Three experimental investigations of individual decision-making behaviour

Inauguraldissertation

zur Erlangung des akademischen Grades eines Doktors
der Wirtschaftswissenschaften des Fachbereichs
Wirtschaftswissenschaften
der Universität Osnabrück

vorgelegt
von

Nadia-Yasmine Kruse

Osnabrück, November 2022

Dekan/in:	Prof. Frank Westermann, Ph.D.
Referenten:	Prof. Steve Humphrey, Ph.D.
	Prof. Dr. Stefanie Engel

Tag der Disputation: 24. November 2022

Acknowledgements

"Life is just a series of peaks and troughs.

And you don't know whether you're in a trough until you're climbing out, or on a peak until you're coming down.

And that's it you know, you never know what's around the corner. But it's all good.

'If you want the rainbow, you've gotta put up with the rain.'

Do you know which 'philosopher' said that?"

David Brent (Ricky Gervais), The Office UK, 2002

First of all, I would like to express my gratitude to Prof. Dr. Stefanie Engel for examining my thesis, Prof. Dr. Bernhard Baumgartner for chairing the examination procedure and Mrs. Birgit Jatzkowski for patiently guiding me through all the administrative procedures during my doctorate.

These last years have been a journey with lots of peaks and troughs. And although I often thought I was completely alone, I never was. At every stage I have been surrounded by incredible people, to whom I would now finally like to express my deepest gratitude.

Before this journey had even started, I faced my first trough – and experienced the most selfless help from Vlad Marincas. Thank you Vlad, for helping me to start this journey in the first place, for impressing me with your ability to derive complex equations in your head and all the fun evenings that followed. During this time, you may not have realised what a great support you have been to me, but you have!

This journey has also ended with a trough – and completely unexpectedly, there was Dr. Sarah Forstinger. Sarah, it's rare to meet people who immediately make you feel like you've known them for a lifetime. Thank you for empathising with my emotional Ph.D. rollercoaster ride, delicious meals, trips in the forest with your dog Kutzu and most of all your warm support of my family during my disputation. I will never forget this!

From beginning to end of this journey, I could also always rely on my wonderful friends and I am grateful that they endured all the positive and negative side effects. Far more personal expressions of gratitude go to my family.

I am beyond grateful for my parents, Soumia and Manfred Kruse, whose endeavour it was to raise me to be a strong and independent woman. Having travelled the world and seen every conceivable form of social inequality, they (and especially my mother) placed special emphasis on our education. My parents made it possible for me to get this far to become a female economics Ph.D. student with a migrant background in Germany. I am aware that they had to make sacrifices for this and I am eternally grateful to them. Thank you Mum and Dad, for never letting me down and always encouraging me not to give up. Thank you, for always letting me know that you love me and 'not my grades'. For the last 5 years, you have been constantly tiptoeing around me, trying to understand my research work and wiping away my tears. Your unconditional love and support have kept me going – all that I am and have, I owe to you!

I also owe a debt of gratitude to my sister Sofia Kruse. Sofi, not only are you my little sister, but also my soul mate. You have been with me through every peak and every trough and I am pretty sure you have experienced all my emotions, too. It was you who always motivated me, hated it all with me, cheered along with me and never stopped believing in me. You have been my anchor in this journey. When one of us runs out of strength, the other has strength for two and that is what I love most about us. You are the glue that holds our family together and I really do appreciate this so much about you. Thank you Sofi, you are my biggest flex!

Relatedly, I would also like to thank the Humphrey family. Thank you Rudi, Mati and Sam, for the picture you drew on the whiteboard in my office. It brought me so much joy every time I came to work and it was so hard (literally) to remove it on my last day. Further, I would like to express my gratitude to Inka Humphrey. Inka, your critical debates on politics, education and feminism have impressed and inspired me beyond all measure. I enjoyed every discussion with you and took away so much from it. Thank you Inka, for criticising me objectively while genuinely caring about me.

This last section of my acknowledgements, I would like to dedicate to my mentor and doctoral advisor Professor Steve Humphrey, Ph.D. When I started working at his chair as a student research assistant, I quickly recognised his exceptional work ethic. This, to be honest, put a bit of pressure on me when I was accepted as his Ph.D. student. It can be intimidating to work for somebody who excels in every way and makes it look easy as pie. However, he patiently taught me everything he knew about structured work and successful teaching and research. Steve, you are the epitome of a doctoral supervisor and your enthusiasm for research is inspiring and contagious. This thesis is the product of your invaluable guidance, your honest criticism of (many of) my ideas and your enthusiastic and empathic encouragement. In these 5 years of my research work, there was not a single day when I could not rely on your constant support and valuable suggestions. By sharing your expertise, I was lucky to benefit from your structure and your dedication to detail while also keeping the 'bigger picture' in mind. I am grateful to have had the good fortune and honour to work for and with you – I learned a great deal from you!

Thank you Steve, for the many hours we spent laughing together in your office, for encouraging me to actually become Dr. Kruse and for the most amazing last 10 years. And to go back to the quote at the beginning: I know you know which 'philosopher' said that.

Contents

List of Figures

List of Tables

Data Statement

The data reported in thesis were collected with the informed consent of subjects according to the ethical guidelines of LaER. Data are stored electronically at Osnabrück University according to the provisions of the GDPR. Enquiries regarding data availability and use should be addressed to the Department of Microeconomics in Economics Faculty at Osnabrück University (www.mikro.uni-osnabrueck.de).

General overview

This thesis consists of three self-contained chapters on individual decision-making behaviour. The chapters address distinct research questions on important features of individual decision-making, which have broad-ranging importance in economc analysis. All three chapters focus on behavioural patterns that deviate from the predictions of standard economic theory. In addition to the common focus on individual decision-making behaviour, all three chapters employ the experimental method of empirical research in economics. The following paragraphs provide a synopsis of each chapter.

Chapter 1

Chapter 1 employs a classroom experiment exploring the impact of a stepwise acquisition of multiple unit bundles on the endowment effect. The endowment effect describes the tendency for the value assigned to a good to depend upon whether that good is owned or not. It suggests that the efficient operation of market trade may be disrupted by the matter of ownership which, according to the insights of the Coase theorem (Coase, 1960), should not be the case. Empirically, the endowment effect is typically studied by eliciting and comparing statements of willingness-to-pay (maximum buying prices) and willingness-to-accept (minimum selling prices) for goods, or by using swapping-tasks. Evidence reveals the endowment effect to be particularly robust. It has been observed for a variety of goods and subjects (Horowitz and McConnell, 2002).

Although the endowment effect has been widely studied, there is still a degree of ambiguity about the factors that affect it. For example, there is growing evidence that the endowment effect is diminished when decisions are made over more than one good (Burson et al., 2013). Other experiments, by contrast, suggest that the presence of an endowment effect might depend upon the acquisition of the traded goods being split into distinct steps (Humphrey et al., 2017).

This recent literature is important because it questions the robustness of much of the established empirical literature, which has been based upon experiments which used single unit of goods as the basis of valuation and swapping decisions. More generally, it suggests that the endowment effect may depend upon the make-up and acquisition of bundles of goods that are traded. This matter is important, because many trading decisions taken outside of the laboratory are not made in respect of single units of goods, but rather bundles of goods which may comprise of multiple units that have been sequentially acquired (e.g. over time, such as a home, a business, an art collection etc.).

The experiment reported in chapter 1 is designed to shed light on these issues by conducting a controlled test of the influence of multi-step acquisition of multiple-unit bundles on the endowment effect. The underlying motivation is that whilst the existing literature suggests that mutli-step bundle acquisition *may* matter to the observation of an endowment effect, this question has not yet been studied in a dedicated and fully controlled trading task experiment. Doing so is important because the results of such a test will shed light on the robustness of the endowment effect to factors which are arguably important to the external validity of standard experimental results.

Specifically, the experiment reported in chapter 1 compares behaviour between treatments where a multiple-unit bundle of goods is acquired in a single step and in multiple steps, respectively. For the goods used in the experiment, the results show evidence of a significant endowment effect manifest in a reluctance to swap a randomly acquired bundle in both treatments. Hence, these findings suggest that existing laboratory results from single-unit experiments are transferable to the multi-unit setting investigated here. There is evidence in the data that the nature of bundle acquisition matters to decision-makers, but it is not a sufficiently strong decision-motive to influence the endowment effects that were observed. In this respect, the results of the existing experimental literature, which typically use single-step endowment acquisition, appear to be robust.

Chapter 2

The experiment reported in Chapter 2 has two related objectives. The first is motivated by the *replication crisis* in experimental psychology, where the failure to replicate a substantial number of important experimentally observed findings has prompted an ongoing debate about the quality of previously published scientific results. This debate has encouraged replication studies to be more commonly conducted in the psychological and social sciences. Despite this, and despite replicability being an often-cited advantage of the experimental approach to empirical work in economics, replication studies in experimental economics remain relatively scarce. In order to address this deficiency, chapter 2 focuses on the replicability of five important patterns of behaviour observed in decision-making experiments: the common consequence effect (Starmer and Sugden, 1991), the common ratio effect (van de Kuilen and Wakker, 2006), preference-reversal choices (Loomes et al. 1991), event-splitting effects (Humphrey, 1995) and cyclical choices (Starmer, 1999).

All of these behavioural patterns have been influential in the evolution of the theoretical literature relating to individual decision-making under risk. Moreover, they have individually been subject to some amount of replication in different experimental laboratories. The experiment reported in chapter 2 extends this literature by conducting a controlled replication study of all five patterns of behaviour in a single *online* experiment. If previous laboratory experiments have successfully elicited genuine features of individual risk preferences, they should also be replicable in an online experiment. i.e. the experiment is a replication study that is conducted in a different environment to that which has been used in previous work.

Following from the 'pure replication motivation', the second objective of the experiment reported in chapter 2 is to provide an 'instrument check' of the online implementation of risky decision-making tasks. If online experiments were unable to replicate findings that have emerged from laboratories, this may mean that either results from laboratories are merely artefacts of the laboratory, or that online experiments are not appropriate instruments to observe genuine risk preferences. Given the increased use of online experiments in the social sciences, it is important to generate data to address this question.

From a practical perspective, most of the research reported in this thesis was conducted within the Covid-19 legal framework, which was in place in Germany at the time. This legal framework restricted laboratory research involving human subjects. It was therefore inevitable that the major study reported in chapter 3 (as well as the experiment reported in chapter 2) would be conducted online. Experimental control was therefore exercised in the selection of behavioural patterns to be studied in the online replication. Given previous work, there would be high confidence in replicating the behavioural patterns studied in this chapter in a laboratory experiment. Therefore, should they be replicated in the online study, it would be reasonable to conclude that the online implementation is a legitimate instrument to observe preferences (or, at least, is comparable to the laboratory for such purposes).

The main finding of the experiment is a replication rate of 82%, with 9 out of 11 replications being highly significant in the same direction as in the original laboratory studies. Each of the five different patterns of preferences is observed with statistical significance at least once. This is a higher replication rate compared to that generally observed in other disciplines (e.g. psychology or medicine). Moreover, although online experiments involve a loss of control relative to laboratory experiments (e.g. choices are made in the physical absence of the experimenter), the replication success in this study suggests that online experiments are appropriate for the elicitation of the genuine preferences of decision-makers.

Chapter 3

Chapter 3 reports the results of two online experiments, which study the normativity of the independence axiom of EUT. The experiments employ a modernised and updated version of a design pioneered by Slovic and Tversky (1974), which connects observed choices to underlying normative judgements. Subjects make choices over decision problems in which they can either be consistent with, or violate the independence axiom of EUT. They are then provided with the opportunity to reflect upon their decisions, in light of normative arguments, which favour the counterfactual behaviour, before being asked to repeat their choices. The repeated choices are therefore made in light of normative arguments and, in this respect, can be interpreted as being based upon a normative judgement.

The motivation for the study is drawn from an argument provided by Friedman and Savage (1952), who defend the axioms of EUT as requirements of rationality by claiming that the normative appeal of those axioms can be considered 'indirect' evidence for them. Starmer (2005) argues that for this proposition to hold, it requires that the normativity of the axioms is empirically valid (rather than asserted on the basis of introspective plausibility). The purpose of the experiments reported in chapter 3 is to investigate whether such validation can be provided in respect of the independence axiom.

The experiments reported here update Slovic and Tverksy's (1974) design such that it meets modern standards of experimental research in terms of (large) sample size and the implementation of incentive compatible decision tasks. The central decision problems are derived from the (first) Allais Paradox (Allais, 1953) and test the implications of the independence axiom. Slovic and Tversky's (1974) design is also extended with additional treatments to control for confounding factors that could be present in their original design: Namely experimenter demand effects and a status quo bias. A final treatment extends the study to the second version of the Allais Paradox (the common ratio problem), to test whether the normative arguments affect behaviour in the same way for both versions of the Allais Paradox. This treatment thereby serves as a robustness check.

The results show that, across all treatments, significant independence violations do not decrease after the presentation of the normative arguments. The control treatments reveal that the findings are not attributable to experimenter demand effects or a status quo bias, and are robust to both versions of the Allais paradox. These findings therefore do not support the normativity of the independence axiom. Consequently, they neither support Friedman and Savage's (1952) proposition that the introspective normative appeal of EUT's axioms counts as evidence of normativity of those axioms.

Summary

The three experiments reported in this thesis are examples of what Colin Camerer (1995, p.656) describes as experimental 'detective work'. That is, they conduct carefully controlled studies that extend the existing literature by identifying and testing factors which may (or may not) influence behaviour in decision-making experiments. In summary, it has been observed that the endowment effect was not significantly influenced by multi-part endowment acquisition, that important risk preferences observed in experimental laboratories can be successfully replicated in an online experiment, and that the independence axiom of expected utility theory is not considered to be normative by decision-makers themselves. These findings should of course be subject the usual robustness checks that should accompany any programme of experimental economics research. Finally, it is hoped that the findings reported in this thesis are fruitful in stimulating future research in the behavioural economics of individual decision-making.

References

Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica, 21*, 503–46.

Burson, K., Faro, D., Rottenstreich, Y. (2013). Multiple-Unit Holdings Yield Attenuated Endowment Effects. *Management Science, 59*(3), 545-555.

Camerer, C.F. (1995). Individual Decision Making in *Handbook of Experimental Economics*. J. Kagel and A. E. Roth, eds. Princeton: Princeton University Press.

Coase, R.H. (1960). The problem of social cost. *Journal of Law and Economics, 3*, 1-44.

Friedman, M., Savage, L. (1952). The expected-utility hypothesis and the measurability of utility. *Journal of Political Economy, 60,* 463–74.

Horowitz, J. K., McConnell, K. E. (2002). A review of WTA/WTP studies. *Journal of Environmental Economics and Management, 44*(3), 426–447.

Humphrey, S.J. (1995). Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *Journal of Risk and Uncertainty, 11*(3), 263–74.

Humphrey, S.J., Lindsay, L., Starmer, C. (2017). Consumption experience, choice experience and the endowment effect. *Journal of the Economic Science Association 3*(2), 109–120.

Loomes, G., Starmer, C., Sugden, R. (1991). Observing Violations of Transitivity by Experimental Methods. *Econometrica, 59*(2), 425–39.

Slovic, P., Tversky, A. (1974). Who Accepts Savage's Axiom? *Behavioral Science, 19,* 368-373.

Starmer, C. (1999). Cycling with Rules of Thumb: An Experimental Test for a New Form of Non-Transitive Behaviour. *Theory and Decision, 46*, 141–58.

Starmer, C. (2005). Normative notions in descriptive dialogues. *Journal of Economic Methodology, 12(2),* 277-289.

Starmer, C., Sugden, R. (1991). Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation. *American Economic Review, 81*, 971-978.

van de Kuilen, G., Wakker, P. (2006). Learning in the Allais Paradox. *Journal of Risk and Uncertinty, 33*, 155-164.

Chapter 1

On the role of sequential multi-unit bundle acquisition in the endowment effect

1 Introduction

This study is concerned with the endowment effect (Thaler, 1980; Kahneman et al. 1990, 1991). This pattern of behaviour is known as the tendency for individuals to assign a higher value to goods they own than if they do not own them. In the economics literature, the endowment effect is tested using ‘valuation tasks’ or ‘swapping tasks’. Studies using ‘valuation tasks’ have shown a significant gap between willingness-to-accept and willingness-to-pay valuations for goods, which has also been considered evidence of the endowment effect (Knetsch and Sinden, 1984). The experiment reported in this chapter uses a variant of Knetsch’s (1989) classic ‘swapping task’ where subjects randomly receive one of two goods and then are asked whether they want to swap the good they own for the alternative good. Evidence of the endowment effect has been found for a variety of subjects (i.e. children, students, non-student adults and traders) and goods (i.e. public or non-market goods, ordinary private goods and lotteries). Because it is not restricted to specific goods or fields, there is still a degree of ambiguity about the determinants of the endowment effect.

As the endowment effect produces irregularities in valuation and disrupts the efficient distribution of entitlements, its economic implications for consumer welfare and the efficient operation of markets are consequential. This is why further investigations of the endowment effect are important and worth undertaking.

This chapter is concerned with recent literature regarding the importance of the composition of bundles of goods that may be traded and how those bundles are acquired. This approach might provide a more realistic picture, as decisions are more likely to be made over bundles rather than single goods in real-life situations. Although research on multiple-unit experiments is relatively scarce, there is growing evidence suggesting a decreasing endowment effect when decisions are based on more than one good (Burson et al., 2013). While experiments on multiple unit holdings indicate that the endowment effect decreases, other experiments show that it re-emerges when bundles are acquired in distinct steps (Humphrey et al., 2017).

This study reports the results of a swapping task experiment using bundles comprising two units of the same good. The experiment involves two treatments, that differ in the acquisition of the bundles. In the first treatment, subjects are randomly assigned a bundle comprising two units of the same good all at once, while in the second treatment, the acquisition of the bundle is divided into two distinct steps. There is evidence that bundles matter, but this study specifically aims to investigate whether acquiring bundles in distinct steps influences the endowment effect. The next section discusses background theory and literature on the endowment effect, multiple-unit holdings and split endowment acquisition, followed by the formal hypotheses in section 3. Section 4 sets out the experimental design and implementation. Section 5 presents results. Section 6 reports the implications of the results and sets out general concluding remarks.

2 Literature discussion

2.1 Classic EE literature

The endowment effect has sustained interest for roughly four decades. Knetsch (1989) reports results from an experiment using a swapping task design where subjects are equally divided and randomly endowed with either a chocolate bar or a coffee mug. Each subject is given the opportunity to swap their initial endowment for the other good. In this case, standard preference theory would predict a 50% trading rate. Assume that a proportion p of the population prefers good x over good y (and so proportion $1-p$ prefers y to x). The population is then randomly allocated one of the goods. For the half endowed with x, the fraction $1-p$ will swap their goods. Of those allocated y, the proportion p will swap. So the predicted trading rate is ½ p + ½ $(1-p)$

= ½. However, the proportion of subjects choosing to swap in Knetsch's (1989) experiment is less than 50%, which differs significantly from the theoretical prediction of a 50% trading rate.

There is a broad range of experimental designs to determine the antecedents of the endowment effect. Swapping tasks as used by Knetsch (1989) generally involve two goods that can either be kept or traded. Experiments of this genre include a variety of different goods and are conducted with different subject types in the laboratory and in the field. For instance, List (2003) examines trading patterns of inexperienced traders using sports memorabilia at a sports card show and collector pins at a Walt Disney World market. In one setting, subjects are randomly endowed with one of two unique pieces of sports memorabilia. After filling in a survey, the other item of sports memorabilia is shown to the subjects and they get the opportunity to trade their initial one for the other. The same procedure is applied in the Walt Disney World market using two different collector pins. Both settings find support for the existence of an endowment effect in the field. Another experiment of similar design is conducted in the laboratory where subjects are asked to trade everyday consumables such as mugs, pens, chocolate bars etc. Again, List (2003) reports significant evidence of an endowment effect. In a follow-up experiment, List (2004) transfers the consumable goods used in the laboratory experiments to the field setting and provides trading results that are consistent with his previous findings. Hence, he shows that endowment effects are found for a wide range of goods in different market settings.

Engelmann and Hollard (2010) use a similar design in a laboratory setting, where students randomly receive one of two goods (i.e. a package of coffee or a package of rice, mugs or chocolate) and are free to trade one good for the other. Here, in line with List's (2003) findings, subjects' reluctance to trade leads to a significant endowment effect.

Another approach to studying the endowment effect is using valuation tasks in experiments. For instance, Kahneman, Knetsch and Thaler (1991) run a number of experiments where they compare the behaviour of buyers and sellers of physical goods. Subjects participate in a series of markets, where half the subjects receive a mug and sell it to those who do not own a mug. The sellers are asked to value the mug they possess by stating a price at which they would be willing to sell the mug. Buyers are asked to set a price at which they would be willing to buy the mug. In another setting, the same experiment is conducted with pens. The prediction of a 50% trading rate does not occur because in either setting, the median selling price is nearly twice as high as the median buying price. Even after replicating this experiment several times, buyers and sellers are unwilling to agree on a price at which to trade.

This discrepancy between the willingness-to-accept (hereafter WTA) and the willingness-to-pay (hereafter WTP) valuations is another interpretation of the endowment effect. A large body of evidence shows that this gap between WTA and WTP is quite robust (Knetsch (1989); Corsey at al. (1987); Bateman et al. (1997)). There is a broad range of explanations of the nature of the endowment effect. A common interpretation of the endowment effect, due to Kahneman, Knetsch and Thaler (1990) is that it stems from loss aversion. Loss aversion is the observation that losses exert a disproportionately large influence on decision-making behaviour than gains of an equivalent magnitude.

Referring to Prospect Theory (Kahnemann and Tversky, 1979), the shape of the utility functions from gains and losses are different. The value function is composed of a concave function for gains and a convex function for losses. Since its slope is steeper for losses than for gains, it serves as a theoretical account of loss aversion. For example, if an individual is selling a mug, that involves the individual “giving it up” or “losing it”. Hence, if individuals are loss averse, this inflates the selling price for the mug relative to what they would be willing to pay for the same mug if they did not own it.

Kahneman, Knetsch and Thaler (1991) report that findings from induced attractiveness rankings in similar experiments indicate that the endowment effect does not arise because of the increased attractiveness of the endowed good, but the pain of giving it up. This ‘giving it all up’ effect also appears when experiments are conducted with real exchanges, hypothetical questions, choices between goods and money and valuations that can be affected by wealth effects (Knetsch, 1989).

Turning to consider the evidence for the endowment effects in valuation tasks in more detail, there is a wide range of goods and subjects that are examined within such experiments. Thus, various explanatory approaches to the observed discrepancy between WTA and WTP are reported.

Horowitz and McConnell (2002) conduct a review of 45 studies which all report a disparity between WTA and WTP. They provide a large record of endowment effect experiments using a variety of goods such as e.g. public or non-market goods, health and safety, ordinary private goods or lotteries. With respect to these categories, they find the gap between WTA and WTP to be the highest for non-ordinary goods (i.e. non-market and public goods) and lower for ordinary goods. Furthermore, they find that results in real experiments do not differ from hypothetical experiments and moving experiments from the laboratory to the field does not lead to a decrease of the endowment effect. As shown, the endowment effect is repeatedly observed in many different settings.

Looking at factors that might erode the endowment effect, Horowitz and McConnell (2002) find mixed but not strong evidence on the effects of repetition. For instance, Shogren et al. (1994) provide strong evidence of a diminishing endowment effect between their first and third round using a repeated auction design. As opposed to this, Kahneman et al. (1990) and Morisson (1997) suggest that repetition is not causing WTA and WTP valuations to converge to predicted levels over time. Thus, Horowitz and McConnell (2002) claim that large discrepancies between WTP and WTA are not due to experimental design features.

Contrary to this, Plott and Zeiler (2005) report that it is important to examine whether the endowment effect arises because of the nature of preferences or simply because of subject misconceptions of experimental tasks. Based on this assumption, they conduct controlled experiments, integrating several procedures found in the literature to test whether an elimination of the gap between WTA and WTP might be viable.

Hypothetically, they claim that if it is possible to turn an endowment effect on and off just by changing experimental procedures, this gap would not be induced by the theoretical assumption of loss aversion. To rule out that either the subject pool or the experimenters may diminish the gap, they replicate Kahneman et al.'s (1990) experiment using the same procedures. The results are consistent with those reported by Kahneman et al. (1990).

In a second step, experiments which control for subject misconceptions are conducted. The complete set of control includes special account of detailed explanation of mechanism and how to arrive at valuations, incentive-compatible mechanisms, paid practice rounds and consideration of anonymity. In these experiments, after 14 paid practice rounds using lotteries, subjects are asked to value a coffee mug in the last round. Plott and Zeiler (2005) report that there is no observable gap between WTA and WTP for the mug and therefore data does not support the hypothesis of WTP valuations being higher than WTA valuations.

They claim that the occurrence of an endowment effect can be reduced or even eliminated by the introduction of market experience through learning mechanisms (in this case the learning experience provided by the practice rounds involving lotteries). They state that the implementation of features, that control for subject misconceptions about the nature of experimental tasks, attenuates the endowment. Hence, they conclude that observed gaps do not reflect a fundamental feature of human preferences.

Isoni, Loomes and Sugden (2011) question Plott and Zeiler's (2005) interpretation of their findings. They replicate Plott and Zeiler's (2005) headline finding of no endowment effect for the task involving mugs. They also find evidence of the endowment effect within the paid practice rounds using lotteries. However, the endowment effect found for lotteries is fully discussed in terms of what it reveals regarding the role of learning in endowment effect experiments. They argue that this evidence does not fully support subject misconceptions as the reason for the emergence of an endowment effect. Rather it casts doubt over whether what Plott and Zeiler (2005) refer to as control mechanisms can be properly considered to be so, and leaves Isoni, Loomes and Sugden (2011) to conclude that they may better be interpreted as treatment variables.

The above debate over the correct interpretation of experimental findings depicts the issue of how experience influences the robustness of the endowment effect to be subject to an ongoing discussion. One part of this debate involves growing evidence suggesting that market experience may diminish the endowment effect.

For instance, List (2003, 2004) and Engelmann and Hollard (2010) observe significant endowment effects with inexperienced traders using variants of classic swapping tasks. However, their research also shows that prior trading experience diminishes the endowment effect. In a market, where experienced traders are used as subjects, List (2003, 2004) shows that there is no endowment effect present. Also, when subjects are "forced" to trade in tasks prior to decisions over endowments, the reluctance to trade in subsequent endowment effect tasks decreases subsequently (Engelmann and Hollard, 2010). Similarly, Loomes, Starmer and

Sugden (2003) report a decline in the disparity between WTA and WTP in repeated median price auctions when lottery values are elicited.[1]

Although the evidence on the robustness of the endowment effect is mixed, all of the experiments discussed above had the use of tasks involving single units of goods in common. When this special feature of experiments is considered in relation to the real world, it becomes obvious that people often hold and trade bundles of goods as opposed to just single units. Consequently, experimental studies have begun to study the role played in the endowment effect by multiple unit holdings, and to investigate the possibility that the endowment effect is magnified by the influence of "giving it all up". Since single-unit experiments of any kind suggest that a subject has to give up their entire holding of a good in exchange for money or another good, low trading rates could be explained by loss aversion induced by the requirement of relinquishing the entire holding of a good.

Assuming that preferences are reference dependent (Thaler, 1980), loss aversion is the notion that negative changes (losses), relative to a subjective reference point, loom larger than positive changes (gains). Attributing the endowment effect to loss aversion, subjects' reluctance to trade may be strongly influenced by the loss of the endowed good outweighing the gain of the alternative good. Because subjects have to give up their only and therefore last good, loss aversion could be considered a strong influence on behaviour and this may explain the robustness of evidence regarding endowment effects.

A natural research question which follows from this possibility, is that if subjects are reluctant to part with their possessions just because they have to "give it all up", will there be a difference in trading and, in particular, a reluctance to trade when they hold more than one

[1] In WTA tasks, an auction imposes market-line conditions on decision-makers by, for example, buying one lottery (demand) from a group of sellers (supply), who therefore make their reservation price decisions whilst facing selling competition from the other sellers (because supply exceeds demand).

good (or more than one unit of a good), and are not necessarily required to give up their entire endowment?

2.2 Multiple-unit holdings

Several studies addressed the question of what happens to the endowment effect when decisions are taken in respect of multiple unit bundles. The essential hypothesis is that if the fear of "giving it all up" drives the endowment effect, trading reluctance should be mitigated in settings where subjects have to part with their possession, but not in its entirety.

For instance, Horowitz, McConnell and Quiggin (1999) report a study involving multiple identical goods. In the experiments, subjects are endowed with three units of a good and asked to state the minimum acceptable price to return one, two or three units. They find that subjects demand significantly more when they are asked to give up the entire endowment than when asked to part with only some of it.

Relatedly, Burson, Faro and Rottenstreich (2013) report experimental evidence on multiple unit holdings using both swapping and valuation tasks. The valuation experiment is composed of two single-unit and two multiple-unit treatments. In the single-unit treatments, subjects are endowed with one piece of chocolate or one box of chocolate, whereas in the multiple-unit treatments subjects encounter 5 or 25 pieces of chocolate. Results from both single-unit treatments are in line with previous work. Here, the median selling price is more than twice as high as the buying price, and strongly supports findings on the endowment effect when only one good is involved. In multiple-unit treatments, the gap between WTA and WTP continues to close the more goods are involved. Within the swapping experiments, subjects are asked to trade chocolate for pens, or vice versa. Again, subjects are endowed with single units, boxes or multiple loose units within the respective treatments.

As in the valuation experiments, the endowment effect diminishes the more units of goods are at stake. The authors suggest the possibility of unit dependence to be one possible explanation for their results. Since a box containing multiple units of chocolate reveals an endowment effect, people might take one box as the relevant unit. With multiple units, the endowment effect is attenuated in both settings. An additional possible explanation of these findings that the authors suggest a stronger attachment to single units compared to multiple goods and that this difference accounts for the declining endowment effect with multiple goods.

Since Burson et al. (2013) only enabled subjects to trade either none of their endowment or all of it, Schurr and Ritov (2014) are led to investigate whether the distinction between partial trades and trades, where the entire endowment is given up, is important to the endowment effect. They present experiments using valuation tasks, where, as in previous classic experiments, subjects are divided into groups of buyers and sellers and asked to state their maximum buying and minimum selling prices for chocolate bars and pens. In the first two experiments, sellers receive different amounts of units of the goods and are asked to state their minimum selling price (WTA) for all the units at once, or a part of the endowed units and buyers are asked to state their maximum buying prices (WTP) correspondingly. When the entire endowment (consisting of one or more units of the goods) is at stake, selling prices exceed buying prices throughout, whereas no significant difference occurs when only parts of the endowment are involved.

In the third experiment, buyers and sellers generally receive the same number of goods in two treatments, but the manner of acquisition is slightly different between treatments. In the first treatment, sellers and buyers receive the same number of goods as an initial endowment after completing a questionnaire. Thereupon, only the sellers receive one additional unit of the good and are asked to state their minimum selling price for this one good. The buyers are asked to state their maximum buying price for one unit of the good.

In the second treatment, sellers initially receive one more unit of the good than buyers after completing a questionnaire. Then, sellers are asked to state their minimum selling price for one unit of the good and buyers are asked to state their maximum buying price for one unit of the good. Although both treatments are identical with respect to the number of goods held, simply the framing is different. The first treatment uses a framing of giving-it-all-up, as the sellers receive the additional unit in a subsequent step and are then asked to state their selling price.

The second treatment uses a framing of giving-some-up as sellers are asked to state their selling price for one unit of their entire initial endowment. Sellers demand significantly higher prices than buyers are willing to pay in the first treatment. By contrast, there is no significant difference between selling and buying prices in the second treatment.

The authors argue that framing the trade of the additional single unit of the good in the first treatment as giving-it-all up yields a significant disparity between WTA and WTP valuations. The endowment effect attenuates when subjects detect the trade as a partial relinquishment of their entire endowment, as in the second treatment. They conclude that the occurrence of the endowment effect depends on the difference between initial and final states. Changes of final states may be regarded as an important reference point when valuations are determined. Therefore, they relate their findings on the giving-it-all-up effect to loss aversion and conclude that this effect is a determinant factor of the endowment effect.

Although there are different explanations of the endowment effect, there is consensus on the role of multiple unit holdings in contributing the endowment effect. The results suggest that people tend to be more prone to trade when multiple units of a good are held. There are different theoretical and empirical interpretations on the determination of the endowment effect and how it can be eroded, but the evidence is still mixed. Compared to single unit experiments, research regarding multiple-unit endowments is relatively scarce. However, the research agenda clearly signifies a trend in which the importance of studying decision-making behaviour in respect of bundles of goods and discussing the implications of findings for real-world decisions-making is growing. The experiment reported in this chapter can be seen as a part of this literature. The experiment will contribute to understanding the importance to the endowment effect of the manner in which endowments comprising multiple units are acquired. The importance of acquiring bundles of goods, over which subsequent trading decisions will be made in a number of distinct stages, will be of specific interest.

2.3 Evidence on split endowment acquisition

Loomes, Orr and Sugden (2009, hereafter LOS) study the influence of taste uncertainty on the endowment effect. They show that the endowment can arise as a consequence of taste uncertainty if decision-makers are loss-averse. Humphrey, Lindsay and Starmer (2017, hereafter HLS) report an experiment testing the key insights of the LOS model. Their experiment is divided into four treatments. All treatments are run with bundles, which consist of unfamiliar consumption goods (premium organic vegetable crisps and handmade organic lemonade). These bundles either comprise of two bottles of lemonade and one packet of crisps (CLL) or the vice versa (CCL).

The treatments BASELINE and TASTING are modified versions of Knetsch's (1989) swapping task experiment. In both treatments, subjects are randomly endowed with either a 'crisps-rich' (CCL) or a 'lemonade-rich' bundle (CLL) and given the opportunity to keep or swap their initial endowment for the other one. The only difference between the BASELINE and TASTING treatments is that subjects in the TASTING treatment consume a small sample of both goods before receiving their bundles. The comparison of the two treatments allows a test of the impact of taste uncertainty on the endowment effect. According to the LOS model, the endowment effect is expected to be weaker in the TASTING treatment relative to the BASELINE treatment.

In the treatments PASSIVE and CHOOSING, the acquisition of the endowed bundles is split into two distinct steps. In the PASSIVE treatment, subjects are initially randomly endowed with one of the two goods, whereas subjects in the CHOOSING treatment choose one of the goods to be part of their endowment. Then, in both treatments, subjects are randomly topped-up with two more goods so that they arrived at a final (randomly determined) bundle of either CCL or CLL. They then face a straightforward swap decision between the bundle they hold and the alternative bundle. The comparison of the two treatments allows a test of the hypothesis that the experience of choosing between the goods prior to the swapping task reduces the endowment effect. Thus, HLS expect to find more swaps in the CHOOSING treatment than in the PASSIVE treatment (and they did).

In the comparison of the BASELINE and TASTING treatments, a test of whether tasting the goods prior to the swapping task matters to the endowment effect is not possible, because no endowment effect has been observed in the BASELINE treatment.[2] Although it is not the primary focus of HLS's study, between-treatment testing reveals a clear difference in the strength of the endowment effect between the treatments where the bundles are acquired in two distinct steps, relative to a single step acquisition. In the PASSIVE and CHOOSING treatments, the endowment effect is stronger than in the other two treatments.

Further, HLS report the results of a follow-up experiment. One purpose of this experiment is to investigate whether the lack of an endowment effect in the BASELINE treatment could be due to the use of unfamiliar goods. Using the same goods as in the original experiment, they replicate their BASELINE treatment and run an additional treatment which is similar to the BASELINE treatment but uses single items of the goods. Again, they find no significant endowment effect in the BASELINE replication. However, there is a significant endowment

[2] Although swap rates in TASTING are higher than in BASELINE, the difference is not statistically significant.

effect in the treatment using single items. This suggests that the nature of the goods used did not lead to the lack of endowment effect in the BASELINE treatment.

Rather, and analogously to the unit dependence approach of Burson et al. (2013) on multiple-unit holdings, the explanation seems to lie within the use of multiple unit-bundles. Specifically, in the HLS experiment, it seems that the use of bundles, coupled with the fact that swapping decisions always allowed subjects to keep at least a single unit of each good (and they therefore did not have to "give it all up"), diminished the endowment effect in the BASELINE session (Humphrey et al., 2017). Consequently, HLS conclude that the strength of the endowment effect may indeed depend on the different acquisition of the bundles: An endowment effect was absent when bundles were acquired in one step, but re-emerged for bundle acquisition in two steps.[3] HLS suggest, that the results driven from their two-step treatments may be related to other work on "splitting-effects".

There is a large body of evidence on features of splitting-effects, being described as a tendency for decisions to be weighed more heavily when the related attributes are unpacked and considered as two sub-categories than as a single category (Starmer and Sugden, 1993). For instance, in lottery choice tasks, Starmer and Sugden (1993) and Humphrey (1995) show that unpacking risks into smaller rises (i.e. a 0.7 chance of 11€ is split into a 0.3 and 0.4 chance of getting 11€) causes those risks to be weighed more heavily in the decision[4].

Weber, Eisenführer and Winter (1988) report evidence on effects on splitting attributes (i.e. splitting the "performance" attribute of a car into "top speed" and "acceleration") in multiattribute utility evaluation studies that are similar to splitting-effects. Tversky and Koehler (1994) interpret the influence of frequent splits of attributes on people's behaviour as evidence

[3] This conclusion could be linked to the presence of an endowment effect in 'the giving it all up' framed experiment in Schurr and Ritov's (2014).
[4]More generally, they claim event-splitting effects to occur if $\pi(p)+\pi(q)>\pi(p+q)$, referring (among others) to subadditivity in prospect theory's probability weighting function $\pi(.)$ (Humphrey, 1995).

of an unpacking effect (Humphrey, 2006). In related work, Bateman et al. (1997) suggest higher valuations for parts of endowment compared to bundles. In contingent valuation studies, they find evidence of part-whole bias for private consumption goods. Since splitting-effects are considered across a broad range of decision contexts, they might also explain why the endowment effect emerges when the acquisition of bundles is split into two steps.

An alternative approach to the determination of the endowment effect is presented by Morewedge et al. (2009) in an experiment, where the acquisition of the endowment is split. They base their assumptions on the occurrence of the endowment effect on a psychological foundation, the so-called mere ownership effect. It implies that the value of a good increases, because owning the good raises an association between the good and the self (Gawronski, Bodenhausen and Becker, 2007). Also, when people have not chosen the good, it becomes extraordinarily attractive to them simply because they own it (Beggan and Scott, 1997). Therefore, psychologists suggest that this specific valuation of ownership might drive the endowment effect. Since the endowment effect is frequently described as a pure and robust finding attributable to loss aversion (Rozin and Royzman, 2001), Morewedge et al. (2009) aim to distinguish between the effect of ownership and loss aversion. This is because both effects are closely related, not only in experimental settings but also in the real word. The authors design an experiment using valuation tasks that allows to test both effects separately. Through a direct comparison, they intend to draw a conclusion about which effect primarily drives the endowment effect.

In their experiment using valuation tasks, subjects are divided into four categories called owner-sellers, nonowner-buyers, owner-buyers and nonowner-pair buyers. Since the endowment effect suggests ownership of a good to increase its valuation relative to non-ownership, owner-sellers are given a mug whereas nonowner-buyers are not. Then, owner-sellers are asked to state their minimum selling price (WTA) and nonowner-buyers are asked to state their maximum buying price (WTP) for the mug. The owner-buyers are initially provided with a mug and are then asked to state their maximum willingness-to-pay (WTP) for a second similar mug. Nonowner-pair buyers are not provided with a mug and are asked to state their maximum willingness-to-pay (WTP) for a pair of two mugs.

In line with expectations and previous evidence, comparing valuations of the mug between owner-sellers and nonowner-buyers reveals a significant endowment effect. Comparing the per-unit valuations of nonowner-buyers and nonowner-pair buyers, they find nonowner-pair buyers to value two mugs twice as high as nonowner buyer value one mug[5].

A critical distinction between ownership effect and loss aversion is drawn from owner-buyers' valuations. Here, owner-buyers have a similar valuation of the second mug (WTP) as owner-sellers have for their mug (WTA) and a higher valuation than the nonowner-buyers (WTP). Thus, as owner-buyers and owner-sellers have similar valuations for a mug they both actually own, the authors conclude that the ownership effect eliminated the endowment effect here. According to their assumptions, loss aversion would have predicted similar valuations from owner-buyers and non-owner buyers because both groups would consider the transaction as a gain (Morewedge et al., 2009).

The conclusion reached by Morewedge et al. (2009) is that the effects they observed are driven solely by ownership. This means that the subjects who own one mug already, are willing to pay more for an additional mug, than subjects who do not own a mug. It is this increase in valuation of the owner-buyers that leads to the elimination of the endowment effect when compared to the valuation of owner-sellers. Morewedge et al. (2009) claim that the ownership effect can also explain the increase in valuation for goods that are owned for a longer period of time (Strahlivetz and Loewenstein, 1998) and can help to explain why the endowment effect

[5] Comparing the per-unit valuations of nonowner-buyers and nonowner-pair buyers served as an additional test to control for diminishing marginal utility and complementarity.

diminishes with additional experience in markets (List 2003, 2004). Moreover, they discuss why ownership effects can lead to stronger endowment effects for goods that one can relate to, such as mugs with a college logo (Tom, 2004) or goods that serve as a reward (Loewenstein and Issacharoff, 1994).

Although ownership seems to offer a straightforward interpretation of their results, the literature on multiple-unit holdings suggest that this explanation may not be as clear-cut as they claim. Evidence from experiments using multiple units of goods suggest that valuations are unit-dependent and selling prices decrease when subjects do not have to give up the last unit of their endowment (Schurr and Ritov, 2014; Burson et al., 2013). Furthermore, the increased valuation of the second mug for the owner-buyers might be related to the procedure of acquisition, rather than the ownership effect. In contrast to the nonowner-pair buyers, the owner-buyers received their total endowment in two distinct steps. The nonowner-pair buyers had a similar valuation of a pair of mugs as owner-buyers had for the additional mug. Thus, for the owner-buyers, the initial mug and the second mug might be unpacked into sub-components, which might have caused an increased valuation of the second mug. This pattern is a common feature of splitting-effects and may provide an alternative contemplation of reported results (Bateman et al. 1997, Humphrey 1995).

3 Hypotheses

The experiment reported in this chapter tests the influence of acquiring an endowment in stages on the strength of the endowment effect. Whilst the literature discussed in the previous section suggest that this is important, it is not the main hypothesis tested in those experiments. Therefore, they do not control for other potentially confounding influences on behaviour necessitated by the primary hypotheses that were tested. In this experiment, controls for these factors are introduced to understand the pure role (if any) of multiple unit endowment acquisition.

In relation to previous experiments on this topic, some important changes are made in this experiment. There are two treatments where the time duration between the two treatments is controlled from initial endowment with the goods and a swapping decision. Thus, the length of time between owning a good (at least one unit) an the subsequent decision is the same for both treatments, assuming that attachment between the good and the self does not depend on the quantity of goods[6]. Contrary to Morewedge et al. (2009), decisions are made at the same final state in both treatments, following the assumption that changes of final states may be regarded an important reference when valuations are determined (Schurr and Ritov, 2014).

Moreover, swapping decisions do not allow for trading parts of the endowment. Subjects can either keep their entire endowment or give it all up. The procedure in this experiment is in a manner analogous to Humphrey et al. (2017) except from the composition of the bundles. They use mixed three-unit bundles with two different goods where subjects therefore hold either a crisps-rich bundle (CCL) or a lemonade-rich bundle (CLL). Therefore, when swapping decisions are made (CCL for CLL or vice versa), subjects do not have to give up their final unit of an endowed good. This may explain why there was no endowment effect in their single step BASELINE treatment and suggest that people are less reluctant to trade when no giving it all up effect is present. However, the endowment re-emerges for decisions on the bundles that are acquired in two steps, although the features of the bundles are similar in the single step and two step treatments. Although subjects do not have to give up their final unit of an endowment in the two step treatment, they may value the bundle more heavily because it is acquired in distinct stages.

[6] Referring to Morewedge et al. (2009), this is a feature of the mere ownership effect.

This experiment is focused on the importance of acquiring bundles of goods in a number of stages. Contrary to Humphrey et al. (2017), subjects are endowed with bundles comprising two units of the same good that they have to give up when they decide to trade. This setting allows for a controlled test whether splitting the acquisition of the endowment matters where duration of ownership (of at least one unit) and final states are the same.

Based on this design and the literature discussed, the hypotheses tested are:

- Incorporating the feature of giving it all up into the swapping decisions, both treatments provide evidence of an endowment effect.
- A stronger endowment effect in the two step treatment supports the importance of splitting effects in the acquisition of the endowment.

The experimental design embodies specific features that have been proven to be relevant to the nature of the endowment effect. Even though the implementation is very similar to Humphrey et al.'s (2017) experiment, relevant changes within the design allow to draw differentiated conclusions. Holding homogenous bundles of two goods in a setting where 'giving it all up' matters, will provide further understanding of the impact of the composition of bundles on the endowment effect. Moreover, testing whether a difference in acquisition of the same bundles matters, this experiment will enhance understanding of the impact of splitting effects on the endowment effect.

4 Experiment

4.1 Design

The hypotheses are tested by comparing behaviour in two treatments, which are labelled BASELINE and TOP-UP.

The BASELINE treatment is a variant of Knetsch's (1989) classic swapping task, where subjects are randomly allocated one of two possible endowments and then asked to either stick or swap. Similar to the classic task, the endowments comprise ordinary goods that have frequently been used in endowment effect experiments. The two goods are king-size chocolate bars and simple grey coffee mugs supplied by IKEA and sell for similar retail prices of approximately 2€ (see A1.3 in the Appendix of this chapter). A distinguishing feature of the BASELINE treatment relative to the classic task is that the endowments are bundles comprising two units of the respective goods.

The TOP-UP treatment is similar to the BASELINE treatment, except for the acquisition procedure of endowments. Here, subjects are initially allocated only one unit of the respective goods at random. In a subsequent step prior to the swap decision, subjects are topped-up with a second unit of the initial good. Following a strict time schedule, it is ensured that time duration between initial endowment with the good (at least one unit) and a swapping decision is the same for subjects in both treatments (see Table 1) to control for the endowment effect being influenced by the length of time goods are held. Therefore, according to standard theory, a swapping rate of 50% is expected in both the BASELINE and the TOP-UP treatment. Therefore, the null hypothesis for the BASELINE treatment is that there is no endowment effect, which implies that the swap rate is not different from 50%. The alternative hypothesis implies that the swap rate is less than 50% and that there is an endowment effect. Similarly, the null hypothesis for the TOP-UP treatment is that the swap rate is not different from 50% and that there is no endowment effect. The alternative hypothesis is that the swap rate is less than 50%, which implies that there is an endowment effect.

Comparing behaviour between both treatments provides a controlled test whether splitting the acquisition of endowments matters. Here, the null hypothesis is that there is no difference in swap rates, which implies that there is no support for splitting-effects. The alternative hypothesis implies that there is support for splitting-effects and that the swap rates are lower in the TOP-UP treatment relative to the BASELINE treatment.

4.2 Implementation

Prior to the experiment, the experimenters were accurately instructed about the procedures and their tasks to assure that each step of the experiment was conducted at the same time in different rooms. Each treatment was run by four experimenters, one of whom was in charge of the procedures.

The experiment was conducted in an undergraduate microeconomics class at Osnabrück University during June 2019. 172 students (69 female) were randomly and anonymously assigned to a treatment without having received any prior instructions on the theories being tested. 89 subjects participated in the BASELINE treatment and 83 subjects in the TOP-UP treatment. Treatment rooms were organised such that they both had a separate anteroom. Before entering the rooms, subjects were randomly allocated a questionnaire which had either a red or a yellow box printed on the front page. One experimenter told subjects with a red box to take a seat on the left side and those with a yellow box on the right side of the room. Subjects were asked not to talk or use their phones during the experiment.

From that point on, the experiment was run according to the following steps (see Table 1). To ensure the simultaneous implementation of each step, the leading experimenters communicated via a messenger with each other. This allowed for accurate control of time duration between initial and final endowment in both treatments. Two identical closed boxes were placed visibly at the front of the room. One box contained a mug and the other contained a single bar of chocolate.

Table 1: Schedule of experiment

	BASELINE	TOP-UP
Step 1	Introduction	Introduction
Step 2	Random allocation and acquisition of goods	Random allocation and acquisition of good
Step 3	Risk and loss aversion questionnaire	Risk and loss aversion questionnaire
Step 4	Demographic questionnaire	Demographic questionnaire
Step 5	Demographic questionnaire	Top-Up acquisition of second good
Step 6	Swapping decision	Swapping decision

After a short introduction, the experimenter asked a volunteer subject to take a red and yellow card and place one on top of each of the boxes at their discretion. The boxes were then opened to reveal which of the goods had been designated "red" and which had been designated "yellow". Subjects were then endowed with the corresponding good, according to whether they had previously been randomly assigned to the red or yellow group.

In the BASELINE treatment, each student received two units of either a mug or a chocolate bar, whereas subjects in the TOP-UP treatment only got one unit of the appropriate good. All subjects were told that the goods they received were theirs to keep but they were asked not to open the chocolate bars because they would be making decisions on the goods later on. Afterwards, subjects were asked to complete the first part of the questionnaire, where decisions on 23 hypothetical lottery-type situations were made. In the meantime, two experimenters prepared swapping stations in the anterooms where final decisions would be made.

In the next step, subjects filled in a second questionnaire that contained questions on individual characteristics. In order to control for the amount of time spent holding at least a single unit of a good, subjects in the BASELINE treatment were given more time to complete this, because immediately after subjects in the TOP-UP treatment had finished the questionnaire, they received a second unit of the good they were initially endowed with. While two experimenters distributed the goods, subjects were told that they could keep their top-up endowment but were asked not to open the chocolate bars (Step 4 and 5). At this point in time, subjects in both treatments were holding a two-unit bundle of the respective goods and had been holding at least one unit of the good for a controlled period of time.

In the last step, subjects of both treatments took their questionnaires and were individually sent out to the swapping stations. In the BASELINE treatment, participants were told that both goods belonged to them and they had the opportunity to either keep or swap them for two units of the other good. In the TOP-UP treatment, participants were told that that both, the initial and the TOP-UP good, belonged to them and they could keep them. Then, they were informed that they could swap both goods for two units of the other good, which meant that they would have to give up both, the initial and the TOP-UP good, for two units of the other good. Consequently, if subjects chose to swap, they immediately received a bundle of the other good and the experimenter highlighted the swap on the subjects' questionnaire. At the exit, students could optionally take along paper bags to facilitate the transport of their chosen goods. Payments involved either two bars of chocolate or two mugs and the experiment lasted for less than 50 minutes.

5 Results

Table 2 reports swap rates by treatment and shows the number of subjects for four possible combinations of initial endowments and final allocations with each of the two bundles. Here, for example, a subject who was initially endowed with a bundle of chocolates but left the experiment with a bundle of mugs has swapped their endowment. Thus, the swap rate is the proportion of subjects who swapped their endowment in this experiment. The *swap rate* column reports the number of total swaps in each treatment and overall swaps. The final column presents *p-values* for one-sided z-tests of difference in sample proportions of the null hypothesis that the final allocation is independent of endowment (i.e. based on standard preference theory, there is a 50% swap rate), against the alternative hypothesis of an endowment effect (i.e. the swap rate is less than 50%). Considering the BASELINE treatment first, the total swap rate is 37% and therefore the null hypothesis can be rejected with significance greater than the 5% level. There is clear evidence of an endowment effect over the goods that have been used in this experiment. The total swap rate of 37% indicates that there are roughly 25% less swaps relative to the standard assumption of a 50% swap rate in this setting.

Table 2: Endowments and trading by treatment

Treatment	Endowment	*n*	Leaves with Chocolate	Leaves with Mug	*Swap rate*	*p-value*
BASELINE	Chocolate	45	25	20	0.44	
	Mug	44	13	31	0.29	
	Overall	*89*	*38*	*51*	*0.37*	*0.04006**
TOP-UP	Chocolate	42	19	23	0.54	
	Mug	41	8	33	0.19	
	Overall	*83*	*27*	*56*	*0.37*	*0.04551**
All		*172*	*65*	*107*	*0.37*	*0.00755***

The BASELINE treatment is a slightly modified replication of classic swapping experiments. Here, although endowments comprise two units of a good, there is a considerable exchange asymmetry. The origin of this asymmetry is the lower swap rate for mugs. Table 2 shows that subjects that are initially endowed with chocolate bars in the BASELINE treatment are 1.5 times more likely to swap than subjects endowed with mugs. This feature will be quantified more detailed in logit regression analysis hereafter.

The experimental design, the implementation and the goods used in this treatment are very close to those in classic swapping experiments (Knetsch (1989), Engelmann and Hollard (2010)). There is no remarkable difference between decisions made in this experiment and other comparable studies, although the endowment comprises two units of a good. In contrast to single-unit endowments, the literature on multiple-unit holdings suggests the endowment effect to attenuate when decisions are made over bundles of goods (e.g. Burson et al. (2013), Humphrey et al. (2017)).

However, the presence of an endowment effect in the BASELINE treatment reveals that if such attenuation was at work in the experiment, it was insufficient to purge behaviour of a significant endowment effect. One reason may be the nature of the bundles. Contrary to Humphrey et al. (2017), clean bundles comprising two units of a good are used in this setting. Moreover, subjects who decide to swap have to give up their total endowment whereas subjects in Humphrey et al.'s (2017) experiment had to give up only one of two units of a good.

There is a possibility that loss aversion is more pervasive in situations where the last unit of a good is given up (Schurr and Ritov, 2014). The observations in this treatment may therefore be explained by loss aversion. Another possibility is that the number of goods in a bundle may influence the magnitude of an endowment effect. According to Burson et al. (2013), the endowment effect declines from single to multiple unit holdings. They report evidence of a smaller endowment effect for bundles comprising multiple units of a goods than in single unit treatments. It is possible that, referring to Burson et al.'s (2013) suggestion of unit dependence, bundles comprising two units of a good may be considered a "pair" rather than a multiple-unit holding. Therefore, the endowment effect in the BASELINE treatment may arise for the same reasons as in single-unit experiments, but there is only modest support for this potential explanation, since research regarding multiple-unit endowments is relatively scarce.

Results in the TOP-UP treatment are effectively very similar to those in the BASELINE treatment. Since the total swap rate is 37%, the null hypothesis that final allocations are independent of endowments can be rejected at the 5% significance level. Hence, there is evidence of an endowment effect in this treatment. As in the BASELINE treatment, the data suggests lower swap rates for subjects endowed with mugs to have an impact on the endowment effect. With swap rates of 54% for subjects initially endowed with chocolate and 19% for those initially endowed with mugs, the distribution is less balanced than in the BASELINE treatment. Taking both treatments together, there is strong evidence of an endowment effect in this experiment (p-value = 0.00755). Although swap rates in both treatments are similar, it is important to look at the disaggregated data to examine even minor changes in behaviour between treatments.

Using a two-sided z-test of the difference in sample proportions to compare total swap rates between treatments, the null-hypotheses that swapping decisions are independent of treatment cannot be rejected. In terms of previous discussion of this hypothesis, a similar swap rate in both treatments contradicts the hypothesis of lower swap rates in the TOP-UP treatment due to splitting effects. The data suggests that the acquisition of endowments in two steps does not influence the endowment effect. This pattern is contrary to the predictions of the impact of splitting effects on the endowment effect (Humphrey et al., 2017). Nevertheless, as in the BASELINE treatment, it is difficult to draw a direct comparison with other experiments, since the bundles comprise two units of one good and trading involves the feature of giving it all up in this experiment. It is possible that splitting effects in the TOP-UP treatment may be undermined by other effects. This feature will be quantified more detailed in the disaggregated data analysis below.

Table 3 presents the results of a logit regression analysis of the decision to swap initial endowments. Observations from both treatments are pooled to increase statistical power of the tests and to isolate a treatment effect by including a treatment dummy variable in models (4) and (5). Further independent variables, such as age and gender, are determined in the demographic questionnaire. A measure of individual-level loss and risk aversion is included from subjects' responses to a series of hypothetical questions[7].

Table 3: Logit analysis of swap decisions

	(1)	(2)	(3)	(4)	(5)
Constant	-0.52*** (0.16)	-1.36 (1.92)	-1.51 (2.16)	1.12*** (0.30)	-1.55 (2.17)
Female		-0.30 (0.34)	-0.32 (0.35)		-0.32 (0.35)
Age		0.02 (0.09)	0.02 (0.09)		0.02 (0.09)
Endowed chocolate		1.07*** (0.33)	1.08*** (0.33)	1.09*** (0.33)	1.08*** (0.33)
Loss Aversion			0.03 (0.07)		0.03 (0.07)
Risk Aversion			-0.02 (0.09)		-0.02 (0.09)
Top-Up				0.01 (0.33)	0.06 (0.33)
n	172	172	172	172	172

The dependent variable takes a value of 1 if the subject swapped their endowment and 0 otherwise. *Female* is 1 for females and 0 for males. *Age* is measured in years. *Endowed chocolate* is 1 if the subject's endowment was chocolate and 0 otherwise. *Loss aversion* was measured using the 13 tasks in the supplementary materials. The earlier in this sequence that the subject switches over from play to not play, the more loss averse they are. The values used in the estimations were transformed by inverting the task order and looking at the last problem in which the subject rejected the gamble. This means that higher values correspond to more loss aversion. *Risk aversion* was measured by the switch-over point from lottery A to B in the sequence of problems in the supplementary materials. *Top-up* is a dummy variable for treatment with stepwise acquisition of endowments. The estimates are logit coefficients with standard errors in parentheses. * Denotes significance at the $p \leq 0.1$ level; ** at the $p \leq 0.05$ level; *** at the $p \leq 0.01$level.

[7] See supplementary material in the appendix for the detailed questionnaires of the experiment.

Model 1 provides a test for the existence of an endowment effect in this experiment. It includes only a constant and reports a highly significant negative coefficient ($p < 0.01$). Therefore, the null-hypothesis of a 50% swap rate can be confidently rejected. This result confirms the presence of an endowment effect in the data and complements the test results reported in Table 2. Tests of the null hypothesis that swap decisions are independent of initial endowment can be rejected at the 1% level.

A general preference for mugs over chocolate is confirmed by models (2), (3), (4) and (5), where the decision to swap is significantly correlated with being endowed with chocolate (p = 0.001 for all models). In all three models that include individual-level characteristics, there is no effect of age or gender. Moreover, the swap decision is not correlated with the included measures of risk and loss aversion. Models 4 and 5 confirm the reported test results that the acquisition of endowments in two steps is not correlated to the decision to swap. Therefore, tests of the null hypothesis that swap rates in both treatments are equal cannot be rejected.

The widely reported contribution of loss aversion to the endowment effect means it is worthwhile considering loss aversion in the data in more detail. The lack of significance means that there is no correlation between loss aversion and the decision to swap. Therefore, in contrast to the discussions in Kahneman et al. (1990), Loomes et al. (2009) and Humphrey et al. (2017) there is no evidence to support the role that loss aversion plays in contributing to an endowment effect.

In order to interpret the lack of significance in this experiment, the procedure of constructing the measure for loss aversion has to be considered. The elicitation of loss aversion measures to use as explanatory variables in descriptive analysis is known to produce mixed results. For example, Humphrey et al. (2017) find a significantly positive relationship between loss aversion and the endowment effect in their initial experiment. However, they do not replicate this relationship in a similar follow-up experiment using the same methods. The latter observation parallels the findings in the BASELINE and TOP-UP treatment. Since the loss aversion measure elicited as part of this experiment used hypothetical tasks, the data do not undermine the conventional wisdom that loss aversion plays an important role in contributing to an endowment effect. However, neither do they provide evidence to support it.

In order to further understand the headline results, some more detailed tests on data disaggregated by initial endowment are conducted. The results in Table 2 suggest a positive correlation between chocolate and the decision to swap and no significant impact of acquiring the bundles in two steps. As shown in other evidence (i.e. Humphrey et al. 2017, Bateman et al. 1997b) it could be that splitting the acquisition of a bundle leads to an increasing valuation. Since this pattern is not emerging as being important in this experiment, it is useful to consider whether other features might have had a strong (or dominant) impact on swapping decisions. One possibility is to look at the goods being used in conjunction with the weather. When the experiment was conducted at the 18th June 2019 it was extremely warm (29 degrees Celsius). The high temperature might have caused subjects who were endowed with chocolate to be more inclined to swap it for the mug because they would not want to leave the laboratory with chocolate that may melt. Therefore, it could be possible that subjects were 'overly' inclined to swap the chocolate than they would have been in average temperatures.

Given that the endowment effect implies a reluctance to trade, the use of chocolate in high temperatures may have had two influences on the data. First, subjects who had been initially endowed with chocolate may have been inclined to trade to a greater extent than they would have been in more usual temperatures. Likewise, those initially endowed with mugs may have been less inclined to trade than they would been in more usual temperatures. The net influence of these factors would have contributed to the overall endowment effect observed and their respective roles are investigated below. Secondly, if the unusually high temperatures rendered these considerations particularly salient, this may have weakened the influence of other decision-motives, such as the role of splitting the endowment acquisition.

It may be possible to gain some insight into whether splitting the endowment acquisition matters by disaggregating the data by initial endowments and treatment, and testing for treatment effects within each group that was initially endowed with chocolate and mugs, respectively. On the assumption that the high temperatures affected the behaviour of each group in the same way, this will allow an understanding of whether the split endowment acquisition affected swap rates. Using two-sided z-tests of difference in sample proportions, table 4 reports test results of swap rates disaggregated by initial endowment and treatment.

Table 4: Swap rates disaggregated by endowment and treatment

Endowment	BASELINE	TOP-UP	*p-value*
Chocolate	0.44 (n=45)	0.54 (n=42)	*0.35224*
Mug	0.29 (n=44)	0.19 (n=41)	*0.28014*
p-value	*0.14156*	*0.0009****	

If the weather matters, given the assumption that the high temperatures affected each treatment in the same way, swap rates for chocolate would be expected to be higher than those for mugs. The swap rates disaggregated by initial endowment in table 4 are consistent with this assumption in both treatments. In the BASELINE treatment, subjects initially endowed with chocolate were 52% more likely to swap their goods than those initially endowed with mugs (0.44 relative to 0.29). Subjects initially endowed with chocolate in the TOP-UP treatment were almost three times more likely to swap their endowment for the alternative (0.54 relative to 0.19), and this difference is highly significant (p=0.0009).

Although there is no difference between treatments, there is some tentative indication that splitting the endowment acquisition matters. Considering whether splitting the endowment acquisition matters within each endowment group, on the assumption that the weather affected each treatment the same, the data show that for those initially endowed with chocolate, splitting the acquisition in the TOP-UP treatment increased the swap rate by 23% relative to the BASELINE treatment (from 0.44 to 0.54). Likewise, for subjects initially endowed with mugs, splitting the acquisition in the TOP-UP treatment reduced the swap rate by 34% relative to the BASELINE treatment (from 0.29 to 0.19). The increased swap rate for chocolate cannot be reconciled with the ownership effect.

According to Morewedge et al. (2009), the mere ownership would have added an association between the good and the self. Thus, the ownership effect would have predicted an increased valuation of chocolate und therefore a decreased swap rate in the TOP-UP treatment.

The data suggest that the weather had an influence on swapping decisions. Firstly, subjects initially endowed with subjects were more inclined to trade than those initially endowed with mugs. Secondly, although there is no clear evidence that splitting effects strengthen the endowment effect, the data suggest that splitting effects might be at work, but less straightforward than expected. It may be possible that the considerations about the use of chocolate in hot temperatures might have drowned out the role of splitting the endowment acquisition.

6 Conclusion

This experiment was set up to explore the impact of a split endowment acquisition on the endowment effect. Comparing behaviour in a one-step and a two-step treatment, there is no clear evidence that a split endowment acquisition influences the endowment effect. There is evidence of a significant endowment effect in both treatments and the results show that the stepwise acquisition of the endowment did not enhance the endowment effect. This finding runs contrary to the theoretical assumption that splitting effects in bundle acquisition promote endowment effects. Moreover, the results of both treatments that incorporated the feature of giving it all up, seem to be the same as in single-unit experiments. Hence, the use of bundles comprising two units of the same good did not attenuate the endowment effect.

The results of this study are different to existing studies that have pointed to the importance of splitting effects over a wide range of decision-making contexts (e.g. Starmer and Sugden (1993); Humphrey (1995); Bateman et al. (1997)). Relative to Humphrey et al. (2017), who related the occurrence of an endowment effect in their two-step treatment to splitting effects, a stepwise endowment acquisition did not influence the strength of the endowment effect in this experiment. It might be possible that the composition the bundles matters for splitting effects to support the endowment effect. As bundles comprised two units of the same good here and choices in Humphrey et al. (2017) were made over mixed bundles, it may be that splitting effects have a stronger impact on the endowment effect in situations where heterogeneous bundles are held.

However, the results of this study indicate that splitting effects might be present, but less straightforward than assumed in the existing literature. The data seem to suggest that splitting the acquisition in the TOP-UP treatment (relative to the BASELINE treatment) might be enhancing the salience of particular features of the bundles that are held by decision-makers. So if the weather caused holding chocolate to have negative aspects (relative to mugs), splitting the endowment acquisition served to increase the swap rate of chocolate. Likewise, if the weather made holding mugs relatively attractive, splitting the endowment acquisition enhanced this aspect and reduced the swap rates for mugs. This interpretation is in line with Humphrey's (2001) considerations in respect of splitting effects. He conjectures that a splitting effect emerges when a situation, that gives a positive outcome in one option but zero under another option, is split into two sub-situations and this enhances the relative attractiveness of the former option. He reports that splitting effects are ascribed either to a preference for more positive outcomes or an aversion to greater numbers of zero (or, more generally, smaller) outcomes - or a combination of both. Consequently, the latter interpretation of splitting effects may apply to both the reduced swap rate for mugs and the increased swap rates for chocolate in the TOP-UP treatment.

Overall, the findings of this study suggest that existing laboratory findings for single-unit experiments spill over to the multiple-unit setting studied in this experiment. This is essentially important because it is a move to a more accurate real-world situation where endowments, that are valued and traded, are built up in stages. Consequently, this may apply to markets for low-value goods (i.e. stamp or coin collection) up to markets involving investments of a large financial size (i.e. housing or business) where endowments are built up in the course of time.

7 Acknowledgements

The assistance of Kristin Arnke, Katharina Beckwermert, Jasmin Bonhage, Marian Eleks, Vanessa Koch, Jon Oodes and Alena Rodda in running the experiment sessions is gratefully acknowledged.

8 Appendix

Appendix A1 Procedures and instructions

Before entering the rooms, subjects were randomly provided with a questionnaire with a red or a yellow box printed on the front page. Subjects with a red box to take a seat on the left side and those with a yellow box on the right side of the room and await further instructions.

Introduction

Welcome to this Experiment. Thank you for helping us with our research.

This is an experiment looking at how people make decisions. During the experiment you will be asked to make decisions and answer some questions on paper forms. Our investigation is divided into multiple parts. It is neither about an ‘examination’ nor about the application of any ‘tricks’. The are no correct or incorrect answers. We just want to determine how people make decisions in particular situations.

During today's experiment, you will be asked to make decisions between lotteries and choose between goods. At the end of the experiment, you get to keep the goods you have chosen. Therefore, it is in your own interest to make your choices carefully.

The payment for your participation includes the goods you will decide on in the course of the experiment.

Please read the materials carefully and listen closely when we explain or read something out. You were all provided with the same materials. We ask you to not communicate with other participants and please be quiet during the entire experiment. If you have a question, please raise your hand. We will come to you and answer your questions personally.

Your answers and decisions will stay absolute anonymously. We will not back up this data associated with your names. All analyses will be performed in anonymized form at Osnabrück University. In order to keep every decision and answer during the investigation private, please try not to look at the materials from the other participants. The payment at the end of the experiment will also be made privately and no one else will know the amount of your payment.

Allocation of goods

Two identical closed boxes were placed visibly at the front of the room. One box contained a mug and the other contained a single bar of chocolate (see Figure A1.1 below). A volunteer subject was asked to take a red and yellow card and place one on top of each of the boxes at their discretion. The boxes were then opened to reveal which of the goods had been designated "red" and which had been designated "yellow". Subjects were then endowed with the corresponding good, according to whether they had previously been randomly assigned to the red or yellow group.

Note: In the BASELINE treatment, each subject received two units of either a mug or a chocolate bar and subjects in the TOP-UP treatment only got one unit of the appropriate good at this stage. All subjects were told that the goods they received were theirs to keep but they were asked not to open the chocolate bars because they would be making decisions on the goods later on.

Risk and loss aversion questionnaire

Introduction

This is the second part of the experiment. You will soon be asked to decide for one of two alternatives in 23 decision situations. These decision situations are hypothetical as they involve large amounts of money. Although they will neither win nor lose money, please think carefully about the decisions and try to make them as if they were not hypothetical situations. Because we are interested in your individual behaviour, you should not communicate while you make your choices and not look at the materials of other participants.

Decision situations 1-10:

In the situations 1 to 10 you have to decide between two hypothetical lotteries.

These lotteries are called „Lottery A“ and „Lottery B“. Every lottery provides a gain with a certain probability. Please write a cross in the box next to your preferred alternative.

The decisions you make will look like this:

Situation	LOTTERY A	**I choose: Lottery A**	LOTTERY B	**I choose: Lottery B**
3	30% probability of winning **45,00 €** 70% probability of winning **35,00€**	❒	30% probability of winning **85,00 €** 70% probability of winning **2,50 €**	❒

When you prefer option A in this example, make a cross in the box under „I choose: option A“.

When you prefer Option B, make a cross in the box under „I choose: option B“.

Are there any questions left?

A1.1: Decision situations 1-10

Situation	LOTTERY A	I choose: Lottery A	LOTTERY B	I choose: Lotteriy B
1	10% probability of winning **45,00 €** 90% probability of winning **35,00 €**	❒	10% probability of winning **85,00 €** 90% probability of winning **2,50 €**	❒
2	20% probability of winning **45,00 €** 80% probability of winning **35,00 €**	❒	20% probability of winning **85,00 €** 80% probability of winning **2,50 €**	❒
3	30% probability of winning **45,00 €** 70% probability of winning **35,00 €**	❒	30% probability of winning **85,00 €** 70% probability of winning **2,50 €**	❒
4	40% probability of winning **45,00 €** 60% probability of winning **35,00 €**	❒	40% probability of winning **85,00 €** 60% probability of winning **2,50 €**	❒
5	50% probability of winning **45,00 €** 50% probability of winning **35,00 €**	❒	50% probability of winning **85,00 €** 50% probability of winning **2,50 €**	❒
6	60% probability of winning **45,00 €** 40% probability of winning **35,00 €**	❒	60% probability of winning **85,00 €** 40% probability of winning **2,50 €**	❒
7	70% probability of winning **45,00 €** 30% probability of winning **35,00 €**	❒	70% probability of winning **85,00 €** 30% probability of winning **2,50 €**	❒
8	80% probability of winning **45,00 €** 20% probability of winning **35,00 €**	❒	80% probability of winning **85,00 €** 20% probability of winning **2,50 €**	❒
9	90% probability of winning **45,00 €** 10% probability of winning **35,00 €**	❒	90% probability of winning **85,00 €** 10% probability of winning **2,50 €**	❒
10	100% probability of winning **45,00 €**	❒	100% probability of winning **85,00 €**	❒

Decision situations 11-23:

You now have to decide whether you want to participate in a hypothetical lottery or not. In this lottery you could win money with a certain probability, but you could also lose money.

As an example:

Situation	LOTTERY		I take part in the lottery	I DO NOT take part in the lottery
11	50% probabilty of winning	**30,00 €**	❐	❐
	50% probability of losing	**5,00 €**		

Please make a cross according to your decision in the box under „I take part in the lottery“, or „I do not take part in the lottery“.

Any questions left?

A1.2: Decision situations 11-23

Situation	LOTTERY		I take part in the lottery	I DO NOT take part in the lottery
11	50% probability of winning 50% probability of losing	**30,00 €** **5,00 €**	❒	❒
12	50% probabilty of winning 50% probability of losing	**30,00 €** **7,50 €**	❒	❒
13	50% probabilty of winning 50% probability of losing	**30,00 €** **10,00 €**	❒	❒
14	50% probabilty of winning 50% probability of losing	**30,00 €** **12,50 €**	❒	❒
15	50% probabilty of winning 50% probability of losing	**30,00 €** **15,00 €**	❒	❒
16	50% probabilty of winning 50% probability of losing	**30,00 €** **17,50 €**	❒	❒
17	50% probabilty of winning 50% probability of losing	**30,00 €** **20,00 €**	❒	❒
18	50% probabilty of winning 50% probability of losing	**30,00 €** **22,50 €**	❒	❒
19	50% probabilty of winning 50% probability of losing	**30,00 €** **25,00 €**	❒	❒
20	50% probabilty of winning 50% probability of losing	**30,00 €** **27,50 €**	❒	❒
21	50% probabilty of winning 50% probability of losing	**30,00 €** **30,00 €**	❒	❒
22	50% probabilty of winning 50% probability of losing	**30,00 €** **32,50 €**	❒	❒
23	50% probabilty of winning 50% probability of losing	**30,00 €** **35,00 €**	❒	❒

Demographic questionnaire and TOP-UP acquisition

Subjects were asked to complete a demographic questionnaire regarding gender, age, their studies and other occupations.

Subjects in the TOP-UP treatment received a second unit of the good they were initially endowed with immediately after they had finished the questionnaire and were told that they could keep their top-up endowment but are asked not to open the chocolate bars.

Subjects in the BASELINE treatment were given more time to complete this questionnaire.

Swapping decision

Subjects took their questionnaires and were individually invited to the swapping stations. Swapping decisions were presented as described in the paper. If subjects chose to keep their goods, the experimenter highlighted this on the subjects' questionnaire. If subjects chose to swap, they immediately received a bundle of the other good and the experimenter highlighted the swap on the subjects' questionnaire.

A1.3: Goods used in the swapping task

9 References

Bateman, I., Munro, A., Rhodes, B., Starmer, C., Sugden, R. (1997). Does part-whole bias exist? An experimental investigation. *Economic Journal, 107*(441), 322–331.

Bateman, I., Munro, A., Rhodes, B., Starmer, C., Sugden, R. (1997b). A theory of reference-dependent preferences. *Quarterly Journal of Economics, 112*, 479–505.

Beggan, J. K., Scott, A. (1997). More than meets their eyes: Support of the mere ownership effect. *Journal of Consumer Psychology, 5*, 285–297.

Burson, K., Faro, D., Rottenstreich, Y. (2013). Multiple-Unit Holdings Yield Attenuated Endowment Effects. *Management Science, 59*(3), 545-555.

Coursey, D.L., Hovis, J.L., Schulze, W.D. (1987). The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value. *The Quarterly Journal of Economics, 102*(3), 679-690.

Engelmann, D., Hollard, G. (2010). Reconsidering the effect of market experience on the ‘‘endowment effect’’. *Econometrica, 78*(6), 2005–2019.

Gawronski, B., Bodenhausen, G. V., Becker, A.P. (2007). I like it, because I like myself: Associative self-anchoring and post-decisional change of implicit evaluations. *Journal of Experimental Social Psychology, 43*, 221–232.

Horowitz, J. K., McConnell, K. E. (2002). A review of WTA/WTP studies. *Journal of Environmental Economics and Management, 44*(3), 426–447.

Horowitz, J., McConnell, K., Quiggin, J. (1999). A test of competing explanations of compensation demanded. *Economic Inquiry 37*(4), 637-46.

Humphrey, S. J. (1995). Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *Journal of Risk and Uncertainty, 11*(3), 263–274.

Humphrey, S.J. (2006). Does learning diminish violations of independence, coalescing and monotonicity? *Theory and Decision, 61*, 93-128.

Humphrey, S.J., Lindsay, L., Starmer, C. (2017). Consumption experience, choice experience and the endowment effect. *Journal of the Economic Science Association 3*(2), 109–120.

Isoni, A., Loomes, G., Sugden, R. (2011). The willingness to pay—Willingness to accept gap, the ''endowment effect,'' subject misconceptions, and experimental procedures for eliciting valuations: Comment. *American Economic Review, 101*(2), 991–1011.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy, 98*(6), 1325–1348.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives, 5*(1), 193–206.

Kahneman, D., Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.

Knetsch, J. L. (1989). The endowment effect and evidence of nonreversible indifference curves. *The American Economic Review, 79*(5), 1277–1284.

Knetsch, J. L., Sinden, J.A. (1984). Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value. *The Quarterly Journal of Economics, 99*(3), 507-521

List, J. A. (2003). Does market experience eliminate market anomalies? *The Quarterly Journal of Economics, 118*(1), 41–71.

List, J. A. (2004). Neoclassical theory versus prospect theory: Evidence from the marketplace. *Econometrica, 72*(2), 615–625.

Loewenstein, G., Issacharoff, S. (1994). Source dependence in the valuation of objects. *Journal of Behavioral Decision Making, 7*, 157–168.

Loomes, G., Orr, S., Sugden, R. (2009). Taste uncertainty and status quo effects in consumer choice. *Journal of Risk and Uncertainty, 39*(2), 113–135.

Loomes, G., Starmer, C., Sugden, R. (2003). Do anomalies disappear in repeated markets? *The Economic Journal, 113*(486), C153–C166.

Morewedge, C. K., Shu, L. L., Gilbert, D. T., Wilson, T. D. (2009). Bad riddance or good rubbish? Ownership and not loss aversion causes the endowment effect. *Journal of Experimental Social Psychology, 45*(4), 947–951.

Morrison, G. (1997). Willingness to Pay and Willingness to Accept: Some Evidence of an Endowment Effect. *Applied Economics, 29*(4), 411-17.

Plott, C. R., Zeiler, K. (2005). The willingness to pay-willingness to accept gap, the ‘‘endowment effect’’, subject misconceptions, and experimental procedures for eliciting valuations. *The American Economic Review, 95*(3), 530–545.

Rozin, P., Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*, 296–320.

Schurr, A., Ritov, I. (2014). The effect of giving it all up on valuation: A new look at the endowment effect. *Management Science, 60*(3), 628–637.

Shogren, J.F., Shin, S.Y., Hayes, D.J., Kliebenstein, J.R. (1994). Resolving Differences in Willingness to Pay and Willingness to Accept. *American Economic Review, 84*(1), 255-70.

Starmer, C., Sugden, R. (1993). Testing for juxtaposition and event-splitting effects. *Journal of Risk and Uncertainty, 6*(3), 235–254.

Strahilevitz, M. A., & Loewenstein, G. (1998). The effect of ownership history on the valuation of objects. *Journal of Consumer Research, 25*, 276–289.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization, 1*(1), 39–60.

Tom, G. (2004). The endowment-institutional affinity effect. Journal of Psychology: *Interdisciplinary and Applied, 138*, 160–170.

Tversky, A., Koehler, D.J. (1994). Support theory. A nonextensional representation of subjective probability. *Psychological Review, 101*, 547-567.

Weber, M., Eisenführ, F., von Winterfeldt, D. (1988). The effects of splitting attributes on weights in multiattribute utility measurement. *Management Science, 34*(4), 431–445.

Chapter 2

Is there a replication crisis in experimental economics? Results from a large-scale risky choice experiment.

1 Introduction

The replicability of scientific findings is the cornerstone of reliability in empirical research. Over the last 15 years, findings that have had a significant influence on the scientific discourse in psychology and behavioural economics have failed to replicate, undermining the output of this discourse. This, now well-known, ‘replication crisis’ initiated an ongoing debate regarding the quality of previously published results and has increased the interest in replication studies.

Relatedly, the Reproducibility Project: Psychology (RPP) (2014) conducted replications of 100 studies from three journals in psychology. Out of these 100 original studies, 97% percent have statistically significant results, but in the replications, only 36% percent of these studies were statistically significant in the same direction (Open Science Collaboration, 2015). The implications of these numbers are clear: If science proceeds on the basis of the original findings, the hindsight provided by the relatively low replication rate suggests that much of this might have been conducted on an unreliable basis. Although replication attempts in psychology and economics have become more common, replication studies in the field of experimental economics are in short supply. According to Duvendack et al. (2017), within the past 50 years, only 188 replication studies have been published in the top 50 journals in economics.

The motivation for this study is two-fold. More and more studies are now being conducted online (i.e. Gächter et al. (2015), Cubitt et al. (2017), Loomes and Pogrebna (2017)) rather than in the laboratory. This raises the question of whether sacrificing the control that is possible in the laboratory matters to the observed behaviour. To investigate this, a replication study is conducted focusing on five important, robust behavioural patterns observed in decision-making experiments. These are common consequence effects (Starmer and Sugden, 1991), common ratio effects (van de Kuilen and Wakker, 2006), preference-reversal choices (Loomes et al. 1991), event-splitting effects (Humphrey, 1995) and cyclical choices (Starmer, 1999). These influential patterns of behaviour have been replicated quite reliably in different laboratories. If online experiments are appropriate to observe genuine preferences (in the same way as in the laboratory), then these behavioural patterns should also be observed in the online implementation of this experiment.

The second purpose of the present study is, that an online implementation will be used for the tests discussed in the next chapter of this thesis. This replication study therefore serves as an 'instrument test' of the online implementation to understand whether it is a reliable method for observing genuine preferences. Should this be the case, it can be expected that the findings studied in this chapter will be replicated in the next chapter.

In order to conduct a thorough replication study that has relevance derived from modern interpretations of the experimental method, this replication study implements a modern experiment design that involves a large sample size and incentive compatible decisions with different levels of financial incentives. Section 2 presents an overview of the related literature on the relevance of replications and the present crisis. The relevant classic tests of risky choices and their theoretical implications are described in section 3. Section 4 outlines the experimental design and implementation, followed by results in Section 5. Section 6 contains the implications of the results and general concluding remarks.

2 The importance of replication and the replication crisis

The ability to independently and directly reproduce findings is an important method to strengthen confidence in scientific results. Although this method is generally known, direct replications of empirical findings are rarely conducted. In experimental economics, replications are commonly included in follow-up studies. These studies directly replicate previous experiments using baseline treatments as the comparators for the novel data generated in these experiments. This approach is mainly used to question findings, rather than verifying them

(Bardsley, 2018). Although direct replications are rarely carried out, their relevance has been widely discussed, with some of the commonly described purposes being as follows.

In a seminal article, Kerr (1998) discussed a research practice called "Hypothesizing after the results are known", or HARKing. This refers to a method, where researchers include or exclude hypotheses after empirical results have been analysed and then mispresenting their post-hoc hypotheses as a priori hypotheses, without mentioning this process in their report. Kerr (1998) argues that the practice of post-hoc hypothesizing unsurprisingly leads to test statistics that support the hypotheses and that this procedure undermines the plausibility of hypothesis testing and thereby obtained results (Bardsley, 2018). Further, it produces hypotheses that are tailored to specific samples, which increases the probability that the findings are not replicable in other samples. This method harms scientific process because it prevents the research community from precisely identifying which hypotheses are false and which are true (Rubin, 2017).

An advantage of conducting replications is to identify possible false-positive results in published scientific research. At times where methods of data collection and analysis have not been controlled, it was relatively easy to publish results that were false positive and not replicable. Simmons et al. (2011) show that selective data analysis, where multiple analyses are conducted but only the ones that yield significant results are reported, might have increased the proportion of false positives in published findings. Although, there is limited empirically documented knowledge about the rate of false positives, simulation and survey-based evidence suggest that it could actually be very high (Greenwald, 1975; Ioannidis, 2005).

Referring to a study from John et al. (2012), about 65% of psychology researchers indicated, that they selectively manipulated their analyses, by i.e. leaving a dependent variable out in their reported studies (Nelson et al. 2018). This approach is called "p-hacking", where researchers modify their data to an extent that it produces statistically significant results according to the desired level of 5%.[8] The aim to achieve the 5% significance level has already been discussed early on in the literature.

Rosenthal (1979) suggests that only the experiments successfully delivering significant results are published, whereas those that fail are stored in the “file drawer”. This file drawer explanation is called into question by Nelson et al. (2018), who believe that researchers would not easily give up a failed study and their beliefs in running it. They claim that researchers would more likely conduct further examinations and explorations on the data, rather than abandoning it. In their opinion, the file drawer is not only filled with unpublished failed studies, but also with failed analyses of studies that have been published. Nelson et al. (2018) suggest that studies are not put in the file drawer when they have failed, but when researchers’ efforts in p-hacking their results failed. This suggestion may not only raise questions about practiced research methods, but also about the amount of false-positive findings in the literature. Independent and successful replications of reported findings are very unlikely to occur when results are false positive, and this strengthens confidence in the original findings (Open Science Collaboration, 2014).

The power of increasing the robustness of reported findings through replication is of great importance in building knowledge. The relatively low rate of published replications might be related to publication requirements that motivate innovation, new methods and new evidence and thereby discourage replications (Schmidt, 2009). Common features of journal publication practices have frequently been called into question. It is not only the case that most of the published results are statistically significant, but also that the majority of published findings are based on underpowered experiments, that are highly unlikely to provide statistically significant results.

[8] p-value indicates the statistical strength of evidence given the null hypothesis

This matter is an ongoing debate regarding published literature (Fanelli, 2012; Nelson et al., 2018).

Greenwald (1975) suggests a perceptible bias against null and negative test results in peer review, that could consequently incentivise scientists to engage in selective data analyses. Hence, it is more likely that null results end up in a file drawer than that an attempt is made to publish these results (Rosenthal, 1979).

On the question of how to report experiments in order to obtain as much information as possible, Roth (1994) proposes that scientific research would require authors to report their data and the process of collecting it. This not only reduces the potential for miscommunication, but facilitates replications of particular findings to examine the robustness of results. Such debates have persuaded some journals to change publication requirements in order to create transparency and to enable reanalyses and replications. Although some changes have been made, there is not yet a central means for scientific policies and procedures that encourages transparency, openness and reproducibility. Ioannidis et al. (2014) report approaches to prevent publication and reporting biases through i.e. study registration, pre-specification and the availability of data, protocols and analyses. These procedures may not only help in creating norms of transparency and openness, but may as well be useful to overcome the replication crisis in many fields.

In psychology, the failure to replicate famous findings has increased scientists' awareness of critically questioning previous findings and the associated methodologies (Bem, 2011; Doyen et al. 2012). Consequently, a vibrant public debate on the relevance of conducting systematic replications for scientific research has encouraged replication attempts to become more common (Nelson et al., 2018). A large collaborative replication project, the RPP (2014), reported relatively low rates of replication success in psychological science. Although this has generated a debate regarding the interpretation of the findings and differences between the replication and the original studies, the amount of failures has attracted attention (Gilbert et al., 2016). The results lend support to the idea that replications are a vital addition to original research.

Awareness of the relevance of replications is also growing in the field of experimental economics. In order to contribute insights into the replicability in economics, Camerer et al. (2016) conducted 18 replications of between-subject laboratory experimental studies that have been published from 2011 through 2014 in the *Quarterly Journal of Economics* and the *American Economic Review*. Based on criteria such as the experimental setup and whether the original studies produced a key result, Camerer et al. (2016) carefully selected the set of studies to avoid the above criticisms made in respect of the RPP. To ensure openness and transparency about their approach, a replication report of each study has been sent to the original authors for feedback before implementing the experiments. Out of these 18 replications, 11 were statistically significant in the same direction as the original study which implies a replication rate of 61% for this sample.

Although this result might be considered a good estimate, there is clearly scope for more replication work to be carried out. However, the academic system does not create incentives for researchers to invest in research that supports verification of reported findings. Moreover, authors whose findings are replicated often feel threatened from such attempts. An approach to cultivating replications as a good and common professional norm may reduce initial authors' impression of feeling attacked and increase their motivation to collaborate with those who try to replicate their work (Camerer et al., 2016; Maniadis et al., 2015). Although economics still lags behind other fields, it has made advances in initiating and motivating the practice of replication studies. Scientists and journals have reacted to the debate on the relevance of replication and the vital features that are necessary to enable it.

3 Expected Utility Theory and classic experimental tests

3.1 Expected Utility Theory

According to von Neumann and Morgenstern (1947), Expected Utility Theory (EUT) can be derived from a set of axioms on preferences. For the purpose of understanding the axiomatic foundation of EUT, some general assumptions of the model will be introduced.

Following Starmer (2000), a rational decision-maker has preferences over prospects, where prospects are defined as a set of consequences that occur with related probabilities. In a situation of risk, when choosing between prospects, it is assumed, that the decision-maker has complete information about the probabilities and consequences. Prospects will be written in bold letters (**q, r, s**) and probabilities will be represented by p, with $1 > p > 0$. Assuming that conventional rules of probability allow compound prospects to be reduced to simple ones, any given prospect can be represented by a probability distribution, such as for example **r** = ($p_1, \ldots, p_n$), over a set of possible consequences $X = (x_1, \ldots, x_n)$, that may come from a particular course of action. Here, p_i denotes the probability of x_i, $p_i \geq 0$ and $\sum_i pi = 1$, for all i. Prospects can be represented as vectors of probabilities but in some cases it is also helpful to include the associated outcomes, for example by writing **r** = $(x_1, p_1; \ldots; x_n, p_n)$.

This notation enables a consistent definition of a set of axioms that EUT can be derived from, which are ordering, continuity, independence and monotonicity. For prospects to be ordered in terms of preferences, the axioms completeness and transitivity are required. Completeness means that for all prospects **q, r**: *either* **r** ≥ **q** *or* **q** ≥ **r** or *both,* which implies that individuals can rank and choose between all prospects. The operator ≥ describes that one prospect is (weakly) preferred to the other prospect. Transitivity entails that for all prospects **q**, **r**, **s**: if **q** ≥ **r** and **r** ≥ **s**, then **q** ≥ **s**. This implies, that individuals are consistent in their preferences. Continuity implies that for all prospects **q**, **r**, **s** where **q** ≥ **r** and **r** ≥ **s**, there is some probability p such that (**q**, p; **s**, $1\text{-}p$) ~ **r**. This relation means that an individual is indifferent between the compound prospect of **q** and **s** and the prospect **r.** The axioms of ordering and continuity imply that individuals' preferences can be represented in a functional form. This function $V(.)$ assigns a real-valued index to any prospect. If the value assigned to **q** by $V(\mathbf{q})$ is not less than the value assigned to **r** by $V(\mathbf{r})$, then $V(\mathbf{q}) \geq V(\mathbf{r}) \leftrightarrow \boldsymbol{q} \geq \boldsymbol{r}$.

The properties of the independence axiom put some restrictions on the form of preferences by requiring that for all prospects **q**, **r**, **s**: if **q** ≥ **r** then (**q**, p; **s**, $1\text{-}p$) ≥ (**r**, p; **s**, $1\text{-}p$), for all p. This relation means, that even if both prospects **q** and **r** entail an additional common prospect **s**, the preference ordering of prospects **q** and **r** should be independent of the value of that common prospect. A further axiom, implied by independence, is monotonicity. It can be described as follows: let a set of consequences $X = (x_1, \ldots, x_n)$ be ordered from the worst consequence (x_1) to the best one (x_n). Then, it can be said that prospect **r** = (p_{r1}, …, p_{rn}) first-order stochastically dominates prospect **s** = (p_{s1}, …, p_{sn}), if for all $i = 1, \ldots, n$:

$$\sum_{j=i}^{n} p_{rj} \geq \sum_{j=i}^{n} p_{sj} \tag{1}$$

with strict inequality for at least one i. Thus, according to monotonicity, the stochastically dominating prospect **r** is preferred to the dominated prospect **s**.

If all axioms defined so far hold, then preferences can be functionally represented by:

$$\mathrm{V}(\mathbf{q}) = \sum_{i} p_i \, . \, u(x_i) \tag{2}$$

where **q** is the prospect and $u(.)$ represents a utility function that is defined on a set of consequences. The shape of the utility function makes some assumptions on the concept of risk. A concave utility function represents risk averse behaviour and implies that an individual prefers a certain outcome x to any risky prospect with an expected value equal to x. The opposite is true for a risk prone individual whose utility function has a convex shape. The axioms imply the existence of a set of indifference curves, that can be illustrated in the unit probability triangle diagram, which is also commonly called the Marschak-Machina Triangle (first used by Marschak (1950), see also Machina (1982)). The triangle shows a class of prospects with three possible outcomes x_H, x_M and x_L, ranked in that order. As already defined, any prospect can be described as a vector of the probabilities, such as (p_H, $1 - p_H - p_L, p_L$). This means, that the prospects can be graphically located in two-dimensional probability space $p_H - p_L$. The vertical axis shows the probability p_h of the best outcome x_h increasing from bottom to top, whereas the horizontal axis measures the probability p_L of the worst outcome x_L, increasing from left to right.

Given the properties of ordering and continuity, the preferences over prospects in a triangle can be represented by indifference curves in this space. According to the independence axiom of EUT, the indifference curves are parallel, upward sloping straight lines, which is considered a strong restriction on the precise form of preferences.

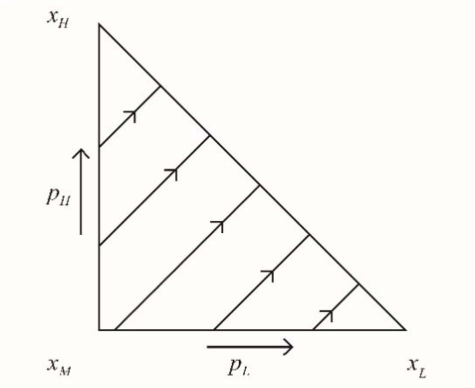


Figure 1a: EUT Indifference curves

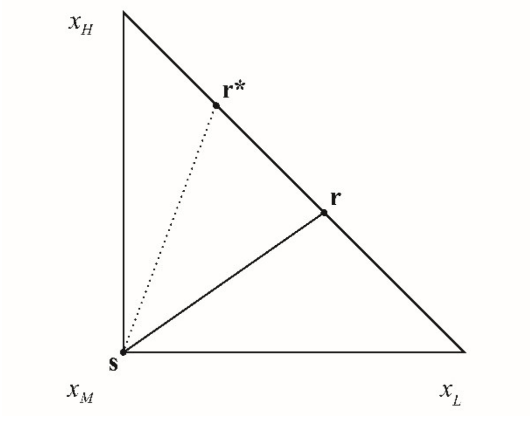


Figure 1b: Indifference curves for different risk attitudes

Figure 1a shows such a set of indifference curves in a probability triangle as predicted by EUT (Camerer and Ho, 1994). The slope of indifference curves in EUT is determined by the attitude towards risk. The steeper the slope of the curves, the more risk averse is the individual. Figure 1b shows the indifference curves for two individuals with different attitudes towards risk. Assume the two individuals can choose between a safe prospect **s** and a riskier prospect **r** (**r*** respectively). Individual 1, who has indifference curves with the slope of the solid line, is less risk averse than individual 2 who has indifference curves with a steeper slope, such as of the dotted line. Being indifferent between the both prospects requires a higher chance of getting the best outcome in the riskier prospect for individual 2 (**s** ~ **r***), compared to individual 1 (**s** ~ **r)**. This means that a movement in the north-west direction along the hypotenuse consequently reflects a higher probability of winning the best possible outcome in the riskier prospect and thus, generates steeper indifference curves for the individual. The simplicity of the theory and the degree of intuitive appeal it has, have made EUT such a powerful modelling tool. (Starmer, 2000).

As the standard theory of individual decision-making under risk in economics, EUT plays a central role both in studying risky choices, but also as a foundational account of individual preferences in broader theoretical analyses. Despite this, a large amount of experimental evidence, primarily produced over the last 60 years, has shown that it does not describe the behaviour of human decision-makers in a variety of situations (and predictably so). A large number of empirical tests using risky choice tasks have revealed patterns of behaviour that are inconsistent with EUT. The descriptive failure of EUT prompted the development of new theories aimed at fixing EUT's descriptive shortcomings. These theories have been induced by the experimentally observed violations of the EUT axioms. Therefore, these theories have to succeed where EUT has failed and that the evidence from which they have been induced, needs to be reliable. The replication crisis can be considered a prima facie reason why this might not be the case. Hence, it is important that this matter is investigated with replication studies.

The purpose of this thesis is not to discriminate between theories, but to investigate whether some alternative theories can explain systematic violations of EUT and thus account for actual decision behaviour in situations of risk. The replication study reported here is based upon the most important and influential observed violations of EUT that are: common consequence effects, common ratio effects, preference-reversal choices, event-splitting effects and cyclical choices. Against this background, the classic tests of risky choice theory, which form the central part of this replication study, are described in more detail below.

3.2 Independence axiom: common consequence effect

The common consequence effect (CCE) has played an important role in the development of alternatives to EUT. This pattern in choice behaviour has first been introduced as a thought experiment during a symposium in Paris in 1952 by Maurice Allais, suggesting that individuals might systematically violate the independence axiom of EUT. The common consequence effect is one of the two so-called Allais Paradoxes, which came in the following form of a pair of hypothetical choice problems: In the first problem, consider choosing between prospect **s** = (1M€, 1) and **r** = (5M€, 0.1; 1M€, 0.89; 0€, 0.01). In the second problem, consider choosing between prospect **s'** = (1M€, 0.11; 0€, 0.89) and **r'** = (5M€, 0.1; 0€, 0.9). Note that both prospects in the second problem give 0€ with roughly the same probability. Under EUT, preference over "**s**" and "**r**" in both problems would be independent of the 0.89 chance of 0€, which is common to both prospects in the second problem. This assumption is based on the implications of the independence axiom.

Allais however, predicted that individuals might choose **s** in the first problem, as the certainty of becoming a millionaire seems very attractive. In the second problem, he expected individuals might choose prospect **r'**, because it offers a higher prize with nearly the same probability of winning as in prospect **s'**.[9] Allais (1953) proposed that individuals might systematically violate the independence axiom of EUT and that the CCE was reasonable. In fact, the CCE became a primer in stimulating the development of alternatives to EUT.

[9] See Allais (1953) for details of the original experiment

Following Humphrey (2000), the more general common consequence problem, as it will be tested in this study is presented in figure 2.

		Probability				
		p	*q*	*r*	*s*	*t*
Option	**S**	b	b	c	b	a
	R	a	c	c	b	a

Figure 2: Generalised common consequence problem

As in Allais' (1953) original version, figure 2 describes a pairwise choice between a "safer" prospect **S** and a "riskier" prospect **R.** For prospect **S**, the chance of getting a positive outcome is higher than for prospect **R**, but the expected value of **S** is lower than that of **R**. The letters a, b, c describe the monetary pay-offs, where a > b > c = 0. The letters p to t represent the probabilities of five possible states of the world, where $0 < p \leq q < 1$, $0 \leq r, s, t < 1$ and $p + q + r + s + t = 1$. Both prospects **R** and **S** have a common consequence in the three states of the world *r, s*, and *t*, which means that under EUT, the outcomes of both prospects are independent of these states. Applying EUT to Figure 2, using *U (.)* to represent the utility of a prospect and each outcome, leads to:

$$U(S) = tU(a) + (p + q + s)U(b) + rU(c)$$

$$U(R) = (p + t)U(a) + sU(b) + (q + r)U(c)$$

Consequently, the following expression denotes the decision rule results from the prediction that, in both problems, the preference ordering of **R** and **S** is independent of *r, s* and *t* :[10]

$$S \overset{\succ}{\underset{\prec}{\sim}} R \Leftrightarrow (p + q)U(b) - pU(a) - qU(c) \overset{>}{\underset{<}{=}} 0 \quad (3)$$

A large body of evidence from experimental studies has shown that individuals do not behave according to this EUT decision rule (Conlisk (1989), Camerer (1989), Starmer and Sugden

[10] Where '∼' denotes indifference and '≻' denotes strict preference.

(1991), Starmer (1992) and Humphrey (2000)). It has been observed that individuals typically violate the independence axiom systematically by choosing **S** over **R** in the first problem when $r = t = 0$ and $s = 1 - p - q$ then **R'** over **S'** in the second problem when $s = t = 0$ and $r = 1 - p - q$. When locating these choice problems within the unit probability triangle, where the probabilities *r, s* and *t* determine the position of problems 1 and 2 in the triangle, the safer option **S** lies on a higher indifference curve than the riskier option **R**. Thus, as illustrated in Figure 3, first choosing **S** and then **R'** causes indifference curves to fan-out (Machina, 1982).

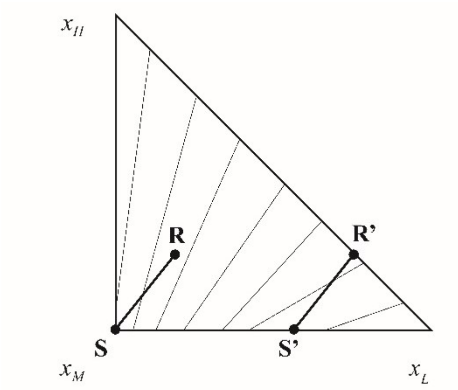


Figure 3: Indifference curves 'fanning-out'

Although the independence axiom provides some flexibility on the slope of the indifference curves, the restriction that they are straight parallel lines has been called into question. Viewed in the context of the triangle, any alternatives seeking to explain this violation of EUT, need to relax the independence axiom to generate indifference curves consistent with the implications of the actual choice behaviour and preferences that violate EUT (Camerer, 1989).

There are now many theories that accommodate the common consequence effect and do not impose the independence axiom, such as Prospect theory (Kahneman and Tversky, 1979), Generalized expected utility theory (Machina, 1982) or Rank-dependent theory (Quiggin, 1982). Since the purpose of this study is not to discriminate between alternative accounts of observed behaviour, for purposes of illustration, only the predictions of one important theory will be used to demonstrate the theoretical explanation of the CCE. For violations of the independence axiom, Prospect theory (Kahneman and Tversky, 1979) will be used and applied to show that it explains observed behaviour.[11] Following Humphrey (2000) and Starmer (1992), prospect theory assumes that individuals make a binary choice between prospects P_i (i=1,2), where each prospect comprises a probability vector p_{ij} (j=1, …, n) which represents the probability that P_i gives consequence x_j. Individuals choose between prospects to maximise overall value *V (.)* according to the expression:

$$\max \sum_{i=1}^{n} \pi\left(p_{ij}\right) v\left(x_j\right) \tag{4}$$

The utility function *v(.)* is assigned to changes in wealth relative to a reference wealth position with *v(0)=0* and *π(.)* assigns subjective decision weight to probabilities with *π(0)=0 and π(1)=1.* When applying prospect theory to the decision problem in Figure 2, the overall values for prospects **S** and **R** are given by:

$$V(S) = \pi(t)v(a) + \pi(p+q+s)v(b) + \pi(r)v(c) \tag{5}$$

$$V(R) = \pi\,(p+t)v(a) + \pi(s)v(b) + \pi(q+r)v(c) \tag{6}$$

As required by prospect theory, setting *v(c) = 0* and so, normalising *v(a) = 1,* the preferences between prospect **S** and **R** are determined by the sign of the decision rule:

$$S \underset{\prec}{\overset{\succ}{\sim}} R \Leftrightarrow [\pi(p+q+s) - \pi(s)]v(b) + [\pi(t) - \pi(p+t)] \gtreqless 0 \tag{7}$$

Furthermore, this expression can be used to evaluate prospect theory's predictions for movements within the unit probability triangle. The common consequence problems considered here involve horizontally moving from left to right across the triangle. This movement is achieved by an increase in r and a corresponding decrease in s. Since the probability r is not included in (7), only the first term in square brackets determines any horizontal movements. Assuming the probability weighting function $\pi(.)$ to be convex, a

[11] Prospect theory is used because it provides an account of three further phenomena with which this study is concerned.

decrease in s will reduce the term $[\pi(p+q+s)-\pi(s)]$ and since v(b) > 0, the left hand side of expression (7) decreases. This means that under prospect theory, a horizontal movement from left to right along the triangle would generate a tendency to switch from **S** ≻ **R** to **R'** ≻ **S'**.[12] Subsequently, this tendency causes indifference curves to horizontally fan-out (HFO). Under the less usual assumption of a concave probability weighting function $\pi(.)$, prospect theory predicts a tendency to switch from **R** ≻ **S** to **S'** ≻ **R'** and the observation of indifference curves that horizontally fan-in (HFI).

This study attempts to replicate patterns of behaviour observed in Starmer and Sugden (1991) and Starmer (1992) studies. Both papers report data from laboratory experiments with university students. Experiments were run using pen-and paper, employed a random-lottery incentive system and involved relatively small real payoffs. Starmer and Sugden (1991) find significant evidence for a systematic violation of EUT. Using within-subject tests they show that the expected choice pattern **S** ≻ **R** to **R'** ≻ **S'** occurs with a greater frequency than **R** ≻ **S** to **S'** ≻ **R'**. These results are consistent with Allais' (1953) predictions and were found in numerous studies on choices with between prospects with this structure (Moskowitz, 1974; Slovic and Tversky, 1974; MacCrimmon and Larsson, 1979).[13]

In Starmer (1992), the relevant problem pairs reveal a significant common consequence in the opposite direction to that predicted. There, the choice pattern **R** ≻ **S** to **S'** ≻ **R'** occurs with a greater frequency than **S** ≻ **R** to **R'** ≻ **S'**, although the structure of the choice problems is similar to other studies in the literature. This choice pattern in the relevant problem pairs generates indifference curves that horizontally fan in, which is inconsistent with findings from Machina (1982). An important feature of Starmer's (1992) study is, that he reports no evidence

[12] The relationship S ≻ R to R' ≻ S' denotes that S is chosen in problem 1 and R' is chosen in problem 2

[13] The relevant choice problems are taken from the original questions 21 and 22. For more details see Starmer and Sugden (1991), p. 974.

for fanning-out, but for vertical, north-west and horizontal fanning-in within the same parameter set which cannot be captured by any of the related theories.[14] Wu and Gonzalez (1998) regard this evidence as anomalous whereas Chew and Waller (1986) report the same results for two problem pairs that were similar in terms of their location in the triangle. Since Starmer's (1992) evidence has been controversially discussed in the literature and raised interest in theories that allow mixed-fanning (Humphrey, 2000), two significant problem pairs that generated horizontal fanning-in (HFI) will be considered and replicated in this study.[15]

3.3 Independence axiom: common ratio effect

The common ratio effect (CRE), which is the second version of the so-called Allais Paradox (1953), is closely related to the common consequence effect. In a similar way, further examinations of this phenomenon provided important evidence against the independence axiom and stimulated the development of alternative theories. Contrary to CCE, the CRE is concerned with the extent to which preferences depend on probabilities. In this context, EUT implies that preferences should be independent of the value of associated probabilities. However, numerous studies found systematic choice patterns that contradict this assumption. Generally, the CRE is observed in choices between problem pairs with the following form: **S** = $(b, p; 0, 1 - p)$ and **R** = $(a, \lambda p; 0, 1 - \lambda p)$, where $a > b > 0$, $0 < p \leq 1$ and (λ) denotes the ratio of "winning" probabilities, which is assumed to be constant. Thus, as shown in figure 4, the "safer" prospect **S** gives a relatively lower outcome b with probability *p* and nothing

[14] The probabilities r, s and *t* determine the position of a choice problem in the triangle. Increasing t and decreasing s, generate vertical movements in the triangle. These changes generate a tendency to switch from **S** $\succ$ **R** to **R'** $\succ$ **S'** and the observation for vertical fanning-in. Increasing t and reducing r, generate north-westerly movements in the triangle. These changes generate a tendency to switch from **S** $\succ$ **R** to **R'** $\succ$ **S'** and the observation for north-west fanning-in. For a more detailed and formal explanation see Humphrey (2000), pp. 242 ff.

[15] The relevant choice problems are taken from the original questions 1 and 4. For more details, see Starmer (1992), p.815.

otherwise, whereas the "riskier" prospect **R** gives a higher outcome a with probability λp and nothing otherwise.

		Probability	
		p	*1-p*
Option	**S**	b	0
		λp	$1-\lambda p$
	R	a	0

Figure 4: Generalised CRE problem

Applying EUT to Figure 4, using *U (.)* (with U(0)= 0) to represent the utility of a prospect and each outcome results in the following decision rule:

$$S \underset{\prec}{\overset{\succ}{\sim}} R \iff pU(b) \underset{<}{\overset{>}{=}} \lambda pU(a) \iff U(b) \underset{<}{\overset{>}{=}} \lambda U(a) \tag{8}$$

Therefore, as it can be derived from expression (8), EUT predicts that for problem pairs of this structure preferences are independent of the value of p. However, evidence from a variety of studies shows that preferences are systematically affected by the value of p (Loomes and Sugden, 1987; Starmer and Sugden, 1989; Camerer, 1995). More specifically, they reveal a tendency for individuals to switch from **S** ≻ **R** to **R'** ≻ **S'** when p changes from 1 to an intermediate value. This choice pattern is characteristic of the CRE and is inconsistent with EUT. Prospect theory, however, accounts for the CRE and is capable of explaining this kind of observed behaviour. For choices between prospects **S** and **R**, prospect theory assumes the following decision rule:

$$S \underset{\prec}{\overset{\succ}{\sim}} R \iff \pi(p)v(b) - \pi(\lambda p)v(a) \underset{<}{\overset{>}{=}} 0 \tag{9}$$

The properties of the probability weighting function *π(.)* include the underweighting of high probabilities and the overweighting of small probabilities. This means that if $p = 1$, prospect **S** – as it involves a certain outcome in this case – is more attractive than the riskier prospect **R**.

But as p is reduced to below 1 (but still high), the decision function under-weights p and prospect **R'** becomes more attractive relative to prospect **S'**, because **S'** is no longer accompanied by the effect of a certain outcome associated with $p=1$**.** (van de Kuilen and Wakker (2006)). Similar to the CCE, this choice pattern under the CRE generates indifference curves which fan out.

This study is attempts to replicate patterns of choice behaviour observed in van de Kuilen and Wakker (2006). In their paper, they report data from a laboratory experiment with university students. The experiment was run using pen and paper, employed a random-lottery incentive system and involved relatively small real payoffs. Using within-subject tests, they find evidence for a CRE in the predicted direction (**S** ≻ **R** to **R'** ≻ **S')**. This pattern of behaviour has been commonly observed for problem pairs of this structure in previous studies and shows that the majority of decision-makers systematically violate EUT (Loomes, 1988; Cubitt et al. 1998; Starmer, 2000). In the context of related literature, the CRE is considered an important finding which has stimulated the development of alternative theories to EUT. One of the fifteen problem pairs from van de Kuilen and Wakker (2006) will be replicated and tested in this study since, for problem pairs of this structure, numerous studies have revealed a significant tendency for individuals to switch choices according to the CRE. [16]

3.4 Transitivity axiom: preference-reversal choices

Transitivity is a fundamental axiom of EUT which is considered a central principle to the notion of optimization. According to transitivity, individuals are consistent in their preferences. This property of the transitivity axiom implies that indifference curves cannot intersect. Although conventional theories agreed on the normative status of transitivity, it seems to be systematically violated in a number of ways (Tversky, 1969; Starmer and Sugden, 1998;

[16] The outcomes for the relevant problem pair are taken from round 12 in the original experiment. For more details, see van de Kuilen and Wakker (2006), p.160.

Humphrey, 2001a). These findings have had a substantial impact on theoretical literature and stimulated the development of so-called nonconventional theories, that cannot be reduced to a single preference function which is defined over individual prospects (Starmer, 2000). One influential violation of the transitivity axiom is the preference reversal (PR), which is known as a challenge for those who seek to explain observed economic behaviour using rational theories of choice (Lichtenstein and Slovic, 1971; Lindman, 1971; Grether and Plott, 1979).

In a classic PR experiment, individuals are asked to carry out two distinct tasks. Following Starmer (2000), the first task involves a choice between two prospects (often called "$-bet" and "P-bet"), where the former prospect leads to a higher prize with a relatively small probability and the latter gives a smaller prize with a larger probability. In the second task, the individuals are asked to assign monetary values to both prospects from the first task, which is usually done by attaching a certainty equivalent to each prospect (often denoted as minimum selling price M($) and M(P)). Various studies have revealed a tendency for choosing the P-bet over the $-bet ($P \succ \$$) in the first task while stating a higher monetary value on the $-bet than on the P-bet ($M(\$) \succ M(P)$) in the second task. This pattern of behaviour is the so-called PR.

In the field of psychology, the occurrence of PRs has been concluded to result from individuals responding differently to valuations and choice tasks, and not having single preferences (Slovic and Lichtenstein, 1983; Tversky et al., 1988, 1990). An interpretation of PRs based on intransitive preferences is provided by regret theory (Bell, 1982; Fishburn, 1982; Loomes and Sugden, 1982), which does not only explain PRs but can actually rationalise them (Bleichrodt and Wakker, 2015)[17].

The experiment described below employs decision problems that have been used in other experiments, observed violations of transitivity that are analogous to PRs, and explained by

[17] Numerous studies have dealt with PRs and provided different attempts at explaining it, i.e. Rubinstein (1988), Starmer (1999), Bateman et al. (2007), Day and Loomes (2010).

regret theory (Loomes and Sugden, 1982). This pattern of behaviour has been extremely important in the risky choice literature, because of the support that it provides for the predictions of regret theory. In turn, this has contributed to the substantial impact of regret theory in the risky choice literature that is discussed by Bleichrodt and Wakker (2015). Following Starmer (2000), a classic PR experiment involves pairwise choices between the following three prospects called Option A, B and C with monetary consequence a > b > c > d ≥ e over three equiprobable states of the world *p, q, r > 0*:

		Probability		
		p	*q*	*r*
	A	a	d	d
Option	**B**	b	b	e
	C	c	c	c

Figure 5: Generalised preference reversal problem

The prospects **A** and **B** have the structure of typical $-bets and P-bets respectively, where **A** involves the higher outcome and **B** the higher probability of winning. Prospect **C** is a certainty equivalent that gives an outcome c for sure. Loomes and Sugden (1983) show that under regret theory, pairwise choices over the three prospects may be cyclical and mostly occur in the following direction: **B** ≻ **A**, **C** ≻ **B** and **A** ≻ **C**. [18] This preference cycle, as predicted by regret theory, is in line with choice behaviour observed in standard PR experiments. It is inconsistent with EUT because it violates the transitivity axiom.

To understand how regret theory (RT) accounts for PRs, some general and relevant assumptions of the theory will be considered first. Following Loomes et al. (1991), RT is concerned with pairwise choices between acts. Let A_i and A_k be any two acts that are defined over a finite set of states of the world with given probabilities. If state S_j occurs, the monetary

[18] ‚≻' stands for the relation 'is chosen over'.

consequences of these acts are denoted by x_{ij} and x_{kj}, and p_j denotes the probability of S_j. Using Loomes and Sugden's (1987) formulation, RT postulates a real-valued function ψ (.) such that:

$$A_i \gtrsim A_k \Leftrightarrow \sum_j p_j \, \psi\,(x_{ij}, x_{kj}) \gtreqless 0. \tag{10}$$

There are three restrictions placed on the function ψ (.): the function is skew-symmetric (i.e. $\psi (x,y) = - \psi (y,x)$ for all x,y, and this implies $\psi (x,x) = 0$, for all x, increasing in its first argument. Further, it involves the property of regret aversion, namely that for any three monetary consequences x, y, z where $x > y > z$, $\psi (x,z) > \psi (x,y) + \psi (y,z)$.[19]

Applying (10) to pairwise choices between the acts represented in Figure 7 will show how PR-cycles can be consistent with RT:

$$\mathrm{A} \gtrsim B \Leftrightarrow p\psi\,(a,b) + q\psi\,(d,b) + r\psi\,(d,e) \gtreqless 0 \tag{11}$$

$$B \gtrsim C \Leftrightarrow p\psi\,(b,c) + q\psi\,(b,c) + r\psi\,(e,c) \gtreqless 0 \tag{12}$$

$$C \gtrsim A \Leftrightarrow p\psi\,(c,a) + q\psi\,(c,d) + r\psi\,(c,d) \gtreqless 0 \tag{13}$$

For a PR-cycle to occur in the predicted direction **B ≻ A**, **C ≻ B** and **A ≻ C,** all three left-hand sides of expressions (11), (12) and (13) must be non-positive. Summing and rearranging these expressions using the skew-symmetry property of ψ (.) yields:

$$\begin{aligned} & p[\psi\,(a,b) + \psi\,(b,c) - \psi\,(a,c)] \\ & + q[\psi\,(b,c) + \psi\,(c,d) - \psi\,(b,d)] \\ & + r[\psi\,(c,d) + \psi\,(d,e) - \psi\,(c,e)]. \end{aligned} \tag{14}$$

Because of the restrictions on the values of the consequences a to e and the regret-aversion property, the terms inside all three sets of square brackets must be non-positive and strictly negative in at least one case. Thus, expression (14) is strictly negative and shows that the PR

[19] In Loomes and Sugden (1987) this property was called 'convexity'

cycle in the predicted direction (*predicted cycle*) is consistent with regret theory. Consequently, a cycle of preferences in the opposite direction (**A** ≻ **B**, **B** ≻ **C** and **C** ≻ **A**), called the *unpredicted cycle*, is not consistent with regret theory. Accordingly, regret theory predicts *predicted cycles* and not *unpredicted cycles*.

The experiment described below uses a set of decision problems taken from Loomes et al. (1991).[20] In this study, a pen-and-paper experiment with university students was used and choices were incentivised according to the random-lottery incentive system. Using within-subject tests, they observed systematic violations of the transitivity axiom of EUT in the direction predicted by regret theory. A notable feature of the Loomes et al. (1991) study is that it controls for a variety of factors discussed by Holt (1986), Karni and Safra (1987) and Segal (1988), that offer alternative explanations for transitivity violations to that provided by regret theory. This allows the evidence reported by Loomes et al. (1991) to be more readily interpreted as supporting regret theory and, given the importance of that theory in the risky choice literature, a natural candidate to be investigated in a replication study.

3.5 Transitivity axiom: event-splitting effects

Many of the studies that have observed violations of transitivity were inspired by regret theory (Loomes et al., 1989, 1991; Loomes and Taylor 1992, Starmer and Sugden, 1998). Although regret theory was said to be the most influential theory of intransitive choice, some studies challenged the predictions of regret theory (Weber et al., 1988; Battalio et al., 1990; Harless, 1992; Bateman et al., 2007). A serious challenge to regret theory, and also to the descriptive performance of rank-dependent theory, which was considered EUT's strongest competitor, came from Starmer and Sugden (1993) and Humphrey (1995, 1996, 2001b). They suggested that a large part of the evidence, that seemed to support the novel predictions of regret theory,

[20] The relevant problems are taken from the original questions 1, 3 and 5. For more details, see Loomes et al. (1991), p.433.

could be explained by event-splitting effects[21]. They have shown that there is a tendency for individuals to assign higher weight to an event (that gives some outcome x) when it is split into two sub-events, even though the total objective probability remains unchanged. This pattern violates most theories of individual choice.

Following Humphrey (2001b), the event-splitting effect (ESE) has been observed for problem pairs of the form as represented in Figure 6, where each prospect represents a probability distribution over monetary outcomes. In this setting, where $p+q+r = 1$ and $a > b > 0$, individuals are asked to choose **R** or **S** in the split problem and then **R'** or **S'** in the combined problem. The split problem separates the event in the combined problem that offers *b* under prospect **S'** and *0* under prospect **R'**, into two sub-events.

Split problem

		Probability		
		p	q	r
Option	**R**	a	0	0
	S	0	b	b

Combined problem

		Probability	
		p	$q+r$
Option	**R'**	a	0
	S'	0	b

Figure 6: Generalised ESE split and combined problem pairs

Apart from the different representation, prospects **R** and **S** offer the same chance of winning the identical outcomes as **R'** and **S'**. Under EUT, both problem pairs are considered to be identical. Thus, an individual with expected utility preferences would either choose **R** and then **R'** or **S** and then **S'** across this pair of problems. The ESE occurs when an individual first chooses **S** in the split problem and then **R'** in the combined problem and thereby violates EUT. This choice pattern can be explained by prospect theory (Kahneman and Tversky, 1979) without the 'editing' stage[22]. The original two-stage model of prospect theory includes an

[21] For similar evidence of ESEs in the context of multi-attribute choice see e.g. Weber et al. (1988).
[22] For other theoretical accounts on explaining ESEs see e.g. Viscusi (1989) and Tversky and Koehler (1994).

editing operation called ‘combination’. This operation entails that individuals add the probabilities of identical outcomes prior to the evaluation of prospects. Combination would therefore make the split problem identical to the combined problem and preclude ESEs. Therefore, it is a ‘stripped-down’ form of prospect theory that can explain ESEs (Starmer and Sugden, 1993). Applying the decision rule and the associated assumptions in expression (3) to the problem pairs in Figure 6, the stripped-down form of prospect theory accounts for ESEs if expression (15) holds:

$$v(b)\ [\ \pi(q) + \pi(r) - \ \pi(q + r)] > 0. \tag{15}$$

The underlying property of sub-additivity for $\pi(.)$ in prospect theory allows expression (15) to hold in the relevant region. Here, the greater number of b consequences as represented in the split problem makes option **S** more attractive than **S’**. In terms of probabilities, the sum of the two decision weights attached to $v(b)$ in the split problem is disproportionally larger than that of the single decision weight in the combined problem.

This study is concerned with replicating ESEs as originally reported by Humphrey (1995). [23] In his paper, he reports data from a laboratory experiment in which university students participated. Problems were presented at computer terminals and subjects received real payoffs, that were determined by the employed random-lottery incentive system, at the end of the experiment. Throughout the experiment, Humphrey (1995) finds significant evidence for a systematic violation of EUT. Using within-subject tests, he also shows that the expected choice pattern **RS’** (choosing **R** in the first problem and **S’** in the second problem) occurs with a greater frequency than the opposite pattern **SR’** and therefore provides strong evidence of ESEs.

[23] The relevant problem pair is taken from the original questions 2 and 3. For more details, see Humphrey (1995), p.269.

3.6 Transitivity axiom: cyclical choices

Starmer (1999) has introduced another observed violation of transitivity, which is consistent with prospect theory. His experiment embodies a feature of monotonicity, which is a property that follows from the independence axiom defined in 3.1, and an indirect test for non-transitive behaviour as recognised in the classic PR experiments. Similar to the classic experiment, it includes pairwise choices between the following three prospects called Option A, B and C:

		Outcome and Probability
	A	$a, p;\ 0, 1-p$
Option	B	$b, q;\ 0, 1-q$
	C	$b, r;\ b-\epsilon, s;\ 0, 1-r-s$

Figure 7: Generalised cyclical choice problem

Where the letters p, q, r, and s denote probabilities with $p < q$ and $q = r + s$. The letters a, b and ϵ represent the monetary outcomes where $a > b > \epsilon > 0$. This implies that Option A offers a chance of winning outcome a with probability p, otherwise nothing. Option B offers a chance of winning a smaller outcome b with a higher probability q, otherwise nothing. Option C offers the same chance of winning as Option B, but the outcome is not always as good. According to monotonicity, Option B dominates Option C. Choice behaviour observed by Starmer (1999) reveals a violation of transitivity of the form A ≻ B, B ≻ C and C ≻ A (called the *predicted cycle*) might appear. The reverse choice behaviour (called the *unpredicted cycle*) B ≻ A, C ≻ B, and A ≻ C implies violations of transitivity and monotonicity. The occurrence of the predicted and unpredicted cycle is inconsistent with EUT.

Following Starmer (1999), in this setting, the original version of prospect theory provides a strategy to avoid general non-monotonic behaviour without ruling out the possibility that choices may not satisfy transitivity. Prospect theory proposes a two-stage theory, where in the first phase individuals 'edit' prospects in various ways.

One of these ways is to look for dominated alternatives within the prospects and reject them without further evaluation. This particular editing operation is called the dominance heuristic, which applies to preferences between B and C in this setting. Consequently, in the second stage of prospect theory, individuals make choices between edited prospects by maximising a preference function of the form defined in expression (4), where the probability weighting function *π(.)* is allowed to be sub-additive over some range and assigns subjective decision weights to probabilities and the utility function v*(.)* assigns subjective values to the outcomes. Applying prospect theory to the situation where A ≻ B in Figure 5 (assuming *v(0)=0*) , it follows:

$$\pi\ (p)v(a) > \ \pi(q)v(b)$$

$$\text{or}\quad \pi\ (p)v(a) - \ \pi(q)v(b) > 0. \tag{16}$$

Then, for the situation where C A, it follows:

$$\pi\ (r)v(b) + \ \pi(s)v(b-\epsilon) > \ \pi\ (p)v(a)$$

$$\text{or}\quad \pi\ (r)v(b) + \ \pi(s)v(b-\epsilon) - \ \pi\ (p)v(a) > 0. \tag{17}$$

Summing expression (16) with (17) gives:

$$\pi\ (r)v(b) + \ \pi(s)v(b-\epsilon) - \ \pi\ (q)v(b) > 0. \tag{18}$$

The underlying property of sub-additivity for *π(.)* in prospect theory allows expression (18) to hold for some *r, s* and *q,* such that $\pi\ (r) + \pi\ (s) > \ \pi\ (q)$. Hence, systematic violations of transitivity by first choosing A ≻ B and then C ≻ A would be consistent with prospect theory for some *ϵ.* Since the dominance heuristic implies that an individual would detect the dominated prospect, choosing B ≻ C would be consistent with prospect theory. Therefore, unpredicted cycles would not be consistent with prospect theory since choosing C ≻ B contradicts the dominance heuristic. According to the implications of prospect theory, if cycles do occur in this setting, it will be predicted cycles and not counter cycles.

The experiment reported below investigates whether the cyclical choices observed by Starmer (1999) can be replicated.[24] Starmer's (1999) experiment was run using pen and paper, employed a random-lottery incentive system and involved real payoffs. Using within-subject tests, he finds systematic violations of transitivity and a significantly higher frequency of predicted cycles than unpredicted cycles. Moreover, his design incorporates a feature to evaluate the rate of monotonicity violation, without playing a role in determining the outcome of the test that is used to detect violations of transitivity. Since the majority of the subjects satisfied monotonicity here, he concludes that the observed pattern of behaviour indicates systematic violations of transitivity.

4 Experiment design

This section describes the design of the experiment based around CCEs, CREs, PR choices, ESEs and cyclical choices. It entails fourteen pairwise choice problems of the general forms illustrated in figures 2, 4, 5, 6 and 7 and two additional problem pairs (see A2.1 in the Appendix of this chapter for the actual decision problems).[25]

The experiment uses a random-lottery incentive system and involves two levels of financial incentives. The low incentive level (set 1) involves sums of money which are consistent with the absolute levels of incentives employed in the original experimental studies. Since these studies were conducted some years ago, the high level of incentives (set 2) accounts for the fact that the value of the original incentives has been eroded by inflation.

[24] The relevant problems are taken from the original questions 1, 2 and 3. For more details, see Starmer (1999), p. 147f.

[25] The two additional problem pairs will be reported in the next chapter of this thesis.

4.1 CCE

The CCE problem pairs entail a choice between options S and R and then options S' and R' with parameter values according to Table 5. Questions 1 and 2, 3 and 4 respectively, originally constructed by Starmer and Sugden (1991) are used to replicate a test for the CCE, that causes indifference curves to HFO.

The null hypothesis, based on EUT, is that the expected frequencies of RS' choices and SR' choices are equal. The alternative hypothesis is that these choice patterns do not occur with the same frequency. If the alternative hypothesis is accepted, the CCE (HFO) consistent violations SR' will occur with a greater frequency than RS' ones. This expected choice pattern would be consistent with prospect theory's prediction under the assumption of a concave probability weighting function as described in section 3.2.

Table 5: CCE parameter values

	Common consequence prob.			*Outcomes*		*Questions*	
Probabilities	*r*	*s*	*t*	*Set 1*	*Set 2*	*Set 1*	*Set 2*
$p = 0.2$; $q=0.05$	0	0.75	0	a=10; b=7; c=0	a=30; b=21, c=0	(1)	(3)
	0.75	0	0			(2)	(4)
$p = 0.1$; $q = 0.1$	0	0.8	0	a=7; b=0; c=0	a=21; b=9, c=0	(5)	(7)
	0.8	0	0			(6)	(8)

Questions 5 and 6, 7 and 8 respectively, originally constructed by Starmer (1992) are used to replicate a test for the CCE that causes indifference curves to HFI. [26] Again, the EUT based null hypothesis is that the expected frequencies of RS' choices and SR' choices are equal and the alternative hypothesis expects these choice patterns not to occur with the same frequency.

[26] For question pair 5 and 6, an error in software coding occurred which changed one of the parameters (outcome value b from 3 to 0 as in table 1). This excludes any further comparison on the CCE with HFI between set 1 and set 2. This test will only be conducted for set 2 in the main analysis.

In this setting however, if the alternative hypothesis is accepted, the CCE (HFI) consistent violations RS' will occur with a greater frequency than the SR' ones. This expected choice pattern would be consistent with prospect theory's prediction under the less common assumption of a convex probability weighting function.

4.2 CRE

The CRE problem pairs entail a choice between options S and R and then options S' and R' with parameter values according to Table 6. Questions 9 and 10, 11 and 12 respectively, originally constructed by van de Kuilen and Wakker (2006), involve two problem pairs, to test for the CRE.

Table 6: CRE parameter values

	Outcomes		*Questions*	
Probabilities	*Set 1*	*Set 2*	*Set 1*	*Set 2*
$p = 1; \lambda = 0.8$	a=15, b=11	a=45; b=33	(9)	(11)
$p = 0.25; \lambda = 0.8$			(10)	(12)

The EUT based null hypothesis expects SR' choices and RS' choices to occur with the same frequency. The alternative hypothesis expects the frequency of these choice patterns not to be equal. If the alternative hypothesis is accepted, the expected frequency of CRE consistent violations SR' will be greater than that of the opposite violation RS'. This expected choice pattern would be consistent with prospect theory's prediction that individuals weigh extreme outcomes higher than relatively moderate ones as described in section 3.3.

4.3 PR choices

The PR choices involve pairwise choices between options A, B and C with parameter values according to Table 7. Questions 23, 24 and 25 (26, 27 and 28 respectively) refer to the choices {A vs B}, {B vs C} and {C vs A}.

Table 7: PR choices parameter values

Probabilities	*Outcomes*		*Question*	
	Set 1	*Set 2*	*Set1*	*Set2*
p=0.3*;* q=0.3*;* r=0.4	a=18, b=8, c=4, d=e=0	a=54, b=24, c=12, d=e=0	(23) (24) (25)	(26) (27) (28)

These questions, originally constructed by Loomes et al. (1991), are used to test for PR cycles. Similar to the test for cyclical choices, there are eight possible choice patterns to the three options: six of which are consistent with EUT and two of which are PR cycles. Cycles of preferences in the direction B ≻ A, C ≻ B, A ≻ C describe predicted cycles, whereas cycles in the direction A ≻ B, B ≻ C, C ≻ A describe unpredicted cycles. The EUT based null hypothesis is that both PR cycles are equally likely to occur. The alternative hypothesis is that the cycles do not occur with the same frequency. If the alternative hypothesis is accepted, the expected frequency of predicted cycles will be greater than that of unpredicted cycles. This expected choice pattern would be consistent with the predictions of regret theory as described in section 3.4.

4.4 ESE

The ESE problem pairs entail pairwise choices between options S and R in the split problem and S' and R' in the combined problem with parameter values according to Table 8.

Table 8: ESE parameter values

Probabilities	*Outcomes*		*Questions*	
	Set 1	*Set 2*	*Set 1*	*Set 2*
p = 0.3, q = 0.4, r = 0.3	a=14, b=5	a=42, b=15	(13)	(15)
p = 0.25, q+r = 0.7			(14)	(16)

The problem pairs in questions 13 und 14 (15 and 16 respectively), originally constructed by Humphrey (1995), offer the same chance of winning the identical outcomes and are used to test for the ESE. Apart from the different representation of the problem pairs, options R and S are identical to options R' and S'. Therefore, the null hypothesis is that there are no systematic violations of EUT. It implies RS' and S'R choices to occur with the same frequency. The alternative hypothesis is that these choice patterns are not equally likely to occur. If the alternative hypothesis is accepted, ESE consistent violations SR' will occur with a greater frequency than the opposite violation RS'. The property of sub-additivity for the probability weighting function would allow this expected choice pattern to be consistent with the predictions of the "stripped-down' version of prospect theory as described in section 3.5.

4.5 Cyclical choices

The cyclical choices involve pairwise choices between options A, B and C with parameter values according to Table 9. Questions 17, 18 and 19 (20, 21 and 22 respectively) refer to the choices {A vs B}, {B vs C} and {C vs A}.

Table 9: Cyclical choices parameter values

	Outcomes		*Question*	
Probabilities	*Set 1*	*Set 2*	*Set1*	*Set2*
			(17)	(20)
p=0.2, q=0.3, r=0.15, s=0.15	a=14, b=8, ϵ =0.25	a=42, b=24, ϵ =0.75	(18)	(21)
			(19)	(22)

These questions, originally constructed by Starmer (1999), are used to test for cyclical choice behaviour. There are eight possible choice patterns to the three questions: six are consistent with EUT and two are cyclical patterns, namely the predicted cycle (A ≻ B, B ≻ C, C ≻ A) and the unpredicted cycle (B ≻ A, C ≻ B, A ≻ C). Question 18 (and 21 respectively) involves a dominant option and will be used to investigate whether subjects obey monotonicity. The EUT based null hypothesis expects both cycles to occur with the same frequency. The alternative hypothesis expects the frequency of both cycles not to be equal. If the alternative hypothesis is accepted, the expected frequency of predicted cycle will be greater than that of the unpredicted cycle. Due to the dominance heuristic and the property of sub-additivity for the probability weighting function, the expected choice pattern would be consistent with the predictions of prospect theory as described in section 3.6.

4.6 Incentives

The primary purpose of using two sets of incentives in this experiment is to off-set inflation. However, a side-benefit of having two levels of incentives is that it allows the experiment to shed light on other matters related to the level of incentives, such as a direct comparison of if/how incentives affect choices. This is especially interesting since the impact of monetary incentives in experiments is still discussed.

Among economists, it has been widely believed that participants in experiments were influenced by monetary incentives. It has been concluded that the higher the amounts and the more direct the link between incentives and performance, the better. This belief was then used to challenge relevant evidence from experiments using hypothetical choices that had shown deviations from standard theory (Beattie and Loomes, 1997).

A wide discussion on this topic in *The Handbook of Experimental Economics* (Kagel and Roth, 1995) generally resulted in the opinion that neither the directness of any links to performance, nor the amount of the incentive seem to have a substantial influence on the observed behaviour in experiments (Beattie and Loomes, 1997). Smith and Walker (1993) have argued that variance around responses reduces when incentives are increased, because a rise in incentives causes participants to exert more cognitive effort in experiments. Hey (1991) has argued that, if incentives are high enough, violations of EUT will decrease, whereas Thaler (1987) has discussed that violations are greater with incentives than without them. In their review of 74 experiments, Camerer and Hogarth (1999) have concluded that the mere increase in incentives has not caused rationality violations to disappear in any replicated study. Both in the psychology and the economics literature, evidence on the impact of monetary incentives and how it may affect behaviour in experiments is mixed. Comparing the two sets of incentives in this experiment contributes to this ongoing debate.

4.7 Random lottery incentive system

Similar to the original studies, a random lottery incentive system is used in this experiment. When all decisions have been made, one of the questions is randomly selected and the participant plays out the chosen lottery of this question for real money. All participants are aware of this. This procedure controls for wealth effects that could arise when participants are paid according to their performance in each question (Starmer, 1992).

Holt (1986) raised theoretical objections to this approach claiming that such experiments fail to elicit true preferences when participants treat random-lottery experiment as one single choice problem, compound lotteries are reduced to simple lotteries (*reduction principle*) and the independence axiom of EUT does not hold (Starmer and Sugden, 1991). However, directly addressing Holt's objections, Starmer and Sugden (1991) provided evidence that there is no difference between a real-choice design and a random lottery incentive system.

4.8 Online design

The experiment was implemented online.[27] As discussed by Arechar et al. (2018), online versions of controlled decision-making experiments are an increasingly employed method of empirical research in the social sciences. They have a number of potential advantages, including offering the possibility of reaching a larger and more representative sample of subjects.

[27] As the experiment was implemented in October 2020, the corona pandemic made the online implementation practically convenient.

A potential disadvantage is that, relative to the laboratory, there is a loss of control and differences in the decision-making environment (e.g. decisions are not made in the physical presence of the experimenter). It is therefore important to understand whether observed patterns of behaviour in the laboratory can be replicated in online experiments because in addition to what this reveals about the replicability of the behaviour in question, it sheds light on the legitimacy of online experiments for the observation of genuine preferences.

If, on the other hand, behaviour does replicate, this counts as primae facie evidence that preferences previously observed in the laboratory studies are robust, and sufficiently for them to survive the transition from the laboratory to an online environment. In this respect, online replications of well-known features of preferences can be used to understand whether the online method is an appropriate instrument for studies of decision-making.

The online implementation of the experiment reported here proceeded as follows. The recruitment process was carried out online. Invitations were sent per email to a pre-existing subject pool and participants could commit to a session. Therefore, the experimenter could specify exclusion criteria i.e. ensure that subjects did not participate in more than one session of the experiment.

Here, instead of physically attending the experiment in a laboratory, the participants were invited to an online meeting room where they could not see other participants, but were able to communicate with the experimenter. There, the experimenter introduced the participants to the experiment and was able to answer questions immediately. Communication with the experimenter was possible at all times. Within the introduction, an example task was presented and explained to the subjects to check if the design was comprehensible.

Moreover, assuming preferences are monotonic, a direct test of monotonicity based on Starmer (1999) was built into the choice tasks of the experiment as a mechanism that indicates whether subjects understood the design. [28]

After the introduction, each participant received a link with a unique participant code which redirected them to the platform where the experiment was carried out. At this point, the online experiment did not differ from a computer-based laboratory experiment as all entries are stored on a server in both settings. At the end of the experiment, where subjects in the laboratory commonly receive their payoffs in cash, subjects in this experiment were asked to enter their IBAN into a field within the platform and the payoffs were transferred to their bank accounts.[29] To assure anonymity, the payment data was stored in a separate file and transfers were made externally by someone who was not involved in the experiment.

4.9 Implementation

The computer-based experiment was conducted online during October 2020. A total of 147 subjects were recruited via e-Mail using the web-based online recruitment system ORSEE (Greiner, 2015) from the Laboratory for Economics Research's (LaER) database of preregistered volunteers. Participants were students from a variety of disciplines at Osnabrück University, the youngest was 18, the oldest 34 years and 62% were female. Participation was restricted to one session, which was controlled by a registration protocol that prevented duplicate participation. Subjects were randomly allocated to one of six pre-arranged sessions on Stud.IP, each of which followed exactly the same procedures.[30]

[28] Responses on questions 18 and 21 indicate whether subjects detected the dominant option and therefore understood the design.
[29] IBAN is the abbreviation for Internal Bank Account number and is a number attached to a bank account.
[30] Stud.IP is a web-based working environment used by Osnabrück University, to support courses at educational institutions, which has been developed as free software to enable and standardise communication between students and teaching staff via the internet.

Stud.IP incorporates a virtual meeting room feature called BigBlueButton (BBB), where participants and the experimenter met at the beginning of the experiment. The experiment was introduced by explaining that the session assisted in research, regarding how individuals make decisions under risk. Subjects were informed that they would be compensated for their participation with a show-up fee of 5€ and the chance to win an additional cash prize, which would be transferred to their bank accounts after the experiment.

For this purpose, subjects were informed in advance during the recruitment process, that they had to provide their name and IBAN within the experiment and that these data were saved separately and not visible to the experimenter. Subjects were told that payments were underwritten by a research fund of Osnabrück University. The experimenter instructed the subjects by explaining the tasks and the progression of the session. Each subject received a unique link to the online experiment platform SoPHIElabs, where each subject responded to thirty-two risky choice questions. At any point in time, subjects had the opportunity to ask questions to the experimenter in a chat window.

Payments were made according to the random lottery incentive system. The subjects were informed about this procedure in the instructions. At the end of the experiment, the random lottery incentive system worked as follows. When all decisions were complete, the computer randomly selected one of the thirty-two questions as the payment-relevant task. The question and the option they chose, were presented to them again on screen. Then, the computer randomly selected a number between 1 and 100, which determined their winnings according to the option they chose.

If, for example, the question in Figure 8 was randomly selected as a subject's payment-relevant question, and Option A was previously chosen, a number between 1 and 20 would yield 3€.

Figure 8: Example of problem display

To control for possible order effects within the experiment, the order of questions was randomized by the computer for each subject. Sessions lasted for approximately 45 minutes, without enforced time limit, and average payoffs were 13.50€ (including show-up fee).

5 Results

The data are analysed using a binominal test based on the null hypothesis that there are no systematic violations of EUT. The null hypothesis implies that violations of EUT are equally likely to occur and therefore occur as a result of random error. This is a neutral stochastic assumption in terms of the process that may generate choices. The alternative, two-tailed hypothesis is that violations are not random, but systematic. The tests for CCEs, CREs, PR choices, ESEs and cyclical choices are within-subject tests, with subjects being drawn from the same population responding to the same set of identical choice problems, for two sets of incentives. Table 10 sets out the number of observed choice patterns for each question pair considered and each parameter set employed. The *p-value* shows, given the null hypothesis, the probability of observing a pattern of violation by chance.

Table 10: Test results

	Choice patterns								
	EUT consistent						*EUT violation*		
Set 1 (Questions)		SS'			*RR'*		*SR'*	*RS'*	*p-value*
CCE (1,2)		34			44		33	36	*0.8099*
CCE (5,6)[31]		-			-		-	-	-
CRE (9,10)		65			22		53	7	*0.0001***
ESE (13,14)		61			41		27	18	*0.2327*
	ABA	*ACA*	*BBA*	*BBC*	*ACC*	*BCC*	*ABC*	*BCA*	
PR choices (23,24,25)	6	28	3	1	13	62	1	33	*0.0001***
Cyclical choices (17,18,19)	49	1	11	45	4	9	26	2	*0.0001***
Set 2 (Questions)		*SS'*			*RR'*		*SR'*	*RS'*	*p-value*
CCE (3,4)		30			50		44	23	*0.0139**
CCE (7,8)		29			55		22	41	*0.0111**
CRE (11,12)		42			25		73	7	*0.0001***
ESE (15,16)		60			35		40	12	*0.0001***
	ABA	*ACA*	*BBA*	*BBC*	*ACC*	*BCC*	*ABC*	*BCA*	
PR choices (26,27,28)	25	4	23	19	7	52	1	16	*0.0003***
Cyclical choices (20,21,22)	48	3	12	35	2	5	40	2	*0.0001***

The choice patterns indicate the frequency of all possible conjunctions of choices over the problem pairs. The *p-value* is calculated for a two-tailed within subject test using the binomial distribution of the hypothesis that violations of EUT consistent with the relevant phenomenon occur with a greater frequency than opposite violations (The null hypothesis is that they occur with equal frequency). * and ** denote rejection of the null hypothesis at the 5% and 1% level respectively.

[31] For question pair 5 and 6, an error in software coding occurred. Thus, the test for the CCE with HFI is only conducted for question pair (7,8) under *set 2*.

5.1 CCE

Table 10 shows that for question pair (1,2) under *set 1*, the null hypothesis cannot be rejected. When incentives are increased, as for question pair (3,4) under set 2, the data show a significant tendency for HFO with 44 SR' responses and 23 RS' ones and thereby a violation of the independence axiom of EUT. Thus, the null hypothesis can be rejected at the 5% level (*p=0.0139*), which shows that a CCE in the predicted direction shows up. This result represents the kind of independence violation that is predicted by prospect theory under the assumption of a concave probability weighting function. Consequently, findings from Starmer and Sugden (1991) are successfully replicated under set 2, but not under set 1.

For question pair (7,8) under set 2, the null hypothesis can be rejected at the 5% level (*p=0.0111*). The data show a significant tendency for HFI with 22 SR' and 41 RS' responses, which indicates a violation of the independence axiom and a CCE in the predicted direction. Based on the assumption of a convex probability weighting function, the significant tendency for HFI consistent choices is in line with predictions from prospect theory. The data suggest a successful replication of the findings reported by Starmer (1992).

5.2 CRE

For question pairs (9,10) and (11,12) the null hypothesis can be confidently rejected at the 1% level (*p<0.0001* for both sets of incentives) in favour of the alternative hypothesis. This suggests, that, as a result of independence violations, there a significant CRE in the predicted direction (SR'>RS') under both sets of incentives, which is in line with findings from van de Kuilen and Wakker (2006). The observed choice pattern is consistent with prospect theory's prediction that subjects weigh extreme outcomes higher than relatively moderate ones.

5.3 PR choices

Responses to questions (23,24,25) and (26,27,28) reveal a tendency for PR cycles to occur under both sets of incentives. The null hypothesis, based on the assumption that violations of EUT occur with the same frequency, can be rejected at the 1% level ($p<0.0001$ and $p=0.0003$ under set 1 and set 2 respectively). The PR cycles clearly do not occur randomly, since 33 out of the 34 observed cycles under set 1 and 16 out of the 17 observed cycles under set 2 are in the predicted direction B ≻ A, C ≻ B, A ≻ C (see BCA responses in table 10). The data support the prediction, that subjects systematically violate transitivity in this particular direction. The observed choice behaviour is therefore consistent with predictions of regret theory and suggests a successful replication of findings from Loomes et al. (1991) under both sets of incentives.

5.4 ESE

Responses to question pair (13,14) indicate that violations of EUT are split in the direction, that is consistent with ESEs (SR'>RS'). However, this effect is not significant and the null hypothesis cannot be rejected under set 1. When incentives are increased, as for question pair (15,16) under set 2, the ESE consistent violations of EUT reach significance. With 40 SR' responses and 12 RS' ones, the null hypothesis can be rejected at the 1% level ($p<0.0001$). The observed choice behaviour implies that subjects systematically violate the transitivity axiom. The property of sub-additivity for the probability weighting function allows this choice pattern to be consistent with predictions of the "stripped-down' version of prospect theory. Taken together, there is a successful replication of findings from Humphrey (1995) under set 2, but not under set 1.

5.5 Cyclical choices

Responses to questions (17,18,19) and (20,21,22) reveal a tendency for cyclical choices to occur under both sets of incentives. The null hypothesis, based on the assumption that violations of EUT occur with the same frequency, can be rejected at the 1% level (*p<0.0001* for both sets of incentives). The patterns of choice are clearly non-random, since 26 out of the 28 observed cycles under set 1 and 40 out of the 42 observed cycles under set 2 are in the predicted direction A ≻ B, B ≻ C, C ≻ A (see ABC responses in table 10). Hence, either a violation of transitivity or monotonicity (or even both) has occurred. Responses on questions 18 and 21 show that subjects detected the dominant option, since an overwhelming majority of 89% and 96% (under set 1 and 2 respectively) chose the dominating option in both questions. This implies that a large part of the subjects obeyed monotonicity. Thus, it can be concluded that the observed choice patterns result from systematic violations of transitivity. Since it embodies the dominance heuristic and the property of sub-additivity for the probability weighting function, prospect theory predicts the observed choice patterns here. Thus, the data suggest a successful replication of findings from Starmer (1999) under both sets of incentives.

5.6 Incentives and online design

For the low incentive set, the CCE test and the test for ESEs fail to reach significance. The remaining tests reveal highly significant effects for the considered patterns of behaviour. [32] For the high incentive set, all results are significant in the predicted directions at least at the 5% level. This means that all tests are successfully replicated when the values of incentives are adjusted to be closest to the parameters of the original experiments. Comparing results across both sets, the CCE and ESE tests turned to become significant under set 2. It might be that choice patterns for both phenomena have strengthened in the predicted direction with increased incentives, leading to significant results. This observation, and the fact that also the remaining tests in the high incentive set revealed highly significant results in favour of the alternative

[32] As, due to a coding error, responses on question pair (5,6) could not be used as CCE test, they can in fact be used as an indirect test of monotonicity. A two-tailed z-test shows that subjects detect the dominating option and do not violate monotonicity (with 84 RS' and 2 SR' responses). This effect is highly significant at the 1% level (p-value < 0.0001).

hypothesis, contradict the assumption that increasing incentives may reduce violations of EUT (Hey, 1991; Smith and Walker, 1993). Similar to incentive comparisons in Humphrey's (2001a) study, it cannot be concluded that the mere increase in incentives caused violations to diminish or even disappear. Although the findings in the present experiment contribute to the conclusions drawn by Camerer and Hogarth (1999) on the impact of monetary incentives in experiments, it cannot be fully ruled out that a further increase in incentives might reduce the occurrence of violations (Humphrey, 2001a).

Furthermore, although this experiment does not claim to provide a fully comparison between a laboratory and an online design, the results are consistent with findings from Arechar et al. (2018) and suggest that behavioural patterns from laboratory experiments are replicable online.

6 Conclusion

Significant violations of EUT have been observed which replicate previous findings and add to the existing evidence with parameters involving low and high incentives. While results under the low incentive set reveal a replication rate of 60%, significant effects have been found for 100% of the replications under the high incentive set. Overall, this experiment finds a replication rate of 82% (9 out of 11), where the replications are highly significant in the same direction as in the original studies.

These results not only speak to the robustness of the patterns of behaviour considered in this experiment, they also contribute data to the more general question of the replicability of empirical findings in the social sciences. Comparing results from their replication study with findings from the Reproducibility Project: Psychology (RPP) (2014), Camerer et al. (2016) suggest that for laboratory experiments in economics published in top journals, the replication success might be higher than for psychology experiments.

Whilst Millroth et al. (2019) find only modest support for the replicability of some of the decision paradoxes that motivated prospect theory, the present study successfully replicates all patterns of behaviour that were considered and which support prospect theory. Although it is not possible to draw a strong conclusion about the disciplinary differences here, methodological research practices in experimental economics might contribute to the measurably higher replicability. Using substantial monetary incentives to motivate subjects might reduce variability in the performance of experiments across research teams and consequently enhance replicability. Moreover, editorial practices amongst experimental economists facilitate replication projects. By providing procedures, instructions and even original data in published studies, original authors make it relatively easy to adopt their methods and thereby improve replicability (Camerer et al., 2016).

Making use of these advantages, the present study finds an above-average replication rate and, whilst acknowledging the small sample of 11 replications, there appears to be no replication crisis in the field of experimental economics in the data reported here. It may very well be that there is a substantial amount of findings across various fields that are difficult to replicate, but as this experiment shows, this is not the case for the most popular and robust findings challenging EUT.

At a local level, the results of the experiment reported here have important implications for other online experiments that are conducted under the auspices of the Laboratory for Economics Research at Osnabrück University and which use similar protocols. The successful replication of previous findings reported here, suggest that these protocols represent an appropriate instrument for the observation of genuine preferences. Therefore, data emerging from similar online experiments Osnabrück University can be regarded as similarly reliable. This assumes, of course, that the laboratory studies in which the patterns of behaviour considered here were originally observed, were appropriately controlled and implemented.

Since the original studies were conducted in widely acknowledged world-leading centres of experimental economics research (the Universities of East Anglia, Nottingham and Rotterdam), and have been peer-reviewed by leading international journals, it seems reasonable to assume that they were.

Relatedly, the findings reported in this chapter have implications for the studies reported in the next chapter. These studies report data generated from online experiments of the type described here (and some of the data were generated as part of the experiment reported here). The studies in the next chapter also require, as a baseline treatment, a replication of a pattern of behaviour observed in an experiment conducted nearly half a century ago (using outdated protocols, such as a small sample size and a lack of incentive compatibility). Given the potential for variance between these studies, a reliable interpretation of the findings reported in the next chapter rests upon the online method being appropriate to elicit genuine preferences of decision-makers. The results of this chapter suggest that it is.

7 Acknowledgements

The assistance of SoPHIE Labs and Felix Meickmann is gratefully acknowledged for programming the experiment software and subject recruitment, respectively.

8 Appendix

Appendix A2 Procedures and instructions

During the recruitment process, subjects were informed that they would need to provide an IBAN because payments would be made by bank transfer. Subjects were informed that to participate in the experiment they would need a device capable of web browsing and with an audio outlet. After the registrations had closed, registered subjects were provided with a link to a BigBlueButton online meeting, which corresponded to their session of the experiment.

In the online meeting room subjects were able to see the experimenter, but no other participant. Subjects were individually provided with a link to the experiment software and a participation code via a chat function. They were then asked to use their code to log into the experiment software and await further instructions.

After all subjects had logged into the experiment software they were informed that the introductory screen would be read aloud by the experimenter and that questions could be asked at any time by sending a message to the experimenter via the chat function of the online meeting room.

Instructions

Introductory screen:

"Welcome to this experiment. Thank you for participating. Before we begin, you will get some general information about the procedure of this experiment.

There are no "tricks" in this experiment. There are no right or wrong answers and the experiment is not a "test". We are only interested in finding out how people make choices in particular situation, which we will then pay you for.

All decisions will be made anonymously, i.e. no other participant will know how you chose and you will not know the choices of the other participants either. All participants will have their payment transferred to their bank account after the experiment.

The funds used for this experiment are drawn from our research account at Osnabrück University. You can therefore be sure that you will be paid for your participation.

Do you have any questions?

In this experiment, you will be asked to make 32 choices between pairs of options, called Option A and Option B. Each option offers a prize with different probabilities between 1-100%. Each of these situations involves real money. The prize therefore depends on a number between 1-100, which will be randomly determined for each of you individually.

You mark your choice by ticking a box next to your preferred option and press "Next". After you press "Next", your choice will be saved and you will be taken to the next question. Please note that you cannot return to the previous questions at this point, so your choice is binding.

The following question serves as an example of the type of choices you will make during the experiment:

Option A

1 … 20	21 … 100
3€	0€
20%	80%

Option B

1 … 10	11 … 100
7€	0€
10%	90%

❐ I choose option A

❐ I choose option B

In this example, if you prefer option A, mark your choice by ticking the box next to "I choose option A". If you prefer option B, mark your choice by ticking the box next to "I choose option B".

The amount of money you receive in this experiment comprises two parts. First, everyone receives a payment of 5 € for participating. Furthermore, you will also receive the prize of the option you selected in one of the 32 questions. You will only find out which of the 32 questions is actually played out for real money once you have made your choices in all of the questions.

At the end of the experiment, a consecutive number field appears, numbered from 1 to 32. The number field is carried out in real time by the computer and stops randomly and individually for each participant at one number. Then, the number that appears is the question that determines your payoff. For example, if the number 6 is chosen, the option you chose in question 6 will determine your payment. A consecutive number field will then appear again, numbered from 1-100. This will determine the outcome of the choice you made in your actual payoff-relevant question.

Imagine that the above example question was randomly selected as your payment-relevant question. If you had chosen option A here, this option would be relevant for your payment. If the second consecutive number field (with numbers from 1 to 100) happened to stop at a number between 1 and 20, your payment would be 3€. If the number field stopped between 21 and 100, your payment would be 0€.

Since you will not know until the end of the experiment which of the 32 questions will eventually determine your payoff, any of the 32 could be the one that determines your payoff. You should therefore make all of your choices as if they are for real money (because one of them will be). Since real money is involved, please consider your choice carefully and make it according to your actual preferences.

Are there any other questions? Then move on to the experiment.”

After the introduction, subjects were directed to the experiment software to start the experiment.

Arguments Screen:

When the 32 decision problems were completed, the initial choices in the CCE problems were put on the screen along with the subjects’ choices and the normative arguments were presented as described in subchapter 3.1 of this chapter. Subjects were then told: “We would now like to ask you to make a decision on these decision problems again. Please remember that one of these choices may determine your actual payout. Since real money is at stake, please consider your choices carefully and decide in the way that you prefer. Note that there are no right or wrong answers.” Subjects then repeated their choices.

Demographic Questionnaire:

After responding to all questions, subjects were asked to complete a demographic questionnaire regarding gender, age and their studies.

Payoff screen:

Subjects were informed of their payoff-relevant question, their resulting payment and when this payment would be issued. Subjects were then asked to provide the IBAN number for the account to which they wanted payment transferred. Transfers were scheduled immediately after the experiment by the experimenter.

A2.1 Decision problems

CCE:

Question 1

A

1	…	100
	7 €	
	100%	

B

1 … 20	21…25	21 … 100
10€	0€	7€
20%	5%	75%

Question 2

A

1 … 25	26 … 100
7€	0€
25%	75%

B

1 … 20	21 … 100
10€	0€
20%	80%

Question 3

A

1	…	100
	21 €	
	100%	

B

1 … 20	21…25	26 … 100
30€	0€	21€
20%	5%	75%

Question 4

A

1 … 25	26 … 100
21€	0€
25%	75%

B

1 … 20	21 … 100
30€	0€
20%	80%

Question 5

A

1 … 100
3 €
100%

B

1 … 10	11 … 20	21 … 100
7€	0€	0€
10%	10%	80%

Question 6

A

1 … 20	21 … 100
3€	0€
20%	80%

B

1 … 10	11 … 100
7€	0€
10%	90%

Question 7

A

1 … 100
9 €
100%

B

1 … 10	11…20	21 … 100
21€	0€	9€
10%	10%	80%

Question 8

1 … 20	21 … 100
9€	0€
20%	80%

B

1 … 10	11 … 100
21€	0€
10%	90%

CRE:

Question 9

A

1	…	100
	11 €	
	100%	

B

1	…	80	81	…	100
	15€			0€	
	80%			20%	

Question 10

A

1	…	25	26	…	100
	11€			0€	
	25%			75%	

B

1	…	20	21	…	100
	15€			0€	
	20%			80%	

Question 11

A

1	…	100
	33 €	
	100%	

B

1	…	80	81	…	100
	45€			0€	
	80%			20%	

Question 12

A

1	…	25	26	…	100
	33€			0€	
	25%			75%	

B

1	…	20	21	…	100
	45€			0€	
	20%			80%	

ESE:

Question 13

A

1 … 30	31 … 70	71 … 100
14€	0€	0€
30%	40%	30%

B

1 … 30	31 … 70	71 … 100
0€	5€	5€
30%	40%	30%

Question 14

A

1 … 30	31 … 100
14€	0€
30%	70%

B

1 … 30	31 … 100
0€	5€
30%	70%

Question 15

A

1 … 30	31 … 70	71 … 100
42€	0€	0€
30%	40%	30%

B

1 … 30	31 … 70	71 … 100
0€	15€	15€
30%	40%	30%

Question 16

A

1 … 30	31 … 100
42€	0€
30%	70%

B

1 … 30	31 … 100
0€	15€
30%	70%

Cyclical choices:

Question 17

A

1 … 20	21 … 100
14€	0€
20%	80%

B

1 … 30	31 … 100
8€	0€
30%	70%

Question 18

A

1 … 30	31 … 100
8€	0€
30%	70%

B

1 … 15	16 … 30	31 … 100
8€	7.75€	0€
15%	15%	70%

Question 19

A

1 … 15	16 … 30	31 … 100
8€	7.75€	0€
15%	15%	70%

B

1 … 20	21 … 100
14€	0€
20%	80%

Question 20

A

1 … 20	21 … 100
42€	0€
20%	80%

B

1 … 30	31 … 100
24€	0€
30%	70%

Question 21

A

1 … 30	31 … 100
24€	0€
30%	70%

B

1 … 15	16 … 30	31 … 100
24€	23.25€	0€
15%	15%	70%

Question 22

A

1 … 15	16 … 30	31 … 100
24€	23.25€	0€
15%	15%	70%

B

1 … 20	21 … 100
42€	0€
20%	80%

PR choices:

Question 23

A

1 … 30	31 … 60	61 … 100
18€	0€	0€
30%	30%	40%

B

1 … 30	31 … 60	61 … 100
8€	8€	0€
30%	30%	40%

Question 24

A

1 … 30	31 … 60	61 … 100
8€	8€	0€
30%	30%	40%

B

1 … 30	31 … 60	61 … 100
4€	4€	4€
30%	30%	40%

Question 25

A

1 … 30	31 … 60	61 … 100
4€	4€	4€
30%	30%	40%

B

1 … 30	31 … 60	61 … 100
18€	0€	0€
30%	30%	40%

Question 26

A

1 … 30	31 … 60	61 … 100
54€	0€	0€
30%	30%	40%

B

1 … 30	31 … 60	61 … 100
24€	24€	0€
30%	30%	40%

Question 27

A

1 … 30	31 … 60	61 … 100
24€	24€	0€
30%	30%	40%

B

1 … 30	31 … 60	61 … 100
12€	12€	12€
30%	30%	40%

Question 28

A

1 … 30	31 … 60	61 … 100
12€	12€	12€
30%	30%	40%

B

1 … 30	31 … 60	61 … 100
54€	0€	0€
30%	30%	40%

CCE pre-argument (analysed in BASELINE treatment chapter 3):

Question 29

A

1 … 100
14 €
100%

B

1 … 20	21…25	26 … 100
20€	0€	14€
20%	5%	75%

Question 30

A

1 … 25	26 … 100
14€	0€
25%	75%

B

1 … 20	21 … 100
20€	0€
20%	80%

CCE post-argument (analysed in BASELINE treatment chapter 3):

Question 31

A

1 … 100
14 €
100%

B

1 … 20	21…25	26 … 100
20€	0€	14€
20%	5%	75%

Question 32

A

1 … 25	26 … 100
14€	0€
25%	75%

B

1 … 20	21 … 100
20€	0€
20%	80%

9 References

Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica, 21*, 503–46.

Arechar, A.A., Gächter, S., Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics, 21*, 99-131.

Bardsley, N. (2018). What lessons does the 'replication crisis' in psychology hold for experimental economics? In A. Lewis (Ed.), *Cambridge handbooks in psychology. The Cambridge handbook of psychology and economic behaviour,* 42–67.

Bateman, I.J., Day, B., Loomes, G. and Sugden, R. (2007). Can ranking techniques elicit robust values? *Journal of Risk and Uncertainty, 34*(1), 49–66.

Battalio, R.C., Kagel, J.H., Jiranyakul, K. (1990). Testing between Alternative Models of Choice under Uncertainty: Some Initial Results. *Journal of Risk and Uncertainty, 3*, 25–50.

Beattie, J., Loomes, G. (1997). The Impact of Incentives Upon Risky Choice Experiments. *Journal of Risk and Uncertainty, 14*(2), 155-168.

Bell, D. (1982). Regret in Decision Making under Uncertainty. *Operations Research, 20*, 961-81.

Bem, D.J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology,100*, 407–25.

Bleichrodt, H., Wakker, P.P. (2015). Regret Theory: A Bold Alternative to the Alternatives. *The Economic Journal, 125*, 493-532.

Camerer, C.F. (1989). An Experimental Test of Several Generalised Utility Theories. *Journal of Risk and Uncertainty, 2*(1), 61–104.

Camerer, C.F. (1995). Individual Decision Making in *Handbook of Experimental Economics*. J. Kagel and A. E. Roth, eds. Princeton: Princeton University Press.

Camerer, C.F., Dreber, A., Forsell, E., Ho, T-H., Huber, J., Johannesson, M., Kirchler, M. et al. (2016). Evaluating Replicability of Laboratory Experiments in Economics. *Science, 351*(6280), 1433-1436.

Camerer, C.F., Ho, T-H. (1994). Violations of the Betweenness Axiom and Nonlinearity in Probability. *Journal of Risk and Uncertainty, 8*, 167–96.

Camerer, C.F., Hogarth, R.M. (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty, 19*, 7-42.

Chew, S.H., Waller, W.S. (1986). Empirical Tests of Weighted Utility Theory. *Journal of Mathematical Psychology, 30*, 55–72.

Conlisk, J. (1989). Three Variants on the Allais Example. *American Economic Review, 79*, 392–407.

Cubitt, R., Gächter, S., Quercia, S. (2017). Conditional Cooperation and Betrayal Aversion. *Journal of Economic Behavior and Organizations, 141*, 110-121.

Cubitt, R., Starmer, C., Sugden, R. (1998). Dynamic Choice and the Common Ratio Effect: An Experimental Investigation. *The Economic Journal, 108*(450), 1362-1380.

Day, B., Loomes, G. (2010). Conflicting violations of transitivity and where they may lead us. *Theory and Decision, 68*(2), 233–42.

Doyen S., Klein O., Pichon, C.L., Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PloS one 7*(1), e29081.

Duvendack, M., Palmer-Jones, R., Reed, W.R. (2017). What Is Meant by 'Replication' and Why Does It Encounter Resistance in Economics? *American Economic Review, 107*(5), 46-51.

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90*, 891–904.

Fishburn, P.C. (1982). Nontransitive Measurable Utility. *Journal of Mathematical Psychology, 26*, 31–67.

Gächter, S., Starmer, C., Tufano, F. (2015). Measuring the closeness of relationships: A comprehensive evaluation of the “Inclusion of the Other in the Self” scale. *PLoS ONE, 10*(6), e0129478. doi.org/10.1371/journal.pone.0129478

Gilbert, D.T., King, G., Pettigrew, S., Wilson, T.D. (2016). Comment on “Estimating the Reproducibility of Psychological Science”. *Science, 351*(6277), 1037-1037.

Greenwald, A.G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*, 1-20.

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

Grether, D.M., Plott, C.R. (1979). Economic Theory of Choice and the Preference Reversal Phenomenon. *American Economic Review, 69*, 623-638.

Harless, D.W. (1992). Predictions about Indifference Curves inside the Unit Triangle: A Test of Variants of Expected Utility. *Journal of Economic Behavior and Organization, 18*, 391–414.

Hey, J.D. (1991). *Experiments in Economics*. Oxford: Blackwell.

Holt, C.A. (1986). Preference Reversals and the Independence Axiom. *American Economic Review, 76*, 508–15.

Humphrey, S.J. (1995). Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *Journal of Risk and Uncertainty, 11*(3), 263–74.

Humphrey, S.J. (1996). Do anchoring Effects underlie event-splitting effects? An experimental test. *Economics Letters, 51*, 303-308.

Humphrey, S.J. (2000). The common consequence effect: testing a unified explanation of recent mixed evidence. *Journal of Economic Behavior and Organization, 41*, 239-262.

Humphrey, S.J. (2001a). Non-Transitive Choice: Event-Splitting Effects or Framing Effects? *Economica, 68*(269), 77-96.

Humphrey, S.J. (2001b). Are Event-Splitting Effects Actually Boundary Effects? *Journal of Risk and Uncertainty, 22*(1), 79-93.

Ioannidis, J.P. (2005). Why Most Published Research Findings Are False. *PLoS Medicine 2*(8): e124.

Ioannidis, J.P., Munafò M.R., Fusar-Poli P., Nosek B.A., David SP. (2014). Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. *Trends in Cognitive Sciences,18*(5), 235-41.

John, L.K., Loewenstein, G., Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices with Incentives for Truth-telling. *Psychological Science 23*(5), 524–532.

Kagel, J., Roth, A. (1995). The Handbook of Experimental Economics. Princeton, NJ: *Princeton University Press*.

Kahneman, D., Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263-291.

Karni, E., Safra, Z. (1987). 'Preference Reversal' and the Observability of Preferences by Experimental Methods. *Econometrica, 55*, 675–85.

Kerr, N.L. (1998). HARKing: hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196-217.

Lichtenstein, S., Slovic, P. (1971). Reversals of Preference between Bids and Choices in Gambling Decisions. *Journal of Experimental Psychology, 89*, 46–55.

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

Loomes, G. (1988). Further Evidence of the Impact of Regret and Disappointment in Choice under Uncertainty. *Economica, 55*(217), 47-62.

Loomes, G., Pogrebna, G. (2017). Do preference reversals disappear when we allow for probabilistic choice? *Management Science, 63*(1), 166–184.

Loomes, G., Starmer, C., Sugden, R. (1989). Preference Reversal: Information-Processing Effect or Rational Non-transitive Choice? *Economic Journal, 99*(395), 140–51.

Loomes, G., Starmer, C., Sugden, R. (1991). Observing Violations of Transitivity by Experimental Methods. *Econometrica, 59*(2), 425–39.

Loomes, G., Sugden, R. (1982). Regret Theory: An Alternative Theory of Rational Choice under Uncertainty. *Economic Journal, 92*, 805–24.

Loomes, G., Sugden, R. (1983). A Rationale for Preference Reversal. *American Economic Review, 73*, 428–432.

Loomes, G., Sugden, R. (1987). Some implications of a more general form of regret theory. *Journal of Economic Theory, 41*, 270-287.

Loomes, G., Taylor, C. (1992). Nontransitive Preferences over Gains and Losses. *Economic Journal, 102*, 357–365.

MacCrimmon, K., Larsson, S. (1979). Utility theory: Axioms versus 'paradoxes', in Allais, M., Hagen, O. (eds.), *Expected utility hypotheses and the Allais paradox* (Reidel, Dordrecht).

Machina, M. (1982). 'Expected utility' theory without the independence axiom. *Econometrica, 50*, 277-323.

Maniadis, Z., Tufano, F, List, J.A. (2015). How to make experimental economics research more reproducible: lessons from other disciplines and a new proposal. (C. Deck, E. Fatas and T. Rosenblat, eds.), *Research in Experimental Economics, 18: Replication in Economic Experiments*, 215– 30, Bingley: Emerald Group Publishing.

Marschak, J. (1950). Rational Behavior, Uncertain Prospects, and Measurable Utility. *Econometrica, 18*, 111-141.

Millroth, P., Nilsson, H., Juslin, P. (2019). The decision paradoxes motivating Prospect Theory: The prevalence of the paradoxes increases with numerical ability. *Judgment and Decision Making, 14*(4), 513-533.

Moskowitz, H. (1974). Effects of Problem Presentation and Feedback on Rational Behavior in Allais and Morlat-Type Problems. *Decision Sciences, 5*, 225–42.

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology, 69,* 511–534.

Open Science Collaboration (2014). The Reproducibility Project: a model of large-scale collaboration for empirical research on reproducibility. In Stodden, V., Leisch, F., Peng, R.D. (eds.) *Implementing Reproducible Research*. Chapman & Hall/CRC the R series. CRC Press, Taylor & Francis Group, 299 – 324.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science 349*(6251), aac4716.

Quiggin, J. (1982) A Theory of Anticipated Utility. *Journal of Economic Behavior and Organization, 3*(4), 323-242.

Rosenthal, R. (1979). The 'file drawer problem' and tolerance for null results. *Psychological Bulletin. 86*(3), 638–641.

Roth, A.E. (1994). Let's Keep the Con out of Experimental Econ.: A Methodological Note. *Empirical Economics*, 19, 279-289. Reprinted in J.D. Hey (ed.), *Experimental Economics*, Physica Verlag, Heidelberg, 99-109.

Rubin, M. (2017). When does HARKing hurt? Identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. *Review of General Psychology, 21*(4), 308–320

Rubinstein, A. (1988). Similarity and decision-making under risk (is there a utility resolution to the Allais paradox? *Journal of Economic Theory, 46*(1), 145–53.

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*, 90-100.

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

Simmons, J.P., Nelson L.D., Simonsohn U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–66.

Slovic, P., Lichtenstein, S. (1983). Preference reversal: a broader perspective. American *Economic Review, 73*(4), 596–605.

Slovic, P., Tversky, A. (1974). Who Accepts Savage's Axiom? *Behavioral Science, 19*, 368-373.

Smith, V., Walker, J. (1993). Monetary Rewards and Decision Cost in Experimental Economics. *Economic Inquiry, 31*, 245-261.

Starmer, C. (1992). Testing New Theories of Choice under Uncertainty Using the Common Consequence Effect. *Review of Economic Studies*, 59, 813-830.

Starmer, C. (1999). Cycling with Rules of Thumb: An Experimental Test for a New Form of Non-Transitive Behaviour. *Theory and Decision, 46*, 141–58.

Starmer, C. (2000). Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk. *Journal of Economic Literature, 38*, 332-382.

Starmer, C., Sugden, R. (1989). Probability and Juxtaposition Effects: An Experimental Investigation of the Common Ratio Effect. *Journal of Risk and Uncertainty, 2*, 159-178.

Starmer, C., Sugden, R. (1991). Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation. *American Economic Review, 81*, 971-978.

Starmer, C., Sugden, R. (1993). Testing for juxtaposition and event-splitting effects. *Journal of Risk and Uncertainty, 6(*1), 235–54.

Starmer, C., Sugden, R. (1998). Testing Alternative Explanations of Cyclical Choices. *Economica, 65*(259), 347–361.

Thaler, R. (1987). The psychology of choice and the assumptions of economics. In A. E. Roth (ed.), *Laboratory Experimentation in Economics: Six Points of View.* Cambridge: Cambridge University Press.

Tversky, A. (1969). Intransitivity of Preferences. *Psychological Review, 76*, 31–48.

Tversky, A., Koehler, D.J. (1994). Support Theory: A Nonextensional Representation of Subjective Probability. *Psychological Review, 101*, 547–67.

Tversky, A., Sattath, S. and Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95*(3), 371–84.

Tversky, A., Slovic, P., Kahneman, D. (1990). The Causes of Preference Reversal. *American Economic Review, 80*(1), 204–17.

van de Kuilen, G., Wakker, P. (2006). Learning in the Allais Paradox. *Journal of Risk and Uncertinty, 33*, 155-164.

Viscusi, W.K. (1989). Prospective Reference Theory: Toward an Explanation of the Paradoxes. *Journal of Risk and Uncertainty, 2*, 235–64.

von Neumann, J., Morgenstern, O. (1947). *The Theory of Games and Economic Behaviour*. 2nd ed. Princeton: Princeton University Press.

Weber, M., Eisenführ, F., von Winterfeldt, D. (1988). The Effects of Splitting Attributes on Weights in Multi Attribute Utility Measurement. *Management Science, 34*, 431–45.

Wu, G., Gonzalez, R. (1998). Common Consequence Conditions in Decision Making under Risk. *Journal of Risk and Uncertainty, 16*, 115-139.

Chapter 3

Who accepts the sure-thing principle now?

1 Introduction

Expected Utility Theory (EUT) is the main model of individual choice in economics. It is proposed as a normative and descriptive model, which can be derived from a set of intuitively appealing axioms on preference.

Allais (1953a, 1953b) contested this as he considered violations of EUT to be reasonable and demonstrated this with his paradoxes (common consequence effect and common ratio effect), proposed as thought experiments on the normativity of the independence axiom at the Paris Symposium in 1952. Savage (1954), on the other hand, was persuaded of the normativity of independence and claimed that individuals would not knowingly violate it. Accordingly, if they had the opportunity to reflect upon a violation, they would realise it was a mistake, endorse the axiom and conform with it.[33]

Slovic and Tversky (1974) proposed a novel approach to resolve the debate that did not merely rely on competing philosophical discussions about the normativity of axioms. They took an empirical approach by conducting an experiment in which subjects had to respond to the decision problems proposed by Allais (1953a). Based on their choices, subjects received a normative argument in favour of the counterfactual choice and were then asked to reconsider the decision problems and make their choices again. This chapter reports the results of an experiment based on Slovic and Tversky's (1974) design, which seeks to replicate their findings and to identify the decision motives that underlie them.

The motivation for doing this is three-fold. Firstly, Slovic and Tversky's (1974) paper has been cited more than 850 times (Google Scholar) and continues to play an important role in the debate over the normativity of EUT (Dietrich et al., 2021). Yet, by present standards, the

[33] There are different axiomatizations of (subjective) expected utility theory, wherein choice principles have different names. For the purposes of the present chapter, "the sure-thing principle" and "the independence axiom" are used interchangeably.

experiment protocols it employs would be considered unsuitable in terms of delivering reliable data. For example, the sample size is small and choices were not incentive compatible. The continued relevance of their findings merits a careful check of whether they are robust to state-of-the-art methods used in the experimental observation of choice behaviour.

Secondly, recent developments in behavioural welfare economics (e.g. Thaler and Sunstein, 2008) have centred around the usefulness of behavioural nudges to allow decision-makers to make welfare enhancing choices (i.e. help individuals to make better decisions in their best interest as well as that of society). According to Thaler and Sunstein (2008), nudges are justified if decision-makers themselves want to be nudged. This argument is used as a defence against accusations that nudges are unduly paternalistic. Slovic and Tversky's (1974) design can be interpreted as investigating whether violators of independence want to be nudged into EUT consistent risky choices. Their results therefore bear upon the question of whether behavioural nudges would be justified in risky decision-making situations where violations of EUT would imply welfare losses. An updated and extended version of Slovic and Tversky's (1974) design is an important contribution to this literature.

Thirdly, the early work of Allais (1953a, 1953b) and Savage (1954) inspired an enormous amount of experimental work, which has documented systematic violations of the axioms of EUT (see Starmer 2000 for a review). Despite this, EUT remains the standard theory of risky choice in economics. One reason for this is that behavioural deviations from the model do not reliably undermine its normative status. The fact that decision-makers do not conform with the model does not mean that they should not conform with it. However, behavioural deviations from expected utility maximisation are consistent with the proposition that decision-makers violate the theory's axioms because they do not believe they should conform to them. The experiments reported here can be understood as investigating whether evidence can be found to support the prima facie consistency between violations of EUT's axioms and an underlying normative rejection of them by decision-makers.

Derived from modern interpretations of the experimental method, this study implements a version of Slovic and Tversky's (1974) design, which incorporates a large sample size and incentive compatible decisions in two experiments. To understand the decision motives, which underlie behaviour in these type of tests, two additional treatments are included to examine the role of experimenter demand effects and status quo bias. In addition, the same experimental design is applied to the common ratio effect, which is done in this form for the first time. Section 2 presents an overview of the related literature. Section 3 describes the experimental design and implementation, followed by results in Section 4. Section 5 includes the implications of the results and general concluding remarks.

2 Motivation

The discovery of robust and reproducible systematic violations of Expected Utility Theory (EUT) has revealed its limitations as a descriptive model in a number of predictable decision situations. Early and pioneering work conducted by Maurice Allais (1953a) not only documented perhaps the most famous of the violations, but also used its introspective appeal as the basis upon which to challenge the notion that EUT could survive descriptive failures because of its claims to being the standard of normative behaviour in risky decision-making. That is, not only did Allais (1953a, 1953b) challenge the descriptive credentials of EUT, he also challenged its normative status by explicitly questioning the normative appeal of the independence axiom. Allais' argument is based upon the (first) 'Allais Paradox' (or common consequence effect) as follows. The 'Allais Paradox', proposed at a conference in 1952, comprised two hypothetical decision problems, that required choices between two gambles.

The decision problems can be shown in the following form in Figure 9. Consider the payoff matrix in Figure 9 with rows showing four risky acts (Gambles A to D) and columns representing three states of the world (with associated probabilities shown at the top of each column). The resulting squares contain the state contingent payoffs to each act in millions of francs.

		Probabilities		
		0.1	*0.01*	*0.89*
Problem 1	Gamble A:	100M	100M	100M
	Gamble B:	500M	0M	100M
Problem 2	Gamble C:	100M	100M	0M
	Gamble D:	500M	0M	0M

Figure 9: Allais Paradox

Here, the payoffs of Gamble A and B (C and D respectively) are identical in the third state. Following the implications of the independence axiom of EUT, the third state has to be excluded from choices, since both gambles contain the same consequence in this state. Thus, if the third state is ignored, both choice problems turn out to be identical. According to the independence axiom, this implies that if Gamble A is chosen over Gamble B, then Gamble C is chosen over Gamble D - or vice-versa. However, there is considerable evidence that a substantial proportion of individuals prefer A over B and D over C in a setting with this general structure, a combination of preferences that contradict the independence axiom of EUT (Starmer, 2005).[34]

The sense in which observed violations of the independence axiom, in line with the Allais Paradox, are regarded by Allais (1953b) as undermining the normative status of the axiom, is

[34] See chapter 2 of this thesis for a more formal explanation of the Allais paradox.

this pattern of preferences in introspectively appealing. Allais (1953a) proposed that it is reasonable to choose the certainty in problem 1 (why take the risk, no matter how small, of missing-out on becoming a millionaire) and alongside choosing the riskier option in problem 2 (if winning nothing is probably going to happen anyway, it makes sense to take on the extra small risk of winning nothing in an attempt to obtain the highest possible outcome).

Prima facie support is lent to Allais' (1953b) argument by the fact that Savage, a strong proponent of the normativity of EUT, himself violated the independence axiom in Allais' (1953a) decision problems in the direction argued by Allais (1953b) as being reasonable. Savage (1954) responded to his own violation of the independence axiom in his seminal book *The Foundations of Statistics.* He explained that, upon reflection, maintaining his initially expressed preferences would have been irrational. Retaining his belief in the normative appeal of the independence axiom, he stated, that his initial preference for gamble 4 over gamble 3 had been erroneous. Savage's (1954) explanation might seem like an attempt to defend his own theory. However, he argued that the main use of his axioms is their normative role in 'policing' decisions for consistency. In this respect, Savage (1954) contends that his own experience is merely an example of the axioms performing their normative role in the manner that they should. On Savage's (1954) view, it is legitimate to conclude that his initial response to Allais' (1953a) decision problems had been a mistake, and that it was the correct use of normative reasoning embedded in the independence axiom had helped him to detect the error and adjust it (Starmer, 2005; Chandler, 2017).

The different views of Allais (1953b) and Savage (1954) regarding the implications of the Allais Paradox for the normative status of the independence axiom have appeared (in various forms) in the broader literature which discusses the normativity of EUT. For example, Allais (1953b) argued that, given that observed independence violations are reasonable, then if individuals who generally behave rationally violate the independence axiom, that case against the normativity of that axiom is strengthened (Chandler, 2017). Friedman and Savage (1952), on the other hand, imply that this simply demonstrates that the decision-maker, is not (yet) normatively committed to the choice principle, and does not mean that it is not normatively persuasive (say, upon reflection). These arguments, in their various forms, are synthesised by Starmer (2005), whose conclusions form the basis of the motivation for the experiment reported below.

Starmer (2005) discusses the problems associated with making normative inferences on the basis of descriptive data. His discussion is based upon a critique of Friedman and Savage's (1952) argument that the introspective normative appeal of EUT's axioms can be considered 'indirect' evidence for those axioms that should be weighed alongside the 'direct' evidence of observed decisions. A central part of Friedman and Savage's (1952) argument (applied to the independence axiom) is that a decision-maker who accepts the axiom will probably adhere to it. Starmer (2005) describes this in terms of two propositions, as follows:

> (P1): a decision-maker accepts the independence axiom as a choice principle.
>
> (P2) this decision-maker probably behaves in accordance with the independence axiom. (Starmer, 2005, p.285).

In these terms, Friedman and Savage's (1952) argument can be interpreted as stating that if P1 holds, P2 will probably be true. Starmer's (2005) argument is that whilst P1 and P2 may be consistent with each other, the former is a proposition about a normative judgement and the latter is a proposition about behaviour. Therefore, the former does not imply the latter without there being an explicit statement which connects them (and in the absence of such, Friedman and Savage's (1952) argument is incomplete). To link the normative beliefs embodied in P1 to the behaviour embodied in P2, Starmer (2005) suggests a third proposition:

(P3) the decision-maker will rarely behave in contradiction with principles they believe they should not deliberately violate (Starmer, 2005, p.285).

Starmer (2005) points out that the mere assumption that P3 is true would not be sufficient to enable the interpretation of observed behaviour which is consistent with the independence axiom as evidence that the axiom is normatively appealing. If such an interpretation is to be made, it needs to be empirically valid. In other words, Starmer (2005) is arguing that if conclusions regarding the normativity of a choice principle, such as the independence axiom, are to be based upon descriptive data, those data need to be observed under conditions where it can be reliably concluded that observed choices are based upon (rather than simply being consistent with) an underlying normative judgement.

Starmer's (2005) critique of Friedman and Savage (1952) provides an ex post motivation for an experimental test of Allais'(1953b) and Savage's (1954) competing arguments regarding the normativity of the independence axiom, reported by Slovic and Tversky (1974).[35] Slovic and Tversky's (1974) experiment design forms the basis of the experiment reported below, and is summarized as follows.

The experiment incorporates three stages. Subjects first respond to Allais-type decision problems of the type described above. Second, they are presented with a normative argument which supports the counterfactual behaviour to that exhibited in their initial decision. Finally, they are asked to respond to the Allais choice problems again. In Slovic and Tversky's (1974) design, the normative arguments presented to subject after their initial decisions reflect the competing claims of Allais (1953b) and Savage (1954) regarding the normativity of the independence axiom. i.e. subjects who make initial decisions consistent with EUT (and the independence axiom) were presented with Allais' (1953b) argument explaining why

[35] A similar interpretation of Starmer (2005) can be found in MacCrimmon (1968), Moskowitz (1974), and MacCrimmon and Larsson (1979).

independence violations are reasonable. Subjects who violated the independence axiom in initial decisions received an explanation of the normative logic of the axiom that underpins Savage's (1954) viewpoint.

Slovic and Tversky (1974) observed that 59% of subjects (17 out of 29) violated independence as suggested by Allais (1953a) in the first stage. After exposure to the counter argument (favouring of Savage's (1954) position), subjects tended to continue to violate the independence axiom in the repeated decisions: 66% of the subjects (19 out of 29) violated independence in the repeated choices. Overall, the majority of subjects chose to violate EUT in both the first and the last stage of the experiment, leading Slovic and Tversky (1974) to conclude that although the normative implications of the independence axiom were presented to subjects in a comprehensible manner, this did not lead to a decreased tendency for that axiom to be violated.

Slovic and Tversky's (1974) design satisfies Starmer's (2005) criterion for evaluating the normativity of choice principles using descriptive data in the following respect. If a decision-maker who violated independence in the initial decisions continued to do so in repeated choices, they had done so having had the logic of the independence axiom explained to them, and having had the opportunity for a Savage-like reflection upon their initial independence-violating decisions. Their repeated choices can in this way be interpreted as reflecting an underlying normative judgement. Similarly, subjects who initially chose in accordance with the independence axiom, but violated it in repeated decisions, did so having had the logic of independence violations explained to them (Allais, 1953b) and having had the opportunity to reflect, in light of this, upon their initial choice.

Although, in light of Starmer (2005), the way in Slovic and Tversky's (1974) design connects underlying normative judgements to observed choices can be considered ahead of its time, it would also be considered deficient by modern standards of experimental research in two important respects. First, the sample size is small, which limits the robustness of the observed data. Second, choices in Slovic and Tversky (1974) were hypothetical.

The lack of a system of genuine financial incentives is potentially problematic in Slovic and Tversky's (1974) design because it implies a lack of control for demand effects. i.e. subjects may have changed their behaviour in accordance with the argument they were presented with, not because this reflects genuine preference, but because they considered to be what the experimenter expected them to do. Doing so would have no implications for payment for taking part in the experiment, because of the lack of incentive compatibility. The experiment reported below addresses both of these issues: it uses a large sample of subjects and implements a system of financial incentives to engender incentive compatible choices. The latter controls for experimenter demand effects by inducing decision-makers to behave in a way that satisfies the objective of maximizing payment. i.e. whatever choices subjects make in the repeated decisions, they do so in light of the *both* the normative argument they have seen *and* knowledge that the choices are relevant for the payment they receive for participation in the experiment.

In addition to the improvements in the experiment design described above, the experiment reported below includes three additional treatments to control for factors that could have influenced behaviour in Slovic and Tversky's (1974) experiment and tests the robustness of their findings (and the behaviour observed in the experiment reported here). These are described in detail in the next section and are summarised as follows.

First, there is an additional control for an experimenter demand effect that may be induced by showing each subject a single normative argument that supports counterfactual choices. This is controlled by showing subjects normative arguments that support both their actual and counterfactual behaviour.

Second, there is a control for a potential status quo bias stemming from the possibility that a reference point could be induced by showing subjects the choices they had previously made (before asking them to repeat those decisions). This may have inclined subjects to stick with those initial choices. This is controlled in a treatment in which subjects are presented with the normative arguments, which respectively support the independence axiom and its violation, before being asked to make their choices in Allais' common consequence effect decision problems. i.e. in this treatment, there are no repeated choices. Subjects simply see both arguments and then decide how to choose.

Thirdly, the experiment extends Slovic and Tversky's (1974) design to the second Allais Paradox (the common ratio effect). The same experimental method is adapted and applied to common ratio choice problems, to investigate whether the normative arguments affect behaviour in the same way for both Allais Paradoxes. This extension serves as a robustness check of the normativity of the independence axiom, because if that axiom has normativity it should apply to both the common consequence and common ratio versions of Allais' decision situations.

The connection that Slovic and Tversky's (1974) design makes between observed choices and underlying normative judgements can be further motivated by the behavioural welfare economics literature (Camerer et al., 2003; Sunstein and Thaler, 2003, 2008). In their book '*Nudge*', Thaler and Sunstein (2008) present an approach that aims at improving individuals' decisions by changing the way options are presented to them. This is done by using so called 'nudges', that exploit or respond to cognitive biases. Their approach is based on the assumption that individuals make 'bad' decisions, when not paying attention and not having complete information, adequate cognitive abilities and self-control. Thus, choices made in the absence of these factors may be reconstructed by using nudges to reveal individuals' underlying (bias-free) preferences, which can be used as an empirical basis for normative economics (Sugden, 2020).

Dold and Schubert (2018) call this approach an application of ‘laundering’ preferences. They emphasize that behavioural welfare economics should respect individuals’ preferences only after they have been ‘laundered’ to remove erroneous effects. Thus, in this approach, it is normally assumed that the bias-free decision-maker is neoclassical. Whilst the experiment described below can be interpreted as an application of the laundered preferences approach (with laundering achieved by the normative arguments, which Sudgen (2018) refers to as a ‘classic nudge’), it is not assumed that laundered preferences are necessarily neoclassical. Hence, this experiment can be interpreted as an investigation of whether decision-makers who violate the independence axiom can be nudged towards EUT consistent behaviour (and vice versa).

3 Experiment

3.1 Design and hypotheses

The experiment is designed around a baseline treatment from a previously implemented experiment and four further treatments, where two of the four treatments control for alternative explanations of Slovic and Tversky’s (1974) results and one deals with the common ratio effect (see A3.1 in the Appendix of this chapter for the complete set of decision problems and A3.1.5 for the test results of the subsidiary decision problems). All treatments will be presented in detail below.

The common consequence effect (CCE), one of the so-called Allais Paradoxes, is probably the most famous violation of EUT’s independence axiom (Conlisk, 1989; Starmer and Sugden, 1991; Starmer, 1992). In this experiment, the CCE will be tested in the *baseline treatment* and *treatments 1, 2* and *3*.[36] Figures 10 and 11 illustrate how the decision problems are presented to the subjects. The decision problems entail pairwise choices between options **A** and **B** in problem 1 and options **A’** and **B’** in problem 2.

Problem 1

Option A

1	…	100
	14 €	
	100%	

Option B

1 … 20	21…25	26 … 100
20€	0€	14€
20%	5%	75%

Figure 10: CCE decision problem 1

Problem 2

Option A’

1 … 25	26 … 100
14€	0€
25%	75%

Option B’

1 … 20	21 … 100
20€	0€
20%	80%

Figure 11: CCE decision problem 2

Each of the four boxes represents a prospect. The numbers in the boxes represent monetary outcomes that can be won and the numbers underneath the boxes describe the chances out of one hundred (in percentages) of winning each amount of money.

EUT consistent choices occur when individuals choose either the safer options **A** and **A’**, or the riskier options **B** and **B’** in both problems. Violations of the independence axiom of EUT, consistent with Allais’ predictions, occur when individuals choose the safer option **A** over the

[36] See subchapter 3.2 of Chapter 2 for an extensive formal representation of the CCE decision problem.

riskier option **B** in the first problem and then choose the riskier option **B'** over the safer option **A'** in the second problem. Choosing option **B** over option **A** in the first problem and then option **A'** over option **B'** in the second problem also violates the independence axiom, but in the opposite direction to that discussed by Allais (1953a).

After stating their initial choices on the Allais-type problem pairs, subjects are given a written normative argument pertaining to the decision problems and are then asked to respond to the same decision problems again. The presentation of arguments varies across treatments. In the *baseline treatment* and *treatment 1*, EUT maximisers receive a version of Allais' (1953b) argument that the CCE is reasonable and EUT violators receive an argument supporting the logic of independence. In *treatment 2*, subjects are presented both arguments, independent of their initial choices. *Treatment 3* differs from the other treatments, since subjects receive both arguments before stating their choices. Thus, subjects in this treatment only respond once to the decision problems. The two arguments involved in these treatments either advocate Allais' (1953) or Savage's (1954) position and read as follows:

Allais' Position

> *"I would choose option **A** over option **B** in problem 1 and option **B** over option **A** in problem 2.*
>
> *In problem 1, I have the choice between €14 for certain or a gamble where I might end up with nothing. Why gamble? The small probability of missing the chance to win something seems very unattractive to me.*
>
> *In problem 2, there is a good chance that I will end up with nothing no matter what I do. The chances of getting 20€ are almost as good as getting 14€, so I might as well go for the 20€ and choose option **B** over option **A**."*

Savage's position

*"In problem 1, if a ticket number between 26 and 100 is drawn, it doesn't matter whether I choose **A** or **B**. I win 14 € irrespectively how I choose. So I will ignore lottery tickets 26 to 100.*

*In problem 2, the same is true. If a ticket number between 26 and 100 is drawn, it doesn't matter whether I choose **A** or **B**. I win 0 € irrespectively of how I choose. So I will ignore lottery tickets 26 to 100.*

*Therefore, because I should always ignore tickets 26 and above, the problems are exactly the same. In both problems, option A tickets 1-25 always pay me 14 €. In both problems, option **B** tickets 1-20 pay 20 € and tickets 21-25 pay nothing.*

*It therefore makes no sense to switch between **A** and **B**. So I would choose either option **A** in both problems or option **B** in both problems."*

3.1.1 Baseline treatment

The *baseline treatment* is run within the replication study in the second chapter of this thesis and also serves as an instrument check for the present indirect replication of Slovic and Tversky's (1974) test. It involves two Allais-type problem pairs (see figures 10 and 11) presented to subjects in an online random lottery experiment within a total of thirty choice problems (see A2.1 in the Appendix of chapter 2). As the experiment is implemented online, this leads to a loss of control compared to the laboratory. Thus, it must be determined whether this is an appropriate and reliable method to observe genuine preferences. Since the replication study contains a number of decision problems which have been robustly observed over thirty years of laboratory experiments, it can be assumed that the online implementation is reliable if the same patterns of behaviour are observed there as in the laboratory. Considering the results of the replication study, this assumption can be confirmed. This means that the tests in the *baseline treatment* can be considered to deliver reliable data, and that the online experiment is an appropriate method for conducting a more detailed investigation and replication of Slovic and Tversky's experiment.

In the *baseline treatment*, after responding to all choice problems, subjects were presented with one of the arguments described above. The argument they received depended on the choices they had made. Subjects who respected the independence axiom and made decisions consistent with EUT, received the argument supporting Allais' position. Subjects who violated the independence axiom received an argument advocating Savage's position and in support of behaving according to EUT. Following the opportunity to reflect upon these arguments, subjects were asked to respond to the same decision problems again.

In initial choices, the null hypothesis is that there are no systematic violations of EUT, and hence, that **AB'** and **BA'** choices occur with equal frequency (i.e. deviations from EUT are random noise). A common consequence effect is where systematically more **AB'** violations are observed than **BA'** violations. The same hypothesis tests will also be conducted on the repeated choice data. This allows an aggregate-level understanding of the net influence of the normative arguments on observed behaviour.

Individual-level analysis will complement the aggregate level analysis to understand how individuals responded (if at all) to the arguments they faced. Here, the null hypothesis is that the arguments do not affect choices. The alternative hypothesis is that behaviour in post-argument decisions differs from that in pre-argument decisions. If the null hypothesis is rejected, and therefore that arguments matter, a comparison of switching rates from respecting to violating EUT, and vice versa, allows an understanding of the normative force of each type of argument. These hypothesis tests apply in each treatment. The purpose and nature of a comparison of behaviour between treatments is explained in the description of treatments below.

3.1.2 Treatment 1

The *treatments 1* to *4* are run in a subsequent online experiment in order to provide a more detailed and extensive investigation of Slovic and Tversky's (1974) findings and those in the *baseline treatment*. *Treatment 1* is identical to the *baseline treatment* in 3.1.1. Although it is implemented in a different experiment, the problem pairs, the respective parameter values and the presented arguments are the exactly the same as in the *baseline treatment* (see figures 10 and 11).

Here, the two relevant Allais-type problem pairs are presented to subjects in an online random lottery experiment within a total of twenty choice problems (see A3.1.1 in the Appendix). Four of these problems were the pre- and post-argument CCE problems and the subsidiary eighteen problems were standard pairwise lottery choice tasks. They were included to retain comparability with the *baseline treatment*, but are not relevant to the hypotheses testing in this chapter, and are therefore not discussed any further. But note that the decision problems were a set of similar pairwise lottery choice tasks with a broadly equivalent expected value.

As *treatment 1* is a direct replication of the *baseline treatment*, the procedure and hypothesis tests in this experiment are the same as described above.

3.1.3 Treatment 2

In this treatment, subjects respond to the twenty-two choice problems, including the decision problems illustrated in figures 10 and 11 (see A3.1.2. in the Appendix). In *treatment 2,* subjects are presented both normative arguments, representing Savage's and Allais' position, after their initial choices, i.e. irrespectively of how they choose in initial choices, they see both arguments. This procedure is implemented to control for potential demand effects that could influence behaviour in the *baseline treatment*, *treatment 1* and in Slovic and Tversky (1974).

It might be that subjects have an inclination to behave in a way that the experimenter expects or which they perceive the experimenter to consider reasonable. Thus, it is conceivable that a reference point for this is determined by the normative argument that is presented to the subject. This means, for example, that subjects confronted with only an argument in favour of the counterfactual might tend to adopt the behaviour advocated in this argument because they believe this will meet the experimenter's expectations.

One experimental method commonly employed to control for demand effects is the introduction of financial incentives. As incentives seem to induce departure from conceivably desirable behaviour they may be useful for minimizing demand effects (Camerer and Hogarth, 1999; Zizzo, 2010). Contrary to Slovic and Tversky's (1974) original study, the present experiment controls for possible demand effects by using financial incentives within a random lottery incentive system. An additional control for demand effects is facilitated by controlling for subjects conditioning their responses to the argument they see, on the perception that this argument represents the behaviour that the experimenter considers reasonable. By presenting subjects with both arguments, which support how they have actually behaved and the counterfactual course of action, it is not possible for subjects to condition responses in this way, i.e. a demand-effect inclination to switch to the single argument presented, is precluded by exposure to both arguments.

Comparing behaviour observed in this treatment to that observed in the previous two (*baseline* and *treatment 1*) will reveal whether demand effects played a role in the data observed in the *baseline treatment*, *treatment 1* and Slovic and Tversky (1974).

3.1.4 Treatment 3

A possible influence on behaviour in Slovic and Tversky's (1974) study is a status quo bias. *Treatment 3* is designed to control for this possibility. The status quo bias is a preference for the current state and is a general experimental finding discovered in a variety of decision-making experiments. It has been found that individuals are disproportionately inclined to stick to the status quo. Kahneman, Knetsch and Thaler (1991) suggest the status quo bias to be prompted by loss aversion, a value asymmetry first introduced by Kahneman and Tversky (1984). In a choice situation between sticking to the status quo or an alternative outcome, individuals take the status quo as the current reference point. According to loss aversion, potential losses by switching from this reference point loom larger than the potential gains. Thus, individuals are biased towards favouring the status quo (Samuelson and Zeckhauser, 1988).

There are lots of different forms a status quo bias could take and various possible explanations. For the purposes of the present experiment, a status quo bias is taken to mean a pre-disposition towards sticking with original choices. This may be supported by a reference point created, or reinforced, by being reminded of initial choices just prior to the presentation of the normative arguments. A status quo bias of this type could deflate switching rates, and thereby lead to an inaccurate reflection of switching attributable to purely normative considerations. In addition, because there should be no presumption that the strength of any status quo bias is independent of the original choices, a differential status quo bias could contribute to the different swap rates observed between original and repeated choices.

A control for this type of status quo bias is implemented in *treatment 3*. This treatment is similar to the previous ones in terms of decision problems and parameter values. However, in this setting, subjects are presented both normative arguments, representing Savage's and Allais' position, without being asked to respond to the CCE decision problems beforehand.

This means that, in contrast to *treatments 1* and *2*, subjects in *treatment 3* respond to a total of twenty choice problems (see A3.1.3 in the Appendix).[37]

Comparing behaviour observed in this treatment to that observed in the previous, the *baseline* and *treatment 1*, will reveal whether a status quo bias played a role in the data observed in the *baseline treatment*, *treatment 1* and Slovic and Tversky (1974).

3.1.5 Treatment 4

In this treatment, subjects are asked to respond to two different Allais-type decision problems, embedded in the same set of questions as in the previous treatments (see A3.1.4 in the Appendix).

The relevant decision problems in *treatment 4* are designed to test for the common ratio effect (CRE), which is the second version of the so-called Allais Paradox (1953). In contrast to the CCE, the CRE is concerned with the extent to which preferences depend on probabilities. In this context, EUT implies that preferences should be independent of the value of associated probabilities. Evidence from a large body of experiments show that individuals exhibit systematic violations of the independence axiom when confronted with common ratio choice problems (Starmer and Sugden, 1989; Camerer, 1995, van de Kuilen and Wakker, 2006).

Figures 12 and 13 illustrate how the decision problems are presented to the subjects; the numbers inside the boxes representing monetary outcomes that can be won and the numbers underneath the boxes describing the chances of winning each sum of money. The decision

[37] This difference in the number of problems implies some loss of control. Alternatively, two additional different problems could have been added to *treatment 3*, but this would have likewise caused a loss of control. The decision against adding two different choice problems to this treatment has resulted from the following considerations: (i) all subjects in this experiments are confronted with the same set of problems and (ii) in response to Holt's (1986) objections to random-lottery incentive systems, Starmer and Sugden (1991) show that individuals treat decision problems independently and that the number of problems does not affect behaviour in individual decisions.

problems involve pairwise choices between options **A** and **B** in problem 1 and options **A’** and **B’** in problem 2.

Problem 1

Option A

1 … 100
11 €

100%

Option B

1 … 80	81 … 100
15€	0€
80%	20%

Figure 12: CRE decision problem 1

Problem 2

Option A’

1 … 25	26 … 100
11€	0€
25%	75%

Option B’

1 … 20	21 … 100
15€	0€
20%	80%

Figure 13: CRE decision problem 2

The probabilities of non-zero outcomes in problem 2 are generated from those in problem 1 by multiplying a common factor. Since this change is applied to both options, the ratio of ‘winning’ probabilities in options **A** and **B** is the same as that in options **A’** and **B’**. Hence, the independence axiom implies choosing either **A,** then **A’** or **B** then **B’**. The CRE involves violating independence by choosing **A** in the first problem and **B’** in the second problem. The less commonly observed violation of independence involves choosing **B** then **A’**.

After stating their initial choices on the Allais-type problem pairs, subjects are presented a written argument pertaining to the decision problems and are then asked to respond to the same decision problems again. The two arguments involved in *treatment 4* differ from the other treatments, but also either advocate Allais' (1953b) or Savage's (1954) position and read as follows:

Allais' position

*"I would choose option **A** over option **B** in problem 1 and option **B** over option **A** in problem 2.*

In problem 1, I have the choice between 11€ for certain or a gamble where I might end up with nothing. Why gamble? The small probability of missing the chance to win something seems very unattractive to me.

*In problem 2, there is a good chance that I will end up with nothing no matter what I do. The chances of getting 15€ are almost as good as getting 11€, so I might as well go for the 15€ and choose option **B** over option **A**."*

Savage's position

*"The two options in both problems 1 and 2 have exactly the same ratio of probabilities of a positive prize. In problem 1, these are probabilities of 80% for option **B** and 100% for option **A**. In problem 2, these are probabilities of 20% for option **B** and 25% for option **A**.*

The ratio of 80/100 is exactly the same as the ratio of 20/25. So these decisions can be considered effectively identical - they involve exactly the same trade-off between risk and return. In both problems, the higher prize of 15€ has a probability which is exactly 80% of the probability of winning 11€.

*So, if I accept this risk-return trade-off, I choose option **B** in both problems. On the other hand, if I reject this risk return trade-off, I choose option **A** in both problems."*

Similar to the CCE tests, the arguments could cause behaviour to go either way. This depends upon whether Allais' or Savage's argument has the greatest normative force. Comparing behaviour in pre-argument decisions and post-argument decisions allows an evaluation of whether the normative arguments affect decisions and, if they do, the relative normative force of each type of argument.

3.2 Implementation

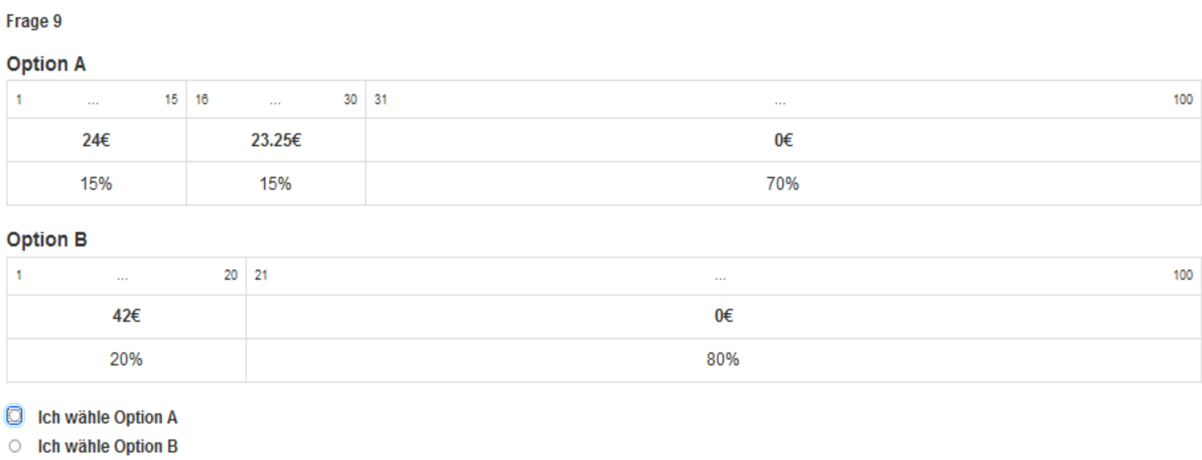
The computer-based experiments were conducted online during October 2020 (hereafter: experiment 1) and August 2021 (hereafter: experiment 2). Subjects were randomly recruited using ORSEE (Greiner, 2015) from the Laboratory of Economic Research's database of pre-registered volunteers at Osnabrück University, and invited to participate in one of a number of pre-arranged sessions. In experiment 1, involving the *baseline treatment*, 147 subjects took part, with an average age of 24 years and 62% of participants being female. Experiment 2 incorporated *treatments 1* to *4.* There, a total of 378 subjects participated of which 55% were female and the average age was 26 years. Participation was restricted to one session per subject, which was controlled by a registration protocol that prevented duplicate participation in both experiments. Subjects in both experiments were randomly allocated to pre-arranged sessions on Stud.IP, each of which followed exactly the same procedures.[38]

In a virtual meeting room called BigBlueButton (BBB), participants and the experimenter met at the beginning of the experiment sessions. In an introduction, it was explained to the subjects that the experiment assisted in research concerning decision-making under risk. Subjects were informed that they would be compensated for their participation with a show-up fee of 5€ and the chance to win an additional cash prize, depending on their decisions, which would be transferred to their bank accounts after the experiment. To this end, they were asked to provide their name and IBAN in the experiment, with the explanation that these data would be saved separately, and not accessible to the experimenter.

[38] Stud.IP is a web-based working environment used by Osnabrück University, to support courses at educational institutions, which has been developed as free software to enable and standardise communication between students and teaching staff via the internet.

Subjects were informed that payments were underwritten by a research fund of Osnabrück University and that transfers would be made by the finance department. Thereafter, the experimenter explained the general form of the tasks and the procedure of the session. Each subject received a unique, anonymised link to the online experiment platform SoPHIElabs. The participants had the possibility to contact the experimenter in a chat window in BBB at any time of the experiment session.

Payments were made according to the random lottery incentive system by informing subjects in the instructions that, when they had responded to all questions, one of those questions would be randomly selected by the computer and played out for real.[39] Note that for all treatments, except *treatment 3*, both the initial and repeated responses on the Allais-type questions were included in the random lottery incentive system. The question and the option chosen by the subjects, were presented to them on screen once again. The computer randomly selected a number between 1 and 100, determining the subject's winnings according to the option chosen in that question.


Frage 9

Option A

1 ... 15	16 ... 30	31 ... 100
24€	23.25€	0€
15%	15%	70%

Option B

1 ... 20	21 ... 100
42€	0€
20%	80%

Ich wähle Option A

Ich wähle Option B

Figure 14: Example of decision problem

[39] Holt (1986) suggests that experiments employing a random-lottery incentive system might not reveal individuals' true preferences. Starmer and Sugden (1991) provide experimental evidence which shows that this objection to be without empirical support.

For example, if the question in Figure 14 was randomly selected as the payment-relevant question and the subject had previously chosen Option B, a number between 1 and 20 would yield 42€ and numbers between 21 and 100 would yield no additional payment. To control for possible order effects within the experiment, the order of questions was randomized by the computer for each subject. Sessions in experiment 1 lasted for approximately 45 minutes, without enforced time limit, and average payoffs were 13.50€, including the show-up fee of 5€. In experiment 2, sessions lasted on average 25 minutes and average winnings were 11,50€ including the show-up fee of 5€.

4 Results

The data are analysed using a two-tailed binominal test based on the null hypothesis that there are no systematic violations of the independence axiom of EUT. The alternative hypothesis, consistent with Allais' (1953b) position, is that violations are not random, but systematic in a predictable direction. The following analysis is within-subjects.

Table 11 reports the observed choices (percentages in parentheses) in the Allais-type decision problems according to the presentation of arguments in each of the treatments. The *p*-value shows, given the null hypothesis, the probability of observing a pattern of violation by chance. The last two rows indicate whether the frequency of choices, consistent with either Allais' (HFO) or Savage's (EUT) position, changed after the presentation of the normative arguments.[40]

[40] Choices consistent with the Allais paradox (here **AB')** cause horizontally fanning-out (HFO) indifference curves (see Machina, 1982), whereas **BA'** choices, which also violate EUT, cause horizontally fanning-in (HFI) indifference curves (see Starmer, 1992).

Table 11:Results

	Treatment								
	Baseline		*1*		*2*		*3*	*4*	
Choice pattern	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	-	*Pre*	*post*
AA'(EUT cons.)	30 (20%)	31 (21%)	19 (16%)	24 (21%)	24 (21%)	22 (19%)	14 (19%)	37 (51%)	23 (31%)
BB'(EUT cons.)	57 (39%)	55 (37%)	41 (35%)	53 (45%)	42 (37%)	54 (47%)	30 (40%)	8 (11%)	10 (14%)
Total (**AA**'+**BB**')	87 (59%)	86 (58%)	60 (51%)	77 (66%)	66 (58%)	76 (66%)	44 (59%)	45 (62%)	33 (45%)
AB'(CCE /CRE) [41]	33 (23%)	50 (33%)	39 (34%)	29 (25%)	27 (23%)	26 (23%)	21 (28%)	24 (33%)	40 (55%)
BA'(EUT viol.)	27 (18%)	11 (8%)	17 (15%)	10 (9%)	22 (19%)	13 (11%)	9 (13%)	4 (5%)	0 (0%)
p-value	*0.519*	*0.000****	*0.005****	*0.003****	*0.568*	*0.053**	*0.004***	*0.000****	*0.000****
n	*147*		*116*		*115*		*74*	*73*	
Frequency of CCE/CRE and EUT consistent choices post-argument									
HFO (CCE/CRE)	Up		Down		Same			Up	
EUT cons.	Same		Up		Up			Down	

The choice patterns indicate the frequency of all possible conjunctions of choices over the problem pairs. The *p-value* is calculated for a two-tailed within subject test using the binomial distribution of the null hypothesis. *, ** and *** denote rejection of the null hypothesis at the 10%, 5% and 1% level respectively.

[41] CCE denotes violations of EUT consistent with Allais' position in the Baseline treatment and treatments 1 to 3. CRE denotes violations of EUT consistent with Allais' position in treatment 4.

For the *baseline treatment*, table 11 shows the null hypothesis, implying that the **AB’** and **BA’** choices occur with the same frequency, cannot be rejected for the initial choices. However, after the presentation of the normative arguments in favour of the counterfactual, there is a significant CCE. For the post-argument decisions, the null hypothesis can be rejected at the 1% level ($p<0.0001$). The aggregate data show that choices consistent with the CCE increased after the presentation of the argument and the amount of EUT consistent choices stayed the same.

In a similar vein to Slovic and Tversky (1974), these data show that the normative arguments do not reduce the CCE. In fact, they increase it and introduce a systematic violation of EUT in the post-argument data that did not exist in the pre-argument data.

In *treatment 1*, there is a significant CCE in both the initial and repeated choices. The null hypothesis can be rejected at the 1% level ($p=0.005$ and $p=0.003$) in both cases. Here, the data show that the frequency of choices consistent with the CCE decreased after the presentation of the argument and EUT consistent choices increased. However, in a similar manner to the *baseline treatment* and Slovic and Tversky (1974), these data show that the normative arguments do not reduce the CCE.

Treatment 2 is controlling for demand effects by presenting subjects with both normative arguments after their initial choices. Here, the null hypothesis cannot be rejected for the pre-argument choices. For the post-argument choices, the null hypothesis can be rejected at the 10% significance level ($p=0.053$), which implies that there is a weakly significant CCE. The aggregate data show that, after the presentation of both normative arguments, EUT consistent choices increased whereas the frequency of choices consistent with the CCE stayed the same. So the significant CCE is driven by a reduction in counter-CCE (**BA’**) violations.

Similar to the previous treatments, these data show that the normative arguments do not reduce the CCE. In fact, they introduce a systematic violation of EUT in the post-argument data, that did not occur in the pre-argument data here. As the presentation of both normative arguments is implemented to control for potential demand effects, the results suggest, that demand effects did not play a role in the data observed in the *baseline treatment*, *treatment 1* and Slovic and Tversky (1974).

Treatment 3, controls for status quo bias and might be considered the 'purest' test of normativity in this study. Here, both normative arguments were presented before responding to the choice problems. The null hypothesis can be rejected at the 5% level (*p=0.004*), which implies that there is a significant CCE. Similar to the *baseline treatment* and *treatment 1*, the aggregate data reveal a significant CCE in the post-argument decisions. Hence, these data show that a status quo bias did not influence results in the *baseline treatment*, *treatment 1* and Slovic and Tversky (1974).

Treatment 4 tested for the CRE and the data indicate that the null hypothesis can be confidently rejected at the 1% significance level (*p<0.000*) for both the initial and the repeated choices. The aggregate data show that the frequency of choices consistent with the CRE increased markedly after the presentation of the normative argument, whereas EUT consistent choices decreased. Similar to the results regarding the CCE observed in the *baseline treatment*, *treatment 1* and Slovic and Tversky (1974), these data show that the normative arguments do not reduce the CRE.

At the aggregate level, the data show that after the presentation of the arguments, the systematic violations of independence do not disappear. Quite the contrary, in the post-argument situations, the CCE reaches significance at least the 10% level in all the treatments considered in the present analysis. This finding contradicts to Savage's (1954) claim, that the use of normative reasoning helps to detect and adjust violations of EUT and that it may be used to 'police' decisions for consistency. In terms of factors that could have influenced results in the *baseline treatment*, *treatment 1* and Slovic and Tversky (1974), the aggregate data reveal that neither demand effects nor a status quo bias played a role in the data observed in these treatments.

As the aggregate level analysis does not contain complete information of how the normative arguments influence the individual's decision to switch or to stick with initial choices, an individual-level analysis is required to fully understand the normative force of Allais' and Savage's arguments.

Thus, in order to compare choice behaviour before and after the presentation of arguments more precisely, the choice behaviour at the individual level is examined using a two-sided McNemar's test of the null hypothesis that there is no difference in pre and post-argument choices.[42] For all treatments, except from treatment 3, the implemented tests assess the EUT-consistency, HFO-consistency and HFI-consistency of choices made before and after the arguments.

The data structure is shown in the tables below, where columns represent choice behaviour before and rows represent choice behaviour after the presentation of the arguments, with percentages in parentheses. Note that the totals do not add up to 100% in all three panels of the tables, as the percentages are expressed relative to the total numbers of subjects. The *p*-value shows, given the null hypothesis, whether the changes in choice behaviour before and after the arguments are significant. Table 12 reports results for the *baseline treatment*. Test results for EUT-consistency in panel 1 show, that 22% of subjects who made choices consistent with EUT

[42] McNemar's test is a statistical test, commonly used in before and after measurements with paired attributes. The test requires the same subject to be included in the before and after measurement. Here, the individual choices of each subject are included, before and after the presentation of the normative arguments.

before the argument, switched to violating EUT (**AB'** or **BA')**. The same proportion of subjects (22%) violated EUT before the argument and made EUT consistent choices after the argument. The null hypothesis, that there is no difference in pre and post-argument choices cannot be rejected (*p=0.901288*).

Table 12: McNemar's test results in *Baseline treatment*[43]

Panel 1: EUT consistency

		Post-argument		
		EUT consistent	Violation EUT	*totals*
Pre-argument	**EUT consistent**	54 (37%)	33 (22%)	*87 (59%)*
	Violation EUT	32 (22%)	28 (19%)	*60 (41%)*
	totals	*86 (59%)*	*61 (41%)*	*147 (100%)*
				p-value = 0.901288

Panel 2: HFO consistency

		Post-argument		
		HFO	**EUT consistent**	*totals*
Pre-argument	HFO	17 (12%)	16 (11%)	*33 (23%)*
	EUT consistent	31 (21%)	54 (37%)	*85 (58%)*
	totals	*48 (33%)*	*70 (48%)*	*118 (80%)*
				p-value = 0.028671*

Panel 3: HFI consistency

		Post-argument		
		HFI	**EUT consistent**	*totals*
Pre-argument	HFI	9 (6%)	16 (11%)	*25 (17%)*
	EUT consistent	2 (1%)	54 (37%)	*56 (38%)*
	totals	*11 (7%)*	*70 (48%)*	*81 (55%)*
				p-value = 0.000967***

A *p-value* less than 0.1, 0.05 and 0.01 is indicated with *, **, and ***, denoting a significant result.

[43] In panel 1, 28 subjects violated EUT in both pre and post argument problems. In panels 2 and 3, 17 subjects were consistent HFO violators in pre and post decisions and 9 subjects were consistent HFI violators. The extra 2 observations in panel 1 were subjects who switched from HFI to HFO here.

Thus, in the *baseline treatment,* the presentation of the normative argument favouring the counterfactual does not affect the overall amounts of EUT-consistent choices and violations of EUT. However, the aggregate data show that there is no CCE in the pre-argument choices, but a significant CCE in the post-argument choices. This means that the systematic post-argument CCE comes from a difference in how arguments influenced HFO relative to HFI (i.e. **AB’** relative to **BA’**), which is investigated in panels 2 and 3.

The second test, evaluating HFO-consistency in panel 2, shows that 11% of subjects who made choices consistent with HFO (**AB’** choices) before the counter-argument, switched to EUT-consistent choices after the argument. 21% of subjects whose choices were consistent with EUT predictions before, switched to HFO consistent choices after the argument. A two-sided McNemar’s test of the null hypothesis that there is no difference in pre and post-argument choices is rejected at the 5% significance level (*p=0.028671**). Thus, in the *baseline treatment,* the normative argument supporting the CCE, systematically nudges EUT maximizers towards HFO.

For completeness, the third test evaluating HFI-consistency in panel 3, illustrates how arguments affect **BA’** choices, that generate horizontally fanning-in (HFI) indifference curves (Starmer, 1992), relative to EUT choices. Here, 11% of subjects who made HFI-consistent choices before the argument, switched to choice patterns consistent with EUT afterwards. Only 1% of subjects whose choices were consistent with EUT before the argument switched to HFI-consistency after it. This low frequency is not surprising, since subjects who initially chose according to EUT’s predictions were provided with the counterargument in favour of Allais’ position (HFO-consistent). In this case, the null hypothesis is rejected with significance greater than 1% (p=0.000967***), indicating that the argument supporting EUT consistent choices, nudges HFI-consistent decision-makers to become consistent with EUT.

The individual-level confirms the patterns of behaviour indicated in the aggregate data. Panel 2 shows the argument supporting the CCE to have greater influence in engendering switches than the argument supporting EUT. Panel 3 shows that the argument advocating EUT consistent choices has normative force in systematically leading HFI-consistent subjects towards EUT maximisation. The combined influence of these individual-level effects leads to highly patterned violations of EUT in the post-argument aggregate data, in the direction of a CCE. The implications of these results are in line with Slovic and Tversky's (1974) finding that the normative arguments do not reduce the CCE, and the results in panel 2 suggest that the normative arguments significantly increased it in this treatment.

Table 13 reports the test results for individual choice behaviour in *treatment 1*. Results from the test on EUT-consistency illustrate that 16% of subjects who initially chose consistently with EUT, violated EUT by choosing **AB'** or **BA'** after the argument. 30% of subjects who violated EUT before the argument, switched to conform to EUT after the arguments were presented. The null hypothesis, that there is no difference in pre and post-argument choices, is rejected at 5% significance (*p=0.019537*). This indicates that in *treatment 1,* the presentation of the normative argument favouring the counterfactual systematically nudges EUT violators towards conforming with EUT. However, the aggregate data show a CCE in pre-argument choices, that persisted in post-argument choices. This means that the post-argument CCE depends on a difference in how arguments influenced HFO relative to HFI which is investigated in panels 2 and 3.

Table 13: McNemar's test results in *Treatment 1*[44]

Panel 1: EUT consistency

		Post-argument		
		EUT consistent	Violation EUT	*totals*
Pre-argument	**EUT consistent**	42 (36%)	18 (16%)	*60 (52%)*
	Violation EUT	35 (30%)	21 (18%)	*56 (48%)*
	totals	*77 (66%)*	*39 (34%)*	*116 (100%)*
				p-value = 0.019537**

Panel 2: HFO consistency

		Post-argument		
		HFO	**EUT consistent**	*totals*
Pre-argument	HFO	11 (9%)	25 (22%)	*36 (31%)*
	EUT consistent	16 (14%)	42 (36%)	*58 (50%)*
	totals	*27 (23%)*	*67 (58%)*	*94 (81%)*
				p-value = 0.159854

Panel 3: HFI consistency

		Post-argument		
		HFI	**EUT consistent**	*totals*
Pre-argument	HFI	5 (4%)	10 (9%)	*15 (13%)*
	EUT consistent	2 (2%)	42 (36%)	*44 (38%)*
	totals	*7 (6%)*	*52 (45%)*	*59 (51%)*
				p-value = 0.020921**

A *p-value* less than 0.1, 0.05 and 0.01 is indicated with *, **, and ***, denoting a significant result.

Results from the test on EUT-consistency illustrate that 16% of subjects who initially chose consistently with EUT, violated EUT by choosing **AB'** or **BA'** after the argument. 30% of subjects who violated EUT before the argument, switched to conform to EUT after the arguments were presented. The null hypothesis, that there is no difference in pre and post-

[44] In panel 1, 21 subjects violated EUT in both pre and post argument problems. In panels 2 and 3, 11 subjects were consistent HFO violators in pre and post decisions and 5 subjects were consistent HFI violators. The extra 5 observations in panel 1 were 3 subjects who switched from HFO to HFI and 2 subjects who switched from HFI to HFO here.

argument choices, is rejected at 5% significance (*p=0.019537*). This indicates that in *treatment 1,* the presentation of the normative argument favouring the counterfactual systematically nudges EUT violators towards conforming with EUT. However, the aggregate data show a CCE in pre-argument choices, that persisted in post-argument choices. This means that the post-argument CCE depends on a difference in how arguments influenced HFO relative to HFI which is investigated in panels 2 and 3.

Testing for HFO-consistency, the results in panel 2 illustrate that 22% of subjects who committed the CCE before the counter-argument, switched to EUT after the argument. 14% of subjects whose choices were consistent with EUT before the argument, committed the CCE after the argument. A two-sided McNemar's test of the null hypothesis that there is no difference in pre and post-argument choices, cannot be rejected. This indicates that the counter-arguments do not significantly affect EUT-consistent and HFO-consistent choice behaviour here.

The third test, evaluating HFI-consistency in panel 3, illustrates that 9% of subjects who made HFI-consistent choices before the argument, switched to choice patterns consistent with EUT afterwards and 2% of subjects who chose in consistency with EUT before, switched to HFI-consistency afterwards. Thus, the null hypothesis is rejected with significance greater than 5% (p=0.020921). This result suggests that the argument advocating EUT consistency nudges HFI-consistent decision-makers to become consistent with EUT. The patterns of behaviour indicated in the aggregate data are mirrored at the individual level. Panel 1 shows that the argument supporting EUT has a greater influence in engendering switches than the argument supporting the CCE. But, as panel 3 shows, this is mainly attributable to the normative force of the argument advocating EUT, which systematically leads HFI-consistent subjects towards EUT maximisation. However, the influence of the individual-level effects did not lead to a reduced CCE after the arguments. This finding is consistent with the results from the *baseline treatment* and Slovic and Tversky (1974).

The test results for individual choice behaviour in *treatment 2*, where subjects received both arguments after their initial choices, are presented in table 14.

Table 14: McNemar's test results in *Treatment 2*[45]

Panel 1: EUT consistency

		Post-argument		
		EUT consistent	Violation EUT	*totals*
Pre-argument	**EUT consistent**	53 (46%)	13 (11%)	*66 (57%)*
	Violation EUT	23 (20%)	26 (23%)	*49 (43%)*
	totals	*76 (66%)*	*39 (34%)*	*115 (100%)*
				p-value = 0.095581*

Panel 2: HFO consistency

		Post-argument		
		HFO	**EUT consistent**	*totals*
Pre-argument	HFO	12 (11%)	14 (12%)	*26 (23%)*
	EUT consistent	9 (8%)	53 (46%)	*62 (54%)*
	totals	*21 (19%)*	*67 (58%)*	*88 (77%)*
				p-value = 0.297147

Panel 3: HFI consistency

		Post-argument		
		HFI	**EUT consistent**	*totals*
Pre-argument	HFI	8 (7%)	9 (8%)	*17 (15%)*
	EUT consistent	4 (3%)	53 (46%)	*57 (49%)*
	totals	*12 (10%)*	*62 (54%)*	*74 (64%)*
				p-value = 0.165518

A *p-value* less than 0.1, 0.05 and 0.01 is indicated with *, **, and ***, denoting a significant result.

[45] In panel 1, 26 subjects violated EUT in both pre and post argument problems. In panels 2 and 3, 12 subjects were consistent HFO violators in pre and post decisions and 8 subjects were consistent HFI violators. The extra 6 observations in panel 1 were 1 subject who switched from HFO to HFI and 5 subjects who switched from HFI to HFO here.

For EUT-consistency, the results show that 11% of subjects whose initial choices were consistent with EUT, switched to violating EUT after the presentation of both normative arguments. In the opposite case, 20% of subjects who violated EUT before the presentation of the arguments, made choices consistent with EUT afterwards. The test results provide weak evidence against the null hypothesis that there is no difference in pre and post-argument (p=0.095581). Thus, in *treatment 2*, there presentation of both normative arguments nudges EUT-violating decision-makers towards consistency with the theory. The aggregate data, however, show no CCE in the pre-argument choices, but a significant CCE in the post-argument choices. This means that the systematic post-argument CCE comes from a difference in how arguments influenced HFO relative to HFI which is investigated in panels 2 and 3.

The second test, evaluating HFO-consistency in panel 2, illustrates that 12% of subjects committed the CCE before the arguments, switched to choices consistent with EUT after the presentation of both arguments. A proportion of 8% of subjects first made EUT consistent choices and switched to commit the CCE after being presented both normative arguments. Here, the p-value fails to reach significance and therefore, the null hypothesis cannot be rejected. Hence, for this case, the arguments do not significantly affect EUT-consistent and HFO-consistent choice behaviour.

In the third test, examining HFI-consistency in panel 3, 8% of subjects whose choices were consistent with HFI before the presentation of arguments, switched to making choices consistent with EUT afterwards. 3% of subjects who initially made EUT consistent choices, switched to making choices consistent with HFI after the presentation of both arguments. The null hypothesis cannot be rejected in this case, indicating that the arguments do not affect EUT-consistent and HFI-consistent choice behaviour. In contrast to the previous treatments, this test fails to reach significance, which may depend on the implemented procedure controlling for a potential demand effects as described above. In the previous treatments, subjects who made choices consistent with HFI before the arguments, were presented with an argument supporting EUT. It may therefore be possible, that demand effects influenced behaviour in those treatments.

Although panel 1 shows the argument supporting EUT to have a slightly greater influence in engendering switches, the results in panels 2 and 3 reveal no significant influence of the arguments on HFO and HFI. Thus, a combination of the individual-level effects in panels 2 and 3 resulted in EUT-consistency becoming weakly significant in panel 1. The patterns of behaviour in the aggregate data are not attributable to a single pattern in the individual-level data, but stem from a combination of individual decision motives. Panels 2 and 3 indicate, that neither of the arguments has a greater normative force. However, the combined influence of the individual-level effects has led to highly patterned violations of EUT in the post-argument aggregate data in the direction of a CCE. Similar to the previous treatments and Slovic and Tversky (1974), there is no reduction of the CCE in post-argument choices. Hence, as this treatment controls for a potential demand effect, the overall behaviour observed in the previous treatments and Slovic and Tversky (1974) cannot be explained by a demand effect.

Table 15 presents individual choices on the CRE problem pairs in *treatment 4*, before and after subjects were confronted with the normative argument favouring the counterfactual depending on their initial choices.

Here, testing for EUT-consistency, the data show that 32% of subjects whose initial choices were consistent with EUT, switched to violate EUT by choosing **AB’** or **BA’** after the presentation of the counter-argument. The null hypothesis, that there is no difference in pre and post-argument choices is rejected at 5% significance (*p=0.039592*). This result indicates that in *treatment 4,* the presentation of the normative argument favouring the counterfactual systematically nudges EUT-maximizers towards choice behaviour consistent with the CRE.

Here, the argument supporting the CRE has a stronger impact in engendering switches, relative to the argument supporting the CCE in the *baseline treatment* and *treatment 1*.

Table 15: McNemar's test results in *Treatment 4*[46]

Panel 1: EUT consistency

		Post-argument		
		EUT consistent	Violation EUT	*totals*
Pre-argument	**EUT consistent**	22 (30%)	23 (32%)	*45 (62%)*
	Violation EUT	11 (15%)	17 (23%)	*28 (38%)*
	totals	*33 (45%)*	*40 (55%)*	*73 (100%)*
				p-value = 0.039592**

Panel 2: HFO consistency

		Post-argument		
		HFO	**EUT consistent**	*totals*
Pre-argument	HFO	14 (19%)	10 (14%)	*24 (33%)*
	EUT consistent	23 (32%)	22 (30%)	*45 (62%)*
	totals	*27 (51%)*	*22 (44%)*	*69 (95%)*
				p-value = 0.023635**

Panel 3: HFI consistency

		Post-argument		
		HFI	**EUT consistent**	*totals*
Pre-argument	HFI	0 (0%)	1 (1%)	*1 (1%)*
	EUT consistent	0 (0%)	22 (30%)	*22 (30%)*
	totals	*0 (0%)*	*23 (31%)*	*23 (31%)*
				p-value = 0.317311

A *p-value* less than 0.1, 0.05 and 0.01 is indicated with *, **, and ***, denoting a significant result.

The second test, evaluating HFO-consistency in panel 2, shows that 14% of subjects who initially made choices consistent with the CRE switched to choices consistent with EUT after

[46] In panel 1, 17 subjects violated EUT in both pre and post argument problems. In panels 2 and 3, 14 subjects were consistent HFO violators in pre and post decisions and 0 subjects were consistent HFI violators. The extra 3 observations in panel 1 were 4 subjects who switched from HFI to HFO here.

the argument. 32% of subjects who first chose according to EUT switched to commit the CRE after the argument. Here, the null hypothesis that there is no difference between pre- and post-argument choices is rejected at the 5% significance level (*p=0.023635*). This result indicates that the argument in support of the CRE systematically nudges EUT-maximizers towards HFO-consistency, similar to the CCE argument in the *baseline treatment*.

In the third test, considering HFI-consistency illustrated in panel 3, the data show that 1% of subjects who initially made choices consistent with HFI switched to EUT consistent behaviour after the presentation of the argument. In the opposite case, no EUT consistent decision-maker switched to HFI-consistency after the argument. This observation is reasonable since subjects who initially behaved consistent with EUT, received an argument in favour of the CRE advocating HFO. The null hypothesis cannot be rejected in this case, indicating that the counter-arguments do not affect EUT-consistent and HFI-consistent choice behaviour. Alongside the aggregate data, these results imply that the normative argument supporting EUT has a weaker impact on choice behaviour of HFI-consistent decision-makers here, compared to those in the *baseline treatment* and *treatment 1*.

In this treatment, the patterns of behaviour indicated in the aggregate data are mirrored at the individual level. Panel 2 shows that the argument in support of the CRE has a strong normative force in systematically leading EUT-maximizers towards CRE consistent behaviour. Compared to the *baseline treatment* and *treatment 1*, the results of this individual-level analysis suggest that the argument in support of the CRE has a stronger influence in engendering switches than that supporting the CCE. However, on the aggregate level, the implications of the arguments are the same for both types of the Allais Paradox. Moreover, it can be observed that the presentation of normative arguments did not reduce or remove systematic violations of the independence axiom of the EUT across all treatments in this experiment.

5 Conclusion

The results presented in this study contribute to both the literature of normative decision theory and behavioural welfare economics. The most striking finding in this updated and properly incentivized version of Slovic and Tversky's (1974) experiment, which also controls for a demand effect and a status quo bias, is that significant violations of EUT either persist or emerge after the presentation of normative arguments, but not disappear. This pattern is observed across all treatments in this experiment which runs contrary to Savage's (1954) argument that violations of the normative axioms disappear when individuals get to reconsider and reflect upon their decisions. The robustness of independence violations in these treatments suggest that the observations in the *baseline treatment*, *treatment 1* and Slovic and Tversky (1974) are not attributable to a demand effect or a status quo bias of the type described.

Recalling Starmer's (2005) critical debate of Friedman and Savage's (1952) premise on connecting actual choice behaviour with underlying normative judgements, he argued that for this premise to hold, it has to be empirically valid. This study provides empirical findings that speak to this issue. A significant proportion of individuals who initially violated EUT and were presented a normative argument in favour of EUT continued to violate the theory. These choices were made in full knowledge of the logic of consistency and the knowledge that the final payments might depend upon them. Thus, in line with Starmer (2005), the findings from this study cast doubt on Friedman and Savage's (1952) premise and its implication that the normative appeal of choice principles implies empirical validity.

As the connection between normative principles and actual choice behaviour played a decisive role in the development of behavioural welfare economics, this study provides data that can be interpreted in terms of nudging individuals' behaviour in a particular direction. The individual data analysis of the *baseline treatment*, *treatment 1* and *treatment 4* is considered in this context, as the presentation of an argument in favour of the counterfactual can be interpreted as a classic nudge. Under the 'laundered preferences' approach to behavioural welfare economics, the role of behavioural nudges is to purge (launder) preferences of cognitive and behavioural biases, in order to identify bias-free preference, the satisfaction of which will increase welfare. On the assumption that the normative arguments presented to subjects in this study constitute an appropriate nudge of this type (Sugden, 2018 suggest they are), the findings of this study show that the presumption, that laundered preference will be neoclassical, cannot be supported. This is because the data are mixed. There is a significant tendency towards HFO-consistency in the *baseline treatment*, whereas significant changes towards EUT-consistency were observed in *treatment 1*. Thus, it can neither be concluded that one argument has been more appealing than the other, nor can it be inferred that individuals' underlying preferences are neoclassical. Results from *treatment 4*, that involved CRE problem pairs, point in the opposite direction. Here, the presentation of the normative argument favouring the counterfactual systematically and significantly nudged EUT-maximizers towards HFO-consistency. In addition, on an aggregate level, the frequency of choices consistent with HFO increased significantly, whereas EUT-consistent choices decreased. These findings and the significant persistence of independence violations after the presentation of normative arguments across all treatments, run contrary to the core element of behavioural welfare economics, assuming that a bias-free individual is neoclassical.

In final conclusion, it is perhaps a fitting testament to pioneering work of Paul Slovic and Amos Tversky (1974) that findings from work they published almost half a century ago are robust to extensive new controls, modern protocols and properly motivated decisions. The ongoing relevance and influence of their work means it was deserving of an update and a robustness-check. Having done this, the data reported here show that decision-makers do not accept the sure-thing principle.

6 Acknowledgements

The assistance of SoPHIE Labs and Felix Meickmann is gratefully acknowledged for programming the experiment software and subject recruitment, respectively.

7 Appendix

Appendix A3 Procedures and instructions

During the recruitment process, subjects were informed that they would need to provide an IBAN because payments would be made by bank transfer. Subjects were informed that to participate in the experiment they would need a device capable of web browsing and with an audio outlet. After the registrations had closed, registered subjects were provided with a link to a BigBlueButton online meeting which corresponded to their session of the experiment.

In the online meeting room subjects were able to see the experimenter, but no other participant. Subjects were individually provided with a link to the experiment software and a participation code via a chat function. They were then asked to use their code to log into the experiment software and await further instructions.

After all subjects had logged into the experiment software they were informed that the introductory screen would be read aloud by the experimenter and that questions could be asked at any time by sending a message to the experimenter via the chat function of the online meeting room.

The experiment comprised four treatments, all of which followed the same procedure, except from treatment 3. In that treatment, subjects were asked to respond to 20 questions, whereas subjects were asked to respond to 22 questions in the remaining treatments. The introductory screen was adapted accordingly to 20 questions for treatment 3.

Instructions

Introductory screen:

"Welcome to this experiment. Thank you for participating. Before we begin, you will get some general information about the procedure of this experiment.

There are no "tricks" in this experiment. There are no right or wrong answers and the experiment is not a "test". We are only interested in finding out how people make choices in particular situation, which we will then pay you for.

All decisions will be made anonymously, i.e. no other participant will know how you chose and you will not know the choices of the other participants either. All participants will have their payment transferred to their bank account after the experiment.

The funds used for this experiment are drawn from our research account at Osnabrück University. You can therefore be sure that you will be paid for your participation.

Do you have any questions?

In this experiment, you will be asked to make 22 choices between pairs of options, called Option A and Option B. Each option offers a prize with different probabilities between 1-100%. Each of these situations involves real money. The prize therefore depends on a number between 1-100, which will be randomly determined for each of you individually.

You mark your choice by ticking a box next to your preferred option and press "Next". After you press "Next", your choice will be saved and you will be taken to the next question. Please note that you cannot return to the previous questions at this point, so your choice is binding. The following question serves as an example of the type of choices you will make during the experiment:

Option A

1 … 20	21 … 100
3€	0€
20%	80%

Option B

1 … 10	11 … 100
7€	0€
10%	90%

❒ I choose option A

❒ I choose option B

In this example, if you prefer option A, mark your choice by ticking the box next to "I choose option A". If you prefer option B, mark your choice by ticking the box next to "I choose option B".

The amount of money you receive in this experiment comprises two parts. First, everyone receives a payment of 5 € for participating. Furthermore, you will also receive the prize of the option you selected in one of the 22 questions. You will only find out which of the 22 questions is actually played out for real money once you have made your choices in all of the questions.

At the end of the experiment, a consecutive number field appears, numbered from 1 to 22. The number field is carried out in real time by the computer and stops randomly and individually for each participant at one number. Then, the number that appears is the question that determines your payoff. For example, if the number 6 is chosen, the option you chose in question 6 will determine your payment. A consecutive number field will then appear again, numbered from 1-100. This will determine the outcome of the choice you made in your actual payoff-relevant question.

Imagine that the above example question was randomly as your payment-relevant question. If you had chosen option A here, this option would be relevant for your payment. If the second consecutive number field (with numbers from 1 to 100) happened to stop at a number between

1 and 20, your payment would be 3€. If the number field stopped between 21 and 100, your payment would be 0€.

Since you will not know until the end of the experiment which of the 22 questions will eventually determine your payoff, any of the 22 could be the one that determines your payoff. You should therefore make all of your choices as if they are for real money (because one of them will be). Since real money is involved, please consider your choice carefully and make it according to your actual preferences.

Are there any other questions? Then move on to the experiment."

After the introduction, subjects were directed to the experiment software to start the experiment.

Arguments Screen:

When the 22 decision problems were completed, the initial choices in the CCE problems were put on the screen along with the subjects' choices and the normative arguments were presented as described in subchapter 3.1 of this chapter. Subjects were then told: "We would now like to ask you to make a decision on these decision problems again. Please remember that one of these choices may determine your actual payout. Since real money is at stake, please consider your choices carefully and decide in the way that you prefer. Note that there are no right or wrong answers." Subjects then repeated their choices.

Demographic Questionnaire:

After responding to all questions, subjects were asked to complete a demographic questionnaire regarding gender, age and their studies.

Payoff screen:

Subjects were informed of their payoff-relevant question, their resulting payment and when this payment would be issued. Subjects were then asked to provide the IBAN number for the account to which they wanted payment transferred. Transfers were scheduled immediately after the experiment by the experimenter.

A3.1 Decision problems

The set of decision problems used for the analysis of the BASELINE treatment is provided in A2.1 of the appendix of chapter 2.

The decision problems used in the treatments 1, 2, 3 and 4 are presented below. In addition, the test results of the subsidiary decision problems not reported in this chapter, are presented thereafter.

A3.1.1 Decision problems treatment 1

CCE pre-argument:

Question 1

A

1 … 100
14 €
100%

B

1 … 20	21…25	26 … 100
20€	0€	14€
20%	5%	75%

Question 2 (CCE pre-argument)

A

1 … 25	26 … 100
14€	0€
25%	75%

B

1 … 20	21 … 100
20€	0€
20%	80%

CRE – HFI (low incentives):

Question 3

A

1 … 100
3 €
100%

B

1 … 10	11 … 20	21 … 100
7€	0€	3€
10%	10%	80%

Question 4

A

1 … 20	21 … 100
3€	0€
20%	80%

B

1 … 10	11 … 100
7€	0€
10%	90%

CRE – HFI (high incentives):

Question 5

A

1 … 100
9 €
100%

B

1 … 10	11 … 20	21 … 100
21€	0€	9€
10%	10%	80%

Question 6

A

1 … 20	21 … 100
9€	0€
20%	80%

B

1 … 10	11 … 100
21€	0€
10%	90%

ESE:

Question 7

A

1 … 30	31 … 70	71 … 100
42€	0€	0€
30%	40%	30%

B

1 … 30	31 … 70	71 … 100
0€	15€	15€
30%	40%	30%

Question 8

A

1 … 30	31 … 100
42€	0€
30%	70%

B

1 … 30	31 … 100
0€	15€
30%	70%

ESE – different matrix order:

Question 9

A

1 … 30	31 … 70	71 … 100
15€	15€	0€
30%	40%	30%

B

1 … 30	31 … 70	71 … 100
0€	0€	42€
30%	40%	30%

Question 10

A

1 … 70	71 … 100
15€	0€
70%	30%

B

1 … 70	71 … 100
0€	42€
70%	30%

Cyclical choices:

Question 11

A

1 … 20	21 … 100
42€	0€
20%	80%

B

1 … 30	31 … 100
24€	0€
30%	70%

Question 12

A

1 … 30	31 … 100
24€	0€
30%	70%

B

1 … 15	16 … 30	31 … 100
24€	23.25€	0€
15%	15%	70%

Question 13

A

1 … 15	16 … 30	31 … 100
24€	23.25€	0€
15%	15%	70%

B

1 … 20	21 … 100
42€	0€
20%	80%

Reversed PR choices:

Question 14

A

1 … 10	11…15	16…20	21 … 100
54€	0€	0€	0€
10%	5%	5%	80%

B

1 … 10	11…15	16…20	21 … 100
24€	24€	0€	0€
10%	5%	5%	80%

Question 15

A

1 … 10	11…15	16…20	21 … 100
24€	24€	0€	0€
10%	5%	5%	80%

B

1 … 10	11…15	16…20	21 … 100
12€	12€	12€	0€
10%	5%	5%	80%

Question 16

A

1 … 0	11…15	16…20	21 … 100
12€	12€	12€	0€
10%	5%	5%	80%

B

1 … 10	11…15	16…20	21 … 100
54€	0€	0€	0€
10%	5%	5%	80%

Juxtaposition effect:

Question 17

A

1 … 55	56 … 100
0€	11€
55%	45%

B

1 … 45	46 … 100
0€	7€
45%	55%

Question 18

A

1 … 55	56 … 100
0€	11€
55%	45%

B

1 … 55	56 … 100
7€	0€
55%	45%

Juxtaposition effect – different matrix order:

Question 19

A

1 … 85	86 … 100
0€	11€
85%	15%

B

1 … 80	81 … 100
0€	7€
80%	20%

Question 20

A

1 … 15	16 … 100
11€	0€
15%	85%

B

1 … 80	81 … 100
0€	7€
80%	20%

CCE post-argument:

Question 21

A

1 … 100
14 €
100%

B

1 … 20	21…25	26 … 100
20€	0€	14€
20%	5%	75%

Question 22

A

1 … 25	26 … 100
14€	0€
25%	75%

B

1 … 20	21 … 100
20€	0€
20%	80%

A3.1.2 Decision problems treatment 2

CCE pre-argument:

Question 1

A

1 … 100
14 €
100%

B

1 … 20	21…25	26 … 100
20€	0€	14€
20%	5%	75%

Question 2

A

1 … 25	26 … 100
14€	0€
25%	75%

B

1 … 20	21 … 100
20€	0€
20%	80%

CRE – HFI (low incentives):

Question 3

A

1 … 100
3 €
100%

B

1 … 10	11 … 20	21 … 100
7€	0€	3€
10%	10%	80%

Question 4

A

1 … 20	21 … 100
3€	0€
20%	80%

B

1 … 10	11 … 100
7€	0€
10%	90%

CRE – HFI (high incentives):

Question 5

A

1 … 100
9 €
100%

B

1 … 10	11 … 20	21 … 100
21€	0€	9€
10%	10%	80%

Question 6

A

1 … 20	21 … 100
9€	0€
20%	80%

B

1 … 10	11 … 100
21€	0€
10%	90%

ESE:

Question 7

A

1 … 30	31 … 70	71 … 100
42€	0€	0€
30%	40%	30%

B

1 … 30	31 … 70	71 … 100
0€	15€	15€
30%	40%	30%

Question 8

A

1 … 30	31 … 100
42€	0€
30%	70%

B

1 … 30	31 … 100
0€	15€
30%	70%

ESE – different matrix order:

Question 9

A

1 … 30	31 … 70	71 … 100
15€	15€	0€
30%	40%	30%

B

1 … 30	31 … 70	71 … 100
0€	0€	42€
30%	40%	30%

Question 10

A

1 … 70	71 … 100
15€	0€
70%	30%

B

1 … 70	71 … 100
0€	42€
70%	30%

Cyclical choices:

Question 11

A

1 … 20	21 … 100
42€	0€
20%	80%

B

1 … 30	31 … 100
24€	0€
30%	70%

Question 12

A

1 … 30	31 … 100
24€	0€
30%	70%

B

1 … 15	16 … 30	31 … 100
24€	23.25€	0€
15%	15%	70%

Question 13

A

1 … 15	16 … 30	31 … 100
24€	23.25€	0€
15%	15%	70%

B

1 … 20	21 … 100
42€	0€
20%	80%

Reversed PR choices:

Question 14

A

1 … 10	11…15	16…20	21 … 100
54€	0€	0€	0€
10%	5%	5%	80%

B

1 … 10	11…15	16…20	21 … 100
24€	24€	0€	0€
10%	5%	5%	80%

Question 15

A

1 … 10	11…15	16…20	21 … 100
24€	24€	0€	0€
10%	5%	5%	80%

B

1 … 10	11…15	16…20	21 … 100
12€	12€	12€	0€
10%	5%	5%	80%

Question 16

A

1 … 0	11…15	16…20	21 … 100
12€	12€	12€	0€
10%	5%	5%	80%

B

1 … 10	11…15	16…20	21 … 100
54€	0€	0€	0€
10%	5%	5%	80%

Juxtaposition effect:

Question 17

A

1 … 55	56 … 100
0€	11€
55%	45%

B

1 … 45	46 … 100
0€	7€
45%	55%

Question 18

A

1 … 55	56 … 100
0€	11€
55%	45%

B

1 … 55	56 … 100
7€	0€
55%	45%

Juxtaposition effect – different matrix order:

Question 19

A

1 … 85	86 … 100
0€	11€
85%	15%

B

1 … 80	81 … 100
0€	7€
80%	20%

Question 20

A

1 … 15	16 … 100
11€	0€
15%	85%

B

1 … 80	81 … 100
0€	7€
80%	20%

CCE post-argument:

Question 21

A

1	…	100
	14 €	
	100%	

B

1 … 20	21…25	26	…	100
20€	0€		14€	
20%	5%		75%	

Question 22

A

1 … 25	26	…	100
14€		0€	
25%		75%	

B

1 … 20	21	…	100
20€		0€	
20%		80%	

A3.1.3 Decision problems treatment 3

CRE – HFI (low incentives):

Question 1

A

1 … 100
3 €
100%

B

1 … 10	11 … 20	21 … 100
7€	0€	3€
10%	10%	80%

Question 2

A

1 … 20	21 … 100
3€	0€
20%	80%

B

1 … 10	11 … 100
7€	0€
10%	90%

CRE – HFI (high incentives):

Question 3

A

1 … 100
9 €
100%

B

1 … 10	11 … 20	21 … 100
21€	0€	9€
10%	10%	80%

Question 4

A

1 … 20	21 … 100
9€	0€
20%	80%

B

1 … 10	11 … 100
21€	0€
10%	90%

ESE:

Question 5

A

1 … 30	31 … 70	71 … 100
42€	0€	0€
30%	40%	30%

B

1 … 30	31 … 70	71 … 100
0€	15€	15€
30%	40%	30%

Question 6

A

1 … 30	31 … 100
42€	0€
30%	70%

B

1 … 30	31 … 100
0€	15€
30%	70%

ESE – different matrix order:

Question 7

A

1 … 30	31 … 70	71 … 100
15€	15€	0€
30%	40%	30%

B

1 … 30	31 … 70	71 … 100
0€	0€	42€
30%	40%	30%

Question 8

A

1 … 70	71 … 100
15€	0€
70%	30%

B

1 … 70	71 … 100
0€	42€
70%	30%

Cyclical choices:

Question 9

A

1 … 20	21 … 100
42€	0€
20%	80%

B

1 … 30	31 … 100
24€	0€
30%	70%

Question 10

A

1 … 30	31 … 100
24€	0€
30%	70%

B

1 … 15	16 … 30	31 … 100
24€	23.25€	0€
15%	15%	70%

Question 11

A

1 … 15	16 … 30	31 … 100
24€	23.25€	0€
15%	15%	70%

B

1 … 20	21 … 100
42€	0€
20%	80%

Reversed PR choices:

Question 12

A

1 … 10	11…15	16…20	21 … 100
54€	0€	0€	0€
10%	5%	5%	80%

B

1 … 10	11…15	16…20	21 … 100
24€	24€	0€	0€
10%	5%	5%	80%

Question 13

A

1 … 10	11…15	16…20	21 … 100
24€	24€	0€	0€
10%	5%	5%	80%

B

1 … 10	11…15	16…20	21 … 100
12€	12€	12€	0€
10%	5%	5%	80%

Question 14

A

1 … 0	11…15	16…20	21 … 100
12€	12€	12€	0€
10%	5%	5%	80%

B

1 … 10	11…15	16…20	21 … 100
54€	0€	0€	0€
10%	5%	5%	80%

Juxtaposition effect:

Question 15

A

1 … 55	56 … 100
0€	11€
55%	45%

B

1 … 45	46 … 100
0€	7€
45%	55%

Question 16

A

1 … 55	56 … 100
0€	11€
55%	45%

B

1 … 55	56 … 100
7€	0€
55%	45%

Juxtaposition effect – different matrix order:

Question 17

A

1 … 85	86 … 100
0€	11€
85%	15%

B

1 … 80	81 … 100
0€	7€
80%	20%

Question 18

A

1 … 15	16 … 100
11€	0€
15%	85%

B

1 … 80	81 … 100
0€	7€
80%	20%

CCE post-argument:

Question 19

A

1 … 100
14 €
100%

B

1 … 20	21…25	26 … 100
20€	0€	14€
20%	5%	75%

Question 20

A

1 … 25	26 … 100
14€	0€
25%	75%

B

1 … 20	21 … 100
20€	0€
20%	80%

A3.1.4 Decision problems treatment 4

CRE pre-argument

Question 1

A

1 … 100
11 €
100%

B

1 … 80	81 … 100
15€	0€
80%	20%

Question 2

A

1 … 25	26 … 100
11€	0€
25%	75%

B

1 … 20	21 … 100
15€	0€
20%	80%

CRE – HFI (low incentives):

Question 3

A

1 … 100
3 €
100%

B

1 … 10	11 … 20	21 … 100
7€	0€	3€
10%	10%	80%

Question 4

A

1 … 20	21 … 100
3€	0€
20%	80%

B

1 … 10	11 … 100
7€	0€
10%	90%

CRE –HFI (high incentives):

Question 5

A

1 … 100
9 €
100%

B

1 … 10	11 … 20	21 … 100
21€	0€	9€
10%	10%	80%

Question 6

A

1 … 20	21 … 100
9€	0€
20%	80%

B

1 … 10	11 … 100
21€	0€
10%	90%

ESE:

Question 7

A

1 … 30	31 … 70	71 … 100
42€	0€	0€
30%	40%	30%

B

1 … 30	31 … 70	71 … 100
0€	15€	15€
30%	40%	30%

Question 8

A

1 … 30	31 … 100
42€	0€
30%	70%

B

1 … 30	31 … 100
0€	15€
30%	70%

ESE – different matrix order:

Question 9

A

1 … 30	31 … 70	71 … 100
15€	15€	0€
30%	40%	30%

B

1 … 30	31 … 70	71 … 100
0€	0€	42€
30%	40%	30%

Question 10

A

1 … 70	71 … 100
15€	0€
70%	30%

B

1 … 70	71 … 100
0€	42€
70%	30%

Cyclical choices:

Question 11

A

1 … 20	21 … 100
42€	0€
20%	80%

B

1 … 30	31 … 100
24€	0€
30%	70%

Question 12

A

1 … 30	31 … 100
24€	0€
30%	70%

B

1 … 15	16 … 30	31 … 100
24€	23.25€	0€
15%	15%	70%

Question 13

A

1 … 15	16 … 30	31 … 100
24€	23.25€	0€
15%	15%	70%

B

1 … 20	21 … 100
42€	0€
20%	80%

Reversed PR choices:

Question 14

A

1 … 10	11…15	16…20	21 … 100
54€	0€	0€	0€
10%	5%	5%	80%

B

1 … 10	11…15	16…20	21 … 100
24€	24€	0€	0€
10%	5%	5%	80%

Question 15

A

1 … 10	11…15	16…20	21 … 100
24€	24€	0€	0€
10%	5%	5%	80%

B

1 … 10	11…15	16…20	21 … 100
12€	12€	12€	0€
10%	5%	5%	80%

Question 16

A

1 … 0	11…15	16…20	21 … 100
12€	12€	12€	0€
10%	5%	5%	80%

B

1 … 10	11…15	16…20	21 … 100
54€	0€	0€	0€
10%	5%	5%	80%

Juxtaposition effect:

Question 17

A

1 … 55	56 … 100
0€	11€
55%	45%

B

1 … 45	46 … 100
0€	7€
45%	55%

Question 18

A

1 … 55	56 … 100
0€	11€
55%	45%

B

1 … 55	56 … 100
7€	0€
55%	45%

Juxtaposition effect – different matrix order:

Question 19

A

1 … 85	86 … 100
0€	11€
85%	15%

B

1 … 80	81 … 100
0€	7€
80%	20%

Question 20

A

1 … 15	16 … 100
11€	0€
15%	85%

B

1 … 80	81 … 100
0€	7€
80%	20%

CRE post-argument:

Question 21

A

1	…	100
	11 €	
	100%	

B

1	…	80	81	…	100
	15€			0€	
	80%			20%	

Question 22

A

1	…	25	26	…	100
	11€			0€	
	25%			75%	

B

1	…	20	21	…	100
	15€			0€	
	20%			80%	

A3.1.5 Test results subsidiary questions

Table 16 (Appendix): Subsidiary results treatment 1

	Choice patterns								
	EUT consistent						*EUT violation*		
Behavioural pattern (Questions)		SS'			*RR'*		*SR'*	*RS'*	*p-value*
CCE – HFI (low) (3,4)		15			53		22	26	*0.6655*
CCE - HFI (high) (5,6)		26			40		18	32	*0.0649**
ESE (7,8)		56			33		22	5	*0.0015****
ESE – diff. order (9,10)		54			20		33	9	*0.0003****
Juxtaposition effect (17,18)		21			71		14	10	*0.5413*
Juxtaposition effect – diff. order (19,20)		35			54		12	15	*0.7011*
	ABA	*ACA*	*BBA*	*BBC*	*ACC*	*BCC*	*ABC*	*BCA*	
Cyclical choices (11,12,13)	36	8	5	18	11	3	24	1	*0.0001****
Reversed PR choices (14,15,16)	48	6	9	11	4	17	20	1	*0.0001****

The choice patterns indicate the frequency of all possible conjunctions of choices over the problem pairs. The *p-value* is calculated for a two-tailed within subject test using the binomial distribution of the hypothesis that violations of EUT consistent with the relevant phenomenon occur with a greater frequency than opposite violations (The null hypothesis is that they occur with equal frequency). *, ** and *** denote rejection of the null hypothesis at the 10%, 5% and 1% level respectively.

Table 17 (Appendix): Subsidiary results treatment 2

	Choice patterns								
	EUT consistent						*EUT violation*		
Behavioural pattern (Questions)		SS'			*RR'*		*SR'*	*RS'*	*p-value*
CCE – HFI (low) (3,4)		21			50		10	34	*0.0004****
CCE - HFI (high) (5,6)		26			36		13	40	*0.0003****
ESE (7,8)		56			32		22	5	*0.0015****
ESE – diff. order (9,10)		65			18		21	2	*0.0001****
Juxtaposition effect (17,18)		18			65		18	14	*0.5966*
Juxtaposition effect – diff. order (19,20)		38			54		12	11	*1.0000*
	ABA	*ACA*	*BBA*	*BBC*	*ACC*	*BCC*	*ABC*	*BCA*	
Cyclical choices (11,12,13)	40	4	10	30	5	8	18	0	*0.0001****
Reversed PR choices (14,15,16)	50	8	9	8	7	15	17	1	*0.0001****

The choice patterns indicate the frequency of all possible conjunctions of choices over the problem pairs. The *p-value* is calculated for a two-tailed within subject test using the binomial distribution of the hypothesis that violations of EUT consistent with the relevant phenomenon occur with a greater frequency than opposite violations (The null hypothesis is that they occur with equal frequency). *, ** and *** denote rejection of the null hypothesis at the 10%, 5% and 1% level respectively.

Table 18 (Appendix) : Subsidiary results treatment 3

	Choice patterns								
	EUT consistent						*EUT violation*		
Behavioural pattern (Questions)		SS'			*RR'*		*SR'*	*RS'*	*p-value*
CCE – HFI (low) (1,2)		14			35		10	15	*0.4244*
CCE - HFI (high) (3,4)		21			24		12	17	*0.4583*
ESE (5,6)		39			16		15	4	*0.0192***
ESE – diff. order (7,8)		38			15		18	3	*0.0015****
Juxtaposition effect (15,16)		12			47		10	5	*0.3018*
Juxtaposition effect – diff. order (17,18)		15			32		13	14	*1.0000*
	ABA	*ACA*	*BBA*	*BBC*	*ACC*	*BCC*	*ABC*	*BCA*	
Cyclical choices (9,10,11)	19	4	7	23	4	2	14	1	*0.0010****
Reversed PR choices (12,13,14)	33	5	6	7	6	10	5	2	*0.4531*

The choice patterns indicate the frequency of all possible conjunctions of choices over the problem pairs. The *p-value* is calculated for a two-tailed within subject test using the binomial distribution of the hypothesis that violations of EUT consistent with the relevant phenomenon occur with a greater frequency than opposite violations (The null hypothesis is that they occur with equal frequency). *, ** and *** denote rejection of the null hypothesis at the 10%, 5% and 1% level respectively.

Table 19 (Appemdix): Subsidiary results treatment 4

	Choice patterns								
	EUT consistent						*EUT violation*		
Behavioural pattern (Questions)		SS’			*RR’*		*SR’*	*RS’*	*p-value*
CCE – HFI (low) (3,4)		17			28		14	14	*1.0000*
CCE - HFI (high) (5,6)		17			23		11	22	*0.0802**
ESE (7,8)		36			13		17	7	*0.0639**
ESE – diff. order (9,10)		40			12		15	6	*0.0784**
Juxtaposition effect (17,18)		18			42		9	*4*	*0.2668*
Juxtaposition effect – diff. order (19,20)		17			37		*12*	*7*	*0.3593*
	ABA	*ACA*	*BBA*	*BBC*	*ACC*	*BCC*	*ABC*	*BCA*	
Cyclical choices (11,12,13)	19	2	3	24	0	6	19	0	*0.0001****
Reversed PR choices (14,15,16)	26	1	10	12	6	12	3	3	*0.3593*

The choice patterns indicate the frequency of all possible conjunctions of choices over the problem pairs. The *p-value* is calculated for a two-tailed within subject test using the binomial distribution of the hypothesis that violations of EUT consistent with the relevant phenomenon occur with a greater frequency than opposite violations (The null hypothesis is that they occur with equal frequency). *, ** and *** denote rejection of the null hypothesis at the 10%, 5% and 1% level respectively.

8 References

Allais, M. (1953a). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica, 21*, 503–46.

Allais, M. (1953b). Fondements d'une théorie positive des choix comportant un risqué et critique des postulats et axioms de l'ecole Americaine. *Economtétrie, Colloques Internationaux de CNRS,* XL, 257-232.

Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., Rabin, M. (2003). Regulation for conservatives: behavioral economics and the case for 'asymmetric paternalism'. *University of Pennsylvania Law Review, 151(3)*, 1211–1254.

Camerer, C.F. (1995). Individual Decision Making in *Handbook of Experimental Economics*. J. Kagel and A. E. Roth, eds. Princeton: Princeton University Press.

Camerer, C.F., Hogarth, R.M. (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty, 19*, 7-42.

Chandler, J. (2017). Descriptive decision theory. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosoph*. Metaphysics Research Lab, Stanford University.

Conlisk, J. (1989). Three Variants on the Allais Example. *American Economic Review, 79*, 392–407.

Dietrich, F., Staras, A., Sugden, R. (2021). Savage's response to Allais as Broomean reasoning. *Journal of Economic Methodology, 28(2)*, 143-164.

Dold, M.F., Schubert, C. (2018). Toward a behavioural foundation of normative economics. *Review of Behavioural Economics, 5*, 221-241.

Friedman, M., Savage, L. (1952). The expected-utility hypothesis and the measurability of utility. *Journal of Political Economy, 60,* 463–74.

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

Holt, C.A. (1986). Preference Reversals and the Independence Axiom. *American Economic Review, 76,* 508–15.

Kahneman, D., Knetsch, J.L., Thaler, R.H. (1991). The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives, 5(1),* 193-206.

Kahneman, D., Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39(4)*, 341-350.

MacCrimmon, K.R. (1968). Descriptive and Normative Implications of the Decision-Theory Postulates. in K. Borch and J. Mossin (ed.), *Risk and Uncertainty*. New York: St. Martins Press, 3–32.

MacCrimmon, K.R., Larsson, S. (1979). Utility Theory: Axioms versus 'Paradoxes'. in Allais, M., Hagen, O. (eds.), *Expected Utility Hypotheses and the Allais Paradox*. Theory and Decision Library, vol. 21, 333–409.

Machina, M. (1982). 'Expected utility' theory without the independence axiom. *Econometrica, 50*, 277-323.

Moskowitz, H. (1974). Effects of Problem Representation and Feedback on Rational Behavior in Allais and Morlat-type Problems. *Decision Sciences, 5,* 225-241.

Samuelson, W., Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1,* 7-59.

Savage, L. (1954). *The Foundations of Statistics*, New York: John Wiley&Sons.

Slovic, P., Tversky, A. (1974). Who Accepts Savage's Axiom? *Behavioral Science, 19,* 368-373.

Starmer, C. (1992). Testing New Theories of Choice under Uncertainty Using the Common Consequence Effect. *Review of Economic Studies, 59,* 813-830.

Starmer, C. (2000). Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk. *Journal of Economic Literature, 38,* 332-382.

Starmer, C. (2005). Normative notions in descriptive dialogues. *Journal of Economic Methodology, 12(2),* 277-289.

Starmer, C., Sugden, R. (1989). Probability and Juxtaposition Effects: An Experimental Investigation of the Common Ratio Effect. *Journal of Risk and Uncertainty, 2,* 159-178.

Starmer, C., Sugden, R. (1991). Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation. *American Economic Review, 81,* 971-978.

Sugden, R. (2018). 'Better off, as judged by themselves': a reply to Cass Sunstein. *International Review of Economics, 65,* 9–13.

Sugden, R. (2020). Normative economics without preferences. *International Review of Economics, 68,* 5-19.

Sunstein, C., Thaler, R. (2003). Libertarian paternalism is not an oxymoron. *University of Chicago Law Review, 70,* 1159–1202

Thaler, R., Sunstein, C. (2008). *Nudge: improving decisions about health, wealth, and happiness.* Yale University Press, New Haven.

van de Kuilen, G., Wakker, P. (2006). Learning in the Allais Paradox. *Journal of Risk and Uncertinty, 33,* 155-164.

Zizzo, D.J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13,* 75-98.