, D. (1992): Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4), 297-323.

Wang, T., Zhao, B. & Thornhill, S. (2015): Pay dispersion and organizational innovation: The mediation effects of employee participation and voluntary turnover, *Human Relations*, 68(7), 1155-1181.

Weitzman, M. L. (1979): Optimal search for the best alternative, *Econometrica*, 47(3), 641-654.

Yanadori, Y. & Cui, V. (2013): Creating incentives for innovation? The relationship between pay dispersion in R&D groups and firm innovation performance, *Strategic Management Journal*, 34(12), 1502-1511.

Appendix A – Parameters of the tasks

Table B1: Task parameters based on Gabaix et al. (2006)

	V_1	p_1	V_2	p_2	V_3	p_3
Task A	21	0.09	10	0.76	1	1
Task B	19	0.11	10	0.79	1	1
Task C	23	0.09	13	0.72	1	1
Task D	18	0.12	10	0.81	1	1
Task E	20	0.12	12	0.85	1	1
Task F	22	0.48	11	0.74	1	1
Task G	24	0.34	9	0.7	1	1
Task H	18	0.52	11	0.74	1	1
Task I	25	0.39	9	0.7	1	1
Task J	10	0.09	8	0.85	1	1

Note: Costs c_i are always 1.

Appendix B – Instructions (translated from German)

Welcome to the experiment!

Foreword

You are taking part in a study of decision behavior in experimental economic research. During the study you and the other participants are asked to make decisions. By doing so you can earn money. How much money you earn, will depend on your decisions. At the end of the experiment you will be paid your total earnings in cash.

None of the participants will be informed about the identity of the other participants during the experiment.

Instructions

Please read the following instructions carefully. About 5 minutes after we hand out the instructions, we will come to you to answer questions. If you have any questions during the experiment, you can raise your hand at any time. We will then come to you.

During the experiment you will take part in 10 lottery rounds.

Assignment of partner and role

At the beginning of the experiment you will get assigned to your partner and randomly receive one of two roles: Participant 1 or Participant 2. During the whole experiment you will play together with the same partner and keep the same role. You will not learn anything about the identity of your partner.

Description of the lottery games

During the lottery rounds of the experiment you will be asked to choose among a group of "lottery tickets". Tickets have different probabilities of paying off. All amounts are given in Experimental Currency (EC).

Participant 1 enters the decisions for himself and Participant 2. Participant 2 only sees these decisions.

Consider the following set of tickets:

Ticket 1:	possible amount (in EC): 20	probability (in %): 50	winner: unknown	actual amount (in EC): unknown	Find Out Whether Ticket 1 is a Winner:	**Find Out**
Ticket 2:	possible amount (in EC): 10	probability (in %): 75	winner: unknown	actual amount (in EC): unknown	Find Out Whether Ticket 2 is a Winner:	**Find Out**
Ticket 3:	possible amount (in EC): 1	probability (in %):100	winner: yes	actual amount (in EC): 1	Take Ticket 3:	**Take**

Here, ticket 1 has a 50 % chance of paying 20 EC. To be more precise, it pays off 20 EC with a 50 % chance and pays off 0 EC with a 50 % chance. Similarly, ticket 2 pays off 10 EC with a 75 % chance and 0 EC with a 25 % chance. Ticket 3 always pays 1 EC.

For 1 EC you can find out whether an "unknown" ticket is a winner or not by clicking on "Find Out" next to "Find Out Whether Ticket is a Winner" in the far right hand column. You may investigate as many tickets as you wish, at a cost of 1 EC each.

At any point you may stop learning and end the round by choosing ONE ticket. In order to do so Participant 1 clicks on "Take Ticket" in the far right hand column. The "Take Ticket" option will be available for all tickets that are known to be a winner or not. Even for those that have been revealed as losing tickets. Once you select a ticket, we will calculate your winnings for that round as the value of the selected ticket minus

1 EC for each "unknown" ticket that you investigated. We will tell you your profit for the round and then reset the screen for the next round.

Payment

At the end of the experiment you will get paid your winnings (net of learning costs) in 1 of the 10 rounds selected at random. [**LOW:** Participant 1 and Participant 2 receive a malus of 30 % in addition to the profit of this round. I.e. their profit is reduced by 30 %. **HIGH:** Participant 1 and Participant 2 receive a bonus of 30 % in addition to the profit of this round. I.e. their profit is increased by 30 %. **UNEQUAL:** Participant 1 receives a malus of 30 % in addition to the profit of this round, while Participant 2 receives a bonus of 30 %. I.e. their profit is decreased by 30 % or increased by 30 %.]

1 EC is equivalent to 1.50 Euro.

Example: *Assuming Participant 1 and Participant 2 have investigated 2 tickets in one round and take a ticket with an actual amount of 10 EC. In this case the profit for this round is for each of the both participants:*

Profit for this round = 10 EC – 2 EC = 8 EC

[**LOW:**

*However, Participant 1 and Participant 2 receive a malus in addition to the profit. The malus is calculated as the absolute value of 30 % of the profit for this round, in this round 8 EC * 30 % = 2.4 EC. That is, the actual profits that are paid are:*

Profit for Participant 1 = 8 EC – 2.4 EC = 5.6 EC

Profit for Participant 2 = 8 EC – 2.4 EC = 5.6 EC

]

[**HIGH:**

*However, Participant 1 and Participant 2 receive a bonus in addition to the profit. The bonus is calculated as the absolute value of 30 % of the profit for this round, in this round 8 EC * 30 % = 2.4 EC. That is, the actual profits that are paid are:*

Profit for Participant 1 = 8 EC + 2.4 EC = 10.4 EC

Profit for Participant 2 = 8 EC + 2.4 EC = 10.4 EC

]

[**UNEQUAL:**

*However, Participant 1 receives a malus in addition to the profit and Participant 2 a bonus. Malus and bonus are calculated as the absolute value of 30 % of the profit for this round, in this round 8 EC * 30% = 2.4 EC. That is, the actual profits that are paid are:*

Profit for Participant 1 = 8 EC – 2.4 EC = 5.6 EC

Profit for Participant 2 = 8 EC + 2.4 EC = 10.4 EC

]

You will start with a practice round that illustrates the process. Then the experiment will begin. The experiment will consist of 10 rounds (in addition to the practice round).

On the next page you will find the screens for one exemplary round.

Screen for Participant 1

remaining time [sec]: 1529

Round: 1

You are Participant 1. You enter the decisions for yourself and Participant 2. Participant 2 sees these decisions.

Ticket 1:	possible amount (in EC):	20	probability (in %):	50	winner:	unknown	actual amount (in EC):	unknown	Find Out Whether Ticket 1 is a Winner:	Find Out
Ticket 2:	possible amount (in EC):	10	probability (in %):	75	winner:	unknown	actual amount (in EC):	unknown	Find Out Whether Ticket 2 is a Winner:	Find Out
Ticket 3:	possible amount (in EC):	1	probability (in %):	100	winner:	yes	actual amount (in EC):	1	Take Ticket 3:	Take Ticket

Clicking on "Find Out" next to "Find Out Whether Ticket x is a Winner" costs 1 EC and shows if the ticket is a winner or not.

Clicking on "Take Ticket" next to "Take Ticket x" ends the round and calculates the profit for this round for you and Participant 2.

Screen for Participant 2

remaining time [sec]: 1541

Round: 1

You are Participant 2. Participant 1 enters the decisions for both of you. You see these decisions.

Ticket 1:	possible amount (in EC):	20	probability (in %):	50	winner:	unknown	actual amount (in EC):	unknown
Ticket 2:	possible amount (in EC):	10	probability (in %):	75	winner:	unknown	actual amount (in EC):	unknown
Ticket 3:	possible amount (in EC):	1	probability (in %):	100	winner:	yes	actual amount (in EC):	1

Participant 1 sees in the far right hand column for each ticket the button "Find Out" for tickets that are unknown to be a winner or the button "Take Ticket" for tickets that are known to be a winner or not.

Clicking on "Find Out" costs 1 EC and shows if the ticket is a winner or not.

Clicking on "Take Ticket" ends the round and calculates the profit for this round for you and Participant 1.

Lottery decisions

In the following we ask you to make 12 lottery decisions for which you will receive additional earnings.

Please read the following instructions carefully. About 5 minutes after we hand out the instructions, we will come to you to answer any questions. If you have any questions during the experiment, you can raise your hand at any time. We will then come to you.

Description of the lottery decisions

One of the 12 lottery decisions will be randomly selected and the corresponding profit will be added to your previous earnings. Please note that your decisions influence your profit only and not that of the other participant.

During the 12 decisions you can choose between a lottery and a secure (or "safe") payment. You will have an additional 2.00 Euro at your disposal. According to your decisions and the lottery results (which is randomly selected by a computerized random number generator), this amount can be reduced or increased.

Example: *You can choose between a secure payment of 0.00 Euro and a lottery. If you choose the lottery you receive a profit of 1.00 Euro with a probability of 50% and with a probability of 50% a loss of -0.50 Euro. Do you prefer the secure payment or the lottery?*

Payment

At the end of the experiment the profit of the lottery decisions will be paid out to you, together with the previous earnings.

The profit of the lottery decisions is composed of the endowment of 2.00 Euro and – depending on your decision – of the secure payment or the lottery result.

Appendix C – Raw data

Table C1a: Optimal choices

Group	Treatment	Task										Age		Female		Loss aversion		Inconsistent	
		A	B	C	D	E	F	G	H	I	J	P1	P2	P1	P2	P1	P2	P1	P2
1	LOW	0	0	0	0	1	0	1	1	1	1	27	23	0	1	1.243	1.250	0	0
2	LOW	0	0	0	0	0	1	1	1	1	1	28	26	0	1	1.734	0.736	0	0
3	LOW	0	0	0	0	0	0	0	0	0	1	24	20	1	0	1.882	2.500	0	0
4	LOW	1	1	1	1	1	1	1	1	1	1	24	21	0	1	2.561	1.667	0	0
5	LOW	1	1	1	1	1	1	1	1	1	0	21	25	0	0	-	39.433	1	0
6	LOW	0	1	0	0	0	1	1	0	1	1	27	22	1	0	-	1.467	1	0
7	LOW	0	1	0	1	0	1	1	1	1	1	24	27	1	0	2.561	0.583	0	0
8	LOW	0	0	0	0	0	1	0	1	0	1	27	22	1	0	11.373	1.667	0	0
9	LOW	0	0	0	0	0	0	0	0	0	1	23	24	0	1	-	1.250	1	0
10	LOW	1	1	1	1	1	1	1	1	1	0	29	23	0	1	3.244	-	0	1
11	LOW	0	0	0	0	0	1	1	1	1	1	26	24	0	1	1.105	6.089	0	0
12	LOW	0	0	0	0	1	1	1	1	1	1	25	24	0	1	1.281	1.667	0	0
13	LOW	1	1	1	1	1	1	1	1	1	1	28	18	1	0	1.250	3.244	0	0
14	LOW	0	0	0	0	0	0	0	0	0	1	24	22	1	0	1.667	1.105	0	0
15	LOW	1	1	0	0	1	0	0	1	1	1	26	25	1	1	1.667	-	0	1
16	LOW	0	1	0	0	0	1	1	1	0	1	19	24	1	0	1.667	74.160	0	0
17	LOW	1	1	1	1	1	1	1	1	1	1	24	22	0	1	1.667	-	0	1
18	LOW	1	0	0	0	1	1	0	0	0	1	25	24	1	1	1.689	1.667	0	0
19	LOW	1	0	1	0	0	1	1	1	1	1	25	26	0	1	2.500	3.244	0	0
20	LOW	0	0	1	0	0	1	1	1	1	1	25	22	0	1	74.160	-	0	1
21	LOW	0	0	0	0	1	1	1	1	1	1	24	26	1	0	1.734	1.105	0	0
22	LOW	0	0	0	0	0	1	1	1	1	1	22	27	0	0	2.500	1.000	0	0
23	LOW	0	1	1	0	0	1	0	1	1	1	27	24	1	0	-	5.000	1	0
24	LOW	0	0	0	0	0	0	0	1	1	1	22	25	0	0	1.105	1.105	0	0
25	LOW	0	0	0	0	0	1	1	1	1	1	18	24	0	1	1.250	1.000	0	0
26	LOW	0	0	0	0	0	1	1	1	1	1	20	27	0	0	2.854	1.250	0	0
27	LOW	0	1	1	0	0	1	1	1	1	1	27	21	0	1	1.250	1.667	0	0
28	LOW	1	1	1	0	0	1	1	1	1	1	25	20	0	1	1.467	12.500	0	0
29	LOW	0	0	0	0	0	0	1	1	1	1	19	17	1	0	3.244	2.561	0	0
30	LOW	0	0	0	0	0	1	1	0	0	1	19	19	1	0	4.440	0.659	0	0
31	LOW	0	0	0	0	0	1	1	1	1	1	18	22	1	0	1.667	1.281	0	0
32	LOW	0	0	0	0	0	1	1	0	1	1	26	25	1	0	2.500	1.734	0	0
33	LOW	1	1	1	0	1	1	1	1	1	1	23	23	1	0	1.734	3.244	0	0
34	LOW	0	0	0	0	1	1	1	0	1	1	28	26	1	0	1.263	2.075	0	0
35	LOW	1	0	1	0	1	1	1	1	1	0	18	23	1	1	1.105	2.561	0	0
36	LOW	0	0	0	0	0	1	0	1	1	1	22	22	0	1	4.440	-	0	1

Note: Inconsistent indicates subjects that switched multiple times in one of the two lottery decisions.

Table C1b: Optimal choices

Group	Treatment	Task										Age		Female		Loss aversion		Inconsistent	
		A	B	C	D	E	F	G	H	I	J	P1	P2	P1	P2	P1	P2	P1	P2
1	HIGH	0	0	1	0	0	1	0	1	1	1	20	24	1	1	1.667	5.000	0	0
2	HIGH	1	0	0	1	0	1	1	1	1	1	25	25	0	1	1.467	1.243	0	0
3	HIGH	1	1	1	1	1	1	1	1	1	1	22	20	0	1	0.736	1.667	0	0
4	HIGH	0	0	0	0	0	1	1	0	1	0	26	25	0	1	2.561	1.667	0	0
5	HIGH	1	0	1	0	1	1	1	1	0	0	20	26	1	0	2.561	11.373	0	0
6	HIGH	1	0	1	1	1	1	1	1	1	1	23	21	1	0	-	17.869	1	0
7	HIGH	1	1	1	1	1	1	1	1	1	1	23	25	1	0	0.833	1.000	0	0
8	HIGH	1	1	1	1	1	1	1	1	1	1	21	28	1	0	1.467	1.250	0	0
9	HIGH	0	0	0	0	0	1	1	1	1	1	25	20	0	1	1.734	3.244	0	0
10	HIGH	1	0	1	1	0	1	1	1	1	1	26	26	0	1	1.000	-	0	1
11	HIGH	1	1	1	1	1	1	1	1	1	1	20	20	0	1	1.667	1.689	0	0
12	HIGH	1	1	0	1	1	1	1	1	1	1	28	23	0	1	2.561	1.667	0	0
13	HIGH	1	0	1	1	0	1	1	1	1	1	21	24	1	0	-	0.833	1	0
14	HIGH	1	1	1	1	0	1	1	1	1	0	23	31	1	0	1.000	2.500	0	0
15	HIGH	0	0	0	0	0	0	0	0	0	1	23	27	1	0	1.243	2.500	0	0
16	HIGH	0	0	0	0	0	1	1	1	1	1	29	25	0	0	2.500	2.561	0	0
17	HIGH	0	0	0	0	0	1	0	0	0	1	24	31	1	1	0.833	1.667	0	0
18	HIGH	0	0	0	0	1	1	1	1	1	1	27	29	0	1	2.561	2.561	0	0
19	HIGH	0	0	0	0	0	0	0	1	1	1	20	22	0	1	1.000	1.734	0	0
20	HIGH	1	1	0	1	0	1	1	1	1	1	26	27	0	1	-	74.160	1	0
21	HIGH	1	1	1	0	1	1	1	1	1	1	30	26	1	0	1.243	2.075	0	0
22	HIGH	0	0	1	1	0	1	0	0	0	1	26	22	1	0	5.708	-	0	1
23	HIGH	1	1	0	0	1	1	1	1	1	1	21	27	1	0	1.000	2.500	0	0
24	HIGH	0	0	0	0	0	1	1	1	1	1	19	21	1	0	3.244	1.734	0	0
25	HIGH	0	0	0	0	0	1	1	1	1	1	29	22	0	1	1.667	1.467	0	0
26	HIGH	0	0	0	0	1	1	1	1	1	1	27	20	0	1	1.281	1.243	0	0
27	HIGH	0	0	0	0	0	1	1	1	1	1	30	21	0	1	17.869	74.160	0	0
28	HIGH	0	0	0	0	0	1	0	0	0	1	23	29	0	1	74.160	5.708	0	0
29	HIGH	1	0	0	1	0	1	1	1	1	1	26	19	1	0	-	1.250	1	0
30	HIGH	0	0	1	1	1	0	0	0	1	1	23	22	1	0	-	0.736	1	0
31	HIGH	0	1	1	0	1	1	0	0	1	1	20	28	1	0	-	39.433	1	0
32	HIGH	0	0	0	0	0	1	1	1	1	1	25	20	0	0	5.708	2.561	0	0
33	HIGH	1	0	1	0	0	1	0	0	1	1	21	22	0	1	-	1.467	1	0
34	HIGH	0	1	1	0	0	1	1	1	1	1	20	25	1	1	39.433	1.667	0	0
35	HIGH	0	0	0	0	0	1	1	1	1	1	27	29	0	0	39.433	17.869	0	0
36	HIGH	0	0	0	0	0	1	0	0	0	1	19	26	0	0		1.250	1	0

Note: Inconsistent indicates subjects that switched multiple times in one of the two lottery decisions.

Table C1c: Optimal choices

Group	Treatment	Task										Age		Female		Loss aversion		Inconsistent	
		A	B	C	D	E	F	G	H	I	J	P1	P2	P1	P2	P1	P2	P1	P2
1	UNEQUAL	0	0	0	0	0	1	1	1	0	1	25	27	0	1	1.105	1.105	0	0
2	UNEQUAL	0	0	0	0	0	1	1	1	1	1	24	27	0	1	1.250	1.667	0	0
3	UNEQUAL	0	1	1	0	0	0	0	1	0	0	23	23	0	0	0.812	2.561	0	0
4	UNEQUAL	0	0	1	0	0	1	1	1	1	1	27	20	0	1	12.500	2.561	0	0
5	UNEQUAL	0	0	0	1	0	1	1	1	1	1	23	24	1	0	5.000	2.561	0	0
6	UNEQUAL	1	1	1	1	0	1	0	1	1	1	20	20	1	0	1.667	5.000	0	0
7	UNEQUAL	1	1	1	1	1	1	1	1	1	1	21	25	1	0	1.281	1.105	0	0
8	UNEQUAL	1	1	1	1	1	1	1	1	1	1	24	25	1	0	2.075	1.667	0	0
9	UNEQUAL	0	1	1	0	0	1	1	1	1	1	24	24	0	1	0.718	74.160	0	0
10	UNEQUAL	1	1	1	1	1	1	1	1	1	1	27	23	0	0	1.105	5.000	0	0
11	UNEQUAL	0	1	0	1	1	0	0	0	1	0	28	31	0	1	-	-	1	1
12	UNEQUAL	1	1	1	1	1	1	1	1	1	0	20	20	0	1	5.000	1.243	0	0
13	UNEQUAL	0	0	0	0	0	0	0	0	0	0	18	18	1	0	1.263	-	0	1
14	UNEQUAL	0	0	1	1	1	1	1	1	1	0	20	24	1	0	1.105	1.734	0	0
15	UNEQUAL	1	1	1	1	0	1	1	1	1	1	35	29	1	0	2.500	2.561	0	0
16	UNEQUAL	0	0	0	0	0	0	0	0	1	1	27	22	1	0	2.075	1.250	0	0
17	UNEQUAL	0	0	1	0	0	1	0	1	1	1	21	20	0	1	1.689	1.689	0	0
18	UNEQUAL	1	1	1	0	1	1	1	1	1	1	20	21	0	1	1.250	12.500	0	0
19	UNEQUAL	0	0	0	0	0	1	1	1	1	1	19	20	0	0	1.667	6.089	0	0
20	UNEQUAL	0	0	0	0	0	1	1	1	1	1	22	21	0	1	1.250	1.467	0	0
21	UNEQUAL	0	0	0	0	0	1	1	0	0	1	21	24	1	0	-	1.000	1	0
22	UNEQUAL	1	1	1	1	0	1	1	1	1	1	25	20	1	0	0.833	2.500	0	0
23	UNEQUAL	0	0	1	1	1	1	1	1	1	1	26	22	1	0	-	1.689	1	0
24	UNEQUAL	0	0	0	0	0	1	1	1	1	1	27	19	1	0	6.089	1.734	0	0
25	UNEQUAL	0	0	1	0	0	1	0	0	0	1	32	23	0	0	74.160	1.105	0	0
26	UNEQUAL	0	1	0	0	1	1	1	1	1	1	22	35	0	1	3.244	3.244	0	0
27	UNEQUAL	0	1	0	1	1	1	1	1	1	1	23	24	0	1	3.244	1.667	0	0
28	UNEQUAL	1	1	1	1	1	1	1	1	1	1	20	20	0	1	1.734	-	0	1
29	UNEQUAL	1	1	1	0	1	1	1	1	1	1	23	19	1	0	4.440	-	0	1
30	UNEQUAL	0	0	0	0	0	1	1	0	1	1	24	25	1	0	1.667	74.160	0	0
31	UNEQUAL	0	0	0	0	1	1	1	1	1	1	21	19	1	0	4.440	1.734	0	0
32	UNEQUAL	1	1	1	1	0	0	1	0	0	0	20	24	1	0	1.000	1.667	0	0
33	UNEQUAL	1	1	1	0	1	1	1	1	1	0	26	20	0	1	74.160	2.075	0	0
34	UNEQUAL	0	0	0	0	1	0	1	1	1	1	28	24	0	1	12.500	1.250	0	0
35	UNEQUAL	1	1	1	1	1	0	1	1	1	1	21	25	1	1	1.250	1.734	0	0
36	UNEQUAL	0	1	0	1	0	1	1	1	1	1	24	30	1	0	1.000	1.243	0	0

Note: Inconsistent indicates subjects that switched multiple times in one of the two lottery decisions.

CHAPTER 5

Experimental Evidence on Innovation Barriers and Collective Innovation Processes in Automotive Industry Value Chains

Unpublished Manuscript:

Buchmann, T., Haering, A., Kudic, M. & Rothgang, M. (2020). Experimental Evidence on Innovation Barriers and Collective Innovation Processes in Automotive Industry Value Chains.

Experimental evidence on innovation barriers and collective innovation processes in automotive industry value chains*

Tobias Buchmann †, Alexander Haering‡, Muhamed Kudic§ and Michael Rothgang**

August 2020

Abstract

The literature on collective R&D processes provides us with important insights on factors that hamper interorganizational knowledge transfer and collective innovation. However, it is still anything but clear how firms embedded in a specific industry context may overcome these barriers. We contribute to this strand of literature by asking under what conditions innovation barriers can be overcome in a setting where a specified number of firms needs to collaborate to ensure the success of a joint innovation effort. We use a laboratory experiment to scrutinize the influence of different situations on the decision to cooperate at the firm level. In our framework, research results represent a public good for the individual firm. By using a linear value chain setting with three suppliers and one OEM, we analyze vertical R&D cooperation. Our results identify certain constellations that support cooperation, and therefore make it easier to overcome innovation barriers. These constellations involve sequential decision-making, which also increase the overall welfare, even in case of unequally distributed R&D budgets.

Keywords: Public goods, value chain, innovation barriers, sequential decision making.

JEL Codes: C92, D79, O31.

---

* We thank Jeannette Brosig-Koch and Michael Fritsch for very helpful comments and Ida Zinke for her research assistance. We have also benefited from the comments of conference participants at the GfeW Annual Meeting 2018 in Paderborn and the 13th Nordic Conference on Behavioral and Experimental Economics in Odense. Financial support from the TÜV Süd foundation is gratefully acknowledged.

† ZSW - Centre for Solar Energy and Hydrogen Research Baden-Württemberg

‡ RWI – Leibniz-Institute for Economic Research, Essen and Fresenius University of Applied Sciences.

§ University of Siegen, Corresponding author: muhamed.kudic@uni-siegen.de

** RWI – Leibniz-Institute for Economic Research, Essen

1 Introduction

In an industrial context, novel technological solutions, products and services are rarely developed by only one inventor or an isolated firm. Instead, in recent decades we see an increase in collective innovation processes in almost all industries and fields of technology. Because collaborative innovation is beset by any number of obstacles, in this paper we attempt to discover specific conditions that may alleviate those barriers. We use a laboratory experiment (Levitt and List 2008, Falck and Heckman 2009, Brüggemann and Bizer 2016) to identify situational characteristics that are either more or less beneficial for R&D cooperation in a multi-actor setting. The experiment enables us to combine an analysis of causality with the possibility of shaping and manipulating specific decision-making situations. Our analysis is inspired by experiences from a practical case study (Rothgang et al. 2018) and contributes to the relevant streams of literature on innovation barriers, innovation cooperation and on public good games. The novelty of our contribution is straightforward: we do not primarily aim at identifying critical innovation barriers, but rather focus on how innovation barriers can be overcome. The added value of using experimental methods in this context lies in the fact that our approach goes beyond a simple correlation analysis allowing us to uncover causal relationships.

Rothgang et al. (2018) conducted a publicly funded research project dealing with innovation cooperation along a part of the automobile value chain. The project focused on a long-term firm effort to develop new **m**assive **l**ightweight **f**orging components (MLF- project) and reveals several interesting insights. Developing lightweight designs for these components (such as gear wheels, wheel hubs, or ball bearings) represent a novel way of reducing the weight of an automobile. To create an innovative final product, intensive cooperation along the value chain is required. Knowledge gained from the MLF project shows that there are multiple factors impeding the actors' engagement in collective R&D. At the same time, findings also show that challenges related to market development and technological progress increase the pressure to cooperate in innovation activities.[1]

The literature on innovation barriers primarily addresses these problems by trying to identify and tackle existing barriers to innovation. This approach has identified multiple innovation barriers in firms' cooperation patterns, as well as certain framework conditions necessary for innovation (see e.g., Hueske and Guenther 2015). However, it is still to be questioned whether a better way to overcome innovation barriers is to focus not on identifying individual innovation barriers but rather on approaches to overcome situations that are characterized by an interdependence of multiple barriers. The reason is that in real-world

[1] Our experiences are substantiated by observations from expert interviews with R&D representatives from different parts of the massive steel value chain (e.g., in the production of transmission parts and transmissions and wheel hubs).

situations we often do not find one single barrier that can be identified but a multitude of interdependent barriers that interact with each other in their influence on innovation processes (Hadjimanolis 2003: 560). Research cooperation has only recently received attention as a possible means to tackle such situations by Antonioli et al. (2017). Their analysis provides an indication that cooperation might be especially suitable if different firms face similar innovation barriers. The results of our case study identify that the core problem is the limited extent of cooperative innovation and information transfer isolated to a specific part of the value chain.

Using the existing research on innovation barriers as the starting point for our analysis, we focus on vertical cooperation between multiple firms along the automotive value chain. The typical structure of automotive value chains is characterized by an Original Equipment Manufacturer (OEM) at the top and a sizable number of suppliers grouped in different tiers. The latter group encompasses component manufactures that can be either large multinational corporations, or more frequently, small and medium sized enterprises (SMEs). In our case, metal forging firms and steel producers play an important role in R&D activities as they possess much of the relevant knowledge.

In our analysis, we attempt to determine those collaboration structures that are more or less suitable to overcoming innovation barriers and focus on the following questions: (i) Does sequentially developed cooperation along the value chain partially solve the public good problem in innovation cooperation? (ii) Do differences in R&D budgets influence collaborative R&D efforts in sequential decision-making? (iii) How do these factors influence the shared welfare generated by cooperative R&D efforts? To this point, these aspects have only been alluded to in extant innovation studies.

In Section 2, we provide a literature review on innovation cooperation and derive our hypotheses based on experiences from public good games. Section 3 briefly outlines the framing of our study and presents the experimental setup. In Section 4, we present our experimental results and findings. Section 5 concludes with some critical reflections and suggestions for future research.

2 Economic Background and Hypotheses

A firm's ability to access and generate new knowledge that is transferred into products and processes is crucial for its innovativeness and hence its economic performance. Previous research shows that different modes of inter-organizational cooperation (Hagedoorn 1990) affect knowledge transfer processes and eventually the success of R&D cooperation (Teichert 1993). The same argument holds for cooperation activities along production value chains. Knowledge transfer among partners through cooperation allows for improving the manufacturing process (Bessant and Francis 1999) and, thus, can be crucial for the competitiveness of the entire value chain. This is the especially the case when innovation barriers can be

overcome by cooperation through knowledge exchange and/or the use of shared knowledge while a lack of knowledge exchange is a barrier to innovation itself (Hadjimanolis 2003). Knowledge gleaned from the MLF-project shows that knowledge flows may be diminished or impeded in many ways along each step of a value chain (Rothgang et al. 2018). What do we know so far of the factors facilitating or hampering collective innovation processes? A look at the interdisciplinary alliance and network literature provides insights.

Experience gained from the MLF project indicates that there is no single factor that creates a barrier to innovation cooperation. Rather, within firms and along the value chain a variety of different factors influence and reinforce each other creating innovation barriers. For example, at the firm level some of the smaller massive forging firms lack the absorptive capacity to cope with the production of new lightweight materials. Alliance capabilities are also often missing along the massive forging value chain because forging firms traditionally do not cooperate with their customers to develop innovative processes. Results of analyses relying on the dyadic level learning concept (an extension of absorptive capacity) imply that a firm's internal logic influences its absorptive capacity and ability to cooperate. The internal logic of OEMs and systems suppliers often divides responsibilities between different departments (such as R&D and purchasing). This prevents new information from flowing into the firm and being taken up by the relevant unit. Governance and trust represent additional important barriers to cooperation in the MLF value chain. For example, lightweight forging firms may hide the production cost of innovative projects from customers in order to maintain a competitive advantage. This lack of trust and transparency makes innovation cooperation difficult.

When multiple barriers to cooperation exist, removing only one barrier does not solve the problem. Practical experience shows, however, that if firms along the value chain are able to overcome the public good problem associated with cooperative processes, other barriers become less problematic. While the public good problem leads to too little investment in innovation and R&D, the barriers to collective innovation processes created by this issue have not been discussed at length. In the MLF case, collaboration by multiple actors was required to reap the benefits of using massive lightweight design thereby potentially creating a public good problem. While the benefit of reducing the weight of individual parts due to lightweight design is rather small, the aggregate weight reduction possible is substantial, amounting to more than 60 kg for an average passenger car. In order to achieve these benefits, however, multiple firms producing different parts must be involved. The MLF project included already 28 partners, even though there were other firms in the value chain that were not included in the project. The observation that collaborative innovation often involves a significant number of partners, has been shown in the findings of other collective industrial research programs (e.g., Rothgang et al. 2011).

An issue that arises in this context is that any actor can simply refuse to contribute its own resources/knowledge to the collaborative project, even though all partners involved would benefit from joint R&D efforts. The reason for this is straightforward: knowledge and research results generated in collective R&D projects – or related joint innovation efforts – have the character of a club good. Firms in the value chain cannot be excluded from using this knowledge even though they did not contribute to the knowledge production process. At the same time the value of this knowledge does not decay no matter how many firms use it. Such situations lead to a free-rider problem that results in a suboptimal level of common research or – more general – innovation activity (Baumol 1952).

Most characteristic features observed in the cooperation setting of the MLF project can be easily transferred to a laboratory experiment. Laboratory experiments have been used to analyze different kinds of market failures in innovation processes (for an overview see Brüggemann and Bizer 2016), such as intellectual property right issues (Buchanan and Wilson 2014), or competition with R&D investment decisions (Cantner et al. 2009). Applications of experimental methods addressing open innovation issues are, in many respects, closely related to the research question raised in this study. By applying a laboratory experiment, we scrutinize the suitability of anchored discussion as a method to structure creative discussions in open innovation processes (Link et al. 2015).

However, to the best of our knowledge, the public good aspect of innovation cooperation has not been addressed in laboratory experiments. We draw upon a public good game to analyze how the free-rider problem inherent to joint cooperation projects can be overcome. The actual situation we are studying here is conceptually substantiated by public good theory (Olson 1967; Hardin 1968). Public good games without punishment are among the first that were analyzed in laboratory experiments. Early work dates back to the 1970s and '80s (Bohm 1972, 1984; Dawes et al. 1977). According to public good theory, nobody makes a contribution to the public good due to possible free-riding behavior of other actors. The common result documented in public good literature suggests that contributions to a public good are generally low (for an overview, see Ledyard 1995). However, Di Cagno et al. (2016) analyze conditions under which contributions are higher than theoretically predicted. Many people are conditionally cooperative (a phenomenon also known as reciprocity), i.e., they contribute to the public good if others do the same. About 50% of all actors typically act as conditional cooperators (Fischbacher et al. 2001). For instance, group members tend to sanction egoistic behavior and reward altruistic behavior (Offerman 2002; Andreoni et al. 2003). Thus, an individual's behavior is dependent on the risk of being sanctioned and/or on the chance of being rewarded.

In the following, we focus on three issues: (i) the effect of sanctions (especially punishment) on the results of public goods games, (ii) how different initial endowments impact the outcomes, and (iii) how the

increased amount of information provided by sequential decision-making influence, both, the outcome of the public good game, and the related welfare effect. While the first two issues are important for forming the assumptions that underlie our experiment, the last issue forms the basis for our hypotheses.

The effect of sanctions and rewards have been scrutinized in several experimental designs. Sefton at al. (2007) employ and specify a public good experiment to study the effects of sanctions and reward institutions on cooperation and economic efficiency. In their setting, sanctions incur costs to both the sanctioning and the sanctioned actor. Their results show that the two institutions differ from each other. Initially, actors tend to choose the "reward" option more often than the "sanctioning" option. However, the decay rate of rewards was faster than that for sanctions. Actors appeared to "give up" more quickly on the use of rewards. In treatments that focus on sanctions, there is a higher commitment to group contributions, even when the sanctions generate costs. The results indicate that sanctions initiate cooperation in the first place. It also appears that a convincing threat of a sanction may be sufficient to induce cooperation. Carpenter et al. (2012) argue that experiments typically assume a situation in which monitoring and punishment takes place in a complete network, i.e., all actors can observe and punish each other. In reality, groups are often formed in a specific architecture where observation and information are not as transparent.

The effects of punishment in public goods games have been analyzed from a variety of angles (for an overview, see: Chaudhuri 2011). The majority of the experiments shows that the possibility of punishment increases the willingness to cooperate. However, the results of the experiments are influenced by how the punishment is implemented. For example, the cost effectiveness of punishment plays an important role. Nikiforakis and Norman (2008) show that the effectiveness of the punishing mechanism impacts the outcome. Highly effective punishing mechanisms[2] have a more significant effect on the overall outcome. Egas and Riedl (2008) show that environments with low cost and highly effective forms of punishment are more successful in increasing contributions to the public good. It has also been shown that the number of rounds conducted in the experiment matters. In an experiment based on a one-round game, Walker and Halloran (2004) find no significant effect on either punishment or reward. By comparing experiments with ten and fifty rounds, Gächter et al. (2008) find that the average contribution increases with the number of rounds.

Based on these findings, we chose a punishment regime similar to Egas and Riedl (2008). Our experiment has 20 repetitions, and a cost/income ratio for the OEM and the punished firm of one third, which should provide an effective punishment regime. This setting reflects the actual situation found in the automobile value chain where OEMs (and, in some cases, systems suppliers) have substantial leverage.

[2] Measured as the relation of effect on recipient's income compared to cost to the punisher.

The results of the experiments on the impact of equal or unequal endowments on the outcome of public good games are rather mixed. While Warr (1982; 1983) and Chan et al. (1999) show that income distribution has no effect on the overall contribution, others find that inequality of endowments leads to a reduction of the contribution (Cherry et al. 2005; Aquino et al. 1992). In our analysis, we compare the results obtained in settings with both equal and unequal R&D budgets. The unequal distribution environment is closer to the real-world situation observed in the MLF project, where we saw large differences in actual R&D budgets.

Several studies analyze whether transparency increases the level of cooperation. In other words, each actor can directly observe the public good contributions of all other actors. Papers in this strand of literature either focus on factors affecting performance within teams (e.g. Winter 2004), or on the common development of public goods (e.g. Chen et al. 2008). The effect of increased transparency of the contribution to a common good depends on reciprocal motivations (see e.g. Coats and Neilson 2005). In our case, transparency is systematically linked with sequentially, i.e., each actor gets a piece of information about the contributions of previous actors. As previous studies show, there is often conditional cooperation as participants respond to the high contributions of others by also making high contributions (Fischbacher and Gächter 2010). This shows that when such conditional cooperation is observed, higher contributions can be expected if the peers of observed individuals also make high contributions. Clark and Sefton (2001) use a prisoner's dilemma to show that in individual situations conditional cooperation can be observed. Nevertheless, the overall cooperation level in their experiment is not higher than in a simultaneous setting without transparency. Others have observed that information about previous contributions increases individual contributions (e.g., Shang and Croson 2009). Related studies find a partly positive effect (Masclet and Willinger 2005; Steiger and Zultan 2014), partly no effect (Figuieres et al. 2012), or even a negative effect (Gächter et al. 2008). Steiger and Zultan (2014) find that partly transparent networks function as well as fully transparent networks.

In a setup where successive actors have information about the contribution of prior actors, Masclet and Willinger (2005) identify a leadership effect in the sense that prior actors try to influence successive actors to contribute more through a higher own contribution. At the same time, individuals who observe defection by other actors could also withhold their contributions, which may lead to negative reciprocity (Steiger and Zultan 2014). This possibility is reflected in our experiment because we vary the sequence of decisions made by the suppliers. Levati et al. (2007) analyze the leadership effect by creating an environment with an equal vs. unequal distribution of assets. They find that leadership influences the outcome in the case of an equal distribution, but that information about the distribution of the assets plays a stronger role in an unequal distribution environment. Although diverging from our analysis with respect to the implementation of leadership, the paper by Levati et al. (2007) is closest to our assumption. Their paper considers leadership

as one actor's decision to be 'leading by example', whereas we focus on the sequential aspect of decision-making. Also, there is no punishment mechanism in their study.

If we assume that a leadership effect exists, an experiment with a sequential decision-making design should have a positive effect on the overall outcome. In addition, framing our experiment as a common R&D effort should reinforce a positive leadership effect. This leads us to the following set of hypotheses:

> *H1: A successive experiment design will increase the average contribution because of higher information content leading to an increase in potential leadership effects.*

> *H2: Higher contributions of one or more actors are associated with higher contributions of all other members along the value chain.*

The rich body of literature on innovation obstacles does not primarily address the assessment of welfare losses, but rather looks at the different barriers as possible obstacles to innovation (e.g Hueske and Guenther (2015), D'Este et al. (2012), and Pellegrino and Savona (2017)). Some studies indicate that it is not clear whether eliminating barriers to innovation (e.g., by increasing information availability) improves innovative performance and thus to increased welfare. Information availability often simply shapes firm-internal decision-making and might not impact innovative performance (for an example, see Tang and Yeo 2003). Welfare losses typically result from a lack of investment in R&D. In our cases, this relates to a lack of contribution to the collaborative R&D project. Due to the public good character of the technological knowledge created during the research project, it is fair to assume that contributions are collectively sub-optimal. This leads us to our final hypothesis:

> *H3: The successive experiment leads to a higher level of economic welfare.*

3 Experimental Design

Experiments are a well-established method used to analysis issues in microeconomics and socialpsychology. A major advantage of the experimental approach is that we can observe behavior in a highly controlled environment, while changing variables of interest (Guala 2005). Not only can we control the variables of interest, we can also control certain conditions that might affect the results (e.g., communication, anonymity, and incentives). Our methodology is similar to a simulation and is situated somewhere between a qualitative and quantitative approach (Roth 1995).

The laboratory allows us to establish decision-making environments that are closely related to existing theoretical models. This allows us to contrast the observed decisions with theoretical predictions, as well as

examine causal links rather than describe relationships (Diekmann 2008). An advantage to the experimental approach is the high level of internal validity due to the controlled environment. A potential shortcoming is external validity, that arises from a neglect of contextual influences and questions concerning time and causality ("cause and effect").

The framing of our experiment is inspired by observations from the MLF project. As such, it is oriented towards characteristics of cooperation along the automobile value chain. Experiences from this project show that core innovative processes along the value chain exhibit the following characteristics (for more details, see Rothgang et al. 2018):

(1) Cooperation has characteristics of a public good as firms from multiple steps along the value chain work together and create re-combinatorial technological knowledge in precompetitive research in order to bring forging innovations to the automotive industry.

(2) OEMs located at the end of the value chain have a special relevance for innovative processes. As system integrators, they can potentially push innovation by creating pressure for other firms in the value chain. This behavior can be modeled as a form of punishment. In practice, the OEMs do not usually punish their component suppliers literally, but they are able to exert pressure through their market power. For example, they may include certain requirements in contracts, or control access to their firm decision-makers. All of these behaviors carry a cost, as resources must be allocated for monitoring the activities and influencing firm-internal decision processes.

(3) One pronounced characteristic of the value chain is that participating firms have unequal R&D budgets to contribute to the common goals. In fact, not all forging firms have an own R&D department. The value chain is also characterized by differing degrees of potential for contributing to the common activities.

(4) As communication is often problematic along the value chain, we increased the transparency of inputs to innovation along the value chain, primarily among the steel producers and forging firms. The question arises whether this will serve to increase the contribution to the public good.

These general characteristics are fully considered in our experimental study that is based on an extended public goods game.[3] In our experiment, actors must decide whether to invest in a joint R&D project or to keep all or part of their endowment/budget for themselves.[4] Participants were placed in groups of four with fixed roles: supplier 1, supplier 2, supplier 3 or OEM. In each of the 20 rounds played, the actors received

[3] Public goods games have also been used in environmental economic experiments, see Sturm and Weimann (2006) for an overview.

[4] It was not possible to transfer (part of) the endowment in the next round.

a personal budget. The monetary values were displayed as Experimental Currency Units (ECU), where one ECU was equal to 0.06 Eurocent. In two treatments the budget was fixed at ECU 1,000 per actor (**_C**: Certain). In two other treatments a budget of ECU 250, ECU 750, ECU 1,250, or ECU 1,750 was randomly assigned in each round to each actor (**_R**: Random).[5]

Beside the budget, we varied the decision-making sequence and the flow of information within the groups. In Sequence 1 (**Seq_1**) suppliers 1 - 3 decide simultaneously. In Sequence 2 (**Seq_2**) supplier 1 makes the first decision, followed by supplier 2 who was informed about supplier 1's contribution before making a decision. Finally, supplier 3 was informed about supplier 2's decision, and so on. In all treatments the OEM was informed about the contributions of each supplier and was given the opportunity to punish. Punishment came at a cost for both the penalized supplier and the OEM. It reduced the payoff of the supplier and of the OEM. The OEM was free to choose any amount for punishment, as long as it does not exceed his endowment. For instance, a punishment of 3 reduced the penalized supplier's payoff by 3 and the OEM's payoff by 1.

Table 1: Treatments

Treatment	Abbr.	Sequence	Endowment
Sequence 1 Certain	**Seq_1_C**	Supplier 1 & Supplier 2 & Supplier 3 >>> OEM	Certain
Sequence 1 Random	**Seq_1_R**	Supplier 1 & Supplier 2 & Supplier 3 >>> OEM	Random
Sequence 2 Certain	**Seq_2_C**	Supplier 1 > Supplier 2 > Supplier 3 >>> OEM	Certain
Sequence 2 Random	**Seq_2_R**	Supplier 1 > Supplier 2 > Supplier 3 >>> OEM	Random

Note:">" represents the information flow

Each player had to decide how much of his endowment he wanted to invest in a collaborative research project. The total payoff equaled the sum of all investments times two. An observation from the MLF project revealed that precompetitive innovation efforts often lead to a higher total return when compared to own R&D. For example, we observed that firms producing ball bearings can profit from projects that focus on gear wheels. In our experiment, the profits of the joint R&D project were divided equally between all group members. Table 1 summarizes our four treatments and Figures 1 and 2 provide screenshots of the actors' decision screens.

[5] Note that the expected value of the initial budget is the same (ECU 1,000) in all treatments and the total budget per group was fixed at ECU 4,000.

We consider the Seq_2_R treatment closest to a real-world situation, because it is characterized by sequential decision making with substantial differences in R&D budgets. Consider a real-world scenario in which an R&D manager must decide whether to invest in a common R&D activity without knowing how that the decision will impact the engagement of other R&D managers. The Seq_2_R treatment also allows us to observe how the other R&D managers react to the initial impulse of leading firm. While the certain budget sequence (Seq_2_C) is used as a reference point, substantially different R&D budgets (Seq_2_R) are a common characteristic in automobile industry as well as in other value chain dominated industries, such as aviation.

Figure 1: Screenshot Supplier 2 (Treatment: Sequence 2 Certain)

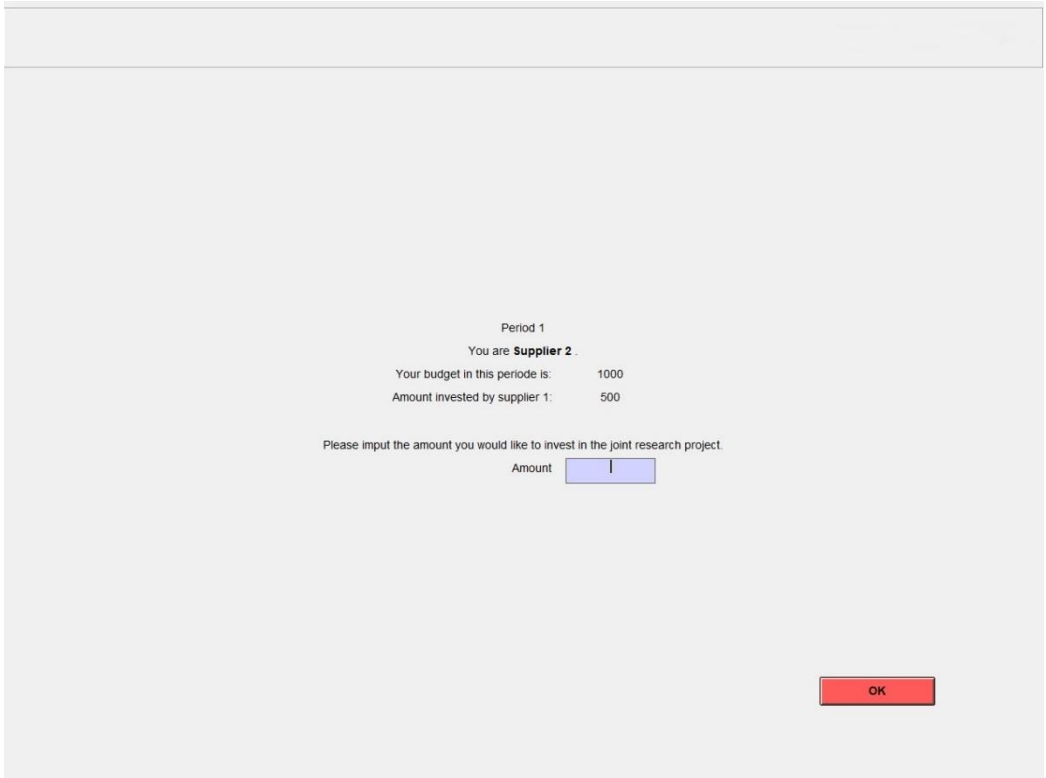
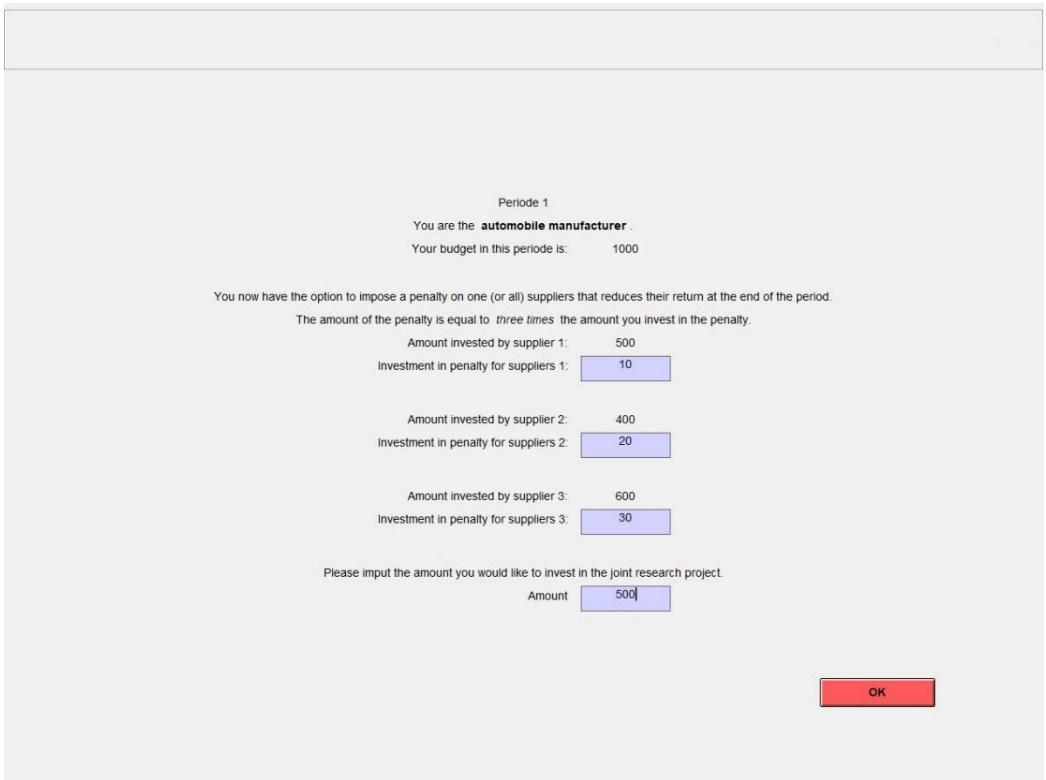


Figure 2: Screenshot OEM (all Treatments)



The actors' roles were randomly assigned by drawing a ball from an urn upon entering the laboratory. We tried to balance the distribution of male and female actors across roles and groups by using two different urns.[6] After taking a seat, the participants read the experimental instructions.[7] Before the beginning the actual experiment, all participants took a test of understanding, which every participant could solve. All remaining questions were answered in private.

Eight sessions (two for each treatment in random order) were conducted at the Essen Laboratory for Experimental Economics (elfe) in Essen, Germany. We implemented a between-subjects design and each participant only participated in one session. A total of 192 economics and natural sciences students[8] took part in our experiment. They were paid 19.40 Euro, on average, for a session that lasted about 75 minutes. The experiment was programmed using z-Tree (Fischbacher 2007) and participants were recruited using ORSEE (Greiner, 2015)

Table 2 summarizes the number of subjects, the number of independent observations (N), the share of female participants and the average age in the different treatments. Although the share of female participant slightly differs between the treatments we do not observe significant differences between treatments ($p \geq 0.414$, Fisher's exact test). The same applies to the participants' age, which is never significantly different ($p \geq 0.507$, two-sided Mann-Whitney-*U* test).

Table 2: Subject pool

Treatment	Subjects	N	Female	Age (SD)
Sequence 1 Certain	48	12	0.563	23.333 (2.838)
Sequence 1 Random	48	12	0.521	22.979 (3.028)
Sequence 2 Certain	48	12	0.542	23.125 (3.324)
Sequence 2 Random	48	12	0.458	23.083 (3.370)

Note: N is the number of independent observations; SD is the standard deviation.

[6] However, we had to deviate from this procedure to some degree in our sessions if participants did not show up.

[7] Instruction are available from the authors upon request.

[8] We focused on this specific subject pool because of our belief that students with a major in economics or natural sciences have the highest probability of facing R&D related decisions in their later careers.

4 Results

In the following section, we analyze firms' behavior in our four treatments.[9] We focus on how differences in the decision-making sequence and endowment influences the level of contribution to the joint research project.[10] We also focus on how sequential decision-making changes the decisions made by R&D managers, and whether and to what extent the more realistic decision-making environments (those with substantially differing R&D budgets) influence the investment in common R&D efforts. We offer a brief discussion of a comparison of the average punishment per treatment, and a brief analysis of the welfare consequences. Where practical, we compare the behaviour of suppliers and OEMs within their respective peer group only.

Share of endowment: suppliers and OEMs

We first observe and analyze the behavior of the entire group (suppliers and the OEMs). Figure 3 shows the average share of endowment invested in the joint R&D project, separated by treatment for all 20 rounds of the game. We observe rising contributions at the beginning of the game, and end-game effects in all four treatments. The highest average contributions are found in **Seq_2_C** (79.1% on average) followed by **Seq_2_R** (71.7%). Lower average contributions were made by group members in **Seq_1_C** (68.4%), while the lowest average could be observed in **Seq_1_R** (59.9%). By looking at the average contribution over all periods, we find significant differences between **Seq_1_R** and **Seq_2_R** (p = 0.038, two-sided Mann-Whitney-U test). All other observable differences are insignificant ($p \geq 0.166$).

Share of endowment: suppliers only

Next, we focus solely on the suppliers, because they differ from the OEMs with regard to two important factors: they cannot punish other group members, and they are always observed by (at least) the OEM. Figure 4 presents the average share of endowment invested in the joint R&D project by the suppliers, separated by treatment in all periods. We see less variation when compared to OEMs (see below). This might indicate that suppliers act in a more stable manner. The level of contribution is always above 50.0% (minimum: 54.0%, period 19, treatment **Seq_1_R**) in all periods.

[9] For our treatments we use abbreviation for ease of reading. "Sequence 1 Certain": **Seq_1_C**, "Sequence 1 Random": **Seq_1_R**, "Sequence 2 Certain": **Seq_2_C** and "Sequence 2 Random": **Seq_2_R**, see Table 1 for details.

[10] Comparing **Seq_1_C** with **Seq_1_R** leads to the result that **Seq_1_C** dominates **Seq_1_R**. The effect is not statistically significant (p = 0.299, two-sided Mann-Whitney-U test) when aggregated over all periods.

Figure 3: Share of endowment, OEMs and suppliers, all treatments

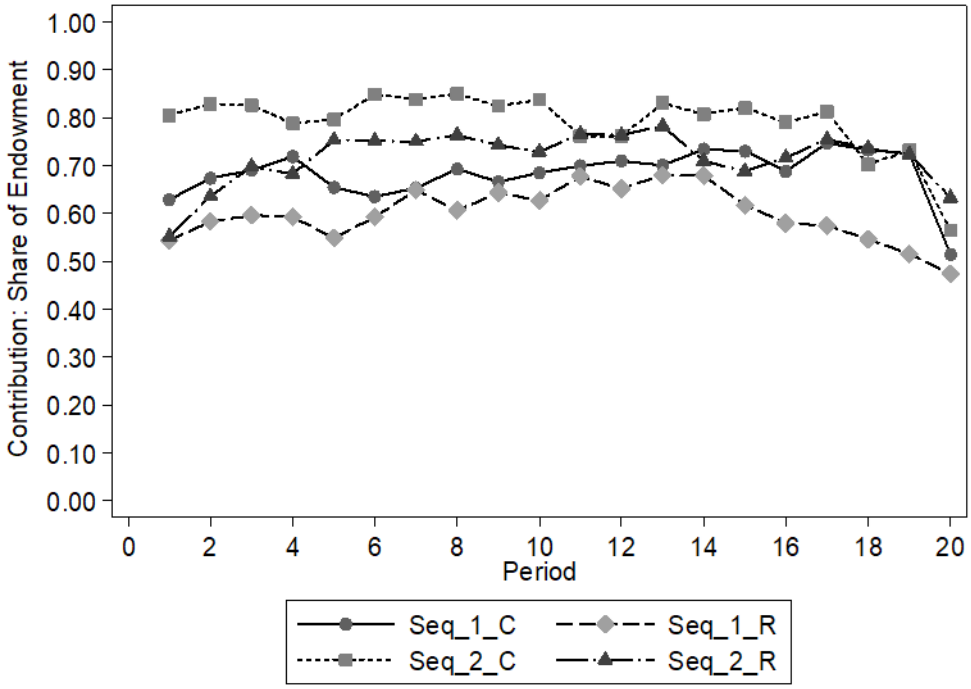
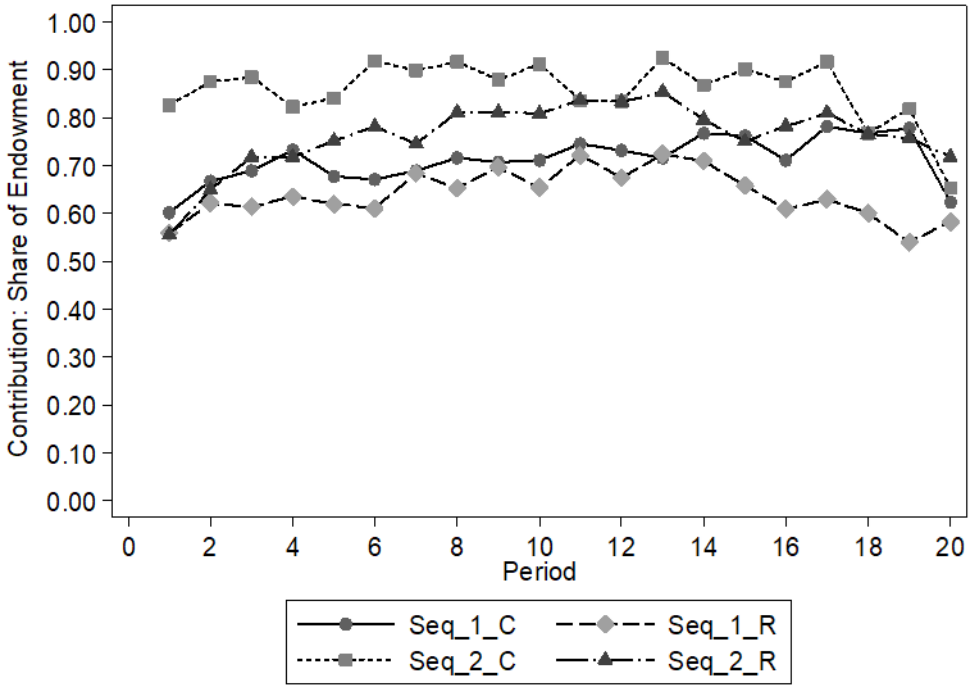
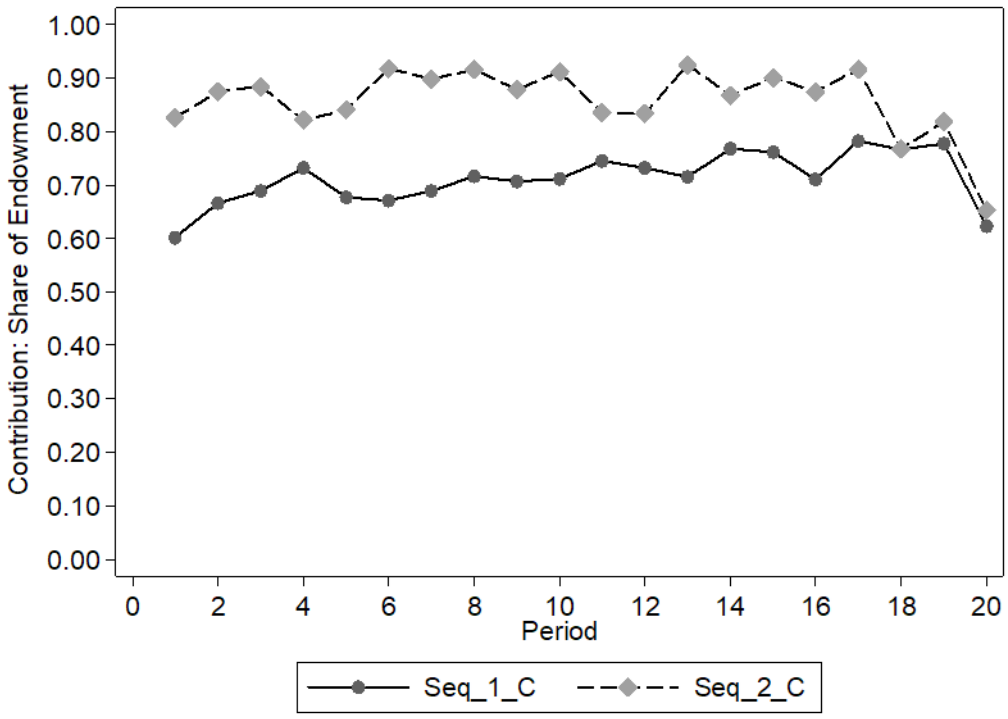


Figure 4: Share of endowment, suppliers only, all treatments



We now turn our focus to the decision-making sequence and the flow of information to determine if these factors affect the level of contribution by the suppliers. To analyze the effects, we separately compare the two different sequences in both endowment types (**_C**: Certain and **_R**: Random). In other words, we hold the endowment mode fixed and vary the sequence. In Figure 5, both sequences are compared under the condition that all team members receive the same initial endowment in each round (**Seq_1_C** compared to **Seq_2_C**).

Figure 5: Share of endowment, suppliers only, certain budget

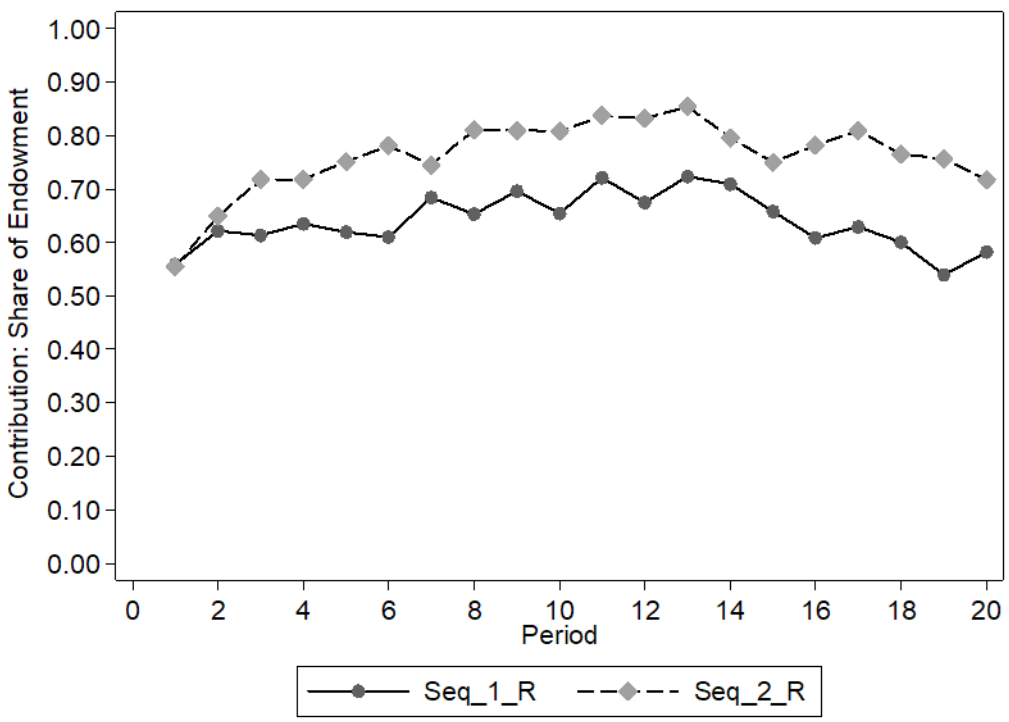


At first glance, **Seq_2_C** results in higher contribution shares, with a weakly significantly higher (p = 0.065, two-sided Mann-Whitney-U test) average share of endowment (85.9%) when compared to **Seq_1_C** (71.2%). Comparing specific period intervals reveals additional information about suppliers' behavior. If we look at only the first three rounds, we observe that in the **Seq_2_C** environment the share of endowment provided to the joint R&D project is significantly higher ($p \leq 0.037$, two-sided Mann-Whitney-U test). If we look at rounds 6 – 17, except for rounds eleven and twelve ($p \geq 0.331$), we notice a (weakly) significant ($p \leq 0.088$) effect. In the last three periods, there are no significant differences in the average contribution ($p \geq 0.495$) between the two environments. This indicates that sequential decision-making fosters

cooperation.[11] In the **Seq_2_C** environment information is more readily available, lowering uncertainty and encouraging suppliers to be more cooperative.

Figure 6 presents our **Seq_1_R** and **Seq_2_R** environments where team members receive a random initial endowment in each round. At first glance, the **Seq_2_R** environment, once again, encourages cooperative decision-making. Looking at the average share of contribution we see a significantly higher ($p = 0.038$) average share in the **Seq_2_R** environment (76.3%) compared to **Seq_1_R** (64.0%). Comparing specific rounds reveals additional information about the suppliers' behavior. When we look at the first five periods, we do not observe a significantly different ($p \geq 0.204$) average share of endowment between the two treatments. From periods 8 - 19, **Seq_2_R** (weakly) significantly ($p \leq 0.093$) dominates **Seq_1_R** (with the expectation of periods 14 and 15, $p \geq 0.260$). Once again, in the final round, we do not observe a difference ($p = 0.222$) between the two environments.

Figure 6 Share of endowment, suppliers only, random budget

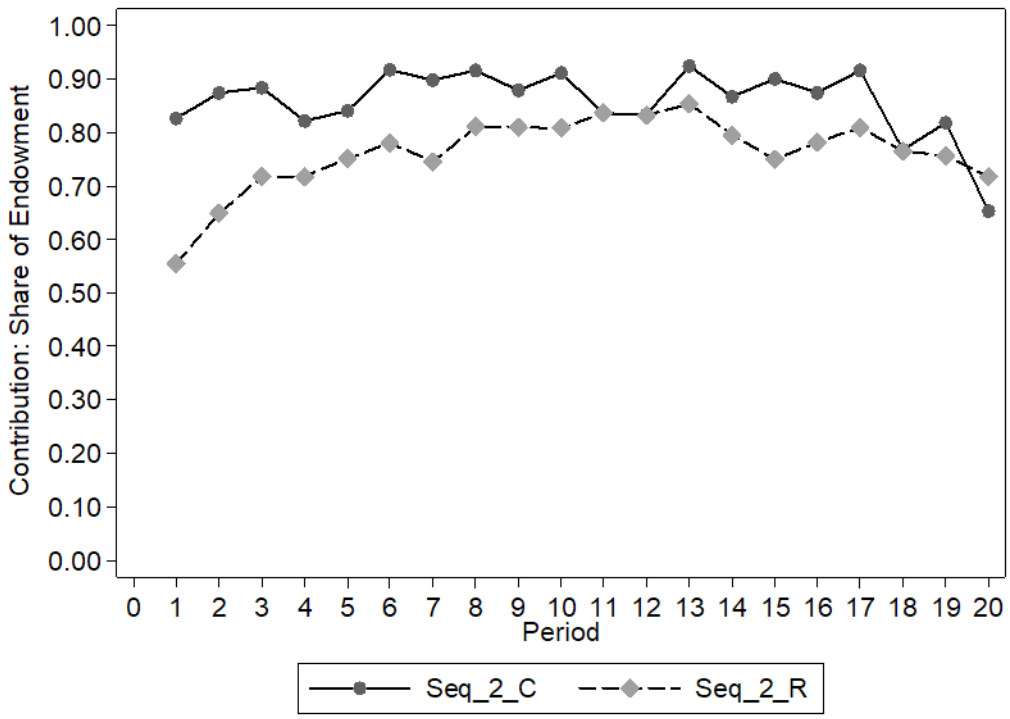


[11] One could assume that these differences are driven by supplier 1, who might invest much more in **Seq_2** in order to give a signal to his fellow team members. However, we do not observe a significantly different share of endowment when comparing suppliers 1, 2 and 3 ($p \geq 0.225$, two-sided Mann-Whitney-*U* test).

We assume that different random budgets lead to different levels of power and commitment within the group, leading to lower contributions. But this setting may also be less vulnerable to free-riding behavior in the final rounds. As can be seen, sequential decision-making foster cooperation even with random budgets. The results of our **Seq_1_C** vs. **Seq_2_C** analysis is robust to a variation in the group members' endowments and, therefore, holds true even with differing levels of power.

To deepen our analysis, we compare **Seq_2_C** and **Seq_2_R** to determine the effect of a random endowment on the average distribution in a sequential decision-making environment. Figure 7 presents our two endowment regimes.

Figure 7: Share of endowment, suppliers only, sequential decision



At first glance, in most of the rounds we see that **Seq_2_C** provides a more cooperative environment than **Seq_2_R**. Looking at the average share of contribution we see a highly significantly higher ($p = 0.013$) average share in case of **Seq_2_C** (85.9%) compared to **Seq_2_R** (76.3%). Looking at the distributions in every period, we see that **Seq_2_C** always dominates **Seq_2_R** (weakly) significantly ($p \leq 0.085$) except in periods 4, 5, 9, 11, 12, 18 to 20. All other observed differences are never significant ($p \geq 0.111$).

As a further robustness check, and to investigate how the invested share by the other group members affects the individual's contribution, we run a Random-Effects[12] General Least Square (GLS) regression with the average share of endowment as dependent variable in periods two[13] to 20. The results are summarized in Table 3.

The dependent variable is a subject's share of endowment in each period. **Seq_1_R**, **Seq_2_C** and **Seq_2_R** are treatment dummies (with **Seq_1_R** as baseline), indicating an individual's treatment. *Period* refers to a specific round, with *Period 19* and *Period 20* as dummies indicating the last two rounds. $AV_{period-1}$ is the lagged average group contribution minus the individual's contribution, *supplier 2* and *supplier 3* are both dummy variables, indicating a subject's role within the group (supplier 1 is the baseline). *Age* and *Female* represent the subject's age and gender. The bottom lines of Table 3 contain *p*-values of the Wald test for equal treatment dummies.

In specification (1) we use a basic parameterization without controlling for role or demographic effects. We observe a positive and highly significant effect of **Seq_*2*_*C*** (with **Seq_1_C** as baseline). This is in line with the results of our non-parametric tests reported above. With fixed budgets, the share of invested endowment increases for a sequential decision-making process. The Wald test *p*-values reveals that this observation significantly holds even for random endowments (**Seq_1_R** = **Seq_2_R**). Our results concerning the effect of a random endowment on the average distribution in case of sequential decision making is also confirmed (**Seq_2_C** = **Seq_2_R**). In addition, we observe a slight increase in the contribution to the joint research project over the time (*Period*), with highly significant negative last round effects (*Period 19* = 1 and *Period 20* =1). Furthermore, we observe a positive and highly significant effect of the lagged share by the other group members to the joint project ($AV_{period-1}$), which is in line with theory. The observed treatment differences are robust when we control for a subject's role in parametrization (2), and when we add additional demographics in Model (3). Here, we observe that female subjects tend to invest significantly less in the joint R&D project compared to their male counterparts, but the significance and direction of our treatment dummies are not affected even once we run our full parameterization regression in specification (4). Accordingly, we conclude that our non-parametric tests are robust to different potential influences.

[12] We choose a random effect approach because our variables of interest are invariant over periods. For a similar approach see e.g., Tan and Bolle (2007).

[13] Period one is missing because of the lagged variable $AV_{period-1}$.

Table 3: Random-effects GLS regression
share of endowment, suppliers

	(1)	(2)	(3)	(4)
Seq_1_R	-0.037	-0.037	-0.038	-0.038
	(0.042)	(0.042)	(0.041)	(0.041)
Seq_2_C	0.103***	0.103***	0.101***	0.101***
	(0.034)	(0.034)	(0.032)	(0.032)
Seq_2_R	0.047	0.047	0.037	0.037
	(0.034)	(0.034)	(0.033)	(0.033)
Period	0.002*	0.002*	0.002*	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)
Period 19 = 1	-0.039**	-0.039**	-0.039**	-0.039**
	(0.018)	(0.018)	(0.018)	(0.018)
Period 20 = 1	-0.118***	-0.118***	-0.118***	-0.118***
	(0.029)	(0.029)	(0.029)	(0.029)
$AV_{period-1}$	0.403***	0.403***	0.405***	0.405***
	(0.039)	(0.039)	(0.038)	(0.038)
Supplier 2 = 1		-0.020		-0.020
		(0.031)		(0.030)
Supplier 3 = 1		-0.024		-0.023
		(0.032)		(0.031)
Age			-0.003	-0.003
			(0.004)	(0.004)
Female = 1			-0.053**	-0.052**
			(0.025)	(0.025)
Constant	0.428***	0.443***	0.521***	0.539***
	(0.044)	(0.048)	(0.112)	(0.110)
N	2,736	2,736	2,736	2,736
N in group	144	144	144	144
Wald Tests:				
p-Value: ***Seq_1_R = Seq_2_R***	0.031	0.030	0.057	0.055
p-Value: ***Seq_2_C = Seq_2_R***	0.064	0.059	0.030	0.027

Note: Robust Standard errors in parentheses, * $p < 0.100$, ** $p < 0.050$, *** $p < 0.010$.

Share of endowment and punishment: OEMs only

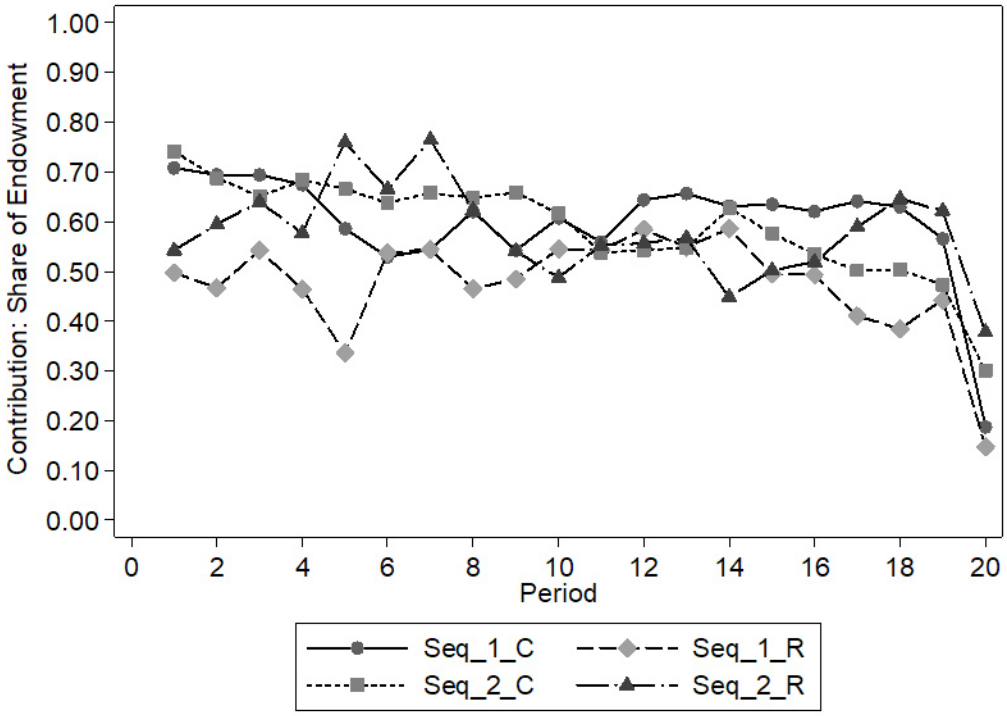
Next, we analyze the OEMs' investment behavior in the joint research project (defined as share of endowment) and the punishment they inflict. Figure 8 shows the average share of endowment invested in the joint R&D project by the OEMs, separated by treatment for all 20 rounds.

We observe a strong end-game effect in round 20, with an average contribution below 40.0%. We also see a higher variation in contribution share compared to the suppliers' behavior.[14] When focusing on the aggregated behavior over all periods we do not observe significant differences ($p \geq 0.356$, two-sided Mann-

[14] This could also be caused by a lower degree of smoothening compared to the suppliers.

Whitney-*U* test), and no treatment seems to dominate. This is also supported conduct a round-by-round analysis. We only observe significant differences between **Seq_1_C** and **Seq_1_R** ($p = 0.049$) in round five, weak significant differences in rounds five and seven between **Seq_1_R** and **Seq_2_R** ($p \leq 0.058$), a weak significant difference between **Seq_2_C** and **Seq_2_R** in the first round ($p = 0.092$).

Figure 8: Share of endowment, OEMs only, all treatments



When we compare the average punishment (as share of endowment) inflicted by OEMs between our four treatments over all periods we do not observe significant differences ($p \geq 0.126$, with shares of 2.5% in **Seq_1_C**, 3.2% in **Seq_1_R**, 1.8% in **Seq_2_C**, and 2.9% in **Seq_2_R**). This indicates that the OEMs' punishment behavior is not influenced by the treatment in general. We do, however, see some differences in specific rounds. If we focus on the effect of the sequential decision-making process, we observe a significantly lower ($p = 0.019$) level of punishment in **Seq_1_C** compared to **Seq_2_C** in round two. In rounds six, nine, ten and 13, this effect (weakly) significantly ($p \leq 0{,}089$) inverts with a higher level of punishment in **Seq_2_C** compared to **Seq_1_C**. In the case of a random budget, we observe a (weakly) significantly ($p \leq 0.072$) higher level of punishment in **Seq_1_R** compared to **Seq_2_R** in periods three, five and six, and a weak significant effect ($p = 0.071$) in the other direction in the last period.

As we did with our analysis of the suppliers' behavior, we run a Random-Effects GLS regression with the average share of endowment as the dependent variable in rounds 2 - 20 to investigate how the invested share by the other group members affects the individual's contribution. The results are shown in Table 4.

Table 4: Random-effects GLS regression
share of endowment, OEMs

	(1)	(2)
Seq_1_R	-0.092	-0.060
	(0.088)	(0.090)
Seq_2_C	-0.066	-0.056
	(0.118)	(0.116)
Seq_2_R	-0.029	-0.048
	(0.101)	(0.099)
Period	-0.007**	-0.007**
	(0.003)	(0.003)
Period 19 = 1	0.021	0.021
	(0.036)	(0.036)
Period 20 = 1	-0.243***	-0.244***
	(0.057)	(0.057)
$AV_{period-1}$	0.359***	0.354***
	(0.083)	(0.083)
Age		0.013
		(0.010)
Female = 1		0.084
		(0.078)
Constant	0.425***	0.063
	(0.079)	(0.242)
N	912	912
N in group	48	48
Wald Tests:		
p-Value: ***Seq_1_R = Seq_2_R***	0.534	0.907
p-Value: ***Seq_2_C = Seq_2_R***	0.777	0.947

Note: Robust Standard errors in parentheses,
* $p < 0.100$, ** $p < 0.050$, *** $p < 0.010$.

Treatment dummies **Seq_1_R**, **Seq_2_C** and **Seq_2_R** (with **Seq_1_R** as the baseline) indicate an individual's treatment. *Period* represents the specific round, and *Period 19* and *Period 20* are dummies indicating the last two rounds of our experiment. *Age* and *Female* are variables capturing potential effects of an individual's age and gender. We observe a positive and highly significant effect of the lagged share by the other group members to the joint project ($AV_{period-1}$), and a significant, but small, negative round effect (*Period*). With regards to the treatment dummies and in line with our non-parametric analysis, we do not observe any significant differences between our four treatments.

In summary, the OEMs' aggregated behavior is not influenced by the treatment. Neither the share of endowment contributed to the joint research project, nor the level of punishment significantly differs between our four treatments. This result comes as no surprise since the OEMs can always observe and punish them all the suppliers, and OEMs' are the last to decide about their contribution regardless of treatment. Therefore, the various decision-making environments do not change the level of information available to the OEMs.

Welfare: suppliers and OEMs

In a final step, we analyze the (net) welfare in our four treatments. The following equation was used to calculate the group welfare in ECU per period:

$$Welfare_g = \sum_{i=1}^{4} SharePoject_i + \sum_{i=1}^{4} Budget_{rem_i} - \sum_{i=1}^{3} Punish_{ef_i} - Punish_{cost_{i=4}} \quad (1)$$

$$with\ g: group\ and\ i: individual\ 1,2,3 = suppliers\ 1,2,3;\ 4 = OEM$$

In equation (1) the welfare for group g ($Welfare_g$) in any round is equal to the sum of the shares contributed to the joint R&D project ($SharePoject_i$) plus the sum of the remaining budget ($Budget_{rem_i}$) minus the effect of punishment ($Punish_{ef_i}$) minus the costs of punishment ($Punish_{cost_{i=4}}$). In other words, the welfare is the project's surplus and the remaining budgets of all team members less the effect and costs of punishment.

Figure 9 shows the average welfare in ECU for certain budget and both decision-making environments for all 20 rounds. At first glance, **Seq_2_C** increases the welfare. Aggregating all rounds, **Seq_2_C** (ECU 6,944.2 on average) yields a weakly significantly (p = 0.094, two-sided Mann-Whitney-U test) higher welfare compared to **Seq_1_C** (ECU 6,432.5). This result seems to be driven by (weak) significant differences in rounds one, two, six, nine and ten only ($p \leq 0.057$). In round 18 and 19, **Seq_1_C** dominates **Seq_2_C**, but this effect remains insignificant.

Figure 10 compares both sequences with random endowments environments (treatment **Seq_1_R** vs. **Seq_2_R**). Again, **Seq_2_R** increases welfare compared to **Seq_1_R**, even in case of random budgets. By comparing the average over all periods, we observe a significant (p = 0.024) lower welfare in **Seq_1_R** (ECU 5,954.3) compared to **Seq_2_R** (ECU 6,586.3). By looking at each period, we observe (weak) significant ($p \leq 0.099$) differences in rounds five, six, eight, nine, 13, 17 and 19.

In Figure 11, we focus on the sequential decision-making environment and compare the average welfare in both the certain and random endowment regimes (Treatment **Seq_2_C** vs. **Seq_2_R**). A comparison of

the average over all periods does not reveal significant differences ($p = 0.119$). Analyzing the distribution in each round, we only observe (weakly) significant differences in periods 1, 2, 10 and 13 ($p \leq 0.076$). Here, **Seq_2_C** dominates **Seq_2_R**.

Hence, we can conclude that the sequential decision-making environment (**Seq_2**) increases the welfare (weakly) significantly in both certain and random endowments.

Figure 9: Welfare in ECU, certain budget

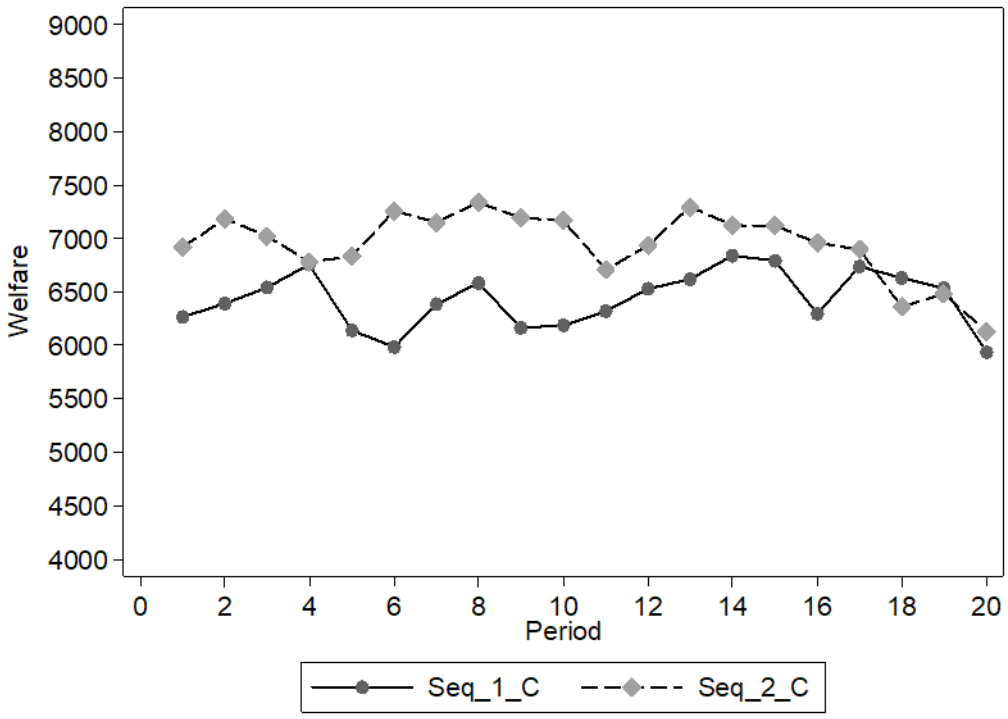


Figure 10: Welfare in ECU, random budget

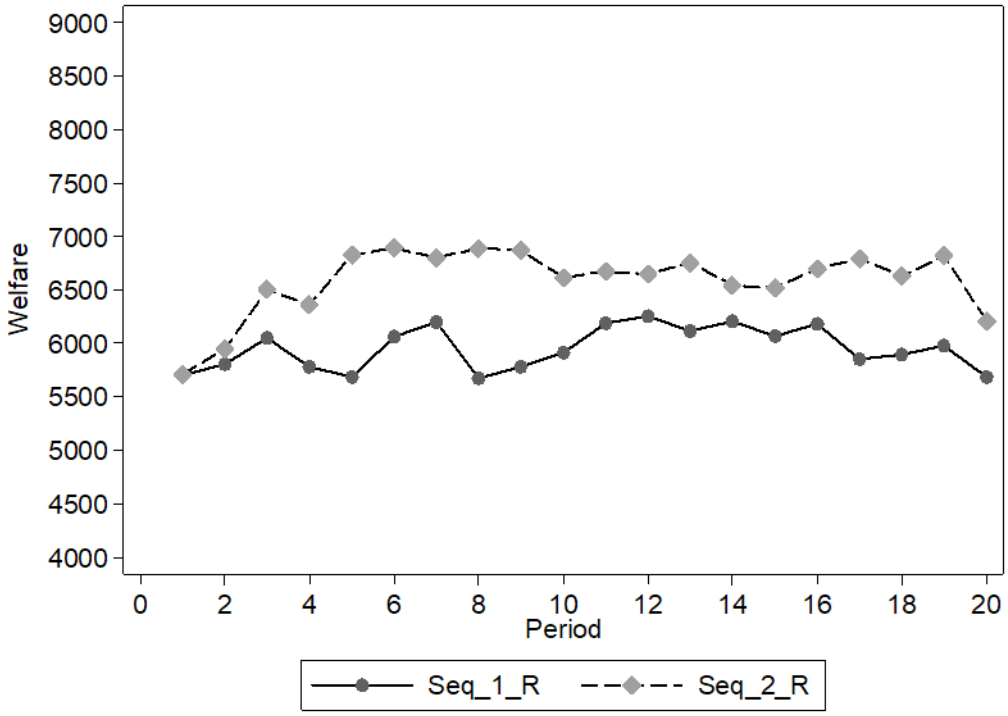
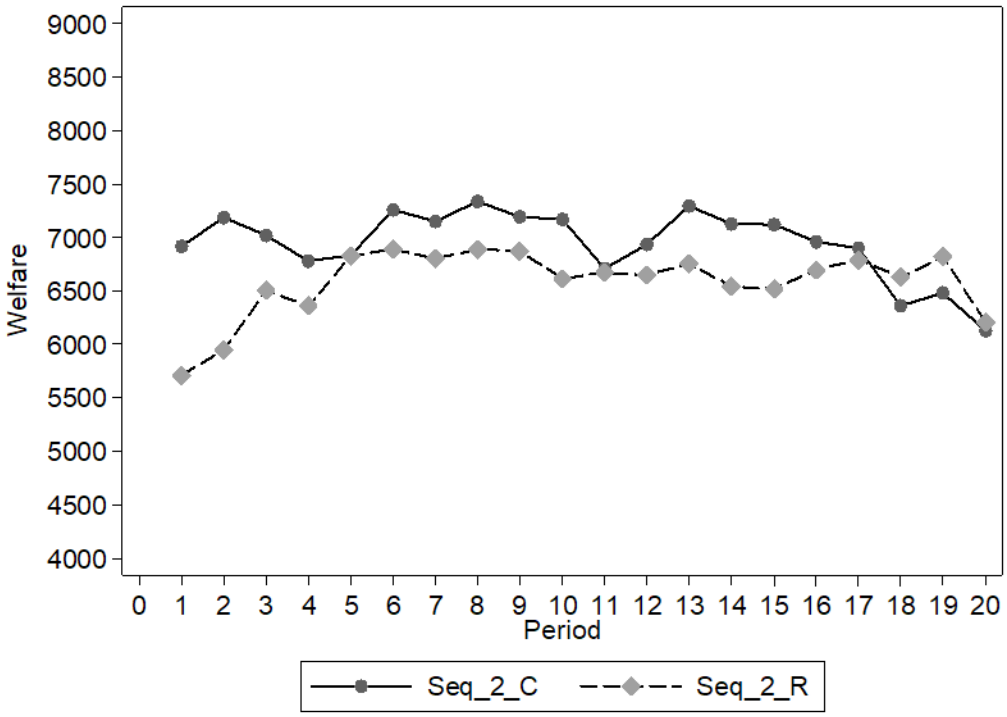


Figure 11:Welfare in ECU, sequential decision



5 Discussion

Our study contributes to the literature on innovation barriers in situations characterized by joint R&D activities. We look at the decision-making environment of R&D managers to determine if these barriers can be overcome. Antonioli et al. (2017) propose cooperation as a possible way of tackling innovation barriers. We conduct a public goods game in a laboratory environment to gain an in-depth understanding of collective investments in joint R&D projects of firms (i.e. suppliers and OEMs) located at different stages along the automotive value chain. We are especially interested in observing the investment behavior of R&D managers located at different stages along the value chain in joint R&D projects when there are changes in key aspects of the decision-making environment.

We focus on the effect of two different decision-making environments in our experimental design: simultaneous decision-making, and sequential decision-making. We assume these different environments will affect the managers' willingness to contribute to joint R&D projects. Because there are substantial differences in R&D budgets in the real world, we also include two different endowment distribution treatments: an equal endowment scenario, and an endowment scenario where all of the actors have differing endowments. We first differentiate by suppliers and OEMs, and then analyze the actors' behavior in our four experimental treatments. We consider the possibility that OEMs might exert pressure on the decision-making process by accounting for the average punishment between the treatments and include these observations in our investigation of welfare consequences.

The results make it clear that certain mechanisms in collaborative decision-making enable the actors to increase their cooperation efforts to overcome existing barriers. This confirms the observation of Antonioli et al. (2017) that cooperation can indeed be a means to overcome innovation barriers. If real-world managers of leading firms increase their contribution to a joint R&D project, they will likely trigger an increase in the contribution of the managers of following firms. In sequential decision-making environments where the actors have an equal budget, the suppliers significantly increase their share of endowment invested in a joint R&D project from 71.2% to 85.9%. Even in the more realistic setting of differing budgets, we observed qualitatively the same results. In this setting, the average share increases from 64.0% to 76.3%. The positive effects of sequential decision-making are underlined by our welfare analysis, which shows that this increased effort also leads to an increase in total welfare. Furthermore, we observe a positive and highly significant effect of the lagged share by the other group members to the joint R&D project in case of both suppliers and OEMs.

Hence, the possibility of the R&D manager in a supply firm to take the lead in the investment process has a significant positive influence on the overall investments in the collaborative R&D project. This holds not only for the case of equal budgets but also in case of varying endowments. In addition, there is some

(although weak) evidence that the share of endowment (and also the welfare effect) of sequential decision-making is somewhat smaller in the more realistic case of varying R&D budgets than with equal R&D budgets. The OEMs exhibit a similar behavior in all scenarios, that is, they are not influenced by varying sequences of decision-making or by budget variations. Similarly, with regard to punishments, the OEMs do not vary significantly and are overall rather moderate. The existence of the punishment option seems to be sufficient for influencing suppliers' behavior. This also stresses the possibility of R&D managers in the supply chain to give an impulse for overcoming innovation barriers. Table 5 summarizes our main findings and compares our observations with the hypotheses presented in Section 2.

Table 5: Summary of observations

	Hypothesis	Observation
H1	The successive experiment leads to an increase in the average contribution	Accepted: suppliers significantly increase their share of endowment in case of sequential decisions
H2	Higher contributions of one or more actors lead to a higher contribution by the other actors	Accepted: we observe a positive and significant effect of the lagged share by the other group members
H3	Sequential decisions lead to a higher level of economic welfare	Accepted: sequential decisions significantly increase the overall welfare

Our results can be compared with some general findings on the dynamics of public goods situations. Steiger and Zultan (2014) show that transparency is important for decision-making, such that in general (though not in all cases) increased transparency leads to an increase in cooperation. Our results show that the successive decision mechanism works in a similar vein and shares features with the leader effect that has been scrutinized by other studies (e.g. Chan et al., 1999). Levati et al. (2007) look at the effect of leadership (similar to sequential decision-making in our case) on contributions to public goods with an unequal endowment distribution. While other studies show that leadership increases contributions in an environment with equal endowments, an unequal distribution only leads to increased contributions in situations with perfect information of the individuals about the distribution of the initial endowment. As our actors have available information about the endowment structure, our results do not contradict their analysis.

6 Summary and conclusions

In summary, the public good characteristics of technological knowledge make collaborative R&D projects a particularly challenging form of organization for many firms. The involved firms must be incentivized to make collectively optimal R&D investment decisions. The sequential nature of observed cooperation patterns provides a promising way to overcome the low contribution problem related to the public good aspect of R&D cooperation and increases the welfare from innovation cooperation. Our results imply that research cooperation can be initiated and intensified by the actor who takes a leading role. This, in turn, encourages others to follow and contribute. The necessary precondition for mitigating the free-rider problem described above is transparency on the efforts of each individual partner to increase the input into common R&D activities. This is especially true for R&D expenditure but can be generalized to promote and organize common R&D related activities. Sequential decision-making leads to higher average and individual contributions to the public good and increased economic welfare. Our results show that overcoming barriers of innovation has a positive effect on economic welfare, which is not automatically the case when it comes to innovation barriers (Hadjimanolis 2003: 560).

This result has important implications for R&D managers who find themselves in a decision-making situation where common research activities of multiple enterprises in an industry are needed in order to overcome industry-wide innovation barriers. If these managers set a good example by investing time and effort to initiate a joint R&D project, and make their initial investment transparent, it is likely that this will increase the incentive of other firm managers to follow their example and increase the overall amount of common R&D-related activities.

At the same time, government programs can play an important role as a facilitator for such processes. By supporting programs that complement the activities of leading firms, innovation policy can facilitate such processes. A possible example are industry specific research programs in a precompetitive setting that allow for larger cooperative efforts to be funded. In Germany, the industrial collective research program enables such activities.

Our analysis also helps to understand the patterns of cooperation formation that have been observed in the MLF project. In this case study, the initiative to cooperate originated with a few large steel producing and metal forming firms at the far end of the value chain. It became obvious that the initial investment from these firms lead to an ever-increasing number of participants, and also to increased contributions.[15] These

[15] With respect to financial contributions, new ideas and time invested in the project.

contributions were also obviously not impeded by rather different research budgets.[16] Information deficits regarding the potential benefits of cooperation were largest among system suppliers and OEMs, who increased their contribution when the information stock was increased during the cooperation. Obviously, that kind of development pattern is consistent with the results from our analysis which implies that the initial initiative of some actors leads to higher investments.

It is important to note that our results were obtained by framing our experiment towards a situation that arises within the value chain of the automobile industry, which is characterized by close R&D cooperation in the production of a complex good. As the public good aspect of cooperative innovation influences the structural characteristics and dynamics of innovation cooperation and network development in a more general way (e.g. Cantner and Graf 2011: 386 f.), it seems plausible that our specific observation corresponds to a wider variety of cases. Similar situations are likely to arise in industries such as aviation or mechanical engineering. In other industries, different patterns of cooperation dynamics might arise that could also be addressed using laboratory experiments. The results of our analysis are at the same time not limited to formalized R&D cooperation. Our observations can be generalized to include a broad range of activities that include: common information collection and knowledge transfer, the financing of R&D projects of research institutes, cooperative activities to develop a pool of innovative ideas, and the dissemination of knowledge to the customers. All of these activities incur costs for firms and are associated with similar decision situations of the individual managers.

There are also implications of our findings for the literature on innovation barriers. Existing studies on often focus on identifying or classifying different kinds of barriers. Our analysis shows that, in order to understand how to overcome barriers, a more processual focus might be helpful. This also points to a need to develop a deeper understanding of time and context dependency of innovation barriers and ways to overcome them. This applies to situations where several interdependent barriers influence innovation processes.

Our study shows that laboratory experiments can be a useful tool to increase our understanding in practical decision-making situations. There are, of course, limitations to any experimental design. For example, our experiment had only one simple monetary sanctioning mechanism. There are also a variety of ways to implement incentives (either positive or negative) that might influence the results of experiments. In a one round experiment, Walker and Halloran (2004) find no significant effect neither of punishment nor of rewards. By comparing experiments with ten and fifty rounds, Gächter et al. (2008) find that the average

[16] There are rather large differences, ranging from metal forming firms with no fixed research budgets at all to system suppliers and OEMs with research budgets of some billion Euro per year.

contribution increases with the number of rounds. There are also a variety of ways to implement a punishing mechanism as a sanctioning tool. Nikiforakis and Norman (2008) show that highly effective punishing mechanism have a more significant effect on the overall outcome. Egas and Riedl (2008) show that environments with a low cost and high impact punishment increase contributions to the public good. An in-depth investigation of sanctioning mechanisms in sequential public good experiments constitutes a highly promising field for future research, and may improve experimental design.

References

Andreoni, J., Harbaugh, W., & Vesterlund, L. (2003). The carrot or the stick: Rewards, punishments, and cooperation. *American Economic Review*, 93(3), 893-902.

Antonioli, D., Marzucchi, A., & Savona, M. (2017). Pain shared, pain halved? Cooperation as a coping strategy for innovation barriers. *The Journal of Technology Transfer*, 42(4), 841-864.

Aquino, K., Steisel, V., & Kay, A. (1992). The effects of resource distribution, voice, and decision framing on the provision of public goods. *Journal of Conflict Resolution*, 36(4), 665-687.

Baumol, W. J. (1952). Welfare Economics and the Theory of the State. Harvard: Harvard University Press.

Bessant, D. & Francis, D. (1999). Using learning networks to help improve manufacturing competitiveness: Technovation, 19(6-7): 373-381.

Bohm, P. (1972). Estimating demand for public goods: An experiment. *European Economic Review*, 3(2), 111-130.

Bohm, P. (1984). Revealing demand for an actual public good. *Journal of Public Economics*, 24(2), 135-151.

Brüggemann, J., Bizer, K. (2016), Laboratory Experiments in Innovaton Research: a Methodological Overview an a Review of the Current Literature. Journal of Innovation and Entrepreneurship 5(24): 1-13.

Buchanan, J.A. and Wilson, B.J. (2014) An Experiment on protecting intellectural property. *Experimental Economics* 17(4): 691-716.

Cantner, U., W. Güth, A. Niklisch, T. Weiland (2009), Competition in product design: an experiment exploring innovation behavior. *Metroeconomica* 60(4): 724-752.

Cantner, U. & Graf H. (2011). Innovation networks: formation, performance and dynamics. In: C. Antonelli (ed.), *Handbook on the Economic Complexity of Technological Change*. Cheltenham, UK, Northampton, MA, USA: Edward Elgar, 366-394.

Carpenter, J., Kariv, S., & Schotter, A. (2012). Network architecture, cooperation and punishment in public good experiments. *Review of Economic Design*, 16(2-3), 93-118.

Chan, K. S., Mestelman, S., Moir, R., & Muller, R. A. (1999). Heterogeneity and the voluntary provision of public goods. *Experimental Economics*, 2(1), 5-30.

Chanaron, J. & Rennard, J. (2007). The automotive industry: a challenge to Schumpeter's innovation theory. In *Rediscovering Schumpeter: creative destruction evolving into "mode 3"*, edited by Carayannis, E. and Ziemnowicz, C., 320-343, New York: Sage.

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

Cherry, T. L., Kroll, S., & Shogren, J. F. (2005). The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab. *Journal of Economic Behavior & Organization*, 57(3), 357-365.

Chen, A. J., Boudreau, M. C., & Watson, R. T. (2008). Information systems and ecological sustainability. *Journal of Systems and Information Technology*, 10(3), 186-201.

Clark, K., & Sefton, M. (2001). Repetition and signalling: experimental evidence from games with efficient equilibria. *Economics Letters*, 70(3), 357-362.

Coats, J. C., & Neilson, W. S. (2005). Beliefs about Other-Regarding Preferences in a Sequential Public Goods Game. *Economic Inquiry*, 43(3), 614-622.

Dawes, R. M., McTavish, J., & Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of personality and social psychology*, 35(1), 1.

D'Este, P., Iammarino, S., Savona, M., & von Tunzelmann, N. (2012). What hampers innovation? Revealed barriers versus deterring barriers. *Research policy*, 41(2), 482-488.

Di Cagno, D., Galliera, A., Güht W., Pannaccione L. (2016), A hybrid public good experiment eliciting multi-dimensional choice data. Journal of Economic Psychology 56: 20-38.

Diekmann, A. (2008). Soziologie und Ökonomie: Der Beitrag experimenteller Wirtschaftsforschung zur Sozialtheorie. *KZfSS - Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 60(3), 528.

Egas, M., & Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1637), 871-878.

Falck, A., J.J. Heckman (2009), Lab experiments are a major source of knowledge in the social sciences. Science 326(5952): 535-538.

Figuieres, C., Masclet, D., & Willinger, M. (2012). Vanishing leadership and declining reciprocity in a sequential contribution experiment. *Economic Inquiry*, 50(3), 567-584.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics letters*, 71(3), 397-404.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2), 171-178.

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American economic review*, 100(1), 541-556.

Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science*, 322(5907), 1510.

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

Guala, F. (2005). The methodology of experimental economics. Cambridge University Press.

Hadjimanolis, A. (2003). The barriers approach to innovation. *The International Handbook on Innovation*, edited by Shavinina, L.V., 559–573, Amsterdam: Pergamon Press.

Hagedoorn, J. (1990). Organizational Modes of Inter-firm Co-operation and Technology Transfer. Technovation, 10(1): 17-30.

Hardin, G. (1968). The Tragedy of the Commons. *Science,* New Series, 162(3859), 1243-1248.

Hueske, A. K., & Guenther, E. (2015). What hampers innovation? External stakeholders, the organization, groups and individuals: a systematic review of empirical barrier research. *Management Review Quarterly*, *65*(2), 113-148.

Ledyard, J. (1995). Public Goods: A Survey of Experimental Research. *The Handbook of Experimental Economics,* edited by Kagel, J. & Roth, A.E., Princeton: Princeton University Press .

Levati, M. V., Sutter, M., & Van der Heijden, E. (2007). Leading by example in a public goods experiment with heterogeneity and incomplete information. *Journal of Conflict Resolution*, 51(5), 793-818.

Levitt, S.D., J.A. List (2008), Homo economicus evolves. Science 319 (5865), 909-910.

Link G.J., Siemon D., de Vreede GJ., Robra-Bissantz S. (2015) Evaluating Anchored Discussion to Foster Creativity in Online Collaboration. In: Baloian N., Zorian Y., Taslakian P., Shoukouryan S. (eds) Collaboration and Technology. CRIWG 2015. Lecture Notes in Computer Science, vol 9334. Springer.

Masclet, D., & Willinger, M. (2005). Does Contributing Sequentially Increase the Level of Cooperation in Public Goods Games? An Experimental Investigation. *LAMENTA Working paper*.

Nikiforakis, N., & Normann, H. T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11(4), 358-369.

Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review*, *46*(8), 1423-1437.

Olson, M. (1967). The Logic of Collective Action. Cambridge: Harvard University Press.

Pellegrino, G., & Savona, M. (2017). No money, no honey? Financial versus knowledge and demand constraints on innovation. *Research Policy*, 46(2), 510-521.

Roth; A.E. (1995). Introduction to experimental economics". *The Handbook of Experimental Economics,* edited by Kagel, J. & Roth, A.E., Princeton: Princeton University Press.

Rothgang, M., Busse, A., Dehio, J. & Dürig, W. (2018), Potentials, Technology Transfer, Innovation Barriers. Final Report of the subproject five of the lightweight forging research network. Düsseldorf, Aachen, Essen. FOSTA – Forschungsvereinigung Stahlanwendung e.V.

Rothgang, M., B. Lageman und M. Peistrup (2011), Industrial Collective Research Networks in Germany: Structure, Firm Involvement, and Use of Results. *Industry and Innovation* 18 (4): 393-414.

Sefton, M., Shupp, R., & Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic inquiry*, 45(4), 671-690.

Shang, J., & Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal*, 119(540), 1422-1439.

Steiger, E. M. & Zultan, R. (2014). See no evil: Information chains and reciprocity. *Journal of Public Economics*, 109(C), 1-12.

Sturm, B., & Weimann, J. (2006). Experiments in environmental economics and some close relatives. *Journal of Economic Surveys*, 20(3), 419-457.

Tan, J. H., & Bolle, F. (2007). Team competition and the public goods game. *Economics Letters*, *96*(1), 133-139.

Tang, H.K., & Yeo, K. T. (2003), Innovation under Constraints: The Case of Singapore, *The International Handbook on Innovation,* edited by Shavinina, L.V., 873-881, Amsterdam: Elsevier.

Teichert, T. (1993). The success potential of international R&D cooperation, *Technovation*, 13(8): 519-532.

Walker, J. M., & Halloran, M. A. (2004). Rewards and sanctions and the provision of public goods in one-shot settings. *Experimental Economics*, 7(3), 235-247.

Warr, P. G. (1982). Pareto optimal redistribution and private charity. *Journal of Public Economics*, 19(1), 131-138.

Warr, P. G. (1983). The private provision of a public good is independent of the distribution of income. *Economics letters*, 13(2-3), 207-211.

Winter, M. F. (2004). Developing a group model for student software engineering teams. *Doctoral dissertation*, University of Saskatchewan.

Zaheer, A. & Venkatraman, N. (1995). Relational governance as an introrganizational strategy: an empirical test of the role of trust in economic exchange. Strategic Management Journal, 16(5), 373-392

Conclusion

This dissertation uses economic laboratory experiments to investigate whether information influences individual decision making focusing on subjects' preferences to take (higher-order) risk and individual behavior in team settings. The results show that this is the case: individual higher-order risk preferences and behavior in team settings are influenced by information.

In case of higher-order risk preferences, the 1st Chapter summarizes that in non-incentivized surveys Chinese subjects are more willing to take risk than Germans and US-Americans, but studies using incentivized experiments suggest that this relationship is less clear. Investigating these mixed results with regards to the homogeneity of risk preferences, Chapter 2 reveals that most choices in the three different countries and after a tenfold increase in the stake size are consistent with a so-called mixed risk-loving or mixed risk averse behavior, i.e. risk loving in odd and risk averse in even orders, or i.e. risk averse in all orders. However, this behavior is strengthened if the lotteries are displayed in compound rather than reduced form. This indicates an information effect in this context. Chapter 3 analyzes further how the way of presentation of these lotteries (framing and lottery display type) affect the degree of higher-order risk preferences. It partly confirms the effect of information by showing that subjects' behavior is influenced by lottery framing (compound or reduced) but not by display format (urn or spinner design).

Chapter 4 and Chapter 5 investigate if information influences individual decision making in teams. The results of Chapter 4 indicate that, compared to a general wage increase, an increasing wage inequality can lead to a steeper rise in agents' performance when the subjects are informed of the wage distribution. This effect is driven by the individual degree of loss-aversion. Chapter 5 uses a framed experiment in a specific industry context. It reveals that when subjects have information on the distribution of others (in a sequential decision-making process), distributions to a joint research and development project increase as well as the overall welfare.

All results presented in this dissertation are obtained by using economic laboratory experiments. A limitation of experimental approaches is the lack of external validity outside the laboratory (see e.g. Levitt & List, 2007, 2008 for a critical discussion about experiments). They are conducted in a controlled environment, mostly with a sample of student subjects and in small samples. Investigating external validity of experimental data, Cox and Oaxaca (1991), for example, compare data obtained in an experiment and apply econometric tools as used in the analysis of field data to estimate supply and demand functions in a labor market experiment.

The authors find mixed results, indicating that their observations in the laboratory are only valid internally.

This raises the question whether the lack of external validity applies to the results of this dissertation as well. Indeed, the investigated samples consist of students and are in general relatively small. Yet, Chapter 3 provides results from a large sample of 827 subjects from three different countries that are consistent with the other findings on risk preferences. And Chapter 5 narrows the gap between economic laboratory experiments and decision making in the field by using a framed laboratory experiment to investigate how firms, occupying a unique position along the automotive industry value chain, may overcome factors hampering innovation processes.

Still, this does not hamper the value of this dissertation as it focuses on how information affects individual risk preferences and behavior in a team setting. To examine these questions, independent of other unobservable factors, economic laboratory experiments are the tool of choice (Smith, 1994). A compelling next step is to use the insights provided by this dissertation and test them with the help of larger (survey-)data: How does the way in which a subject is informed about the riskiness of investments influence her decisions. Do we observe mixed risk-loving and mixed risk-averse behavior regarding precautionary saving and/or portfolio choice? And does information about other team members' wage influence performance? But even in case of larger survey data, the question remains, whether results are generalizable or if they are just valid for a specific group (Bardsley, 2010).

References

Bardsley, N. (2010). Sociality and external validity in experimental economics. Mind & Society, 9(2), 119-138.

Cox, J. C., & Oaxaca, R. L. (1992). Direct tests of the reservation wage property. The Economic Journal, 102(415), 1423-1432.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world?. Journal of Economic perspectives, 21(2), 153-174.

Levitt, S. D., & List, J. A. (2008). Homo economicus evolves. Science, 319(5865), 909-910.

Smith, V. L. (1994). Economics in the Laboratory. Journal of economic perspectives, 8(1), 113-131.

Eidesstattliche Erklärung zu § 14 Abs. 1 Nr. 6

Ich gebe folgende eidesstattliche Erklärung ab:

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig ohne unzulässige Hilfe Dritter verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und alle wörtlich oder inhaltlich übernommenen Stellen unter der Angabe der Quelle als solche gekennzeichnet habe.

Die Grundsätze für die Sicherung guter wissenschaftlicher Praxis an der Universität Duisburg-Essen sind beachtet worden.

Ich habe die Arbeit keiner anderen Stelle zu Prüfungszwecken vorgelegt.

Ratingen, der 12.10.2020
Ort, Datum

Alexander Haering