Quntitative Analysis of Risky Decision Making in Economic Environments

Inauguraldissertation
zur
Erlangung der Würde
einer Doktorin der Philosophie
vorgelegt der
Fakultät für Psychologie
der Universität Basel

von

Sandra Andraszewicz

aus Gdynia, Polen

Basel, 2014

UNI
BASEL

Genehmigt von der Fakultät für Psychologie

auf Antrag von

Prof. Dr. Jörg Rieskamp
Prof. Dr. Eric-Jan Wagenmakers

Basel, den ________________________

____
Prof. Dr. Rosenlind Lieb

UNI
BASEL

Declaration

I, Sandra Andraszewicz, born on 7th July, 1987 in Gdynia, Poland, hereby declare the following:

(i) My cumulative dissertation is based on five manuscripts, of which four are submitted (Andraszewicz, Rieskamp, & Scheibehenne, 2013; Andraszewicz, & Rieskamp, 2013; Andraszewicz, Scheibehenne, Rieskamp, Grasman, Verhagen, & Wagenmakers, 2013; Klucharev, Andraszewicz, & Rieskamp, 2013) and one will be submitted soon (Andraszewicz, von Helversen, & Rieskamp, 2013). I have been responsible for the ideas, data collection and writing of the manuscripts. I contributed to these manuscripts in the following way.

1. Andraszewicz, & Rieskamp (2013): Primarily responsible for the development of the idea and proposed measure, simulation, data collection and writing of the paper.
2. Andraszewicz, Rieskamp & Scheibehenne (2013): Jointly responsible for the idea. Primarily responsible for the development of the experimental paradigm and the formulation of the hypothesis, data collection, analysis and writing of the paper.
3. Andraszewicz, von Helversen, & Rieskamp (2013): Jointly responsible for the idea. Primarily responsible for the data collection, analysis and writing of the paper.
4. Klucharev, Andraszewicz, & Rieskamp (2013): Primarily responsible for developing models proposed in the paper and the data analysis related to the modeling. Partially responsible for writing the paper.
5. Andraszewicz, Scheibehenne, Rieskamp, Grasman, Verhagen, & Wagenmakers (2013): Primarily responsible for data analysis and writing the paper.

(ii) I only used the resources indicated.

(iii) I marked all the citations.

Basel, 6th June 2014

Sandra Andraszewicz

Acknowledgments

Doing a PhD is an unforgettable experience. I already knew during my second year at university that I would like to go through this experience. Apart from acquiring useful skills and academic knowledge, I have learned to be persistent. This is because the PhD time is a roller coaster with many ups and downs, and crazy unexpected turns.

I would like to thank several people who have traveled with me on this roller coaster. First of all, I would like to thank Jörg Rieskamp for his supervision, criticism and time. Secondly, I would like to thank my co-authors, Benjamin Schiebehenne, Vasily Klucharev, and Bettina von Helversen. In particular, I would like to thank E-J Wagenmakers - my co-author and the second reviewer of my dissertation thanks to whom I learned a lot about Bayesian statistics. I would like to thank Reiner Greifenender for being the chairperson during my defense.

Many thanks to all the research assistants, in particular to Regina Weilbächers who worked hard to help me collect the final data for this dissertation, and Kirsten Hard and Stefan Thommen who assisted with my experiments.

I would like to thank Laura Wiles for correcting my manuscripts, and Mirjam Jenny, Nicolas Berkowitsch and Janina Hoffmann for giving me useful feedback on my dissertation framework. I give special thanks to Janina Hoffmann for being a wonderful office mate. Further, I would like to give special thanks to the Economic Psychology and Cognitive and Decision Sciences research teams for great times at the institute and outside of it.

Finally, I would like to thank my friends for distracting me from time to time from my PhD work and I would especially like to thank my parents, who gave me a lot of support during the roller coaster ride. At times, it would have been hard to survive the ups and downs without them.

Abstract

Risky economic decisions play an important role in everyone's life. This dissertation presents mathematical approaches to the analysis of these decisions. It discusses how statistical measures can describe properties of choice options, and how these properties can be used to describe the decision context. Also, this dissertation includes a practical tutorial on a Bayesian approach to the hierarchical regression analysis in management science. Therefore, the combined dissertation presents mathematical and statistical tools in, and for better research of, decision making under risk.

The first manuscript proposes *standardized covariance*, a measure that can quantitatively describe the strength of the association and similarity between choice options' outcomes. The standardized covariance can also describe how risky one option is with respect to another. It can influence predictions of choice models. The second manuscript shows experimentally how association measured with the standardized covariance can influence people's choices. The third manuscript proposes applying the expected shortfall of an option's outcomes as a measure of risk in the standard risk-value models. In an experiment, the *risk-value shortfall* model successfully predicted people's preference for options with higher expected value, lower variance and more positively skewed distributions of outcomes, and outperformed competing models.

The fourth manuscript proposes a new version of a reinforcement learning model, which can be applied in a social context. The proposed model can account for the behavior of other people competing for a common pool resource. As experimentally tested, the model could successfully predict human behavior and correlated with the brain activity measured with an fMRI method.

The last manuscript outlines advantages of using Bayes factors instead of p-values for interpretation of results from hierarchical regression analysis. As the results in the manuscript show, the Bayesian approach and the standard null-hypothesis statistical testing can lead to different conclusions.

General Introduction

"Mathematics is the queen of all sciences."

Carl Friedrich Gauss

In this brief framework, I present a summary and discussion of my dissertation titled *Quantitative analysis of risky decision making in economic environments.* This dissertation applies mathematical tools and methods to the field of decision making under risk. It shows how mathematical and statistical measures can be used to better describe decision environments and how mathematical models can describe human decision-making behavior and its underlying cognitive processes. Also, it describes how statistical properties of numerical choice options, such as covariance and skewness, influence people's decisions. Additionally, it presents theoretical and practical aspects (i.e., a tutorial) of applying Bayesian statistics in management science, to conduct more accurate data analyses. Thus, this dissertation combines statistical, mathematical and computational techniques used in, and for investigating, decision making under risk.

Models of Economic Decision Making under Risk

People face risky decisions involving monetary consequences on an everyday basis. For example, should one choose a well-paid job far from home, or a badly-paid job in one's home town? Which insurance provider should one choose to insure one's family? Should one buy real estate now, or should one wait to see if the price drops? Should one invest in risky stocks, or rather choose safer, but less profitable bonds? Every person faces similar questions during their lifetime. Therefore, understanding how people make decisions is a valid and important research topic. Various models of decision making can help us better understand the invisible cognitive decision process, which cannot be measured otherwise.

Standard economic theory of decision making (von Neumann & Morgenstern, 1944) assumes that people choose the option with the highest expected utility. However, many studies have shown that people do not always follow this rule (e.g. Gigerenzer & Goldstein, 1996; Kahneman & Tversky, 1979; Rieskamp, Busemeyer, & Mellers, 2006). Various theories of decision making have been proposed to explain how people make decisions and what the cognitive processes underlying these decisions are. In this section, I provide a brief overview of selected groups of models used in decision making research, to which I will later refer in the article summaries. The choice of models was motivated by the fact that different groups of models are sensitive to different characteristics of the choice environment. Therefore, different models focus on different properties of choice options and depending on the property of interest, the models propose different process underlying the decision.

The group of *fixed utility* models consists of theories that assume that people under- or overweight monetary values of their choices. Therefore, their expected gain is expressed as an expected utility, rather than the expected value (von Neumann & Morgenstern, 1944). This utility can be defined as a utility function $u(.)$, where a convex utility function for gains and a concave function for losses implies that a person is risk averse, underweighting gains and overweighting losses. Opposite shapes imply risk-loving decision makers. Further extensions of the *expected utility* theory (von Neumann & Morgenstern, 1944), such as *rank-dependent utility models* (Green & Jullien, 1989; Luce, 1990), assume that people order options' outcomes according to the probability of their occurence. The most prominent theory from this group, *prospect theory* (Kahneman & Tversky, 1979), proposed that apart from subjectively evaluating the options' outcomes, people overweight low probabilities of the occurrence of an outcome and underweight high probabilities. This can be mathematically described with a probability weighting function. This group of models can explain how people respond to risk and provide explanation of how they value monetary outcomes. However, it does not assume a context-dependent evaluation of choice options.

Context-dependent models, such as *regret theory* (Loomes & Sugden, 1982), the *proportional difference model* (González-Vallejo, 2002) and *decision field theory* (Busemeyer & Townsend, 1993) assume that people compare options' outcomes with each other. Regret theory (Loomes & Sugden, 1982) proposes that the subjective utility of option A is measured as the perceived regret with respect to the forgone option B. During a decision process, people would aim to reduce their regret, by comparing options attribute-wise and estimating the expected loss for each outcome of option A as compared to option B. The proportional difference model and decision field theory assume that a decision process is stochastic. The proportional difference model further assumes that the difference between respective outcomes of options A and B is relative to the maximum outcome of the two. Decision field theory is a sequential sampling model, according to which a decision maker repeatedly samples information about options A and B, stochastically switching attention from one option to another. By this means, evidence in favor of one option or another is accumulated over time. A decision is made either when the accumulated information exceeds a predefined threshold, or when the decision time is over. In the second case, the option for which more evidence is accumulated would be chosen.

The first manuscript shows how predictions of context-dependent models are influenced by the strength of the association and similarity between two choice options. Fixed-utility models cannot account for that. In the second manuscript, two models from the group of fixed-utility models were be tested against two context-dependent models. The results provide a clear evidence that the context of how options are presented influences people's decisions.

Risk-value models (see Sarin & Weber, 1993) take a different approach on evaluating choice options. Namely, this group of models proposes that a decision process is based on a trade-off between the expected gain of an option and the risk it carries. Various versions of these models provide different interpretations of the risk component. In the third manuscript, we propose *expected shortfall* as a measure of risk and rigorously test the model against other risk-value models and the expected utility model.

The fourth manuscript proposes two new versions of a reinforcement learning model. *Reinforcement learning models* (Sutton & Barto, 1998), assume that people learn to make decisions over time, based on received feedback. Reinforcement learning models are based on Markov decision process, meaning that a person's choice at time t depends on their choice at time $t-1$, but not at times $t-n$, $n \geq 2$. At each time t, a person's choice i from a set of choices I is assigned a subjective value $Q_{i,t}$. This value is expressed as a subjective value of the same option in the previous trial updated with the prediction error:

$$Q_{i,t} = Q_{i,t-1} + \alpha(R_{i,t} - Q_{i,t-1}), \tag{1}$$

where the prediction error is the difference between the reinforcement and the actual choice in the previous trial ($R_{i,t} - Q_{i,t-1}$). α is a learning rate parameter, where the higher values of *alpha* indicate quicker learning. Different versions of reinforcement-learning models provide different interpretations of the reinforcement component. The fourth manuscript proposes a new version of the reinforcement learning model that is sensitive to the depletion of common resources in social and private situations. Importantly, the prediction of this model correlate with the activity of the ventral striatum, which is responsible for the monitoring of the resource depletion.

Characteristics of Choice Options

As mentioned in the previous section, various models of decision making focus on different aspects of choice options. For example, context-dependent models emphasize the association between two choice options, whereas risk-value models focus on the expected gain and riskiness of options. These aspects have specific mathematical representations. Table 1 outlines a systematic classification of the statistical properties of choice options and their interpretations in the light of decision making under risk.

Many models consider expected value as an interpretation of an expected gain from the option (e.g. Busemeyer & Townsend, 1993; Sarin & Weber, 1993). The expected value refers to the statistical mean of an option's possible outcomes. Variance has been the most popular measurement of risk, not only in decision making research (e.g. Sarin & Weber, 1993), but also in the area of finance (see Fishburn, 1977; Markowitz, 1959; Weber, Shafir, & Blais, 2004). The higher the variance, the greater the range of possible outcomes implying that both very high and very low outcomes can occur. When variance is low, the option's possible outcomes fall within a narrow range, which means a safer option.

As listed in Table 1, there are more statistical properties that can describe riskiness of a choice option. Previous literature points out skewness (see Burke & Tobler, 2011; Samuelson, 1970; Symmonds, Wright, Bach, & Dolan, 2011), which defines the distribution of an option's outcomes. A more positively skewed distribution indicates that an option has more moderate-valued outcomes occurring with a relatively high probability and a few very high outcomes occurring with a very low probability. A non-skewed distribution means outcomes are evenly distributed. A negatively skewed distribution is characterized by moderate outcomes occurring with high probabilities and some very low outcomes occurring with low probabilities. Another measure of risk is expected shortfall (Acerbi & Tasche, 2002), which can account for both the spread of outcomes and their distribution. The third manuscript makes a novel contribution of applying this measure in risk-value models.

Finally, covariance between the outcomes of choice options reflects the relationship between the options. This characteristic of choice options is investigated in manuscripts one and two. The first manuscript proposes the *standardized covariance*, a measure of association, similarity, and relationship between risk of one option with respect to the other. Therefore, the second manuscript elaborates on expected value, variance and covariance, and points out the relationships between these three measures.

Table 1

Relationhip between mathematical properties, characteristics of risky choice options and their cognitive interpretation.

Mathematical Representation	Notation	Characteristic	Interpretation
Expected value	$E[A]$	Gain/Loss	Mean outcome
Variance	σ_A^2	Risk	Spread of outcomes
Covariance	σ_{AB}	Association	Relationship between outcomes
Skewness	γ_A	Risk	Distribution of outcomes
Expected shortfall	$ES_{c,A}$	Risk	Outcomes below expected level

Statistical Analysis

The last manuscript of this dissertation is a practical tutorial on conducting hierarchical Bayesian regression. Hierarchical regression is one of the most popular statistical methods in the field of management science, to which the tutorial is addressed. The name "hierarchical" is based on the fact that a researcher decides about the hierarchy of the independent variables and adds them the analysis according to this hierarchy in a stepwise fashion. In the standard null-hypothesis significance testing (NHST), the statistical evidence is measured with the significance of the change in the explained variance (Cohen, Cohen, West, & Aiken, 2003), indicated by a p-value.

However, the NHST regression analysis can lead to substantially different conclusions than the analysis conducted using the Bayesian approach. Recent literature points out flaws relying on p-values (c.f. Wagenmakers, 2007). Bayes factors (BFs, Jeffreys, 1961; Kass & Raftery, 1995), provide an alternative way of testing hypotheses in an objective way. Thus, the article outlines advantages of using Bayes factors for hypothesis testing in the regression analysis, and points out the differences between the Bayesian and NHST approaches. It explains on a conceptual and practical level, how to conduct a hierarchical regression analysis, with the use of BFs.

Standardized Covariance

Andraszewicz, S. & Rieskamp, J. (2013). *Standardized Covariance - A Measure of Association, Similarity and Co-Riskiness between Choice Options.* Manuscript submitted for publication.

This work proposes standardized covariance between two choice options A and B, denoted σ^*_{AB}, as an easy-to-interpret measure that can reflect the association and similarity between two risky choice options, and can reflect a ratio of how risky one option is with respect to the other. We define the ratio of how risky one option is relative to the other as the *co-riskiness* between two options. *Association* is a relationship between two options and it implies statistical dependence between them. *Similarity* defines how the features of one object are related to the features of another object (Tversky, 2004).

Many prominent theories of decision making, such as *priority heuristic* (Brandstätter, Gigerenzer, & Hertwig, 2006), *the proportional difference model* (González-Vallejo, 2002), *regret theory* (Loomes & Sugden, 1982) and *decision field theory* (Busemeyer & Townsend, 1993), assume that people compare options in an attribute-wise fashion, indicating that the association between the options' outcomes should influence people's choices. In the second manuscript, we further experimentally showed that the association systematically influences people's preferences, even when controlling for the difference between expected values.

There have been multiple attempts to measure similarity between choice options (c.f. Tversky, 2004). For example, the *contrast model* (Tversky, 1977) measures the amount of features that are the same for two objects (rather than risky choice options) with respect to the number of features that are different. Along similar lines, the *similarity model* (Leland, 1994, 1998; Rubinstein, 1988) assumes that when options are similar on one dimension (attribute) but different on another, a decision maker should choose the option that is better in the dissimilar attribute. Busemeyer and Townsend (1993) proposed covariance as a measure of similarity. However, covariance depends on the range of possible outcomes which makes it hard to interpret.

Alternatively, one could use a correlation measure. Unfortunately, a large part of the research on decision making is conducted using two-outcome gambles (e.g. Birnbaum, 2008; González-Vallejo, 2002), for which the correlation is either 1 or -1 (see Rodgers & Nicewander, 1988). Therefore, we proposed *standardized covariance*, which equals twice the non-standardized covariance between options A and B, divided by the sum of variances of each option:

$$\sigma^*_{AB} = \frac{2\sigma_{AB}}{\sigma^2_A + \sigma^2_B}. \tag{2}$$

The standardized covariance varies between -1 and 1, where $\sigma^*_{AB} = 1$ means strong positive association, high similarity and high co-riskiness. In contrast, $\sigma^*_{AB} = -1$ implies strong negative association, high dissimilarity and high co-riskiness. When $\sigma^*_{AB} = 0$, there is no association and similarity between choice options and the co-riskiness between them is minimal. When σ^*_{AB} is positive, but close to 0, there is a weak positive association, low similarity and low co-riskiness. Analogously, for negative σ^*_{AB}, values close to 0 imply weak negative association, low dissimilarity and low co-riskiness. The association of statistically independent options (i.e. options depending on different external events) is 0. However, their outcomes can be quite similar to each other. Therefore, for statistically independent options, we proposed a similarity measure S_{AB}.

Covariance between two choice options depends on both variances in a non-linear fashion. Therefore, these two measures cannot be disentangled from each other. When the standardized covariance of two stochastically non-dominant options is high and positive (i.e. $\sigma^*_{AB} \to 1$), variances of both options are similarly large (i.e. $\sigma^2_A \approx \sigma^2_B$)

and the covariance between them is high because the outcomes a_i and b_i of options A and B corresponding to the same probabilities p_i are similar to each other (i.e. $(a_i - b_i) \to 0$). Therefore, the options that carry a similar level of risk (i.e. high co-riskiness), measured with variance, are strongly associated and similar to each other. Swapping the outcomes of one option (i.e. $a_1^* = a_2 \wedge a_2^* = a_1$, where $I \in \{1, 2\}$) results in the same co-riskiness level, but in dissimilarity of the same strength and the opposite association.

In contrast, when $\sigma_{AB}^* \to 0$ the variance of one option is substantially larger than the variance of the other option (i.e. $\sigma_A^2 > \sigma_B^2$) and one option is almost a sure option. This implies that the outcomes corresponding to the same probabilities are not similar to each other and the covariance between the options' outcomes is low. Therefore, the co-riskiness between the options is low (i.e. one option is significantly more risky than the other) and association and similarity between the options are low. When variances of both options are equal, $\sigma_{AB}^* = 1$, which implies that the options are identical and $2\sigma_{AB} = \sigma_A^2 + \sigma_B^2$. Therefore, the standardized covariance is related to the difference between expected values (ΔEV). However, for every ΔEV there is a wide range of options with various σ_{AB}^*.

Further, in a simulation, we showed that the standardized covariance and the correlation measure are very strongly correlated with each other, $r = .98$, $p < .001$. However, in some cases, the two measures would indicate different strengths of association and similarity between the options (i.e. $r = .95$ and $\sigma_{AB}^* = .32$). This is because correlation is a special case of association that measures the strength of linear relationship between choice options. In contrast, the standardized covariance reflects the distance of outcomes corresponding to the same probabilities.

Therefore, we conducted an experiment in which we asked 20 participants to choose between 120 pairs of four-outcome options for which the difference between their expected values was constant (i.e. $\Delta EV = 15$), presented in a random order. In half of the trials, the standardized covariance between the options was low, $\sigma^*_{AB} \leq .2$, whereas the correlation was high, $r \geq .8$ ("low covariance" condition). In the second half of the trials, the standardized covariance was similarly high as the correlation measure, $\sigma^*_{AB} \geq .8$ and $r \geq .8$ ("high covariance" condition). Based on a simulation of model predictions, we hypothesized that if people's choices are influenced only by the statistical correlation between the choice outcomes, then we should observe similar choices in both conditions. However, if people's choices are influenced by the standardized covariance between options, then people should choose the options with the larger expected value more frequently in the condition with the high standardized covariance. The participants chose the option with the larger expected value more frequently in the high covariance condition ($Me = 90\%$, $SE = 9\%$) than in the low covariance condition ($Me = 80\%$, $SE = 18\%$). We fitted three models assuming interdependent evaluation of choice options: *regret theory*, *decision field theory* and the *proportional difference model* to the behavioral data using a maximum-likelihood approach. All models predicted higher choice probability in the high covariance condition than in the low covariance condition and decision field theory fitted the data best. Additionally, we asked the participants to indicate whether options within a pair from the "low covariance" or "high covariance" condition are more similar to each other. All participants pointed to the options from the high covariance condition. This implies that the participants perceived the association and similarity between choice options above the linear correlation and the standardized covariance could successfully describe this association and similarity.

In sum, the standardized covariance is an easy-to-interpret measure that can quantify the choice environment. This especially useful in testing models of decision making because the standardized covariance influences models' predictions, despite constant difference between the expected values. Therefore, controlling for σ^*_{AB} apart from the ΔEV can result in a more accurate experimental setting.

Influence of Covariance on Choice Preference

Andraszewicz, S., Rieskamp, J., & Scheibehenne, B. (2013). *How Associations between Consequences of Choice Options Affect Decisions Under Risk*. Manuscript re-submitted for publication.

In the first manuscript, we proposed a measure of association between choice options. Here, we investigated how the association between the consequences of two risky choice options influences people's choices. We tested various models of decision making against each other to see, which model can predict how people's choices are influenced by this association best.

The strength of the association can be illustrated with a choice between two monetary gambles whose outcomes depend on the throw of a die, as shown in Figure 1 in the manuscript. In case 1, gamble A leads to substantially higher payoffs than gamble B if the die lands on 1, 2, or 3, whereas for the numbers 4 to 6 gamble B has a small advantage over A. Presumably, most people would prefer gamble A over B due to the large advantage of A for numbers 1-3 and disregard the small disadvantage for numbers 4-6. In contrast, in case 2, gamble A is worse than B* when throwing 1, 2, or 3, whereas it is better when throwing 4, 5 or 6, making the choice much more complicated. However, gamble B and B* are identical as they both result in the same outcomes with identical probabilities.

Standard economic theory (von Neumann & Morgenstern, 1944), as well as more cognitively inspired *fixed-utility theories*, such as rank-dependent utility (see e.g. Green & Jullien, 1989; Luce, 1990) and their most prominent version, *cumulative prospect theory* (CPT, Kahneman & Tversky, 1979), assume that people evaluate options independently of each other. However, previous literature has shown that people compare options' outcomes with each other (c.f. Rieskamp et al., 2006).

Context-dependent decision theories, such as the *lexicographic semiorder heuristic*

(Tversky, 1969), *priority heuristic* (Brandstätter et al., 2006), *elimination-by-aspects* (González-Vallejo, 2002), *regret theory* (RT, Loomes & Sugden, 1982) or *decision field theory* (DFT, Busemeyer & Townsend, 1993), assume interdependent evaluations of choice options. Previous research has indicated that the context in which choice options are presented influences people's choices. For example, Mellers and Cooke (1994) and Mellers, Schwartz, and Cooke (1998) discussed the influence of other options, whereas Train (2009) proposed extending the *probit model* by incorporating possible relationships between attributes of choice options. However, to our knowledge, little work has been done to quantitatively define the influence of context on decision making with respect to monetary gambles.

Therefore, the aim of this article was to quantify the choice context using the standardized covariance between the options' outcomes (σ^*_{AB}, Andraszewicz & Rieskamp, under review). In two behavioral experiments ($N = 39$ and 24 consecutively), we examined how different levels of association influence people's choices and predictions of four models of decision making: *expected utility* (EU, von Neumann & Morgenstern, 1944), CPT, RT and DFT. At the same time, we kept the difference between expected values constant, such that $\Delta EV = 15$. Given simulation results, we hypothesized that the higher the standardized covariance between two choice options, the higher the preference for the option with the larger expected value.

In both experiments, people repeatedly chose between 180 pairs of two stochastically non-dominant, statistically dependent gambles whose outcomes varied between -100 and 100 points, presented in a random order. In the first experiment, the gambles had two outcomes with their corresponding probabilities of either .4, .5 or .6, whereas in the second experiment, they had four outcomes, with their corresponding probabilities of either .1, .2, .3, or .4. Half of the gambles had only negative outcomes and the other half, only positive outcomes. We classified the gambles into three equal groups, according to the standardized covariance between the gambles:

1) small ($\sigma^*_{AB} \leq .1$), 2) medium ($.4 \leq \sigma^*_{AB} \leq .5$) and 3) large ($.8 \leq \sigma^*_{AB} \leq .95$).

As expected, in both experiments, participants chose more frequently the gamble with the larger expected value when the standardized covariance was higher (Friedman's test, experiment 1: $p < .001$, $\chi^2(2) = 31.32$, $CI = 95\%$, experiment 2: $p < .001$, $\chi^2(2) = 15.48$, $CI = 95\%$). Also, the variance of these choices systematically decreased for larger σ^*_{AB}. We fitted the four models to the data, using a maximum likelihood approach. DFT and RT successfully predicted the effect observed in the data, while EU and CPT could not. According to the Bayesian information criterion, DFT had the best fit out of all models. In both experiments we observed the same effect, which shows that the association between the choice options influences people's preferences independently of the amount of outcomes.

Our results indicate that the association between choice options measured with covariance systematically influences people's choices, even when the difference between the expected values of the options is kept constant. The choices between the options that are highly associated with each other are presumably easier because when the association is high, some outcomes of each option, corresponding to the same probabilities become very similar to each other, whereas for the remaining outcomes the advantage of one option over another can be easily indicated. In contrast, when the association is low, the advantage of one option over another is hard to classify. We show that the standardized covariance of the options' outcomes provides a useful description of the choice context as it defines the difficulty of the choice situation.

Expected Shortfall and Skewness of Choice Options

Andraszewicz, S., von Helversen, B., & Rieskamp, J. (2013). *Expected Shortfall as a Measure of Risk in Risk-Value Models.* Working paper.

In this article, we propose *expected shortfall* as a measure of risk in the application of risk-value models. Risk-value models (Sarin & Weber, 1993) assume that a subjective value of option A is expressed as a trade-off between the expected gain of the option and the risk carried by this option, such that

$$SV(A) = E[A] - \beta R(A), \tag{3}$$

where β is a free parameter measuring weight which a person assigns to the risk.

Variance has been a commonly used measure of risk (e.g. Fishburn, 1977; Markowitz, 1959; Tobler, O'Doherty, Dolan, & Schultz, 2007; Weber et al., 2004). However, a substantial amount of research indicates that the risk of a choice option is defined not only by the range of the option's outcomes (i.e. variance), but also by the outcomes' distribution (i.e. Chiu, 2005; Li, Qin, & Kar, 2010; Samuelson, 1970). The distribution of an option's outcomes is described by their *skewness*.

Previous research has shown that people are sensitive to skewness of choice options' outcomes and that they prefer options that are more positively skewed (Burke & Tobler, 2011; Chunhachinda, Dandapani, Hamid, & Prakash, 1997; Symmonds et al., 2011). However, the standard two-element risk-value model (Sarin & Weber, 1993) assumes that $R(A) = \sigma_A^2$. As a consequence, some research included skewness (γ_A) as another component of risk, resulting in a three-element risk-value model (i.e. Post, van Villet, & H., 2006; Symmonds et al., 2011). Therefore, the subjective value of option A could be expressed as $SV(A) = E[A] - \beta_1\sigma_A^2 + \beta_2\gamma_A$. In this form, $SV(A)$ is convex for gains and concave for losses.

Another prominent measure of risk, used in financial Value-at-Risk model (VaR Acerbi & Tasche, 2002), is the *expected shortfall*. Despite its popularity in the finance area, to our knowledge, expected shortfall has not been applied as a psychological measure of risk. It measures how much decision makers fall short in their expectations about the possible gain from option A with I outcomes a_i and their corresponding probabilities p_i:

$$R(A) = E_{c,A} = \sum_{i=1}^{I} p_i[max(c - a_i, 0)]. \tag{4}$$

Parameter c is a person's individual threshold below which outcomes are undesired. The expected shortfall can be large, either because the outcomes deviate substantially from the threshold c and/or because these outcomes occur with a high probability. The expected shortfall has a psychological interpretation, because decision makers in general want to avoid outcomes that fall short of their expectation or "aspiration levels" (Lopes & Oden, 1999), in particular when the deviations are large. This measure of risk can account for both the range of the options' outcomes (variance) and their distribution (skewness). We call the model incorporating this form of risk the *risk-value shortfall* model.

We experimentally tested the proposed risk-value shortfall model by asking 24 participants to repeatedly choose between two risky options. The experiment consisted of two blocks: 1) three-outcome options, and 2) five-outcome options, each containing 100 randomly presented pairs of options. The 100 pairs were divided into 5 conditions: 1) low vs. high mean, 2) low vs. high variance, 3) positively skewed vs. negatively skewed, 4) positively skewed vs. not-skewed, and 5) negatively skewed vs. not-skewed. In the last three conditions the expected values and variances of both options in each pair were the same and the options differed only with respect to their skewness levels. Based on previous findings (Burke & Tobler, 2011) and simulations, we hypothesized that participants would prefer options with higher expected value, lower variance and more positively skewed outcomes.

We found positive evidence supporting our hypothesis. This effect was consistent, independent of the amount of options' outcomes, which provides a solid evidence for the positive skewness preference. Further, we fitted four models: *mean-variance* (Sarin & Weber, 1993), *mean-variance-skewness* (Symmonds et al., 2011), *expected utility theory* (von Neumann & Morgenstern, 1944) and the *risk-value shortfall* model. We included the expected utility model because depending on the value of its α-parameter, the shape of the utility function could account for the expected value, variance and skewness (see Kroll, Levy, & Markowitz, 1984). Therefore, we considered it as a competitive model to the risk-value shortfall model. According to the Bayesian information criterion (BIC), for 96% of the participants, all models fitted the data better than the naïve baseline model assuming random choice, with the probability of .5. According to the median and total BIC, the risk-value shortfall model fitted the data best. Model comparison for individual participants, based on the Bayes factor, showed strong and very strong evidence supporting the risk-value shortfall model and against the expected utility model.

In sum, this study makes a novel contribution by proposing expected shortfall as a measure of risk on a cognitive dimension and integrating it in a standard risk-value model. It provides empirical evidence for the positive skewness preference. The proposed model was rigorously tested against existing competitive models and fitted the data best.

The Neuroscience of the Tragady of the Commons

Klucharev, V., Andraszewicz, S., & Rieskamp, J. (2013). *Neural Underpinnings of Exploitation of Common Goods.* Manuscript re-submitted for publication.

The 21st century's big concern is the preservation of natural resources. *Common-pool resources* (CPRs) are goods that can be accessed by every individual in society (Ostrom, 1990). Economic theory predicts overexploitation of these resources by self-interested people. The "tragedy of the commons" (Hardin, 1968) is a social dilemma in which individuals acting independently and rationally, will ultimately deplete a shared limited resource, even if it is in their long-term interest to preserve the resource. One explanation for a tendency to overharvest CPRs refers to people's social preference for equity and reciprocal cooperation (Falk & Fischbacher, 2006; Fehr & Schmidt, 1999): If others are cooperative, then people act cooperatively, but if others free ride, people retaliate.

In this article, we hypothesized that the brain's dopaminergic system monitors not only a person's own reward in the CPR games but also the behavior of others who are using the same CPRs. In contrast, we hypothesized that when dealing with private resources, the system monitors long-term sustainable use of the resource, which results in resource preservation behavior. Here, we proposed a reinforcement learning model, which predicts people's behavior, as outlined in the hypotheses; its parameters correlate with the neural activity of the brain structure that monitors the use of the resource.

We conducted an fMRI experiment with 50 participants. In the scanner, the first half of the participants played a game dealing with CPRs ("social" condition), while the other half with the private resources ("private" condition). In 16 games, each containing a maximum of 8 trials, the participants had to fish from a lake, where they could fish either 1, 2 or 3 fish (net size). Initially, the lake contained 18 fish and the proliferation rate was 1.5. Each game ends prematurely if the resource becomes repleted. In the social condition, participants were told that they play against two other players (pre-recorded, competitive players), whereas in the private condition, they were told that the resource depletes due to the outflow of the fish to another lake.

On average, participants depleted the resource significantly faster in the social condition ($M_{Ntrials} = 7.0$) than in the private condition ($M_{Ntrials} = 6.3$), $t(1, 46) = 4.89$, $p < .001$. The participants showed different harvesting styles in the two conditions such that the smallest net size was used less frequently in the social than in the private condition. The fMRI analysis revealed that depletion of resources resulted in a stronger deactivation of the ventral striatum (part of the dopaminergic system) in the social than in the private condition. The deactivation of the ventral striatum was positively correlated with the resource preservation in the private condition and negatively correlated in the social condition. This indicates that the ventral striatum monitors reward (resource) differently in the private as compared to the social condition. To provide a computational explanation of these neurophysiological data, we proposed a cognitive model whose parameters we used as predictors the parametric analysis.

We created two versions of a reinforcement learning model (Sutton & Barto, 1998): *social* and *non-social*. The *social model* assumes that the reinforcement of a decision was defined as a weighted sum of the participant's private reward and the reward resulting from the social comparison with the competitors:

$$R = \beta Payoff + (1 - \beta) Comparison, \tag{5}$$

where β is the free weighting parameter of the model and $Comparison$ is defined as the difference between the players' choices and the average choices of the competitors:

$$Comparison = Payoff - \langle Others \rangle. \tag{6}$$

According to this model, participants are punished when they harvest less than the competitors implying self-interested behavior.

The *non-social model* assumes that reinforcement is defined as a weighted sum of the participant's private reward and a sustainability component:

$$R = \beta Payoff + (1 - \beta) Sustain, \tag{7}$$

where the *Sustain* component is defined as minus the absolute value of the difference between the sustainable number of fish that can leave the lake (i.e. $Sustainability = 6$) to keep the resource level constant over all trials (Note: $(18 - 6) \cdot 1.5 = 18$):

$$Sustain = -|Sustainability - Payoff - FishOutflow|. \tag{8}$$

Sustainability is a fixed task-dependent parameter. This model simultaneously punishes people for overexploiting the resources and not taking enough resources for themselves, which is a trade-off between one's short-term and long-term reward.

We fitted each model to its corresponding task using a maximum likelihood approach. According to the Bayesian information criterion (BIC), the social model fitted the data better than the baseline model assuming random choices for 75% of the participants, and the non-social model was better for the 69% of the participants. Paired sample t-tests confirmed a significant difference between the learning models and the baseline model ($p < .001$). On the qualitative level, average model predictions were in line with the observed choices. Also, we conducted a reverse fitting (i.e. fitting social model to the private condition) and according to the BIC, the social model fitted the data in the private condition better than the baseline model for only 50% of the participants. The non-social model fitted the data from the social condition for only 54% of the participants. Additionally, we fitted to both conditions two models widely used in the previous literature: the Rescorla-Wegner model (Sutton & Barto, 1998) and the Fehr-Schmidt inequity aversion model (Fehr & Schmidt, 1999). In both conditions, these models had a worse fit than the proposed social and non-social learning models. Using parametric fMRI analyses, we found modulation of the right ventrial striatum activity by the reward prediction error signal of the social model in the social condition ($p < .005$, uncorrected). Similar analysis with the non-social model did not reveal a significant modulation of the ventral striatum in the private condition.

In sum, our results indicate that people use different harvesting strategies in social and private situations. The resource depletion in each condition is monitored by different strengths of deactivation of the ventral striatum. The two versions of the social and non-social reinforcement learning model can effectively predict people's behavior in each situation and are correlated with the neural activity in the ventral striatum.

Bayesian Hierarchical Regression

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen J., & Wagenmakers E-J. (2013). *A Pracical Tutorial on Bayesian Hierarchical Regression in Management Science*. Manuscript submitted for publication.

Regression analysis is one of the most popular statistical methods used in the empirical sciences, in particular in management science. Usually, to judge the statistical support of each independent variable researchers use p-values obtained in the course of the null-hypothesis significance testing (NHST). This procedure will be referred to as either classical, orthodox or frequentist. The p-value is the probability of encountering a test statistic at least as extreme as the one that was observed, given that the null-hypothesis is true (Schervish, 1996). Unfortunately, p-values have a number of limitations (c.f. Wagenmakers, 2007): 1) they overstate the evidence against the null-hypothesis (e.g. Berger & Delampady, 1987; Sellke, Bayarri, & Berger, 2001), 2) they cannot quantify the evidence in favor of a null-hypothesis (e.g. Gallistel, 2009; Rouder, Speckman, Sun, Morey, & Iverson, 2009), and 3) they depend on the sampling plan and on the researcher's intention with respect to the data collection (Berger & Wolpert, 1988).

Here, we outline the theoretical and practical advantages of using Bayes factors (Jeffreys, 1961; Kass & Raftery, 1995) as an alternative to p-values. Bayes factors (BFs) quantify the support that the data provide for one hypothesis versus another, for example when the Bayes factor of hypothesis 1 (H_1) to the null hypothesis (H_0) is $\text{BF}_{10} = 5$, it means that H_1 is five times more likely than H_0 and alternatively, $\text{BF}_{10} = .2$ means that H_0 is five times more likely than H_1. Therefore, BF can quantify evidence for any hypothesis including the null. Further advantages of Bayes factors include independence of the result of the sampling plan.

BF is the ratio of the marginal likelihoods that the data were observed under hypothesis H_1 and hypothesis H_0: $\text{BF}_{10} = \text{Pr}(D|H_1)/\text{Pr}(D|H_0)$. The marginal likelihoods are obtained by integrating or averaging the likelihood over a model's prior parameter space; this way, all predictions that the model makes are taken into account. Flexible models (i.e. models with many parameters, such as regression models with many predictors) make many different types of predictions, and if most of these predictions are incorrect the average likelihood is low (Lee & Wagenmakers, in press). This way BFs implement Occam's razor or the principle of parsimony, which states that a model should be as complex as necessary to explain the data and as sparse as possible to avoid redundancies (Myung, Forster, & Browne, 2000; Wagenmakers & Waldorp, 2006).

In Bayesian inference, uncertainty about a model's parameter values is expressed by the prior belief about these parameter values, rather than by the confidence interval. For example, when testing a correlation between the proportion of the popular votes in the US presidential elections and the height ratio (i.e., height of the president divided by the height of his closest competitor, see Stulp, Buunk, Verhulst, & Pollet, 2013), one could assume that all values between -1 and 1 of the correlation coefficient ρ are equally likely, which implies an uninformed prior $\rho \sim \text{Uniform}(-1, 1)$. This prior belief would be later updated with evidence from the collected data. The resulting posterior probability distribution tells us how likely it is that, for example ρ ranges between .2 and .4. Alternatively, one could define a 95% credible interval for ρ . In contrast to the frequentist 95% interval, it provides a "plausible interval" for ρ, rather than an interval based on the normal distribution with the mean equal 0 (Hoekstra, Morey, Rouder, & Wagenmakers, in press). Importantly, Bayesian hypothesis testing answers the question "To what extent do the data support the presence of correlation?" rather than providing a significant, or not-significant result.

We used an example of a recent study by Dierdorff, Rubin, and Bachrach (2012) from the field of management science to illustrate Bayesian hierarchical regression.

Using a self-reported questionnaire ($N = 198$), Dierdorff et al. (2012) investigated the influence of five factors on *citizenship behavior. Citizenship* is defined as a "cooperative, helpful behavior extending beyond job requirements" (Barnard, 1938; Katz, 1964). Possible influences of these five factors were formulated as hypotheses H_{1-5}. Additionally, the authors expected that other factors, not included in any of the hypotheses, could influence the citizenship behavior. In the original study, the authors conducted a three-step hierarchical regression analysis, where in step 1 they included irrelevant factors, in step 2, they tested H_1 and in step 3, they tested hypotheses 2-5 (see Table 3 of the manuscript). According to p-values obtained for each independent variable, the authors found statistical evidence in favor of hypotheses 1 and 3-5.

Rouder and Morey (2012) proposed two ways of Bayesian hypothesis testing for regression analysis: 1) *covariance testing* and 2) *model comparison. Covariance testing* is based on building a full model (i.e. a model including all independent variables) and excluding one independent variable at a time to test each hypothesis. Then, the BF of the full model against the model without the variable testing the hypothesis is computed. If the BF is substantially high, one could assume that there is strong evidence in favor of the model including an additional predictor measuring the hypothesis. We repeated this analysis to test hypotheses 2-5 and we found positive evidence for hypothesis 5, negative evidence against hypothesis 2 and mixed evidence in favor of hypotheses 3-4. *Model comparison* is based on comparing models containing different combinations of predictors. This method, despite its greater complexity, allows controlling for possible correlations between independent variables. Using this method, we again concluded did not find evidence in favor of these hypothesis 2, found mixed evidence for hypotheses 3-4 and found strong evidence supporting hypothesis 5.

Therefore, Bayesian and classical inference can lead to different conclusions and these differences cannot be neglected. This can have important implications for management science research. BFs can be computed from R^2, which can be obtained from standard statistical software (see Rouder & Morey, 2012). Therefore, this manuscript makes an important contribution to the area of management science by providing a tutorial on how to easily apply Bayesian hypothesis testing in regression analysis, in order to provide a more accurate hypothesis testing.

Conclusion

In this dissertation, I use and develop three aspects of quantitative analysis of decision making: 1) modeling the decision process, 2) quantitative description of choice options, and 3) statistical analysis of behavioral data. I argue that various models of decision making can successfully describe the decision process, which cannot be measured and observed otherwise, and can predict people's behavior. Also, different models react differently to various decision environments described by statistical properties of the choice options. I argue that the decision environment can be described with the statistical and mathematical properties of choice options.

I make a novel contribution to the research of judgment and decision making by proposing *standardized covariance*, a measure of association, similarity and coriskiness between choice options. In three experiments, I show that association measured with standardized covariance can describe the decision environment. It can influence people's preferences and predictions of models of decision making. Despite its similarities to the correlation measure, standardized covariance can quantify similarity between two choice options above the linear relationship. Therefore, the proposed mathematical measure can relate to people's perceived association and similarity between choice options.

I also argue that context-dependent evaluation measured with the proposed measure of association between choice options plays an important role in people's decisions. This is in contrast to fixed-utility theories, including their most prominent version, CPT (Kahneman & Tversky, 1979). I show sufficient evidence favoring context-dependent theories, in particular the DFT (Busemeyer & Townsend, 1993), which predicted people's behavior best. These results show, how the context-dependent models are influenced by different strengths of the association.

Further, I provide evidence the risk carried by the outcomes' distribution can be described by the expected shortfall, which accounts for outcomes that fall below a person's expectations, by measuring the distance of each outcome from the expectation (Lopes & Oden, 1999). Therefore, this measure can account for both the range of the outcomes' values and their distribution. This measure can be successfully used as a measure of risk in standard risk-value models (Sarin & Weber, 1993).

Whereas the large part of this dissertation deals with human individual decisions, in one paper, I develop a version of a reinforcement learning model that can account for feedback resulting from the social environment. The model assumes that an individual competing with others for common pool resources learns to behave in a competitive and selfish way. That is, when others extensively harvest the resources, a person learns to harvest a lot as well. In contrast, when in a private, non-competitive environment, a decision maker learns to preserve the resource in order to sustain the sufficient level of the resource. Importantly, this model not only predicted people's behavior but also correlated with neural activity, which supports the explanation the neurophysiological data.

Finally, I argue in favor of Bayesian hypothesis testing in the regression analysis. I show the differences and flaws of the NHST approach, in comparison to using BFs. Re-analysis of the data of a prominent study in the area of management science resulted in conclusions contrasting the original study. This, however, was not due to the researchers' mistake, but due to the inaccuracy of the commonly used statistical method.

In sum, the overarching goal of my dissertation is to provide ways and measures of quantitative analysis to improve and facilitate the research of decision making under risk. Every person has to take risky decisions in their lives, especially those involving monetary consequences. Therefore, I believe that this dissertation significantly contributes not only to the research community, but also provides an important insight on people's economic decisions.

References

Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. *Jourenal of Banking and Finance*, *26*(7), 1487–1503. doi: 10.1016/S0378-4266(02)00283-2

Andraszewicz, S., & Rieskamp, J. (under review). *Standardized covariance - a measure of association, similarity and co-riskiness between choice options.*

Barnard, C. (1938). *The functions of the executive.* Cambridge, MA: Harvard University Press.

Berger, J. O., & Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, *2*(3), 317–352. doi: 10.1214/ss/1177013238

Berger, J. O., & Wolpert, R. L. (1988). *The likelihood principle* (2nd ed.). Hayward, CA: Institute of Mathematical Statistics.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, *115*(2), 463-501. doi: 10.1037/0033-295X.115.2.463

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*(2), 409-432. doi: 10.1037/0033-295X.113.2.409

Burke, C., & Tobler, P. (2011). Reward skewness coding in the insula independent of probability and loss. *Journal of Neurophysiology*, *106*(1), 2415–2422. doi: 10.1152/jn.00471.2011

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432-459. doi: 10.1037/0033-295X.100.3.432

Chiu, W. (2005). Skewness preference, risk aversion, and the precedence relations on stochastic changes. *Management Science*, *51*(12), 1816–1828. doi: 10.1287/mnsc.1050.0431

Chunhachinda, P., Dandapani, K., Hamid, S., & Prakash, A. J. (1997). Portfolio selection and skewness: Evidence from international stock markets. *Journal of Banking & Finance*, *21*(2), 143–167. doi: 10.1016/S0378-4266(96)00032-5

Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Dierdorff, E. C., Rubin, R. S., & Bachrach, D. G. (2012). Role expectations as antecedents of citizenship and the moderating effects of work context. *Journal of Management*, *38*(2), 573–598. doi: 10.1177/0149206309359199

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*(2), 293–315. doi: 10.1016/j.geb.2005.03.001

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, *114*(3), 817-868. doi: 10.1162/003355399556151

Fishburn, P. (1977). Mean-risk analysis with risk associated with below-target returns. *The American Economic Review*, *67*(2), 116–126. Retrieved from http://www.jstor.org/stable/1807225

Gallistel, C. (2009). The importance of proving the null. *Psychological Review*, *116*(2), 439–453. doi: 10.1037/a0015251

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way:

Models of bounded rationality. *Psychological Review*, *103*(4), 650-669. doi: 10.1037/0033-295X.103.4.650

González-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review*, *109*(1), 137–155. doi: 10.1037//0033-295X.109.1.137

Green, J., & Jullien, B. (1989). Ordinal independence in nonlinear utility. *Journal of Risk and Uncertainty*, *2*(1), 355–387. doi: 10.1007/BF00055713

Hardin, G. (1968). Tragedy of the commons. *Science*, *162*(3859), 1243–1248. doi: 10.1126/science.162.3859.1243

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (in press). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). New York: Oxford University Press.

Kahneman, D., & Tversky, A. (1979, March). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263-91. doi: 10.2307/1914185

Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of American Statistical Association*, *90*(430), 773-785. doi: 10.1.1.19.6821

Katz, D. (1964). The motivational basis of organizational behavior. *Behavioral Sciences*, *9*(2), 131-146. doi: 10.1002/bs.3830090206

Kroll, Y., Levy, H., & Markowitz, H. (1984). Mean-variance versus direct utility maximization. *The Journal of Finance*, *39*(1), 47–61. doi: 10.1111/j.1540-6261.1984.tb03859.x

Lee, M. D., & Wagenmakers, E.-J. (in press). *Bayesian modeling for cognitive science:*

A practical course. Cambridge MA: Cambridge University Press.

Leland, J. W. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty*, *9*(2), 151–172. doi: 10.1007/BF01064183

Leland, J. W. (1998). Similarity judgments in choice under uncertainty: a reinterpretation of the predictions of regret theory. *Management Science*, *44*(5), 659–672. doi: 10.1287/mnsc.44.5.659

Li, X., Qin, Z., & Kar, S. (2010). Mean-variance-skewness model for portfolio selection with fuzzy returns. *European Journal of Operational Research*, *202*(1), 239–247. doi: 10.1016/j.ejor.2009.05.00

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, *92*(368), 805-24. doi: 10.2307/2232669

Lopes, L. L., & Oden, G. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and sp/a theory. *Journal of Mathematical Psychology*, *43*(2), 286–313. doi: 10.1006/jmps.1999.1259

Luce, R. (1990). Rational versus plausible accounting equivalents in preference judgements. *Psychological Science*, *1*(4), 225–234. doi: 10.1111/j.1467-9280.1990.tb00205.x

Markowitz, H. (1959). *Portfolio selection: Efficient diversification of investment*. New York: John Wiley.

Mellers, B., & Cooke, A. (1994). Trade-offs depend on attribute range. *Journal of Experimental Psychology*, *20*(5), 1055–1067. doi: 0.1037/0096-1523.20.5.1055

Mellers, B., Schwartz, A., & Cooke, A. (1998). Judgement and decision making. *Annual Review of Psychology*, *49*(1), 447–477. doi: 10.1146/annurev.psych.49.1.447

Myung, I. J., Forster, M. R., & Browne, M. W. (2000). Model selection [Special issue]. *Journal of Mathematical Psychology*, *44*(1), 1–2. doi: 10.1006/jmps.1999.1273

Ostrom, E. (1990). *Governing the commons: the evolution of institutions of collective action.* Cambridge: Cambridge University Press.

Post, T., van Villet, P., & H., L. (2006). Risk aversion and skewness preference. *Journal of Banking and Finance*, *32*(1), 1178–1187. doi: 10.1016/j.jbankfin.2006.02.08

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, *44*(3), 631-661. doi: 10.1257/jel.44.3.631

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*, 59–66. Retrieved from `http://www.jstor.org/stable/2685263`

Rouder, J. N., & Morey, R. (2012). Default bayes factors for model selection in regression. *Multivariate Behavioral Research*, *47*(6), 877-903. doi: 10.1080/00273171.2012.734737

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. doi: 10.3758/PBR.16.2.225

Rubinstein, A. (1988). Similarity and decision-making under risk (is there a utility theory resolution to the allais paradox?). *Journal of Economic Theory*, *46*(1), 145–153. doi: 10.1016/0022-0531(88)90154-8

Samuelson, P. (1970). The fundamental approximation theorem of portfolio analysis in terms of means, variances and higher moments. *The Review of Economic Studies*, *37*(4), 537–542. doi: http://www.jstor.org/stable/2296483

Sarin, R. K., & Weber, M. (1993). Risk-value models. *European Journal of Operational Research*, *70*(2), 135-149. doi: 10.1016/0377-2217(93)90033-J

Schervish, M. J. (1996). P values: What they are and what they are not. *The American Statistician*, *50*(3), 203206. doi: 10.1080/00031305.1996.10474380

Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician*, *55*(1), 62-71. doi: 10.2307/2685531

Stulp, G., Buunk, A. P., Verhulst, S., & Pollet, T. V. (2013). Tall claims? Sense and nonsense about the importance of height of US presidents. *The Leadership Quarterly*, *24*(1), 159–171. doi: 10.1016/j.leaqua.2012.09.002

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction.* Cambridge MA: Cambridge MA: MIT Press.

Symmonds, M., Wright, N., Bach, D., & Dolan, R. (2011). Deconstructing risk: Separable encoding of variance and skewness in the brain. *NeuroImage*, *58*(1), 1139–1149. doi: 10.1016/j.neuroimage.2011.06.087

Tobler, P., O'Doherty, J., Dolan, R., & Schultz, W. (2007). Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *Journal of Neurophysiology*, *97*(2), 1621–1632. doi: 10.1152/jn.00745.2006

Train, K. (2009). *Discrete choice methods with simulation*. New York: Cambridge University Press.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*, 31–48. doi: 10.1037/h0026750

Tversky, A. (1977). Intransitivity of preferences. *Psychological Review*, *84*(4), 327–352. doi: 10.1037/0033-295X.84.4.327

Tversky, A. (2004). *Preference, belief and similarity: Selected writings* (1st ed.; E. Shafir, Ed.). Cambridge, MA: MIT Press. (Chapters 1-6)

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulleting & Review*, *14*(5), 779–804. doi: 10.3758/BF03194105

Wagenmakers, E.-J., & Waldorp, L. (2006). Model selection: Theoretical developments and applications [Special issue]. *Journal of Mathematical Psychology*, *50*(2), 99-214. doi: 10.1016/j.jmp.2005.01.005

Weber, E., Shafir, S., & Blais, A.-R. (2004). Predicting risk-sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review*, *111*(2), 430–445. doi: 10.1037/0033-295X.111.2.430

Standardized Covariance - A Measure of Association, Similarity and Co-Riskiness between Choice Options

Sandra Andraszewicz*, Jörg Rieskamp

University of Basel
Center for Economic Psychology
Missionsstrasse 62A, 4055 Basel, Switzerland

Abstract

Predictions of prominent theories of decision making, such as the *proportional difference model*, the *priority heuristic*, *decision field theory* and *regret theory*, strongly depend on the association between outcomes of choice options. In the present work, we show that these associations reflect the similarity of two choice options and riskiness of one option with respect to the other. We propose a measure labeled standardized covariance that can capture the strength of the association, similarity and co-riskiness between two choice options. We describe the properties and interpretation of this measure and show its similarities to and differences from the correlation measure. Finally, we show how the predictions of different models of decision making vary depending on the value of the standardized covariance, which can have implications for research on decision making under risk.

Keywords: covariance, association, similarity, risk, decision making

---

*Corresponding author's contact information
e-mail: s.andraszewicz@unibas.ch
tel.: +41 (0)61 267 06 57

[1]Acknowledgments: We would like to thank Gilles Dutilh and Bettina von Helversen for very useful remarks and comments.

1. Introduction

People face risky decisions in their everyday lives. For example, a choice between two car insurance offers is a choice between risky options with payouts depending on the occurrence of an accident. An accident can occur with a certain probability that can be estimated based on the driver's age, years of experience in driving a car and history of previous accidents. A person choosing between the two insurance offers would probably compare the coverage and conditions of both insurances with regard to specific situations such as a broken window, towing, help abroad etc., rather than evaluate each option independently of the other.

Many models of decision making assume that during a decision process, people compare the options' outcomes with each other, in an attribute-wise fashion. For instance, the *priority heuristic* (Brandstätter et al., 2006) assumes that people first compare all options with respect to their minimum outcomes. If these outcomes do not allow for discrimination between the options, the options are compared with respect to the probability of the minimum outcomes, and finally with respect to the highest outcomes. *Regret theory* (Loomes & Sugden, 1982), the *proportional difference model* (González-Vallejo, 2002), and *decision field theory* (Busemeyer & Townsend, 1993) are three other prominent computational models of decision making, which assume that decision makers compare outcomes of the choice options with one another. These comparisons are then accumulated to form an overall preference.

Hence, all these models predict that the choice preference depends on the association between the options' outcomes. *Association* is a relationship between two variables and it implies statistical dependence between them. Andraszewicz et al. (under review) experimentally showed that the strength of association between choice options influences people's decisions, such that the stronger the association, the higher the probability that a decision maker chooses an option with the larger expected value, even when the difference between the expected values is the same for options with various strengths of the association.

Studies on decision making under risk often overlook the association between choice options and put the main focus on the difference between their expected values. However, the way choice options are selected has a crucial influence on testing choice models. As highlighted in the work on optimal experimental design, selecting gambles for discriminating between various models of decision making is an essential issue that determines the effectiveness of an experiment (see Cavagnaro et al. 2013; Myung & Pitt 2009; Zhang & Lee 2010). Although optimal experimental design is still hard to apply in a simple experimental setting, one could easily control for the association between options to eliminate possible confound variables.

The concept of associations between outcomes of choice options is related to the concept of similarity between options. *Similarity* defines how features of one object are related to the features of another object (Tversky, 2004, pp. 3). Therefore, both *association* and *similarity* depend on the comparison of two options' attributes with each other. However, in case of completely independent options the association between the options outcomes is zero, while the outcomes of the options could still be quite similar to each other.

The literature on similarity (c.f. Tversky, 2004), reports some ways to describe similarity. For example, one measure is a metric of dissimilarity between objects' features. This metric, which ranges between 0 (no dissimilarity) and 1 (maximal dissimilarity), is based on calculating the distance between values of objects' features in a coordinate space. Also, Tversky (1977) proposed a *con-*

trast model, which is based on a ratio of the number of features that are the same for both objects compared to the features which are different. Along similar lines, the *similarity model* (Rubinstein, 1988; Leland, 1994, 1998) assumes that when options are similar in one dimension (attribute) but different in another, a decision maker should choose the option that is better in the dissimilar attribute.

A few measures of similarity have been proposed. For example, Tversky (1977) proposed that similarity can be measured with the probability judgment of how much one object is similar to another. Busemeyer and Townsend (1993), indicated covariance as a measure of similarity between choice options. In portfolio theory, covariance between financial assets is used as a measurement of association between two assets (i.e. Pafka & Kondor 2003; Disatnik & Benninga 2007). Therefore, covariance could be a measure that reflects both the association and similarity of options. Unfortunately, covariance measure depends on the range of the outcome values, which makes it hard to interpret.

A correlation measure would be an alternative. Tversky (2004) lists correlation as a possible measure of similarity. However, a large part of the research on decision making is conducted with two-outcome gambles (e.g. González-Vallejo 2002; Birnbaum 2008), for which the correlation is either 1 or -1 (see Rodgers & Nicewander, 1988). As a consequence, the correlation measure is unable to capture the strength of the association between pairs of two-outcome options. Also, correlation is a special case of the association measure, which indicates the linear relationship between two variables. Consequently, for cases with small amounts of data (i.e. only two data points), correlation does not provide a useful approach.

Therefore, we propose the *standardized covariance*, as a measure of the strength of the association and similarity between choice options. This measure is meant for the application in risky numerical choices[1]. The main advantages of the standardized covariance is that 1) it ranges between -1 and 1, therefore its values can be interpreted similar to the correlation values; 2) it can be applied to choice options with only two outcomes, for which correlation does not provide a meaningful solution. The specific standardization procedure applied in the standardized covariance also allows for measuring how risky is one option with respect to the other option in a pair. We call this as *co-riskiness* between two options.

Co-riskiness defines how risky one choice option is relative to the risk of another option. When both options have a similar level of risk, then co-riskiness is high. When one option is not very risky (a safe option) and the other is very risky, co-riskiness is low. Co-riskiness should have an influence on decision process because when one option is substantially more risky than the other, the difference between riskiness should have a greater impact on a person's decision, that is a person might prefer the safer option. However, if the co-riskiness is high (i.e. same level of risk), a decision maker would rely on other choice criteria, for instance giving more importance to the expected value of the options.

In short, $\sigma^*_{AB} = 1$ means strong positive association, high similarity and high co-riskiness. In contrast, $\sigma^*_{AB} = -1$ implies a strong negative association, high dissimilarity and high co-riskiness. When $\sigma^*_{AB} = 0$, there is no association and similarity between choice options and the co-riskiness between them is minimal. When σ^*_{AB} is positive, but close to 0, there is a weak positive association, low similarity and low co-riskiness. Analogically, for negative σ^*_{AB} whose value is close to 0 we observe a weak negative association, low dissimilarity and low co-riskiness.

[1]Most of the existing measures describe similarity between objects.

In the following sections, we provide an in-depth description and analysis of properties of the standardized covariance, supported by examples and simulations. Importantly, we show the implications of controlling for the standardized covariance for the decision making research. We demonstrate applications of the standardized covariance not only as an alternative to the correlation measure between variables with two data points, but also as a stand-alone concept. We start the analysis with two-outcome options and later extend it for the application in options with many outcomes. We show the similarities and differences between the standardized covariance and the correlation measure. Finally, we test empirically whether people's choices are influenced by the association between the choice options measured solely by the correlation measure or by the standardized covariance.

2. Properties of Standardized Covariance

The standardized covariance between a pair of options A and B, labeled σ^*_{AB}, where the non-standardized covariance is denoted as σ_{AB}, is a ratio between twice the covariance and the sum of variances σ_A and σ_B of options A and B, respectively:

$$\sigma^*_{AB} = \frac{2\sigma_{AB}}{\sigma^2_A + \sigma^2_B}. \tag{1}$$

σ^*_{AB} is a continuous variable ranging from -1 to 1. σ^*_{AB} is not higher than 1 or lower than -1 because $2\sigma_{AB} \leq (\sigma^2_A + \sigma^2_B)$. In Appendix A, we provide the mathematical proof. When $\sigma^*_{AB} = 0$, either the options are completely unrelated (i.e. they are statistically independent) or the covariance between the options' outcomes is equal to 0. The second case occurs when one of the options is a sure option. When σ^*_{AB} approaches 0, variance of one option is low, and variance of the other option is high. Then, the association between the options is low. In

Table 1, we show eight examples of statistically dependent pairs of options to demonstrate properties of σ^*_{AB}.

Example 1

Stochastically non-dominant options are identical only when

$$\sigma^*_{AB} = 1 \iff 2\sigma_{AB} = \sigma^2_A + \sigma^2_B.$$

Example 2

Options become dissimilar by interchanging the outcomes of one option:

$$\sigma^*_{AB} = -1 \iff -2\sigma_{AB} = \sigma^2_A + \sigma^2_B.$$

We call this property *symmetry* for positively and negatively associated options. For any pair of options with $\sigma^*_{AB} = x$, interchanging the outcomes of one option results in $\sigma^*_{AB} = -x$.

Example 3

σ^*_{AB} decreases when one outcome of one option is altered so that the options are no longer the same.

Examples 3 and 5

The probabilities of the outcomes do not influence σ^*_{AB}. Interchanging the probabilities does not change the value of σ^*_{AB}.

Example 6

$\sigma^*_{AB} = 1$ when the difference between I outcomes a_i of option A and b_i of option B is the same and this difference equals the difference between expected values:

$$\sigma^*_{AB} = 1 \iff (a_i - b_i) = (a_{i+1} - b_{i+1}) = \Delta EV.$$

Examples 1, 3 and 7

The less similar the outcomes a_i and b_i corresponding to the same probabilities p_i, the smaller σ^*_{AB}:

$$(a_i - b_i) \nearrow \Longleftrightarrow \sigma^*_{AB} \searrow .$$

Example 8

When the outcomes of one option are almost the same, $\sigma^*_{AB} \to 0$

Examples 1, 3, 7 and 8

σ^*_{AB} is a measure of how large the variance of outcomes (riskiness) of one option is with respect to the other option. We define this property as *co-riskiness*, such that

$$|\ \sigma^*_{AB}\ | \nearrow \Longleftrightarrow co-riskiness \nearrow .$$

3. Standardized Covariance vs. Correlation

Examples presented in Table 1 indicate that the correlation and the standardized covariance are similar measures. Options in Example 1 are identical, in Example 2, opposite and in Example 6, all outcomes of option A are better than outcomes of option B, by the same amount of points. These examples are characterized by both the "perfect" correlation (i.e. $r = 1$) and the "perfect" standardized covariance. Correlation coefficient r equals

$$r = \frac{\sigma_{AB}}{\sigma_A \sigma_B}. \tag{2}$$

Therefore, the relationship between the correlation coefficient and the standardized covariance is

$$\sigma^*_{AB} = \frac{2r\sigma_A\sigma_B}{\sigma_A^2 + \sigma_B^2} \tag{3}$$

and the correlation is equal to the standardized covariance when

$$2\sigma_A\sigma_B = \sigma_A^2 + \sigma_B^2. \tag{4}$$

The correlation and the standardized covariance have exactly the same values in only two cases, such that $r = \sigma^*_{AB} = -1 \cup r = \sigma^*_{AB} = 1$. Given that for options with two outcomes, the correlation is always either -1 or 1, then $\sigma_{AB} = \sigma_A\sigma_B$. Thus, when $r = 1$, the standardized covariance could also be written as

$$\sigma^*_{AB} = \frac{2\sigma_A\sigma_B}{\sigma_A^2 + \sigma_B^2}, \tag{5}$$

when the association between the options' outcomes is positive. When the association between the options' outcomes is negative, covariance equals minus the product of variances (i.e. $\sigma_{AB} = -\sigma_A\sigma_B$).

4. Similarity of Statistically Independent Options

Until now, we have been discussing choice options that depend on the same external events. However, researchers also consider statistically independent options whose outcomes depend on different external events. In such cases, the association between the choice outcomes is zero and the covariance between them is 0, which results from

$$\sigma_{AB} = \sum_{i=1}^{I}\sum_{j=1}^{J} p_i \cdot p_j(a_i - E[A])(b_j - E[B]), \tag{6}$$

where p_i and p_j refer to the probabilities of occurrence of the respective outcomes a_i and b_j.

However, it is still reasonable to consider how similar the two independent choice options are, using a similarity measure S_{AB}. By modifying Equation 6, we define the strength of the similarity between the outcomes of two options as s_{AB}, such that

$$s_{AB} = \sum_{i=1}^{I} \sum_{j=1}^{J} p_i \cdot p_j \sqrt{(a_i - E[A])^2 (b_j - E[B])^2}. \tag{7}$$

Because Equation 7 always returns a positive value, one needs another parameter, which defines whether the options are similar ($S_{AB} > 0$) or dissimilar ($S_{AB} < 0$). As a consequence, the strength of the similarity s_{AB} should be multiplied by a direction parameter d_{AB} defined as

$$d_{AB} = \begin{cases} -1 & \textit{if } \sum_{i=1}^{I} (p_{Ai} + p_{Bi})/2 \cdot (a_i - E[A])(b_i - E[B]) < 0 \\ 1 & \textit{if } \sum_{i=1}^{I} (p_{Ai} + p_{Bi})/2 \cdot (a_i - E[A])(b_i - E[B]) > 0 \end{cases}. \tag{8}$$

Therefore, for the statistically independent gambles we define *similarity* as

$$S_{AB} = \frac{2 d_{AB} \cdot s_{AB}}{\sigma_A^2 + \sigma_B^2}. \tag{9}$$

The proposed S_{AB} and σ^*_{AB} measures lead to similar descriptions of the similarity of statistically dependent and independent options. Using four examples in Table 2, we demonstrate properties of similarity between two statistically independent choice options.

Example 9

Probabilities of outcomes influence similarity of statistically independent gambles. Options A and B have the same outcomes, but different probabilities of occurrence of these outcomes. Therefore, the expected values and variances of the two options differ, such that $E[A] = 20$ and $E[B] = 25$, $\sigma_A^2 = 100$ and $\sigma_B^2 = 75$.

Examples 9 and 10

Similarity decreases when the outcome of one option is changed. When one option is a sure option, $S_{AB} = 0$, which results from Equation 7.

Examples 9 and 11

Similar and dissimilar statistically independent options are symmetrical (see also Section 2).

Example 12

When the difference between the riskiness of options increases (i.e. one option has a substantially lower variance than the other), S_{AB} decreases.

Example 12

$E[A] - E[B] = 0$ does not imply that $S_{AB} = 1$. In contrast, for statistically dependent options, reducing the difference between expected values implies making the outcomes more similar to each other because the probabilities cannot be manipulated (compare with Section 2).

5. Statistical Properties of Choice Options

In the previous sections, we noticed that association, similarity and riskiness of choice options are related to the difference between expected values, variances and covariance between them. Hence, understanding the relationships among these measures helps to understand the properties of the standardized covariance. Variances of both options and the covariance between them depend on the same components, (Note: $\sigma_{AB} = E[(a - E[A])(b - E[B])]$ and $\sigma_A^2 = E[(a - E[A])^2]$), where these components define the distance of the outcome values from the expected value of an option. Therefore, this section clarifies why the standardized covariance can explain the association, similarity and co-riskiness between options and how it depends on the expected value difference.

5.1. Expected Value and Variance of Two-Outcome Options

When two options are stochastically non-dominant, one option has higher variance than the other (compare range of outcomes of options A and B in Table 1 in Examples 3, 4, 5, 7 and 8, to Example 6). Therefore the sum of variances is composed of a smaller and larger variance. To investigate the relationship between variances of two options and the differences between their expected values, we generated pairs of stochastically non-dominant pairs of non-identical two-outcome options with outcomes ranging between 0 and 100 points, and probabilities of these outcomes equal to 40%, 50% or 60%. The sample consisted of seven sets, where each set contained all possible pairs of gambles with the expected value difference (ΔEV) of 0, 5, 10, 15, 20, 25 and 30 points.[2]

As shown in Figure 1 (left panel), when the options' outcome values are defined within a fixed range (i.e. $range \in [0, 100]$), the relationship between all possible values of variances is symmetric with respect to the diagonal. The smaller the ΔEV, the greater the possible range of both variances (Figure 1, right panel). Therefore, when $\Delta EV = 0$, the variances of both options can become very similar to each other, whereas they cannot when the difference between ΔEV is large. This provides important information for the research in decision making under risk, because variance is the most common risk measurement of choice options (Sarin & Weber, 1993). The greater the ΔEV, the greater the difference between riskiness of one option in comparison to the other. Therefore, high co-riskiness can only occur in options with low ΔEV and for choices with large ΔEV, the co-riskiness is low.

[2] We later created options with the same properties, but having 1) only negative outcomes, and 2) one positive and one negative outcome. These two groups of options exhibit the same properties as the options with only positive outcomes.

5.2. Covariance of Two-Outcome Options

Covariance between two options is related to both variances in a non-linear fashion. This means that it is impossible to keep covariance constant while manipulating variances, and the other way round. This motivates incorporating variances as the standardization component in standardized covariance and the correlation measure. Covariance is symmetric for positively and negatively related options, i.e. in Examples 3 and 4, we observe that $\sigma^2_{A_{Example3}} = \sigma^2_{A_{Example4}} \wedge \sigma^2_{B_{Example3}} = \sigma^2_{B_{Example4}}$, but $\sigma_{AB_{Example3}} = -\sigma_{AB_{Example4}}$. Due to this property both the standardized covariance and the correlation measure can indicate relationship of the same strength for positively and negatively related options.

Figure 2 (left panel) shows the area of possible relationships between covariance and the sum of the variances of two stochastically non-dominant choice options. The smaller the difference between expected values, the greater the area (Figure 2, right panel). When $\Delta EV > 0$, standardized covariance does not overlap with the correlation measure. The gray line in the left panel is outside the black area and it overlaps with black dashed line on the right panel. Although controlling for the difference between expected values reduces the amount of possible pairs of options, the range of properties of the options is still large.

To further investigate the space of statistical properties of choice options when ΔEV is fixed, we measured the range of the standardized covariance, the ratio of the smaller to the larger variance in the pair and the number of possible pairs of options that can be found. According to Table 3, for any of the seven listed ΔEVs, there is a very large range of possible options, with various riskiness levels and strength of the association between the options in a pair. This shows that for any chosen difference between expected values, pairs of gambles of various strengths of association and riskiness can be generated. This points out the limitation of studies on decision making, which only control for the difference between expected values and do not control for variances and covariance.

6. Influence of σ^*_{AB} on Models of Decision Making

As described in section 1, prominent models of decision making emphasize the association between options' outcomes. As a consequence, we tested how the strength of the association, expressed by σ^*_{AB}, influences predictions of three models: *regret theory* (Loomes & Sugden, 1982), *decision field theory* (Busemeyer & Townsend, 1993), and the *proportional difference model* (González-Vallejo, 2002). We generated model predictions for pairs of stochastically non-dominant options with a fixed expected values difference ($\Delta EV = 15$). We used all 377500 pairs of two-outcome options with outcomes ranging between 1 and 100 points and probabilities either 40%, 50% or 60%. At the same time, we manipulated the standardized covariance and we separated all pairs of options into three groups: 1) small, $\sigma^*_{AB} \leq .2$ (21.2%), 2) medium, $.2 < \sigma^*_{AB} \leq .5$ (34.7%), and 3) large $.5 < \sigma^*_{AB}$ (44.2%). The model specifications are outlined in Appendix B.

We averaged the probabilities of choosing the option with the larger expected value for each of the three groups. As shown in Figure 3, all three models predict higher probabilities when the association between the options is higher. This effect can be best observed in the predictions of regret theory and decision field theory. The variances of predictions of regret theory, decision field theory and the proportional difference model also differ, where the first two models predict smaller variance of choices when the association is higher, whereas the proportional difference model predicts an increase of variance.

Further, we checked the interaction between σ^*_{AB} and ΔEV, using seven populations of options with the properties described above, but with various expected value differences (i.e. $\Delta EV \in \{0, 5, 10, 15, 20, 25, 30\}$[3]). Figure 4 shows that the models on average make different predictions depending on σ^*_{AB}, for every ΔEV level. Also, the variance of these predictions changes depending on σ^*_{AB}. The exception are predictions of decision field theory when $\Delta EV = 0$, which are always $Pr(A|A,B) = .5$. Overall, there are systematic differences in average and variance of predictions of the models, depending on the association, similarity and co-riskiness of the options, for every ΔEV level. For regret theory and decision field theory these differences decrease with the ΔEV increase. For the proportional difference model, it is the other way round.

It is beyond the scope of this paper to investigate predictions of all existing models whose predictions could depend on the standardized covariance. However, we selected models that reflect different aspects of σ^*_{AB}. Regret theory, in the regret function (see Equation B.1), incorporates the difference between options' outcomes corresponding to the same probabilities $(a_i - b_i)$. A similar assumption is made by the proportional difference of the proportional difference model (see Equation B.8). Decision field theory directly incorporates variances and covariance between options in its mathematical specification (see Equation B.6), where Busemeyer & Townsend (1993) interpret covariance as a measure of similarity between two choice options. Therefore, this model accounts for both riskiness of each option (expressed by variances) and relationship between choice options (expressed by covariance).

In sum, the predictions of different choice models depend on an interaction between the difference between expected values and the strength of the association, similarity and co-riskiness between the choice options. This finding is very important, as it shows that results of studies that focus on differences between

[3]The sizes of the populations are listed in Table 3.

expected values can be confounded with unmeasured association between choice options.

7. Options with More than Two Outcomes

σ^*_{AB} is a stable measure and properties of the standardized covariance hold for options with more than two outcomes[4]. Therefore, we used four-outcome options, for which we were able to compute a meaningful correlation, to compare the standardized covariance with the correlation measure. Figure 5 shows a scatter plot between the two measures. The gray line indicates the regression line. As Figure 5 shows, the two measures are very strongly correlated with each other, $r = .98, p < .001$. Also, the slope of the regression line is high and the intercept is very small (see caption of Figure 5). Thus, the standardized covariance is a similar measure as correlation, but it has the advantage that it can be applied to both two-outcome options and options with several outcomes.

The difference between the correlation measure and the standardized covariance is that the first describes a linear relationship between two variables, rather than how similar the values of two variables are to each other. In contrast, the standardized covariance measures the size of the distance between the outcomes' values of both variables. For demonstration, we selected one case from Figure 5 for which correlation is much higher than standardized covariance, that is $r = .95$ and $\sigma^*_{AB} = .32$. In Figure 6, we plotted the outcome values of both options as a function of four independent events (left panel) and the outcomes of option B as a function of the outcomes of option A (right panel). Despite the fact that in the right panel the outcomes lie very close to the regression line, their values are not

[4]We tested these properties in options with more than two outcomes.

close to each other on the left panel. Therefore, the correlation is a measure of a linear relationship between options, whereas standardized covariance measures association as a similarity between options' outcomes.

7.1. Behavioral Experiment

We have discussed differences and similarities between the standardized covariance and the correlation. Andraszewicz et al. (under review) showed that choice situations that differ by σ^*_{AB} or r influence people's preferences. However, their study tested the influence of the standardized covariance independently of the correlation, where cases similar to the one presented in Figure 6 did not occur. Therefore we conducted a behavioral study to examine whether differences in the standardized covariance affect peoples decisions under risk beyond the association measured by the correlation. We have also asked participants to judge similarity of gambles expressed by the correlation and the standardized covariance.

7.1.1. Method

20 students ($N_{female} = 15$, $M_{age} = 21$) of the University of Basel participated in the experiment and received course credit for compensation. During 120 trials presented in a random order, they repeatedly chose between two four-outcome pairs of gambles that were stochastically non-dominant and positively related to each other. The pairs were presented graphically on a screen as hypothetical stocks (see Figure 7). Sixty trials contained pairs of gambles for which $\sigma^*_{AB} \leq .2$ and $r \geq .8$ ("low covariance" condition). The other 60 trials contained pairs of options for which $\sigma^*_{AB} \geq .8$ and $r \geq .8$ ("high covariance" condition). At the end of the experiment, one gamble was chosen and played out, where the outcome was paid to the participants as a bonus of 0-2 Swiss Francs ($\approx 0-2$ US Dollars).

In both conditions, $\Delta EV = 15$. Further specification of the gambles' generation can be found in Appendix C. If people's choices are influenced only by the statistical correlation between the choice outcomes, then we should observe similar choices in both conditions. However, if people's choices are influenced by the standardized covariance between options, then people should choose the options with the larger expected value more frequently in the condition with the high standardized covariance.

Afterwards, the participants were asked to fill out a short demographics questionnaire and answer in which condition stock A is more similar to stock B. Two examples of pairs of gambles, each from each condition, were presented on paper and for the first half of the participants, the left example contained gambles from low covariance condition and right picture contained an example from high covariance condition. For the other half of the participants, it was the other way round. The aim of this last question was to explicitly ask participants whether they find gambles with low standardized covariance less similar to each other than the gambles with high standardized covariance. Due to the similar correlation value both situation should be judged identically whereas they should be judged differently according to the standardized covariance, that is $r_{LowCovariance} = .97$ and $r_{HighCovariance} = .81$, whereas $\sigma^*_{AB_{LowCovariance}} = .08$ and $\sigma^*_{AB_{HighCovariance}} = .80$.

7.1.2. Results

For each participant, we calculated the average frequency of choices of the gamble with the larger expected value for the high and low covariance conditions. According to the Wilcoxon summed-rank test[5], the frequency was higher in high covariance condition ($Me = 90\%$, $SE = 9\%$) than in the low covariance condition ($Me = 80\%$, $SE = 18\%$), $p < .001$. We obtained similar results in a simulation study, where we found that decision field theory, regret theory and the proportional difference model predict higher choice probabilities of the gambles with the larger expected value in the high covariance condition than in low covariance condition[6].

Therefore, in situations with low standardized covariance, despite the high correlation, choices of the option with the larger expected value were less frequent as compared to a situation in which the standardized covariance was high. These results clearly indicate that people's choices are affected by the similarity of choice options as described by the standardized covariance. Also, in the final question, all participants indicated that the gambles in the pair from the high covariance condition were more similar to each other than the gambles in the pair from low covariance condition.

Further, we fitted *regret theory*, *decision field theory* and the *proportional difference model* to the behavioral data using a maximum-likelihood approach. According to the Bayesian information criterion (BIC), all models performed better than a naïve baseline model assuming random choices, for the majority of participants (95% for decision field theory and regret theory and 65% for the proportional difference model). As displayed in Figure 8, on average all three models predict higher probability (Wilcoxon summed-rank test: $p < .005$ for decision field theory and regret theory, $p < .001$ for the proportional difference

[5]According to Kolmogorov-Smirnov test, the mean frequencies in each condition were not normally distributed.

[6]Parameter values were the same as in section 6.

model) in the high covariance condition as compared to the low covariance condition. Variances of predictions of decision field theory and regret theory were smaller in the high covariance condition than in the low covariance condition. It was the other way round for the proportional difference model. According to median and total BIC, decision field theory predicted the data best ($BIC = 1928.60$, $Me_{BIC} = 94.18$), followed by regret theory ($BIC = 2175.40$, $Me_{BIC} = 107.34$) and the proportional difference model ($BIC = 3228.20$, $Me_{BIC} = 163.03$)[7]. All three models make different predictions for the high and low covariance condition, given the same set of parameters used in both conditions. Despite its relatively bad fit in comparison to decision field theory and regret theory, the proportional difference model also predicts higher choice probabilities in the high than the in low covariance condition.

In sum, we replicated the results of Andraszewicz et al. (under review), such that the higher the association between the outcomes of the choice options, the higher the probability of choosing the gamble with the larger expected value. In their study, the strength of the association could be described equally well by the correlation and the standardized covariance. In the present study, we illustrated the usefulness of the standardized covariance measure. In situation with identical statistical correlation between the outcomes of the choice options, the standardized covariance, could be varied. In cases with the high standardized covariance, the choices became easier and the options with the larger expected value was chosen more often as compared to the low covariance condition. This effect was predicted by all three choice theories, in particular decision field theory which also described the data best.

[7]Estimated models' parameters, individual model fits and data are available online on http://psycho.unibas.ch/fakultaet/personen/profil/person/andraszewicz/

8. Discussion

In the current paper, we have shown that the strength in the association and similarity between risky choice options are related quantities that can be measured with the use of the *standardized covariance*. We have also shown that standardized covariance reflects how risky one choice option is relative to the second choice option. We call this property of choice options the *co-riskiness*. The standardized covariance and the correlation are related measures. However, correlation is a special case of association measure, which describes a linear relationship between choice options. As we have experimentally shown, people's choices between options were substantially affected by the differences in the standardized covariance. Likewise options with high standardized covariance were perceived as more similar than options with low standardized covariance besides similar levels of correlation values.

The proposed measure is related to theories and models of similarity. For example, when $\sigma^*_{AB} \to 1$, some attributes of both options are almost the same, whereas for others there is a substantial difference. Therefore, according to the similarity model (Rubinstein, 1988; Leland, 1994, 1998), attributes that are almost the same can be discarded and the option that is better for the remaining attributes should be chosen. Also, when attributes are almost the same, the distance function between them is very small and these attributes could be accounted as "shared" by two options (Tversky, 2004). Therefore, the standardized covariance provides an objective quantitative dimension of aspects of similarity that appeared in the previous literature.

In section 6, we showed that different models of decision making make different predictions depending on various levels of σ^*_{AB}, while ΔEV was constant. Therefore, in research involving decision making models, not controlling for the association between choice options may include additional noise in model predictions. Andraszewicz et al. (under review) showed that this noise reduction can be observed in a decrease of variability of human choices of option with higher σ^*_{AB}. The level of this noise depends on the model assumptions. For example, decision field theory and regret theory react very strongly to the association and similarity of choice options. In contrast, expected utility-based models (e.g. cumulative prospect theory, Kahneman & Tversky (1979)) would make the same predictions independently of the association between choice options because this group of models assumes that options are evaluated independently of each other.

The standardized covariance is not robust against using different measurement scales of two variables. However, this can be solved in the process of normalization of the variables' scales first. This solution makes the standardized covariance applicable in more domains than decision making research. Measurement of non-linear association, similarity and co-riskiness between two variables could be used in other fields of research. Also, the standardized covariance can be used as a measurement of association of any two variables with only two data points.

Interestingly, although the standardized covariance is not robust against the different measurement scales of the variables, the value of the standardized covariance would fall within the range $\mathbb{R} \cap [-1,1]$ (see proof in Appendix A), and for positively related variables the standardized covariance would never fall below 0, and above 0 for the negatively related variables. This is because the direction of the relation between variables is determined by the covariance. Parameter d_{AB} (see Equation 8) is robust against the scales inconsistency because the variances and covariance are related to each other in a non-linear fashion.

In sum, this work presents an easy-to-interpret measure that can describe important properties of choice options, namely the associations between the consequences of the options, similarity between them, and a relationship of riskiness of one option to the other. These properties are often neglected in the decision making literature. As a consequence, we describe how one could easily use this measure to characterize decision making situations to create more accurate experimental designs. The standardized covariance can be especially useful to measure association of options with only two outcomes (i.e. two-outcome gambles), for which the correlation measure does not provide a meaningful interpretation. Also, we have shown that in specific cases, standardized covariance can more effectively describe similarity between risky choice options than the correlation measure. Thus, the standardized covariance should be of special interest to the researchers in the field of judgment and decision making.

References

Andraszewicz, S., Rieskamp, J., & Scheibehenne, B. (under review). How associations between consequences of hoice options affects decision making under risk.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, *115*, 463–501. doi:10.1037/0033-295X.115.2.463.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*, 409–432. doi:10.1037/0033-295X.113.2.409.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459. doi:10.1037/0033-295X.100.3.432.

Cavagnaro, D. R., Gonzalez, R., Myung, J. I., & Pitt, M. A. (2013). Optimal decision stimuli for risky choice experiments: An adaptive approach. *Management Science*, *59*, 358–375. doi:10.1287/mnsc.1120.1558.

Disatnik, D., & Benninga, S. (2007). Shrinking the covariance matrix. *Journal of Portfolio Management*, *33*, 55–63. doi:10.3905/jpm.2007.690606.

González-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review*, *109*, 137–155. doi:10.1037//0033-295X.109.1.137.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–91. doi:10.2307/1914185.

Leland, J. W. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty*, *9*, 151–172. doi:10.1007/BF01064183.

Leland, J. W. (1998). Similarity judgments in choice under uncertainty: a reinterpretation of the predictions of regret theory. *Management Science*, *44*, 659–672. doi:10.1287/mnsc.44.5.659.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, *92*, 805–24. doi:10.2307/2232669.

Myung, J. I., & Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychological Review*, *116*, 499–518. doi:10.1037/a0016104.

Pafka, S., & Kondor, I. (2003). Noisy covariance matrices and portfolio optimization ii. *Physica A: Statistical Mechanics and its Applications*, *319*, 487–494. doi:10.1140/epjb/e20020153.

Pathan, A., Bonsall, P., & de Jong, G. (2011). How the choice of mode is conditioned by information source. In *International Choice Modelling Conference, Leeds, UK*.

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1446–1465. doi:10.1037/a0013646.

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*, 59–66. doi:http://www.jstor.org/stable/2685263.

Rubinstein, A. (1988). Similarity and decision-making under risk (is there a utility theory resolution to the allais paradox?). *Journal of Economic Theory*, *46*, 145–153. doi:10.1016/0022-0531(88)90154-8.

Sarin, R. K., & Weber, M. (1993). Risk-value models. *European Journal of Operational Research*, *70*, 135–149. doi:10.1016/0377-2217(93)90033-J.

Tversky, A. (1977). Intransitivity of preferences. *Psychological Review*, *84*, 327–352. doi:10.1037/0033-295X.84.4.327.

Tversky, A. (2004). *Preference, Belief and Similarity: Selected Writings*. (1st ed.). Cambridge, MA: MIT Press. (Chapters 1-6).

Zhang, S., & Lee, M. D. (2010). Optimal experimental design for a class of bandit problems. *Journal of Mathematical Psychology*, *54*, 499–508. doi:10.1016/j.jmp.2010.08.002.

Table 1: Eight examples of options with different standardized covariance. In each example, the top row indicates the probability of the occurrence of two outcomes. Two consecutive rows display the possible outcomes of option A and option B.

$\sigma^*_{AB} = 1$			$\sigma^*_{AB} = -1$		
Example 1	60%	40%	Example 2	60%	40%
A	80	55	A	80	55
B	80	55	B	55	80
$\sigma^*_{AB} = .80$			$\sigma^*_{AB} = -.80$		
Example 3	60%	40%	Example 4	60%	40%
A	80	55	A	80	55
B	80	30	B	30	80
$\sigma^*_{AB} = .80$			$\sigma^*_{AB} = 1$		
Example 5	40%	60%	Example 6	40%	60%
A	80	55	A	80	55
B	80	30	B	70	45
$\sigma^*_{AB} = .32$			$\sigma^*_{AB} = .05$		
Example 7	60%	40%	Example 8	60%	40%
A	80	20	A	42	40
B	50	40	B	80	6

Table 2: Examples of statistically independent pairs of two-outcome options. The percentages correspond to the probability of occurrence of outcomes listed for options A and B.

$S_{AB} = .86$			$S_{AB} = .79$		
Example 9	50%	50%	Example 10	50%	50%
A	10	30	A	10	30
	25%	75%		25%	75%
B	10	30	B	15	30
$S_{AB} = -.86$			$S_{AB} = .54$		
Example 11	50%	50%	Example 12	50%	50%
A	10	30	A	10	30
	25%	75%		25%	75%
B	30	10	B	14	22

Table 3: Ranges of values of the standardized covariance, ratio of the smaller to the larger variance and the amount of pairs of options generated for each of the seven differences between expected values of the options.

ΔEV	σ^*_{AB}	$\frac{min(\sigma^2_A,\sigma^2_B)}{max(\sigma^2_A,\sigma^2_B)}$	N
0	.02-1	.00-.96	570400
5	.02-.99	.00-.81	859052
10	.02-.99	.00-.72	975021
15	.02-.94	.00-.49	377750
20	.02-.94	.00-.49	418251
25	.02-.94	.00-.49	244734
30	.02-.89	.00-.36	119241

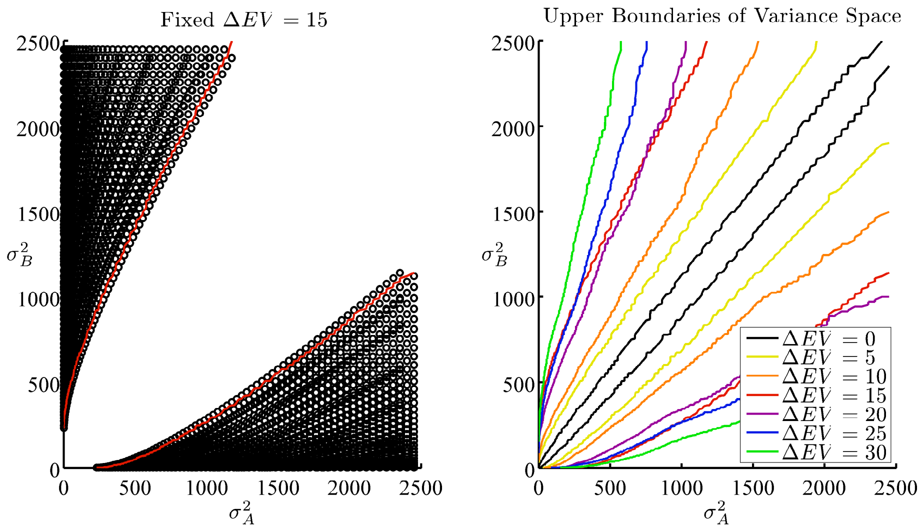


Figure 1: Left: Relationship between variances of two two-outcome options. Each point defines one pair of options ($N = 377750$). Empty space around the diagonal of the graph and symmetry of the distribution of both variances indicates a systematic relationship between variances and difference between expected values. Red lines correspond to upper boundaries of variances of each option in a pair; Right: Upper boundaries of variances for various differences between expected values. The smaller the difference between expected values, the smaller the gap between the lines of the same color. Therefore, when the difference between expected values decreases, the variances of two options can be either very similar or very different. When ΔEV is high, one option has substantially higher variance than the other option.

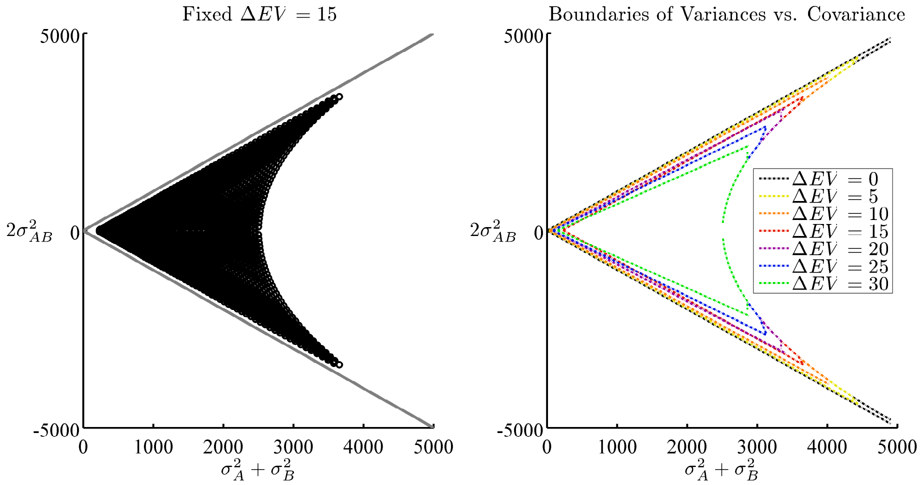


Figure 2: Left: Relationship of the sum of variances to twice the covariance. Black area shows the space of possible combinations of sum of variances with respect to covariance. The gray line indicates the cases for which the standardized covariance overlaps with the correlation measure, such that $\sigma^*_{AB} = r = 1$ or $\sigma^*_{AB} = r = -1$; Right: Boundaries of covariance and sum of variances depending on the difference between expected values. Therefore, the greater the ΔEV, the smaller the area of possible relationships of variances to covariance (i.e. the area inside the boundaries).

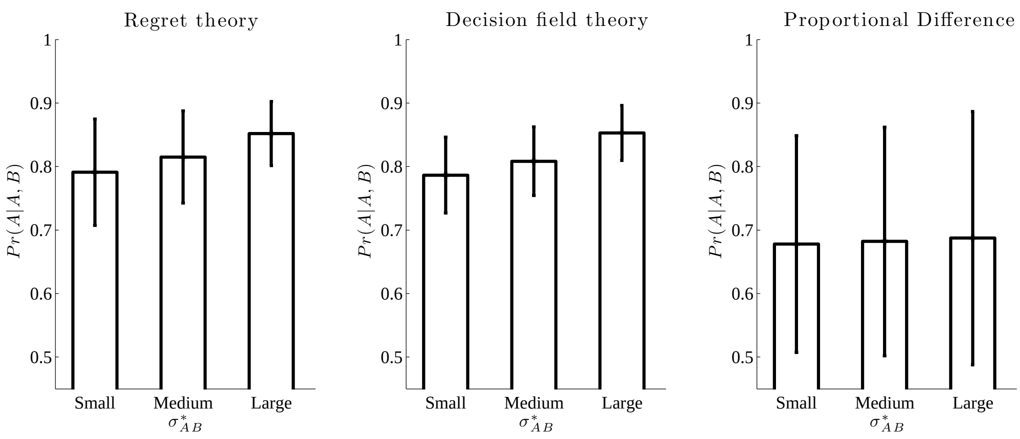


Figure 3: Average predictions of *regret theory* and *decision field theory* and the *proportional difference model*. To generate predictions the following parameters were used: *regret theory* $\beta = .05$, $\theta = 4.6$, *decision field theory* $\theta = 1.19$, *proportional difference model* $\gamma = .12$, $\sigma_{PD} = .63$. The parameter of *decision field theory* was based on Rieskamp (2008) and the parameters of *regret theory* and the *proportional difference model* were adjusted so that the predictions of all models are comparable (i.e. all predict $Pr(A|A,B) > .65$). Error bars indicate standard deviation.

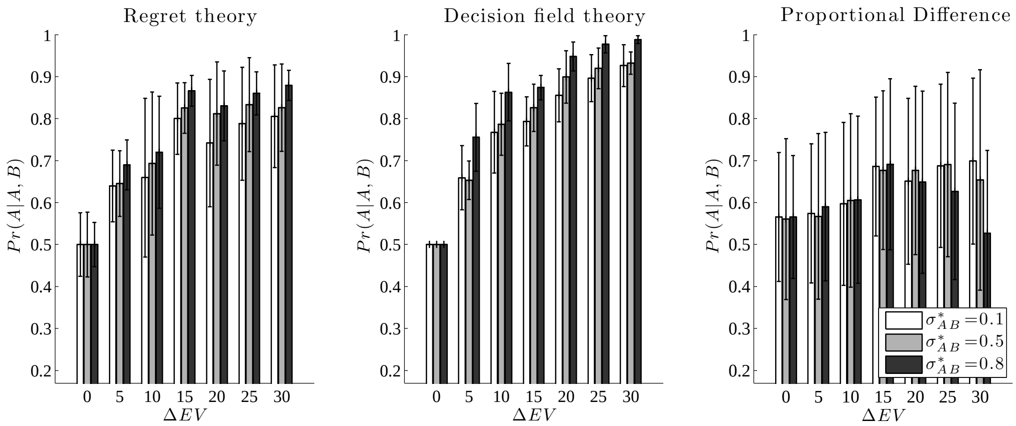


Figure 4: Average predictions of *regret theory*, *decision field theory* and the *proportional difference model* for pairs of options with three levels of σ^*_{AB} and various expected values. Error bars indicate standard deviation. The parameter values are the same as in Figure 3. The populations of pairs of options are described in section 5.1

.

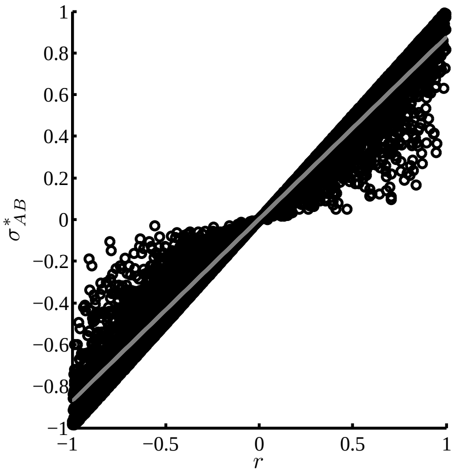


Figure 5: Relationship between the correlation coefficient and the standardized covariance between options with four outcomes. Each point corresponds to one of 10000 randomly generated pairs of stochastically non-dominant statistically dependent pairs of four-outcome options. Their outcomes range between 0 and 100 points and the corresponding probabilities range between 1% and 40%. The gray line indicates the regression line, with a slope of .87 and intercept .0027.

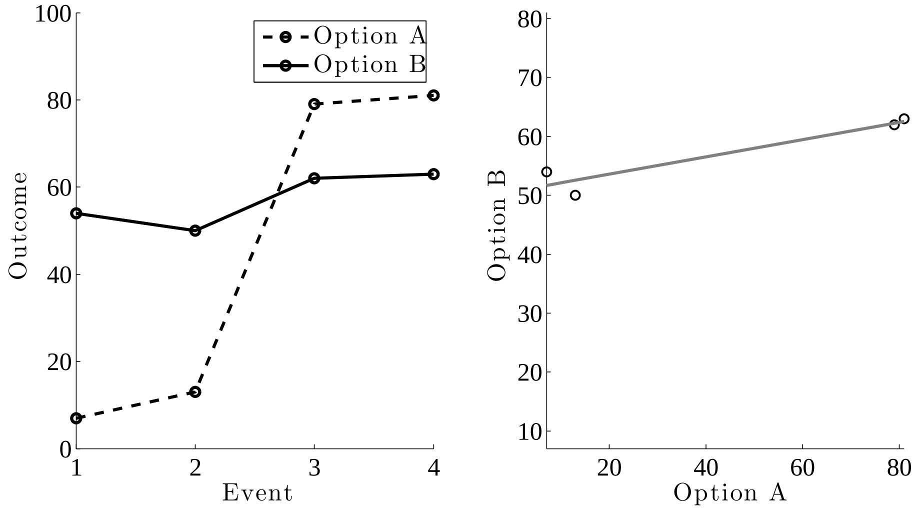


Figure 6: Left: Graphical representation of outcomes of options A and B for four independent events; Right: Scatter plot of outcomes of option A against option B. The gray line indicates the regression line. In both panels the same pair of options is shown, for which $r = .95$ and $\sigma^*_{AB} = .32$

	Yellow: 10%	Blue: 40%	Red: 30%	Green: 20%
Stock A	4	14	89	75
Stock B	26	30	35	38

	Yellow: 40%	Blue: 10%	Red: 30%	Green: 20%
Stock A	1	14	93	22
Stock B	27	70	91	20

Figure 7: Left: Example of a stimulus from the “low covariance” condition; Right: Example of a stimulus from the “high covariance” condition.

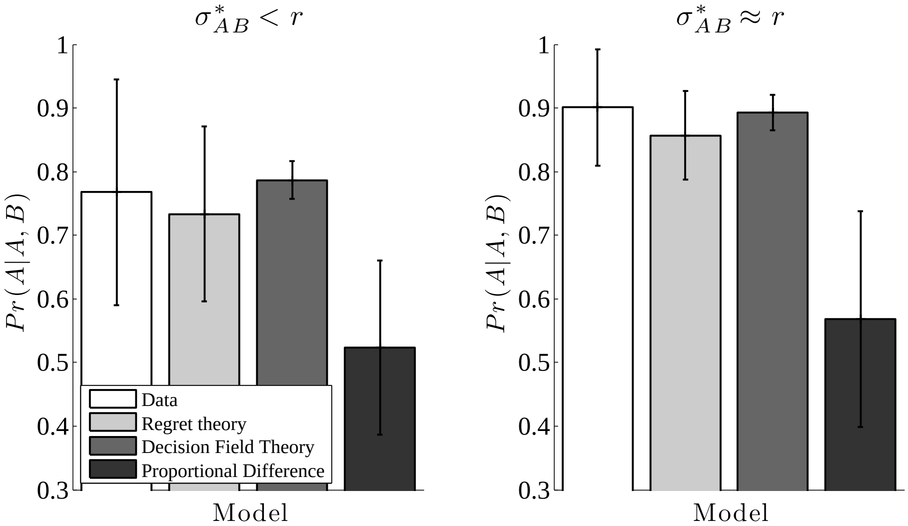


Figure 8: Average frequencies and predicted probabilities of choices of the gamble with the larger expected value (i.e. $E[A] > E[B]$) in the low covariace condition (left panel) and the high covariance condition (right panel). Probabilities correspond to predictions of three models: *regret theory*, *decision field theory* and the *proportional difference model*. Error bars correspond to bootstrapped standard error.

Appendix A: Mathematical Proof

Proof. $2\sigma_{AB} \leq (\sigma_A^2 + \sigma_B^2)$

From definitions of variance and covariance,

$2E\left[(a-\mu_A)(b-\mu_B)\right] \leq E\left[(a-\mu_A)^2\right] + E\left[(b-\mu_B)^2\right]$,

where a and b are outcomes of options A and B respectively, and μ_A and μ_B are their expected values. Given that both options have exactly I outcomes, with their corresponding probabilities p_i,

$0 \leq \Sigma_{i=1}^{I}\left[p_i(a_i-\mu_A)^2 + p_i(b_i-\mu_B)^2 - 2p_i(a_i-\mu_A)(b_i-\mu_B)\right]$

$0 \leq ((a_i-b_i)-(\mu_A-\mu_B))^2$

Let $(a_i-b_i)-(\mu_A-\mu_B) = m.$

Then $0 \leq m^2$.

$\lim_{m\to\pm\infty} m^2 = 0 \iff m \in \mathbb{R}$ ☐

Appendix B: Models Specification

Regret Theory

Following Pathan et al. (2011), we defined the regret function of choosing option A over option B with I outcomes a_i and b_i respectively as

$$R_{iA} = ln(1+exp(\beta(a_i - max(a_i,b_i)))). \tag{B.1}$$

The total regret from choosing option A equals to

$$R_A = \Sigma_{i=1}^{2} R_{iA}. \tag{B.2}$$

Further, the probability of choosing option A over option B is estimated using the softmax rule

$$Pr(\{A|A,B\}) = \frac{1}{1+exp(\theta(R_B - R_A))}. \tag{B.3}$$

β and θ are free parameters of the model. More details regarding regret theory is provided in Loomes & Sugden (1982).

Decision Field Theory

We implemented a parsimonious version of decision field theory (Busemeyer & Townsend, 1993). The expected value of option A is calculated as

$$E[A] = \sum_{i=1}^{N} = p_i a_i. \tag{B.4}$$

Decision field theory assumes that the decision maker compares two options with each other. The difference between options A and B is defined as

$$d_{DFT} = E[A] - E[B]. \tag{B.5}$$

This theory also assumes that the comparison of two options is a dynamic process, where the variance of the difference d_{DFT} defined as

$$\sigma^2_{DFT} = \sigma^2_A + \sigma^2_B - 2\sigma_{AB}. \tag{B.6}$$

σ^2_A and σ^2_B are variances of options A and B correspondingly, while σ_{AB} is the covariance of two options' outcomes. They are defined as $\sigma^2_A = E[(A - E[A])^2]$ and $\sigma_{AB} = E[(A - E[A])(B - E[B])]$.

The probability of choosing option A over option B equals

$$Pr(A|\{A,B\}) = \frac{1}{1 + exp[-\theta(2d_{DFT}/\sigma_{DFT})]}. \tag{B.7}$$

Proportional Difference Model

We implemented the proportional difference model according to (González-Vallejo, 2002). The proportional difference between the outcomes of options A and B corresponding to the same probabilities is defined as

$$\pi(a_i, b_i) = \frac{max\{|a_i|, |b_i|\} - min\{|a_i|, |b_i|\}}{max\{|a_i|, |b_i|\}}. \tag{B.8}$$

The decision threshold is defined as a difference between proportional differences such that

$$d_{PD} = \begin{cases} \pi(a_i, b_i) - \pi(a_j, b_j) & \iff a_i > b_i \\ \pi(a_j, b_j) - \pi(a_i, b_i) & \iff a_i < b_i \end{cases} \tag{B.9}$$

where i and j are two different events. The probability that a decision maker chooses option A over option B is defined by a cumulative normal distribution defined by the function

$$Pr(A|\{A,B\}) = f\left(\frac{d_{PD} - \delta}{\sigma_{PD}}\right). \tag{B.10}$$

δ and σ_{PD} are free parameters of the model, where δ represents how much a decision maker weights attribute differences. σ_{PD} is a variance of the trade-off process.

Appendix C: Description of Gambles in the Behavioral Experiment

In the behavioral experiment, we randomly generated four-outcome, stochastically non-dominant pairs of gambles with possible outcomes ranging between 0 and 100 points and their corresponding probabilities equal to either 10%, 20%, 20% or 40%. Gambles were displayed as hypothetical stocks, upper Stock A and lower Stock B. In each pair, gamble A had two better outcomes than gamble B. The order of the better outcomes was randomly chosen for each gamble from a predefined list. Outcomes corresponding to the same external events were marked with the same color and the order of the colors was the same for all gamble pairs. The difference between expected values of the gambles was constant, $\Delta EV = 15$. For 57% of gamble pairs, upper gamble A had a larger expected value than lower gamble B. For 54% of gamble pairs, the gamble with the larger expected value had the lower variance and for the remaining pairs it was the other way round. Half of the pairs of gambles were assigned to the "low covariance" condition, for which $Mean_r = .89, SD_r = .05$ and $Mean_{\sigma^*_{AB}} = .15, SD_{\sigma^*_{AB}} = .04$, and the other half to the "high covariance" condition, for which $Mean_r = .89, SD_r = .05$ and $Mean_{\sigma^*_{AB}} = .84, SD_{\sigma^*_{AB}} = .03$.

How Associations between the Consequences of Choice Options Affect Decisions Under Risk

Sandra Andraszewicz, Jörg Rieskamp, and Benjamin Scheibehenne

University of Basel

Department of Psychology

Basel, Switzerland

Author Note: Correspondence concerning this article should be addressed to Sandra Andraszewicz, University of Basel, Department of Psychology, Center for Economic Psychology, Missionsstrasse 62a, 4055 Basel. E-mail: s.andraszewicz@unibas.ch. This work was supported by a research grant from the Swiss National Science Foundation (SNSF 100014_130149) to the second and third authors.

Abstract

Many economic theories of decision making assume that people evaluate options independently of other available options. However, recent cognitive models suggest that people's evaluations rely on a comparison of the options' potential consequences with each other such that the subjective value of an option critically depends on the context in which it is presented. To test these opposing predictions, we examined pairwise choices between monetary gambles and varied the degree to which the gambles' outcomes covaried with one another. When people evaluate options based on comparisons of their consequences then a high covariance between the outcomes should make a decision easier. In line with this prediction, the observed choice proportions in two experiments (N = 39 and 24, respectively) were influenced by magnitude of the covariance. These results confirm that interdependent evaluations of options play an important role in human decision making under risk and show that covariance can quantitatively describe the choice context.

Keywords: decision making under risk, cognitive modeling, outcome associations, covariance, sequential sampling models

How Associations between the Consequences of Choice Options Affect Decisions Under Risk

Decisions under risk and uncertainty play an important role in daily life. Explaining and predicting risky decisions is an important area of research in psychology, economics, and cognitive science. Many cognitive models of decision making predict that people compare the potential consequences of options with each other, yielding context-dependent evaluations. For example, when choosing between health insurances, decision makers might compare the coverage for different illnesses against each other. However, standard economic theory of decision making such as *expected utility theory* (von Neumann & Morgenstern, 1974) assumes that the subjective value of a single offer is independent of its alternatives.

Theories of Decision Making

Previous literature has shown that people compare options' outcomes with each other, rather than evaluating each option independently of other available options (cf. Rieskamp, Busemeyer, & Mellers, 2006). For example, Tversky and Shafir (1992) found that people's preferences depend on the context of other choice options in which an option is presented. According to Tversky's (1969) *lexicographic semiorder heuristic*, a person choosing between two health insurances might order attributes (i.e. coverage in case of different illnesses) of the two insurance offers according to their importance for the person. If one of the offers is substantially better than the other for the most important attribute, this offer should be chosen. If both offers are comparable for the most important attribute, the decision should be based on the second best attribute, and so on.

Context-dependent decision theories, such as Tversky's (1969) *lexicographic semiorder heuristic*, Tversky's (1972) *elimination-by-aspects* theory, Gonzàlez-Vallejo's (2002) *stochastic difference model*, or Brandstätter, Gigerenzer, and Hertwig's (2006) *priority heuristic*, predict that options are compared relative to each other. A further example is *regret theory* (RT, Loomes & Sugden, 1982), which assumes that decision makers anticipate feelings of regret when obtaining lower outcomes relative to forgone outcomes of alternative options. The association between choice options is represented by the regret utility function that expresses the summed regret of each attribute of an option compared to the regret of the forgone option.

Another distinguished theory is *decision field theory* (DFT, Busemeyer & Townsend, 1993), which assumes that preferences accumulate over time by comparing the outcomes of the options with each other one at a time and by accumulating the differences between the outcomes. The attention to the different outcomes shifts stochastically. A decision is reached when the accumulated differences exceed a predefined threshold, or when a predefined time limit has been reached. Both RT and DFT will be of special interest in the current work because they not only provide two possible mechanisms of the decision process, but also allow quantitative predictions. RT and DFT directly stress the importance of the association between choice options.

In contrast, the class of *fixed-utility theories* assumes that each option can be assigned a value representing the option's subjective value to the decision maker and this value is independent of other options in the choice set. Standard *expected utility theory* (EU) represents the most eminent fixed-utility theory. However, similar to EU theory, the assumption of independence can also be traced in more descriptively inspired decision theories such as *rank dependent utility* theories (see e.g., Green & Jullien, 1989; Luce, 1990). A prominent example, *cumulative prospect theory* (CPT, Tversky & Kahneman, 1992), distinguishes gains and losses and includes a weighting function to represent subjective probabilities, but in the end still assigns a context-independent value to each option.

The idea of context-dependent evaluations can be illustrated with a choice between two monetary gambles whose outcomes depend on the throw of a die, as shown in Figure 1.

In case 1, gamble A leads to substantially higher payoffs than gamble B if the die lands on 1, 2, or 3, whereas for the numbers 4 to 6 gamble B has a small advantage over A. Presumably, most people would prefer gamble A over B due to the large advantage of A for numbers 1-3 and disregard the small disadvantage for numbers 4-6. In contrast, in case 2, gamble A is worse than B* when throwing 1, 2, or 3, whereas it is better when throwing 4, 5 or 6, making the choice much more complicated. However, gamble B and B* are identical as they both result in the same outcomes with identical probabilities. Thus, EU and CPT cannot predict any difference for the two choice situations, whereas context-dependent theories can.

Research Goal

Past research has provided substantial evidence that the context in which the choice options are presented influences people's preferences. Nevertheless, so far only a few attempts have been made to quantify the relevant characteristics of the choice environment that govern these changes. For example, Mellers and Cooke (1994), and Mellers, Schwartz and Cooke (1998) discussed the influence of other available options. Other attempts to incorporate possible relations between attributes of choice options include extensions of the *probit model* in the area of consumer choice (e.g. Train, 2009). Also, Tversky (1977) discussed the influence of similarity between two items such as countries, faces, letters and shapes, on the choice context. However, to our knowledge, little work has been done to quantitatively define the context and its' influences on decision making, with respect to monetary gambles.

In the present work, we make the novel contribution of quantifying the influence of an important aspect of the choice context, namely the covariance between the choice outcomes, and we analyze how the size of the covariance affects people's preferences. In particular, we will 1) quantify the strength of the associations between options' outcomes, 2) examine by simulation how the predictions of different decision models are affected by the strengths of the associations 3) examine on empirical grounds how different association strengths actually influence people's decisions, and 4) rigorously compare models of decision making against each other on qualitative and quantitative basis, to see what cognitive processes could explain context-dependent choice comparison. In sum, our work should lead to a better understanding of the cognitive process underlying decisions under risk.

Characterizing the Association Between Choice Options

Figure 1 illustrates how different degrees of covariance may influence people's choices between gambles. In case 1, the covariance of 78.7 is much higher than in case 2 with 24.7. However, the main drawback of covariance as an association measure is that its scale depends on the range of the outcomes. Using correlations instead also does not provide a feasible solution because for gambles with only two outcomes, it equals either -1 or 1 (see Rodgers & Nicewander, 1988). As an alternative scaling, we use a standardized covariance σ^*_{AB} (Andraszewicz & Rieskamp, 2013, in preparation), which equals twice the original covariance σ_{AB} divided by the two gambles' sum of variances:

$$\sigma^*_{AB} = \frac{2\sigma_{AB}}{\sigma^2_A + \sigma^2_B} \tag{1}$$

σ^*_{AB} is a continuous variable that ranges from 1 (maximum positive association) to -1 (maximum negative association), whereas $\sigma^*_{AB} \to 0$ characterizes options with low association or low internal variability. When $\sigma^*_{AB} = 0$, the options are statistically independent of each other or one option is a sure option whose variance is 0. Appendix A and Andraszewicz and Rieskamp (2013, in preparation) provide an elaborated description of the standardized covariance's properties.

Figure 2 plots the variances of two gambles against their covariance and shows that σ^*_{AB} is high when the covariance between the options is high and both variances are high (upper right corner of Figure 2). In this case, each gamble has both a large advantageous outcome and a very disadvantageous outcome. The advantageous outcomes of both options occur at the same event with the same probability, and so do the disadvantageous outcomes. Outcomes of options with medium σ^*_{AB} (center of Figure 2) vary less compared to gambles with high σ^*_{AB}. Options with low σ^*_{AB} have low covariance (lower left corner of Figure 2). This means that the outcomes of one gamble are very similar and the other gamble has one lower and one higher outcome.

The idea of incorporating both the covariance and variances of two choice options in one measure follows from the normalization of covariance, in computing the correlation measure, by division of the product of variances. Variances and covariance are non-linearly related measures, such that it is impossible to manipulate one while keeping the other constant. Therefore, covariance carries some information about variances that cannot be analytically classified. As a consequence, a normalized measure of covariance relative to the variances is needed, to effectively quantify the strength of the association between choice options.

Associations Between Options' Outcomes Affect the Predictions of Decision Theories

To explore how choice probabilities are influenced by the association between the consequences of the choice options as measured by the standardized covariance, we ran a simulation study. This simulation study also allows us to test possible differences in model predictions, depending on the standardized covariance. It is difficult to analytically derive the exact impact of the standardized covariance on the models' predictions. We chose RT and DFT to represent two *context-dependent decision theories.* As a comparison benchmark, we include EU and CPT to represent the *fixed-utility theories*, whose predictions should not be influenced by standardized covariance. Appendix B provides mathematical specifications of the respective models. All models make a probabilistic prediction that a decision maker chooses one gamble over another one in a pair.

Materials

We used pairs of stochastically non-dominant gambles whose outcomes varied between 0 and 100 integer points with outcome probabilities equal to either .4, .5, or .6. The absolute difference in expected value (ΔEV) within each pair of gambles was kept constant at 15 points. Within these bounds, we created all 377,750 possible gamble pairs. We grouped each pair of gambles into one of three categories indicating small ($\sigma^*_{AB} \leq 0.2$, 21.2% cases), medium ($0.2 < \sigma^*_{AB} \leq 0.5$, 34.7% cases), or large ($0.5 < \sigma^*_{AB}$, 44.1% cases) σ^*_{AB} . Finally, we converted all the outcomes from gains to losses to create the set of negative gambles.

Results and Discussion

Results of the simulation indicate that σ^*_{AB} had a strong effect on the average predictions of DFT and RT, whereas, as expected, no effect was observed for EU and CPT. As shown in Figure 3, for DFT and RT the predicted choice shares of the gambles with the larger expected value increased with standardized covariance. Presumably, higher σ^*_{AB} makes the comparison process easier. This positive relation is systematic and non-linear, such that there is a greater difference between large and medium covariance conditions, than between medium and small.

Interestingly, the standardized covariance also affects the variability of the models' predictions. Figure 3 shows that the larger σ^*_{AB} , the smaller is the variability of DFT's and RT's predictions, whereas again no effect was observed for EU and CPT. Thus, according to DFT and RT, choices with small σ^*_{AB} should be more difficult and the models predict less consistent choices; that is, larger variability in the predicted choice probabilities. Across the whole range of gambles, RT predicts more variability than DFT.

Study 1

The simulation illustrates that the two context-dependent theories, DFT and RT, but not EU and CPT, predict that a stronger association between two choice options leads to a stronger preference for one option. In Study 1, we tested whether the systematic differences observed in the simulation are reflected by observed human choice behavior.

Participants

A total of 39 people (16 male), aged 19-52 years (M_{age} = 25 years), mainly students from the University of Basel, participated in the study. Four participants were excluded from further analyses because they made purely random choices, which resulted in data outliers.

Material

We randomly selected three times 60 pairs of two-outcome gambles from the simulation study for three sets with small, medium, and large positive covariances, with the following constraints: half of the gambles had only positive outcomes and the other half only negative (with all outcomes smaller than 0). Gambles were randomly assigned as the upper gamble A or the lower gamble B on the screen, so that in 53% of all pairs of gambles, gamble A had a larger expected value. In the simulation, we used the whole population of gambles with the predefined described properties. In Study 1, we narrowed down the values of σ^*_{AB} ranges to obtain a clear-cut distinction among small, medium and large σ^*_{AB}, given the limited amount of stimuli we could present to the participants. Therefore, 60 pairs of gambles had a standardized covariance such that $\sigma^*_{AB} \le .1$, another 60 pairs $.4 \le \sigma^*_{AB} \le .5$, and a third set $.8 \le \sigma^*_{AB} \le .95$. For half of the gamble pairs, the outcomes of the gamble with the larger expected value varied less, whereas in the remaining half of pairs it was the other way round.

Procedure

In the main part of the experiment, participants repeatedly chose from each pair of gambles on a computer screen presented in a random order, as shown in Figure 4. Gambles were framed as hypothetical stocks with outcomes representing the return on investment along the corresponding probabilities of occurrence. The outcomes were matched by color with their corresponding probabilities and the same colors were used for both gambles to indicate that the outcomes of both were dependent on the same external event. At the end of the experiment, one gamble was randomly chosen and played out. Two percent of the gamble's outcome was added to or subtracted from the initial endowment of 8 Swiss Francs (≈ 8 USD). Participants were informed about the payment procedure before they started the experiment. The experiment was completely self-paced.

The main part of the experiment was preceded by six practice trials presented in a fixed order. During the practice trials, the gambles were played out and the participants saw the results. After the experiment, the participants were asked to fill out a short questionnaire regarding their demographic data.

Results and Discussion

Figure 5 shows that the observed choice proportions were systematically influenced by the strength of the association between the gambles: The share of choices of gambles with the larger expected value systematically increased in a non-linear fashion when σ^*_{AB} increased. We computed for each participant an average proportion of choices of the gamble with the higher expected value, separately for the three conditions. There were significant differences in the choice proportions, $p < .001$, χ^2 (2) = 15.48, CI = 95% according to a Friedman's test[1]. We applied a series of post-hoc paired-comparisons using one-sided Wilcoxon signed-rank tests with Bonferroni correction of α. There was a significant difference between the large and small ($p < .001$), and large and medium ($p < .001$) covariance conditions, but no differences between the medium and small covariance

[1] The data untransformed and transformed were not normally distributed, thus we relied on non-parametric tests.

conditions (p = .18). There was a greater difference between medium and large conditions than between medium and small, which confirms the relation found in the simulation study.

To compare the predictions of EU, CPT, RT and DFT we estimated their parameters using maximum log-likelihood approach. Table C1 in Appendix C presents individual parameter estimates and model fits. We compared the models using their Bayesian Information Criterion (BIC, Kaas & Raftery, 1995), which takes the models' fit and complexity into account. Table 1 shows that on average all models describe the observed decisions substantially better than a baseline model[2]. According to BIC, DFT provides the best description of the observed data, followed by RT, EU, and CPT (see Table 1). EU predicted the behavior relatively well. This result seems mainly due to EU's small number of parameters and the ability to account for people's risk attitudes with a corresponding utility function[3]. When comparing DFT with EU on an individual level by their BICs, 74% of the participants are better described by DFT. Figure 6 shows the corresponding Bayes factors (Kaas & Raftery, 1995) when comparing DFT with EU for each participant separated for the three covariance conditions. In each condition, the majority of participants are better described by DFT than EU and the evidence is strong or very strong (compare upright to downright bars in Figure 6).

Additionally, we tested the models' predictions (using the models' estimated parameters) on a qualitative level by comparing them with the observed data. We averaged the predictions generated for each participant across each of the 180 pairs of gambles. RT predicts increasing choice proportions for the three covariance conditions, whereas DFT predicts a stronger increase between the medium and large covariance conditions and almost

[2] Baseline model predicts all choices with the probability of .5.
[3] Thus, we implemented DFT with the utility function defined the same as in the expected utility theory (Equation B2). The fit measured with the log-likelihood function improved; however, BIC was worse than for the DFT with the utility function defined as $u(x) = x$, indicating that the additional parameter α did not improve the model's predictive power.

no differences between the small and medium covariance condition, which is consistent with the observed choice patterns (Figure 7).

Study 2

The association between choice options might systematically influence people's preferences in a non-linear fashion for very simple choice problems such as gambles with two outcomes. Study 2 examines whether the findings from Study 1 can be generalized to more complex gambles with more than two outcomes. Also, by using gambles with more than two outcomes, we are able to compare σ^*_{AB} with the correlation measure.

Participants

Twenty-four people (7 male), aged between 19 and 52 years (M_{age} = 28 years), recruited through the participant database at the University of Basel, participated in the study.

Materials and Procedure

The gambles were again presented in pairs. Each gamble had four possible outcomes that varied between -100 and 100 points, with outcome probabilities of .1, .2, .3, or .4. In each gamble pair, gamble A had two outcomes that were higher and two outcomes that were lower than the respective outcome of gamble B. The order of the better outcomes was randomized. Due to the large number of possible combinations of gambles, we randomly generated exactly 180 different pairs of gambles with the properties described here. Ninety pairs had positive outcomes, while the other 90 pairs had only negative outcomes. Within each of these, there were 30 gambles with small ($\sigma^*_{AB} \leq .1$), 30 with medium ($.4 \leq \sigma^*_{AB} \leq .5$), and 30 with large ($.8 \leq \sigma^*_{AB} \leq .95$) standardized covariance. For 53% of the gamble pairs, the outcomes of the gamble with the larger expected value varied less, whereas for the remaining 43%, it was the other way round. The procedure was exactly the same as in Study 1. The base payment for participants was 15 Swiss Francs, with the bonus equal to 20% of the outcome of a randomly chosen and played out gamble.

To test the validity of the standardized covariance, we computed the correlation coefficient of the gamble pairs. The mean correlation coefficient in each of the three groups of gambles was similar to the standardized covariance, such that small $M_{r_{small}} = .06(SD = .05)$, $M_{r_{medium}} = .56(SD = .14)$, $M_{r_{large}} = .88(SD = .06)$. Only six pairs of gambles in the medium condition were correlated with the strength $r > .8$. The correlation coefficients of pairs in the small and large conditions did not exceed the ranges of standardized covariance defined for each group. The two measures were very strongly correlated to each other ($r = .98$, $p < .001$). Thus, the standardized covariance is a very similar measure to the correlation.

Results and Discussion

The larger σ^*_{AB} between the outcomes of the gambles, the higher the choice proportions of the option with the larger expected value (see Figure 8), which replicates the results from Study 1. People's preferences differed among the three conditions, as indicated by a Friedman's test ($p < .001$, χ^2 (2) = 31.32). A series of Wilcoxon signed-rank tests with Bonferroni correction (CI=95%) indicated significant differences among all covariance conditions ($p < .001$ for all comparisons)[4]. The variance of these people's choices systematically decreased when σ^*_{AB} increased.

To compare the models' ability to predict the observed choices we again estimated the models' parameters (see Table C2 of Appendix C). We compared parameters estimated for the first and second experiment with Wilcoxon rank sum tests (the parameters were not normally distributed). Parameters for all models apart from CPT did not differ between the two experiments[5]. According to BIC, all models predicted the data better than the baseline model, but only DFT did so for all participants (see Table 2). Similar to Study 1, DFT was

[4] We relied on non-parametric tests, because neither the untransformed nor the log-, reciprocal- and square-root-transformed data were normally distributed.
[5] According to the estimated parameters of CPT, participants in Study 1 were risk-averse, whereas participants in Study 2 were risk-seeking

the best model to predict the data, followed by EU. For 79% of the participants the BF was in favor of DFT compared to EU. Figure 9 shows this advantage separately for each condition.

Again, only, RT and DFT correctly predict the increased choice proportions for gambles with larger σ^*_{AB}. However, RT overall predicts less extreme choice probabilities and greater variance of these probabilities than observed in the data. This is due to the averaging of the predictions across the participants. In contrast, DFT predicts the observed differences between the conditions and observed choice proportions more accurately. Similar to Study 1, there is a bigger difference between medium and large conditions than between small and medium. This finding holds for observed choice proportions and RT and DFT predictions. However, in Study 2, the difference between small and medium conditions was more similar to the difference between medium and large conditions, as compared to Study 1. Importantly, RT's and DFT's predictions matched this pattern.

General Discussion

The present work examines how much the association between choice options influences people's preferences. To characterize the association between the consequences of options, we used the standardized covariance as an easy-to-interpret measure. We showed that this measure systematically influences predictions of context-dependent models consistent with people's decisions under risk. To explain the specific choice effects, we tested different context-dependent choice models against fixed-utility models. The four models were rigorously tested against each other. In two behavioral experiments, *decision field theory* predicted the data best.

Our results provide clear evidence that the context in which options are presented, can be quantitatively described by the strength of the association between choice options. The larger the association, the higher the chances that people choose the option with the larger expected value. To characterize this association, we used the standardized covariance. It is a ratio of how much two choice options co-vary with each other to how risky each of them is, where risk is defined by variances of the options' outcomes. This measure can readily be applied for options with only two outcomes where the conventional correlation measure does not lead to meaningful descriptions. When examining options with more than two outcomes, the standardized covariance highly overlaps with the correlation measure.

The standardized covariance and, if applicable, the correlation apparently provide a feasible measurement that quantifies the difficulty of the choice. If the standardized covariance is large, the choice gets easier, because when comparing the outcomes of the choice options they all point in the same direction. In contrast, if the standardized covariance is small, comparing the different outcomes yields conflicting information. Here, people have to make real trade-offs and identifying the option that is most preferable to them gets more difficult. For example, when choosing between investments in two stocks with similar price listings over the last month (indicating high correlation), it is easier to indicate the one whose ratings increased more over the last week than if the ratings of the stocks are completely uncorrelated.

Theory Based Explanation of Results

How can the observed effects of associations between the consequences of choice options be explained? Decision field theory assumes that people while comparing options accumulate differences between the outcomes over time until a decision threshold is reached. The smaller the variance of the differences the more likely that the threshold for the option with the larger expected value will be passed. This variance of the difference incorporates variances (riskiness) of both choice options and the covariance (association) between them. Busemeyer and Towsend (1993, p. 439) noted that increasing the similarity, expressed by the covariance between outcomes, makes the better choice option easier to discriminate. They name the valence difference divided by the variance of this difference the "discriminability ratio”. The closed form representation of DFT reflects the process of accumulation of evidence by this discriminability ratio, such that the smaller the discriminability ratio, the less evidence is accumulated. This ratio is multiplied by the decision threshold, which reflects how much information has to be accumulated for the decision to be made.

Our study has shown that this discriminability ratio can reflect the choice difficulty. Because in this paper we kept the valence difference constant, the discriminability ratio depended only on the variance of the differences. When the variance of the difference is large, the choice is difficult and the decision maker accumulates only little evidence. In contrast, when this variance is low, the choice is easy and sufficient evidence is accumulated in favor of the preferred option.

Regret theory (Loomes & Sugden, 1982) incorporates the variance of the valence difference in the regret function. When for some events two choice options are similar, the regret of not having chosen the slightly better option is very low, whereas it is high for the events for which the options differ substantially. When two options are dissimilar for all events, the total regret will be large, which makes the choice difficult.

DFT and regret theory are not the only theories that can explain how the covariance of the options’ outcomes can affect people’s preferences. For example, the *similarity model* described by Rubinstein (1988) or Leland (1994, 1998) provides an alternative account. The similarity model assumes that attributes for which the two choice options have similar values are disregarded when comparing the options with each other. Therefore the decision is then based on the attributes for which one of the options has a clear advantage.

Gonzàllez-Vallejo’s (2002) *proportional difference model* proposes that options are compared attribute-wise and for each attribute the proportional differences between options are accumulated to favor one or the other option (see also Scheibehenne, Rieskamp, & Gonzàllez-Vallejo, 2002). Analogically, Payne, Bettman, and Johnson (1988) investigated a series of decision strategies, including *lexicographic rules*, assuming a comparison of the options' outcomes with each other (see Rieskamp & Hoffrage, 1999; 2008).

The *transfer of attention exchange* (TAX) model (see Birnbaum, Patton & Lott, 1999; Birnbaum, 2008) proposes that more important attributes should receive more attention. Attention switches between the two options. If one attribute receives significantly more attention than others and option A is worse than option B for this attribute, the advantage of option A with respect to option B for the remaining attributes would be ignored. This is because the remaining attributes would not receive sufficient attention. Therefore, this model could also account for the effect of covaraiance with the assumption that attention is shifted to the attributes for which the options differ the most.

In contrast, *fixed utility theories* including *expected utility theory* and *cumulative prospect theory* that assume independent option evaluations, cannot explain why people's preferences differ for situation with different associations of the choice outcomes. However, we might speculate whether one could modify the standard EU and CPT approach by adding an error model to the theory. Accordingly one could specify a choice rule that determines the choice probabilities by taking the covariance into account (see Appendix B). Despite the fact that such an amendment might improve the prediction of the theory, it lacks the psychological explanation that RT and DFT provide. Nevertheless we also estimated EU model with a choice rule that included the error σ_d defined as in Equation B14.

Here, the only mathematical difference between EU with the error model and DFT was that EU defined the difference between utilities rather than between the valences. EU with the error model successfully predicted the pattern observed in the data and improved log-likelihood as compared to DFT. However, the fit measured with the BIC was worse than of DFT and the estimated α parameters in both experiments oscillated around 1 for most of the participants, indicating that the variance of the difference already accounts for both the association and riskiness of the options.

In sum, there is a large body of research showing that people make choices by comparing options against each other. The present work shows that the associations between the options' outcomes, as measured by the standardized covariance, affect people's decision. The larger the covariance, the easier the choice becomes. The effect can be explained by various theories assuming interdependent evaluations of choice options. When the association is high, the options are more similar to each other and the choice becomes easier. When the association is low, the choice becomes more difficult, which introduces high noise levels in the decision process. DFT and RT propose different cognitive explanations for introducing the noise. We show that the covariance of the options' outcomes provides a useful description of the choice context as it defines the difficulty of the choice situation. The strength of the association has a systematic, non-linear influence on people's preferences.

References

Andraszewicz, S., Rieskamp, J. (2013). Standardized covariance – A new measure of association, similarity and co-riskiness of choice options. (in preparation)

Birnbaum, M.H. (2008). New paradoxes of Risky decision making. *Psychological Review, 115*(2), 463-501. doi: 10.1037/0033-295X.115.2.463

Birnbaum, M.H., Patton, J.N., Lott, M.K. (1999). Evidence against rank-dependent utility theories: Tests of cumulative independence, interval independence, stochastic dominance, and transitivity. *Organizational Behavior and Human Decision Process, 77*(1), 44-83. doi: http://dx.doi.org/10.1006/obhd.1998.2816

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review, 113*(2), 409-432. doi:10.1037/0033-295X.113.2.409

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic cognition approach to decision making. *Psychological Review*, *100*(3), 432-459. doi:10.1037/0033-295X.100.3.432

Gonzàlez-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review*, *109*(1), 137-155. doi:10.1037//0033-295X.109.1.137

Green, J. R., & Jullien, B. (1989). Ordinal independence in nonlinear utility theory. *Journal of Risk & Uncertainty*, *1*, 355-387. doi:10.1007/BF00055713

Harrison, G. W., & Ruström, E. E. (2009). Expected utility theory and prospect theory: One wedding and a decent funeral. *Experimental Economics, 12*(2), 133-158. doi:10.1007//s10683-008-9203-7

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430)*,* 773-795. doi:10.1.1.19.6821

Leland, J. W. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk & Uncertainty*, *9*(2), 151-172. doi:10.1007/BF01064183

Leland, J. W. (1998). Similarity judgments in choice under uncertainty: A reinterpretation of the predictions of regret theory. *Management Science*, *44*(5), 659-672. doi:10.1287/mnsc.44.5.659

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, *92*(368), 805-824. doi:10.2307/2232669

Luce, R. D. (1990). Rational versus plausible accounting equivalents in preference judgments. *Psychological Science*, *1*(4), 225-234. doi:10.1111/j.1467-9280.1990.tb00205.x

Mellers, B.A., Cooke, A.D.J. (1994). Trade-offs depend on attribute range. *Journal of Experimental Psychology. 20*(5), 1055-1067, doi: 10.1037/0096-1523.20.5.1055

Mellers, B.A., Schwartz, A., Cooke, A.D.J. (1998). Judgement and decision making. *Annual Review of Psychology. 49*(1), 447-477, doi: 10.1146/annurev.psych.49.1.447

Pathan, A. F. H., Bonsall, P., & de Jong, G. (2011, June). How the choice of mode is conditioned by information source. In S. Hess (Chair), *Plenary session IIIA(4)*. Symposium Conducted at the International Choice Modelling Conference. Leeds, UK.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 14*(3), 534-552. doi:10.1037/0278-7393.14.3.534

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1446-1465. doi:10.1037/a0013646

Rieskamp, J., Busemeyer, J. R., & Mellers, B. (2006). Extending the bounds of rationality:

Evidence and theories of preferential choice. *Journal of Economic Literature*, *44*(3), 631-661. doi:10.1257/jel.44.3.63

Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In: G. Gigerenzer & P. M. Todd (Eds.), *Simple heuristics that make us smart* (pp. 141-167). New York: Oxford University Press.

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, *127*(2), 258-276. doi: http://dx.doi.org/10.1016/j.actpsy.2007.05.004

Rubinstein, A. (1988). Similarity and decision-making under risk (Is there a utility theory resolution to the Allais paradox?). *Journal of Economic Theory*, *40*(1), 145-153. doi:10.1016/0022-0531(88)90154-8

Scheibehenne, B., Rieskamp, J., & González-Vallejo, C. (2009). Models of preferential choice: Comparing the decision field theory with the proportional difference model. *Cognitive Science, 33*(5), 911-939. doi:10.1111/j.1551-6709.2009.01034.x.

Train, K. (2009). *Discrete choice methods with simulation*. New York: Cambridge University Press.

Tversky, A. (1969). The intransitivity of preferences. *Psychological Review. 79*(1), 281-299, doi: 10.1037/h0026750

Tversky, A. (1972). Choice by elimination. *Journal of Mathematical Psychology. 9*(4), 341-367. doi:10.1060/0022-2496(72)90011-9

Tversky, A. (1977). Features of similarity. *Psychological Review. 84*(1), 327-352, doi: 10.1037/0033-295X.84.4.327

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297-323. doi:10.1007/BF00122574

Tversky, A., & Shafir, E. (1992). The Disjunction effect in choice under uncertainty. *Psychological Science, 3*(5), 305-209, doi: 10.1111/j.1467-9280.1992.tb00678.x

von Neumann, J., & Morgenstern, O. (1947). *Theory of economic games and economic behavior.* Princeton, N.J.: Princeton University Press.

Table 1

Estimated parameter values (standard deviations) and results of model comparison (BIC) for each of the four models.[6]

Model	*Median Parameters (SD)*	*BIC*	*Individual BIC<Baseline*
Baseline	*None*	5489.7	-
EU	α=.90 (.25)	4897.3	89%
	θ= .24 (.70)		
CPT	α=.82 (.38)	5361.8	83%
	β=.87 (.52)		
	γ=.89 (.37)		
	δ=.42 (.41)		
	ϕ=.27 (1.8)		
RT	β=.04 (.05)	4881.4	89%
	θ=7.16 (10.60)		
DFT	θ= 1.48 (.92)	4702.9	91%

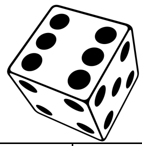
[6] EU: expected utility theory, CPT: cumulative prospect theory, RT: regret theory, DFT: decision field theory

Table 2

Estimated parameter values (standard deviation) and results of model comparison (BIC) for each of the four models.

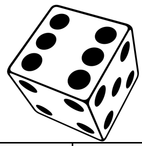
Model	*Median Parameters (SD)*	*BIC*	*BIC<Baseline*
Baseline	*None*	5988.8	-
EU	α=.78 (.37) θ= .42 (1.16)	2865.9	96%
CPT	α=1.29 (.4) β=1.35 (.36) γ=1.0 (.26) δ=1.0 (.25) ϕ=.04 (1.03)	3616.3	92%
RT	β=.04 (.03) θ=5.66 (2.75)	3502.4	92%
DFT	θ= 2.00 (1.14)	2856.7	100%

Case 1)



Event	*1*	*2*	*3*	*4*	*5*	*6*
Gamble A	20	10	25	30	15	30
A - B	5	5	5	-1	-1	-1
Gamble B	15	5	20	31	16	31

Case 2)



Event	*1*	*2*	*3*	*4*	*5*	*6*
Gamble A	20	10	25	30	15	30
A - B	-11	-6	-6	15	10	10
Gamble B*	31	16	31	15	5	20

Figure 1. Choice between two monetary gambles A and B whose outcomes depend on the throw of a die. In case 1, the gambles appear similar and in case 2, they appear dissimilar. Gamble B is identical to gamble B*, with the only difference that the outcomes have been swapped for the events 1-3 and 4-6. For both pairs of gambles, the expected value difference is 4, but the covariance of the outcomes is 78.67 for the first and 24.67 for the second case.

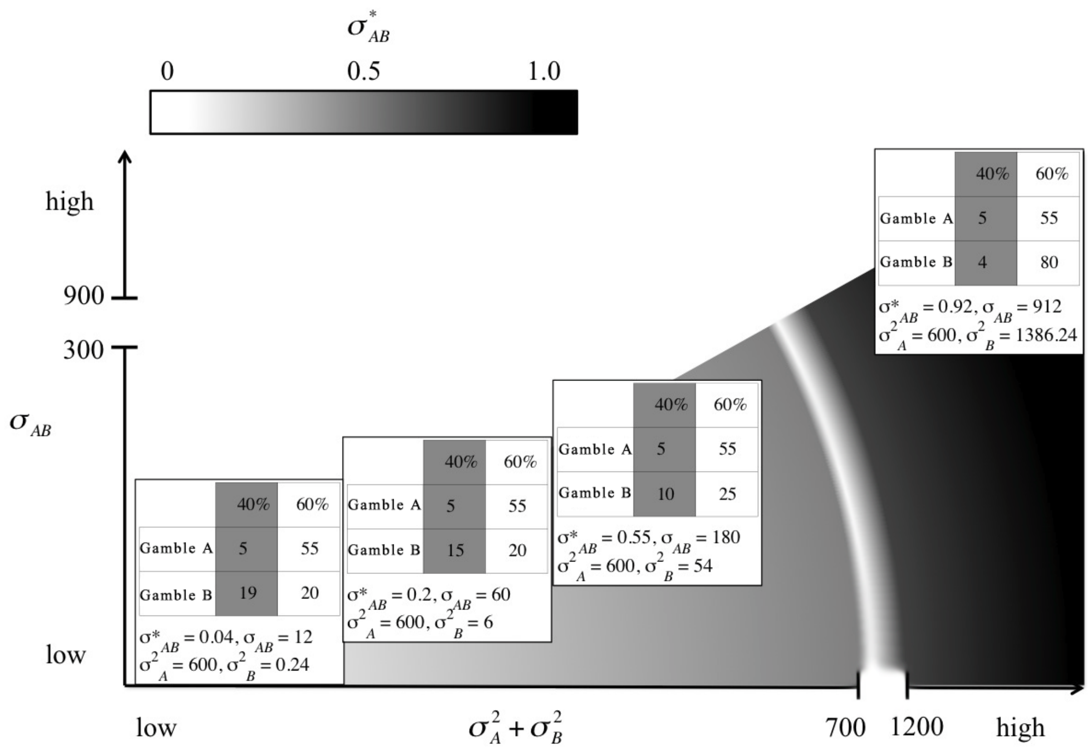


Figure 2. Landscape with examples of pairs of gambles, arranged on a plane with respect to the size of their variances and covariance. The difference in expected values ranges between 14.6 and 17 points. Gamble A is more advantageous in the first three examples. The gradient background indicates the strength of the standardized covariance. The gradient is triangular because the covariance values do not exceed the sum of the variances.

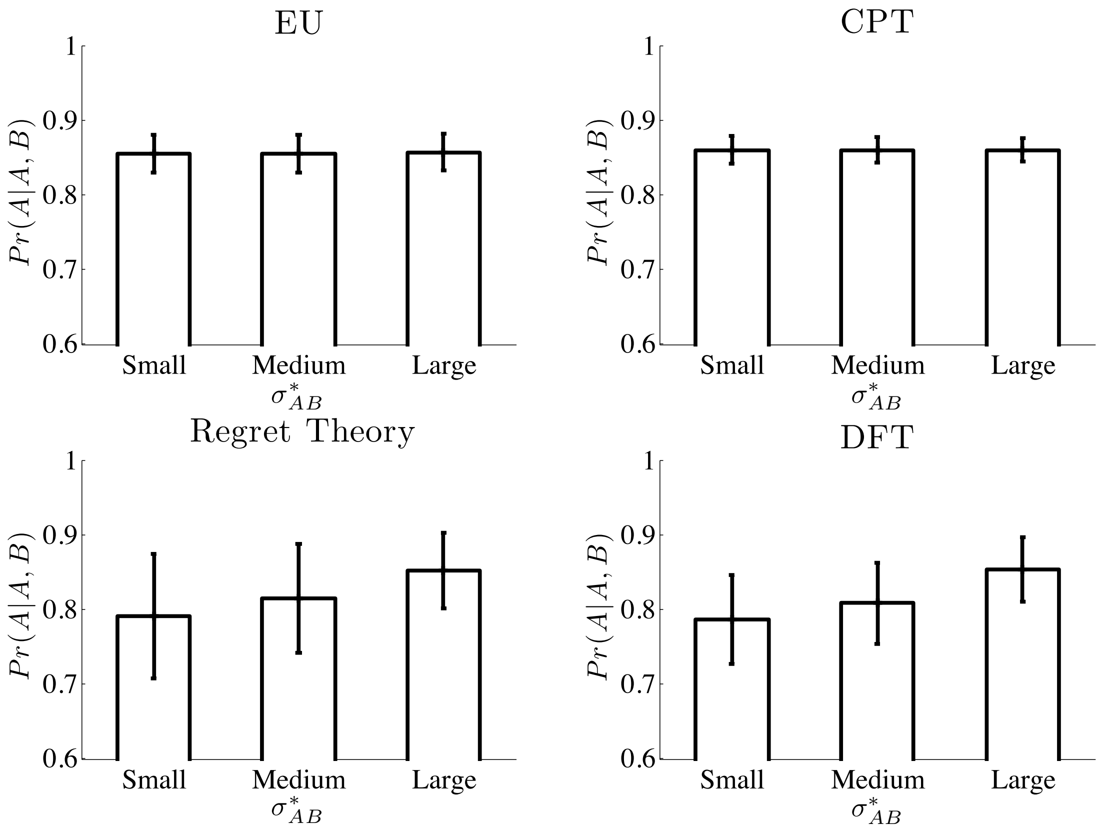


Figure 3. Mean predictions of EU, CPT, RT, and DFT depending on the size of the standardized covariance. The error bars depict population standard deviations. *Pr*(*A*|*A,B*) refers to the probability of choosing the gamble with the larger expected value.

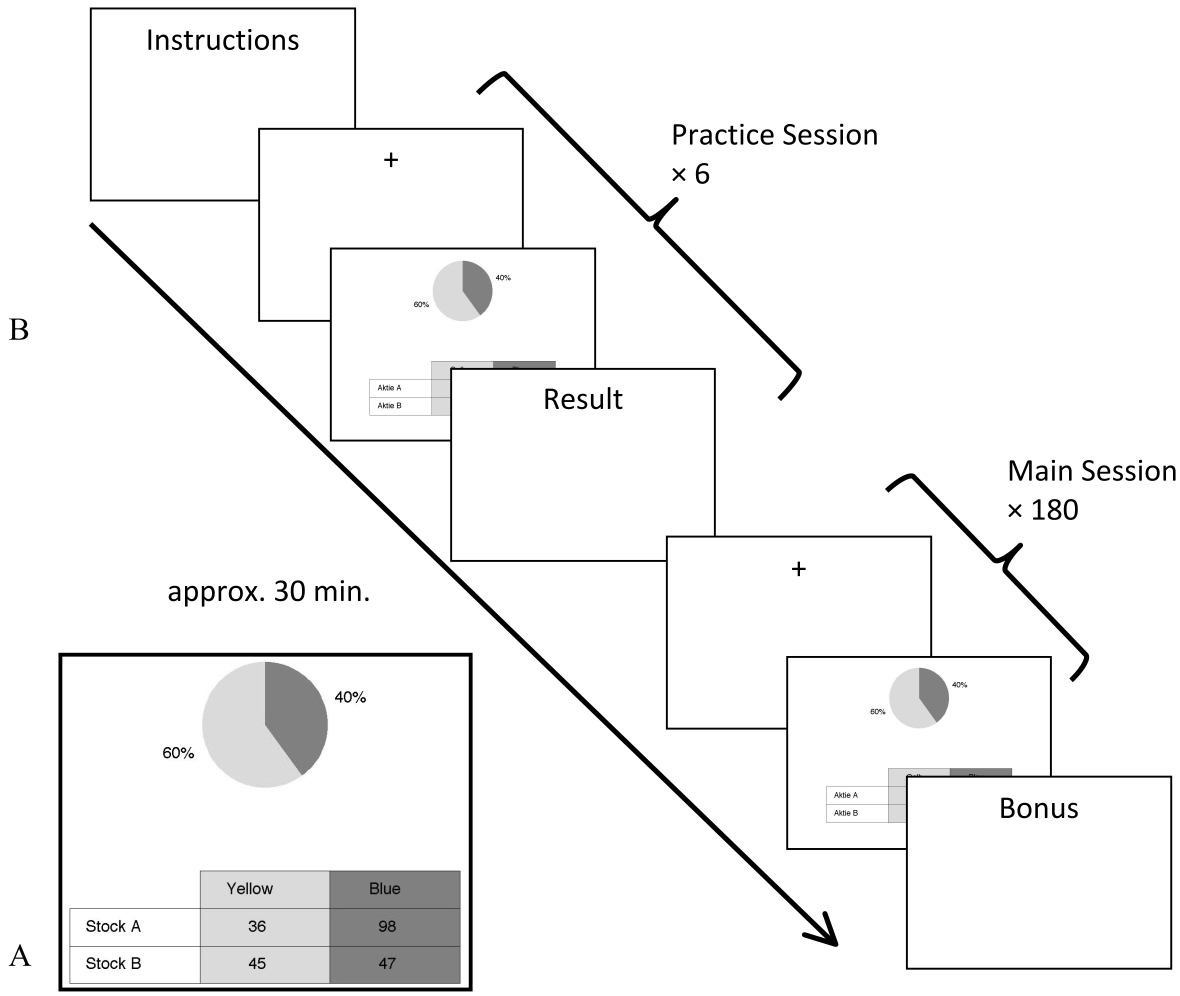


Figure 4. A) Example of a stimulus used in the experiment; B) Experimental paradigm.

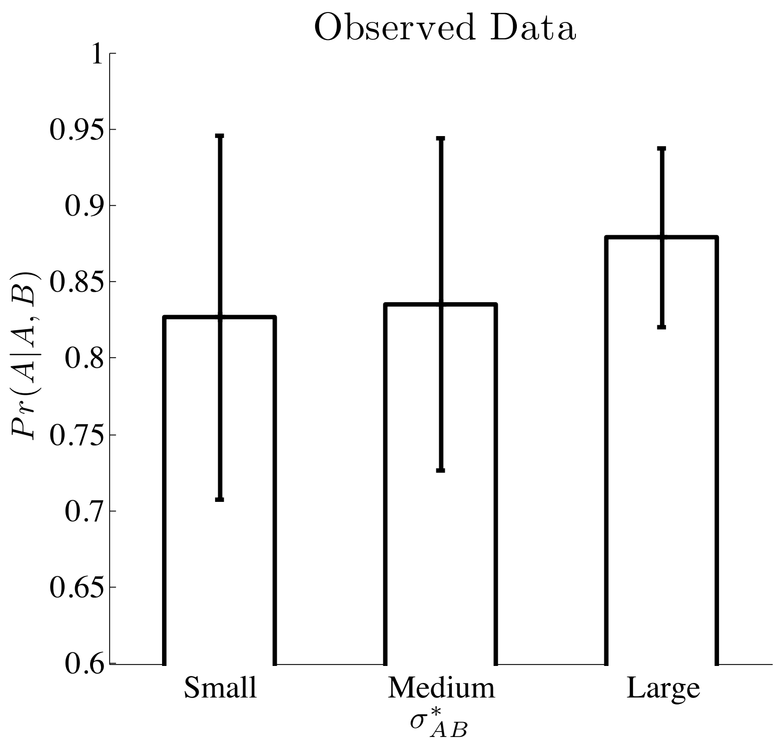


Figure 5. Average observed probabilities of choosing the gamble with the larger expected value. Error bars indicate standard error obtained with a bootstrap method.

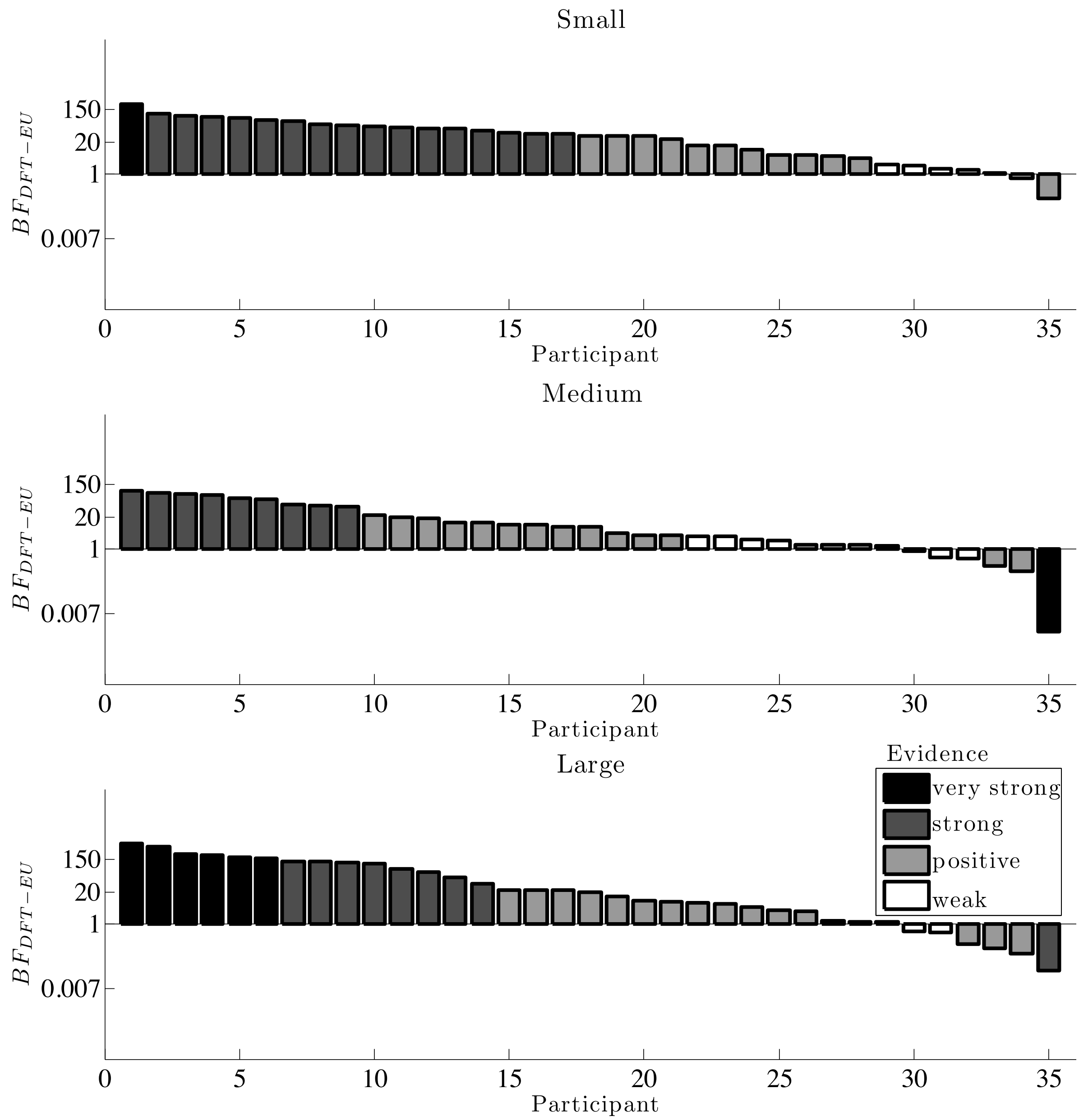


Figure 6. Evidence of DFT as compared to EU expressed by the logarithm of the Bayes factor. The strength of the evidence is categorized following Kass and Raftery (1995).

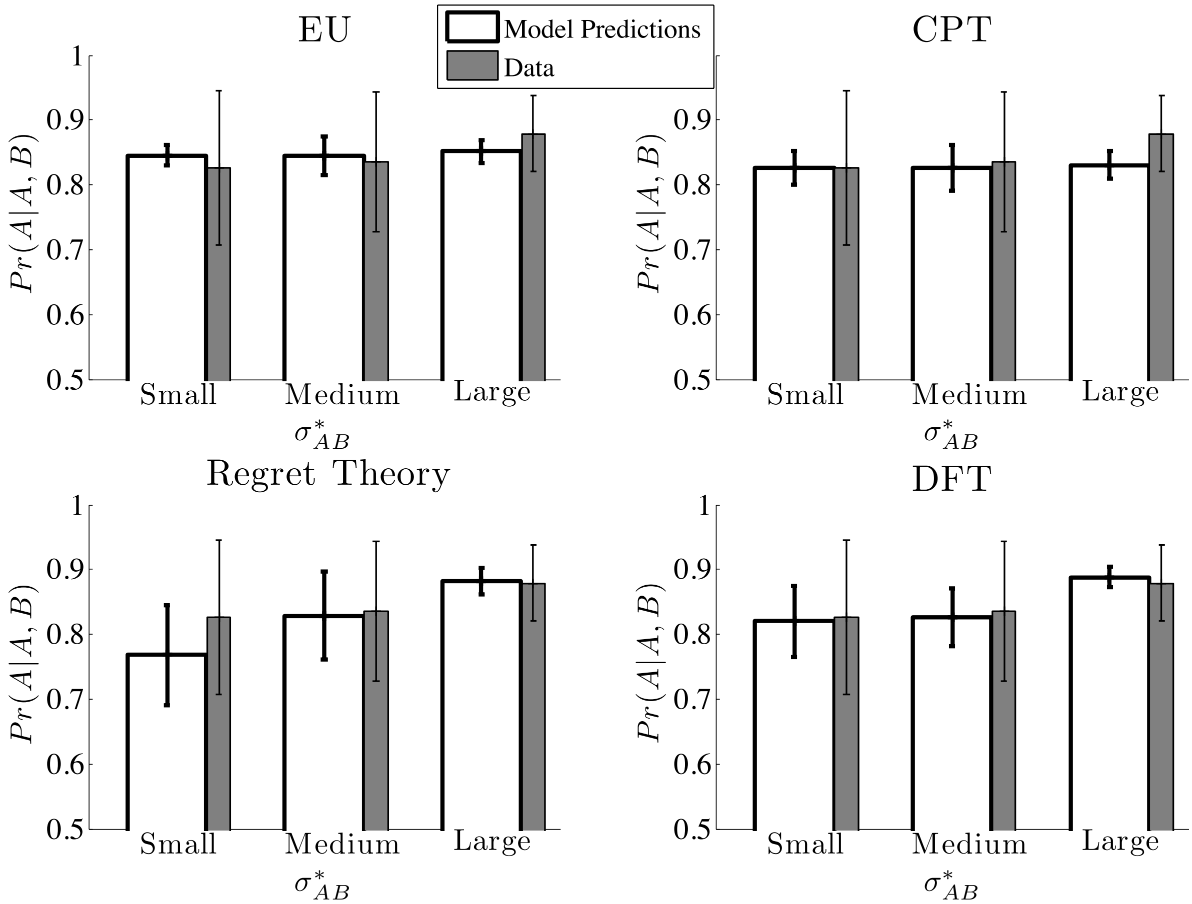


Figure 7. Mean predictions of EU, CPT, RT, and DFT for each covariance condition. The error bars indicate the bootstrapped standard errors. Gray bars indicate the observed data (as in Figure 5) and are the same in all four cells.

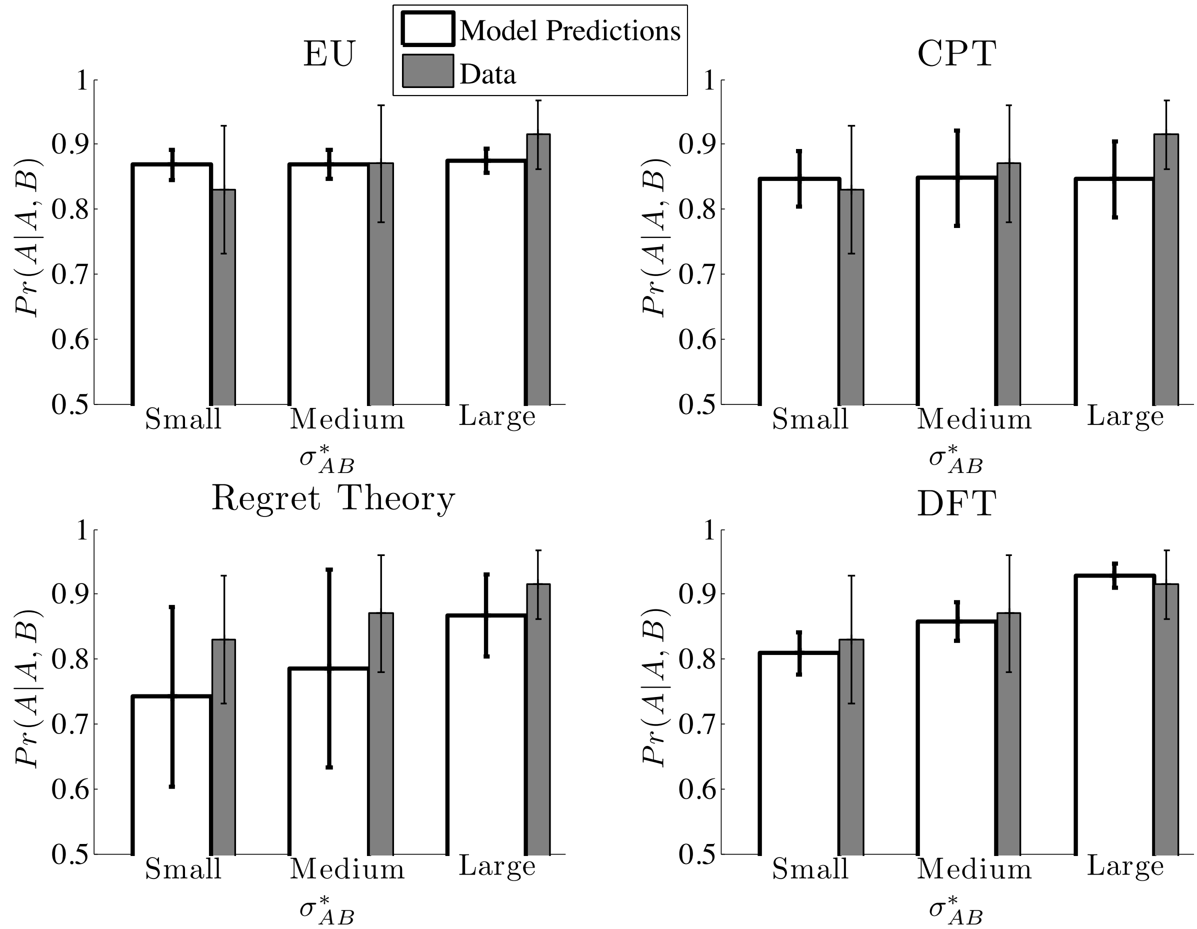


Figure 8. Mean predictions of EU, CPT, RT, and DFT for each covariance condition. The error bars indicate the bootstrapped standard errors. Gray bars (identical for all panels) show the observed data.

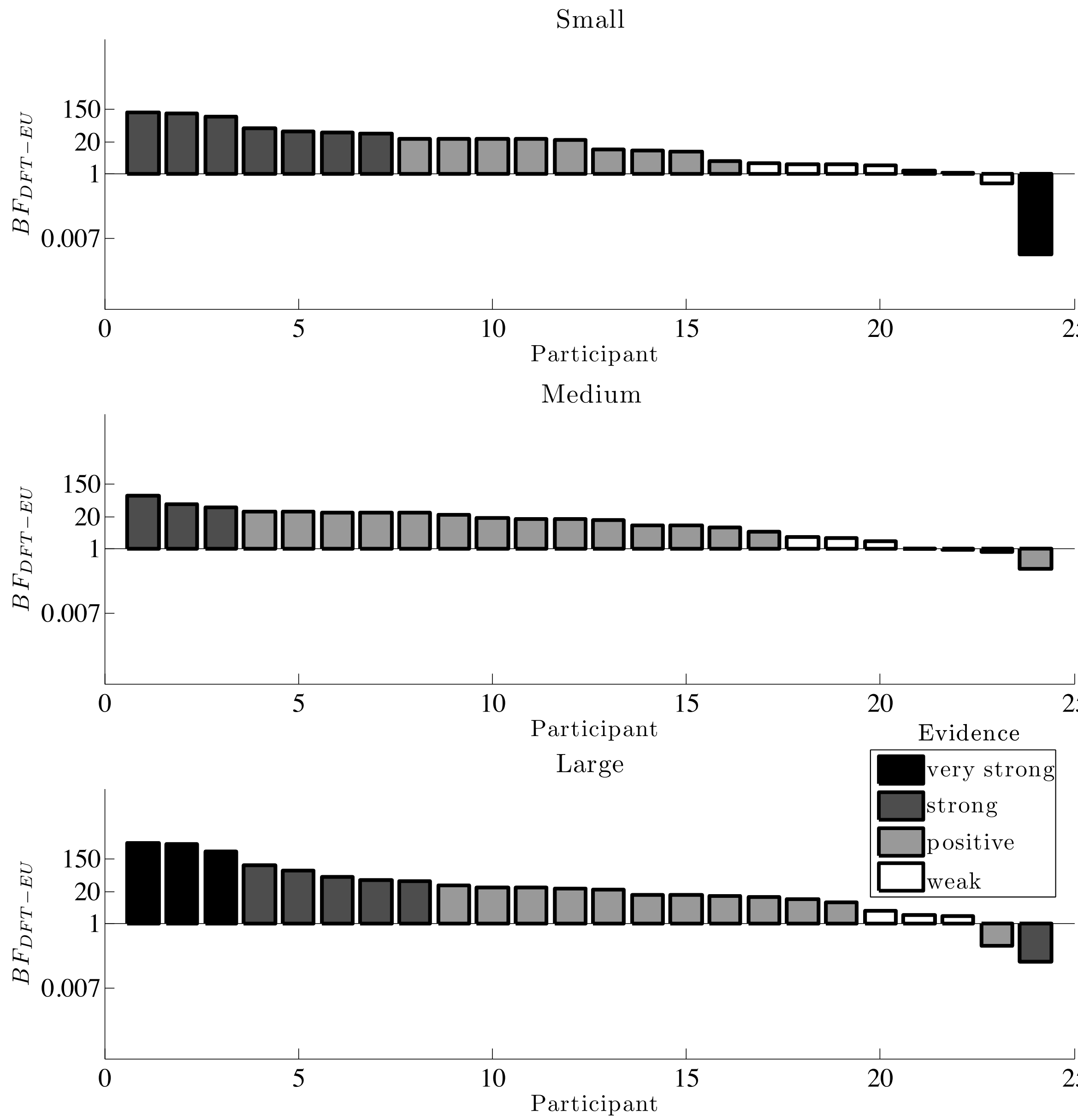


Figure 9. Evidence in favor of DFT over EU expressed by the logarithm of the Bayes factor. The strength of the evidence is categorized following Kass and Raftery (1995).

Appendix A – Properties of the Standardized Covariance

The value of σ^*_{AB} is defined by the value of the covariance between options relative to the options' variances. For positively correlated options, σ^*_{AB} takes values between 0 and 1. When twice the covariance is similar to the sum of variances, then σ^*_{AB} is approximately 1:

$$2\sigma_{AB} \approx \left[\sigma_A^2 + \sigma_B^2\right] \Leftrightarrow \sigma^*_{AB} \approx 1 .$$

When twice the covariance is smaller than the sum of variances, σ^*_{AB} is smaller than 1 minus the value of h

$$2\sigma_{AB} < \left[\sigma_A^2 + \sigma_B^2\right] \Leftrightarrow \sigma^*_{AB} \approx 1 - h \qquad \wedge \qquad h \in \Re^+ .$$

Note that twice the covariance cannot be greater than the sum of the variances, as the following proof shows. In stochastically non-dominant options the variances of options' outcomes are unequal, thus

$$\sigma_A^2 < \sigma_B^2 \qquad \vee \qquad \sigma_A^2 > \sigma_B^2 .$$

As a consequence, we can write that

$$\sigma_A^2 < \sigma_B^2 \Leftrightarrow \sigma_B^2 = \sigma_A^2 + s \qquad \wedge \qquad s \in \Re^+ .$$

Then,

$$2\sigma_{AB} > \sigma_A^2 + \sigma_A^2 + s$$

$$0 > \sigma_A^2 + \frac{s}{2} - \sigma_{AB}$$

$$0 > E[a^2] + \frac{E[b^2] - E[a^2]}{2} - E[a \cdot b] \quad ,$$

where a and b are the expectations of gambles A and B minus values of outcomes of A and B. From $\sigma_A^2 < \sigma_B^2$ we derive that

$$E[a^2] < E[b^2] \Leftrightarrow b = a + g \qquad \wedge \qquad g \in \Re^+ ,$$

such that g is the difference between a and b. By expanding the inequality we get

$$0 > \frac{g^2}{2}.$$

Because $g^2 > 0$, the inequality is false.

$$0 \le \frac{g^2}{2}$$ is true.

Further, the σ^*_{AB} is sensitive to the difference between expected values of the choice options. As shown in the Figure A1, the bigger the expected value difference the more narrow the range of the σ^*_{AB} values. The following mathematical proof shows that the greater the expected value difference, the smaller the maximum value of σ^*_{AB}.

Mathematical proof that $\Delta EV \uparrow \Leftrightarrow \sigma^*_{AB} \downarrow$.

For two options with expected values of $E[A]$ and $E[B]$, such that $E[A] \neq E[B]$, we can define the expected value of option A as

$$E[A] = E[B] + l, \text{ where } l\text{=const., } l \in \Re.$$

If the options are not identical, we define that their outcomes x_{Ai} and x_{Bi} for each event $i \in I$, are in the relation $x_{Ai} \neq x_{Bi}$. We can also define the difference between the outcomes of the options, corresponding to the same external event as

$$x_{Ai} = x_{Bi} + m_i \text{, where } m_i = \text{const., so}$$

$$a = x_A - E[A] = x_B + m - E[B] - l = b + (m - l).$$

In sum, m is the difference between outcome values of option A and B, whereas l is the difference between the expected value of the options. We can define the difference between the outcomes' difference and the expected value difference as not negative value k, such that

$$k = m - l.$$

We can rewrite the variances and the covariance of the options as

$$\sigma_A^2 = E\left[a^2\right] = E\left[(b+k)^2\right] = E\left[b^2\right] + 2kE\left[b\right] + k^2$$
$$\sigma_B^2 = E\left[b^2\right]$$

$$\sigma_{AB} = E\left[ab\right] = E\left[b^2\right] + kE\left[b\right].$$

We rewrite σ_{AB}^* as

$$\sigma_{AB}^* = \frac{2\sigma_{AB}}{\sigma_A^2 + \sigma_B^2} = \frac{u}{u+k^2} \quad \wedge \quad u = 2E\left[b^2\right] + 2kE\left[b\right].$$

Thus, the relation between σ_{AB}^* and *k* is as follows

$$k \uparrow \Leftrightarrow \sigma_{AB}^* \downarrow.$$

Further, the difference *l* between the expected values of the options with two outcomes *x* and probabilities *p* depends on the difference *m* of the outcomes' values. If we write *l* as a

$$\begin{aligned} l &= E\left[A\right] - E\left[B\right] \\ &= px_{A1} + (1-p)x_{A2} - \left(px_{B1} - (1-p)x_{B2}\right) \\ &= p\left(x_{A1} - x_{B1} - (x_{A2} - x_{B2})\right) + x_{A2} - x_{B2} \\ &= p\left(m_1 - m_2\right) + m_2 \\ &= pm_1 + (1-p)m_2 \\ &= E\left[m\right] \qquad \wedge \qquad m_i = x_{Ai} - x_{Bi}, \end{aligned}$$

then the value *k* on which σ_{AB}^* depends equals

$$k = m - E\left[m\right].$$

Finally, the value of σ_{AB}^* depends on the difference between values of outcomes of option A and B, such that the higher the difference, the lower σ_{AB}^*:

$m \uparrow \Leftrightarrow \sigma^{*}_{AB} \downarrow$.

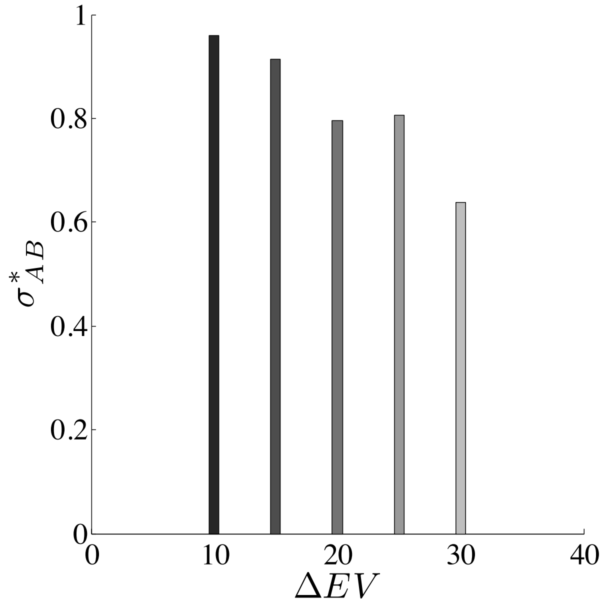


Figure A1. Relationship between the options' expected value difference and the range of values of σ^{*}_{AB}. For each expected value difference, a sample of 1000 pairs is presented. The bigger the difference between expected values of two gambles, the smaller the range of the σ^{*}_{AB} and the smaller the maximum value of the σ^{*}_{AB}.

The stimuli created for each of the three conditions in Study 1 had specific properties regarding the relationship between their variances and covariance. Figure A2 shows these relations for each of the groups. The mean distance is greater between the medium and large group than between the small and medium group. Also, the coverage of the sum of variances almost completely overlaps for the small and medium group, whereas it only partially overlaps for the medium and large group. Nevertheless, the covariance values do not overlap for any of the groups.

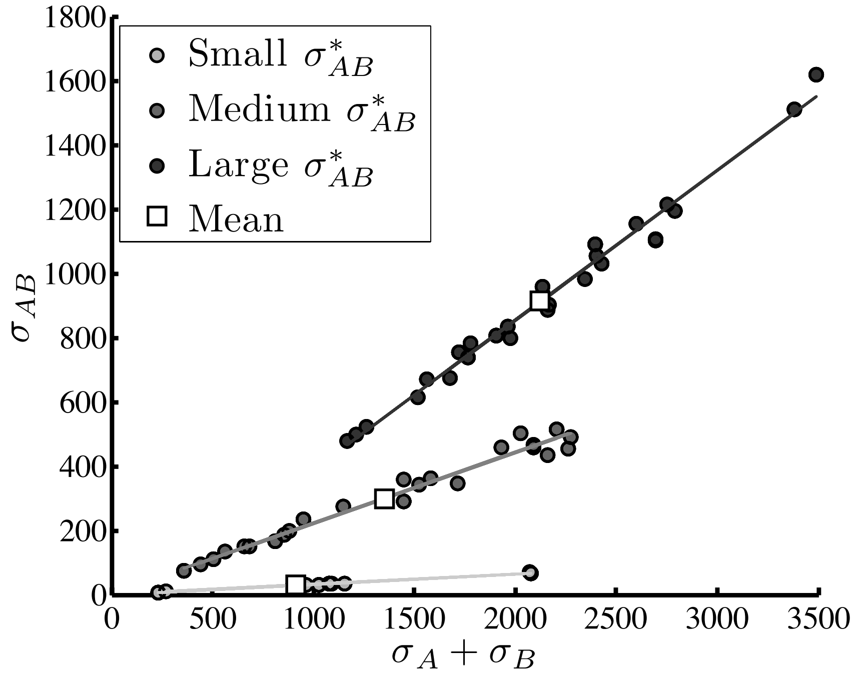


Figure A2. Variances and covariances of gambles in each condition of Study 1. Straight lines represent regression lines. The higher σ^*_{AB}, the steeper the slope is, which indicates that the higher σ^*_{AB}, the stronger the relation between variance and covariance within each gamble. The Euclidean distance from the mean in condition large to the mean of condition medium is larger than the distance between condition medium and small, which is in line with the expectation that the difference in preferences will be greater.

Appendix B – Specification of Selected Models of Decision Making

Expected Utility Theory

EU defines the expected utility of an option *A* with *I* outcomes by:

$$EU(A) = \sum_{i=1}^{I} p_i u(x_i), \tag{B1}$$

where *p* represents the probability that outcome *i* will occur and *x* is the outcome. We defined the utility of an outcome x_i (i.e., a monetary payoff) by a power function:

$$u(x_i) = \begin{cases} x_i^{\alpha}, & x_i \geq 0 \\ -(-x_i)^{\alpha}, & x_i < 0 \end{cases} \tag{B2}$$

where the parameter α determines a person's risk attitude. The probability with which an option is chosen is defined by an exponential choice rule:

$$\Pr(A \mid A, B) \frac{1}{1 + \exp\left[\theta\left(EU(B) - EU(A)\right)\right]} \tag{B3}$$

Cumulative Prospect Theory

The overall subjective value of option *A* is defined as

$$V(A) = \sum_{i=1}^{I} \pi(p_i) \cdot v(x_i) \tag{B4}$$

where the subjective value of an outcome is defined as:

$$v(x_i) = \begin{cases} x_i^{\alpha}, & x_i \geq 0 \\ -\lambda(-x_i)^{\beta}, & x_i < 0 \end{cases} \tag{B5}$$

where α and β define the curvature of the utility function of gains and losses correspondingly; $\lambda \geq 1$ specifies loss aversion.

The probability weighting function $\pi(p_i)$ accounts for the individual perception of the outcomes' probabilities and is defined as:

$$\pi(p_i) = \begin{cases} \sum_{x_i \geq x}^{I} w(p_i,\gamma) - \sum_{x_i > x}^{I} w(p_i,\gamma), & x_i \geq 0 \\ \sum_{x_i \leq x}^{I} w(p_i,\delta) - \sum_{x_i < x}^{I} w(p_i,\delta), & x_i < 0 \end{cases} \tag{B6}$$

$$w(p_i) = \frac{p_i^c}{\left(p_i^c + (1-p_i)^c\right)^{1/c}} \tag{B7}$$

with $c = \gamma$ for positive and zero payoffs and $c = \delta$ for negative payoffs. The choice probability of CPT can be defined by Equation B3.

Regret Theory

In the current work, following Pathan et al. (2011), we define the regret function R_i of choosing option A with outcomes x_i, $i \in \{1,...,I\}$ and probabilities p_i over option B with outcomes y_i and probabilities p_i as

$$R_{iA} = \ln\left(1 + \exp\left(\beta \cdot \left(x_i - \max(x_i, y_i)\right)\right)\right) \tag{B8}$$

where β is a parameter of the sensitivity to the losses and corresponds to the curvature steepness of the exponential function. The total regret of choosing an option with several possible outcomes is

$$R_A = \sum_{i=1}^{I} R_{iA} \quad . \tag{B9}$$

The probability of choosing A over B is estimated using an exponential choice rule:

$$\Pr(A \mid A,B) = \frac{1}{1 + \exp\left[\theta\left(R_B - R_A\right)\right]} \tag{B10}$$

with θ as a free sensitivity parameter of the model (in contrast Pathan, et al., 2011 used a constant sensitivity parameter of $\theta = 1$).

Decision Field Theory

Assuming a predefined threshold, the probability of choosing option A over B can be approximated by

$$\Pr(A \mid A, B) = \frac{1}{1 + \exp\left[-2\left(\frac{d}{\sigma_d} \cdot \theta_{DFT}\right)\right]} \tag{B11}$$

where d represents the expected difference between the two options, σ_d represents the variance of that difference, and θ_{DFT} is the decision threshold. DFT is similar to the probabilistic versions of the RT such that the regret effects results from dividing the mean valence difference by the standard deviation of the valence difference (Busemeyer and Towsend, 1993). The difference between the options can be determined by

$$d = v(A) - v(B) \tag{B12}$$

where *v* is an option's subjective value defined as

$$v = \sum_{i=1}^{I} p_i u(x_i) \tag{B13}$$

and I is the number of possible outcomes, W is a continuous random variable representing attention weight assigned to each possible outcome of an option and $u(.)$ represents the utility of outcome *x*. In the current study, we assume that $W(p_i) = p_i$ and $u(x_i) = x_i$. The variance of the difference σ_d is defined as

$$\sigma_d = \sqrt{\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB}} \tag{B14}$$

where σ_{AB} defines the strength of relationship between the options' outcomes, such that when $\sigma_{AB} = 0$ the options' outcomes are independent and the higher σ_{AB} the stronger the statistical relationship between the options' outcomes. The decision threshold chosen by the decision maker θ_{DFT}^* is proportional to the standard deviation of the differences and the threshold θ is equal to $\theta_{DFT}^* / \sigma_d$.

Expected Utility Theory with an Error Model

The probability of choosing gamble A over gamble B is expressed as a choice rule of EU model (see Equation B3) with an error model σ_d defined in Equation B14, is defined as

$$\Pr(A \mid A,B) \frac{1}{1+\exp\left[\frac{\theta\left(EU(B)-EU(A)\right)}{\sigma_d}\right]} . \quad \text{(B15)}$$

Simulation and Parameter Estimation

Following Harrison and Rutström (2009) and Rieskamp (2008), the predictions in the simulation are based on the parameters adjusted to generate probability predictions at the same level: EU: $\alpha = .867$, $\theta = .23$, CPT: $\alpha = .93$, $\beta = .89$, $\gamma = .77$, $\delta = .76$, $\lambda = 1$, $\phi = .18$, RT: $\beta = .05$, $\theta = 4.6$, DFT: $\theta = 1.19$.

In Study 1 and Study 2, we estimated parameters using maximum log-likelihood approach. The allowable parameter estimates were as follows. The parameter space was restricted to reasonable ranges: EU: $\alpha \in [0,3]$, CPT: $\alpha \in [0,3]$, $\beta \in [0,3]$, $\delta \in [0,1]$, $\gamma \in [0,1]$, RT: $\beta \in [0,1]$. The loss aversion parameter λ of CPT was irrelevant as no mixed gambles were included (i.e., $\lambda = 1$). The sensitivity parameters (θ or ϕ for CPT) ranged between 0 and 40.

Appendix C – Additional Tables and Figures

Table C1

Fit and the parameter estimates for expected utility theory, cumulative prospect theory, regret theory and decision field theory for the two-outcome gambles.

	EU			CPT						Regret			DFT	
Ppt	BIC	α	θ	BIC	α	β	δ	γ	ϕ	BIC	β	θ	BIC	θ
1	171.14	0.72	0.42	185.61	0.78	0.82	0.77	1	0.29	165.81	0.08	4.77	165.49	1.16
2	122.54	1.09	0.1	134.12	0.98	1.21	0.69	0.33	0.18	122.53	0.01	18.61	130.74	1.53
3	75.838	0.96	0.25	82.594	0.99	0.87	0.23	0.4	1.18	73.339	0.01	26.55	81.294	2.21
4	164.97	0.58	0.92	178.61	0.58	0.56	0.57	0.76	1.1	175.3	0.03	5.61	176.41	1.06
5	171.62	0.86	0.21	157.39	0.79	1.84	1	0.12	0.26	147.77	0.11	6.06	150.33	1.32
6	55.316	1.15	0.12	67.921	1.05	1.1	0.67	0.39	0.26	57.065	0.01	30.5	69.881	2.43
7	122.96	1	0.15	128.14	0.87	1.36	1	0.21	0.25	102.85	0.1	8.83	103.59	1.86
8	106.18	0.79	0.47	121.45	0.83	0.86	1	1	0.35	112.87	0.08	7.35	90.214	2.06
9	258.42	1.06	0.01	275.5	0.01	0.01	0.01	0.01	0.01	257.88	0.01	1.65	254.11	0.08
10	32.076	0.9	0.49	44.965	0.82	1.68	1	1	0.63	36.872	0.01	37.94	33.132	3.54
11	238.89	1.33	0.01	275.5	0.01	0.01	0.01	0.01	0.01	229.82	0.19	2.89	222.68	0.61
12	208.26	0.97	0.09	223.46	0.91	0.95	1	1	0.11	220.81	0.1	2.7	191.02	0.92
13	40.432	1.13	0.15	54.535	1.1	1.11	0.45	0.39	0.31	39.577	0.02	21.12	39.014	3.29
14	96.037	0.7	0.82	107.8	0.62	0.77	0.89	0.42	1.17	98.291	0.04	8.61	98.496	1.94
15	163.18	0.77	0.35	178.97	0.83	0.84	1	1	0.25	166.91	0.1	4.93	146.72	1.35
16	131.58	0.74	0.51	135.47	0.56	1.26	1	0.15	1.22	139.41	0.03	7.35	135.2	1.48
17	86.963	0.44	4.07	94.743	0.4	0.52	0.75	0.69	4.38	106.16	0.09	8.11	104.21	1.86
18	40.607	1.09	0.18	95.046	1.45	1.42	0.83	0.82	2.64	40.821	0.07	15.35	29.839	3.71
19	254.95	0.22	1.16	275.5	0.01	0.01	0.01	0.01	0.01	256.08	0.01	2.27	254.2	0.07
20	239.17	0.66	0.24	275.5	0.01	0.01	0.01	0.01	0.01	239.72	0.04	2.09	235.59	0.46
21	146.45	1.16	0.06	156.62	1.21	1.11	0.61	0.6	0.07	146.88	0.03	6.97	144.93	1.37
22	146.98	0.87	0.24	162.37	0.86	0.83	1	1	0.27	146.88	0.07	5.5	138.16	1.44
23	253.07	1.21	0.01	275.5	0.01	0.01	0.01	0.01	0.01	254.9	0.02	1.51	248.49	0.25
24	126.34	1.13	0.08	121.28	0.66	1.74	1	0.12	0.85	112.07	0.09	7.69	106.02	1.83
25	174.62	0.83	0.24	190.22	0.84	0.81	1	1	0.23	167.62	0.1	4.9	157.44	1.24
26	48.552	1.06	0.19	64.528	1.01	1	1	1	0.23	58.857	0.04	11.88	44.375	3.09
27	111.48	0.81	0.41	126.36	0.8	0.84	1	1	0.39	118.69	0.06	6.82	103.63	1.86
28	157.89	0.65	0.68	127.41	0.26	1.62	0.7	0.09	10	136.85	0.13	7.35	141.86	1.4
29	169.41	0.57	0.94	185.32	0.6	0.59	1	1	0.76	166.4	0.11	5.11	157.56	1.24
30	62.662	0.92	0.33	76.644	0.95	0.94	0.58	0.85	0.33	64.123	0.01	28.64	73.238	2.36
31	89.714	0.77	0.59	101.11	0.68	1.01	1	0.26	0.87	91.416	0.06	8.42	82.529	2.19
32	122.87	1.05	0.12	138.11	1.01	1.04	0.96	1	0.13	126.42	0.05	6.6	117.31	1.69
33	88.082	1.3	0.05	92.073	0.93	1.66	1	0.14	0.33	83.148	0.08	9.6	76.468	2.3
34	160.91	0.84	0.25	175.95	0.83	0.86	1	1	0.25	161.22	0.08	4.96	148.2	1.34
35	257.12	1.11	0.01	275.5	0.01	0.01	0.01	0.01	0.01	256.07	0.18	0.92	250.53	0.21

Table C2

Fit and the parameter estimates for expected utility theory, cumulative prospect theory, regret theory and decision field theory for the four-outcome gambles.

	EU			CPT						Regret			DFT	
Ppt	BIC	α	θ	BIC	α	β	δ	γ	ϕ	BIC	β	θ	BIC	θ
1	180.92	0.76	0.32	184.05	0.78	0.94	1	0.9	0.19	184.18	0.01	6.52	178.72	1.19
2	69.50	0.55	2.23	97.078	1.17	1.26	0.97	0.96	0.07	96.91	0.03	7.51	82.86	2.61
3	243.22	1.3	0.01	257.35	0.73	0.98	1	1	0.06	275.45	0.12	0.84	234.04	0.53
4	68.84	0.82	0.52	91.259	1.37	1.44	0.85	1	0.03	91.66	0.03	7.71	65.87	3.02
5	75.76	0.95	0.26	89.672	1.29	1.3	1	1	0.05	89.36	0.05	6.05	58.44	3.24
6	40.24	0.78	0.82	95.046	1.97	1.94	0.89	0.88	5.09	95.05	0.03	9.93	30.17	4.47
7	169.92	1.52	0.01	185.41	1.44	1.57	1	1	0.01	187.18	0.09	2.58	150.48	1.52
8	252.13	0.15	3.07	275.5	0.01	0.01	0.01	0.01	0.01	275.5	0.01	1.83	248.42	0.28
9	48.58	1.11	0.15	67.053	1.57	1.53	0.78	1	0.02	67.25	0.04	8.77	44.54	3.73
10	198.68	0.44	0.44	203.78	1.26	1.34	0.29	0.4	0.04	204.13	0.01	7.33	198.93	0.96
11	209.50	0.71	0.31	233.97	1.29	1.38	1	1	0.01	232.44	0.08	1.75	196.85	0.98
12	120.38	0.72	0.61	144.4	1.14	1.13	1	1	0.08	145.36	0.06	3.68	104.47	2.19
13	106.63	0.71	0.7	121.07	1.19	1.38	0.9	1	0.05	126.24	0.04	5.12	100.20	2.27
14	46.32	1.36	0.05	60.323	1.72	1.6	0.54	1	0.02	80.66	0.04	9.69	38.486	4.01
15	73.74	0.73	0.79	100.8	1.3	1.33	0.76	1	0.04	105.28	0.04	6.98	66.047	3.02
16	80.62	0.78	0.58	96.678	1.48	1.53	0.51	0.71	0.02	106.38	0.03	9.86	71.06	2.89
17	153.98	0.3	5	192.17	0.79	0.91	1	0.45	0.23	192.22	0.01	7.88	172.93	1.25
18	69.47	0.93	0.3	84.526	1.49	1.31	1	1	0.04	97.61	0.04	6.49	60.70	3.17
19	131.06	0.84	0.31	160.65	1.26	1.28	1	0.99	0.04	163.88	0.03	4.29	131.90	1.77
20	135.08	0.59	1.09	162.8	1.08	1.25	1	1	0.05	165.07	0.04	3.89	129.45	1.81
21	176.85	0.55	0.98	198.16	0.88	1.03	1	1	0.11	198.25	0.04	2.88	176.21	1.22
22	162.51	0.96	0.14	185.65	1.48	1.55	0.6	0.65	0.01	196.50	0.02	5.28	159.82	1.41
23	100.87	1.47	0.02	124.04	1.55	1.56	1	1	0.01	132.58	0.06	4.65	83.69	2.59
24	183.61	1.49	0.01	204.91	1.39	1.54	1	1	0.01	214.31	0.05	2.01	188.72	1.07

Expected Shortfall as a Measure of Risk in Risk-Value Models

Sandra Andraszewicz, Bettina von Helversen, Jörg Rieskamp

University of Basel

Center for Economic Psychology

Abstract

Models of decision making under risk propose different definitions of risk. The most commonly used definition is variance of choice option's outcomes. Further developments of the standard risk-value models incorporate skewness in the model specification. In contrast, expected utility theory assumes people's risk aversion results from the curvature of the utility curve. Here, we propose expected shortfall, as an established in finance measure of risk, which provides a plausible psychological interpretation of risk, where a decision maker falls short of their aspiration expected outcome. We integrate this measure in the standard risk-value models assume a trade-off between the expected gain and expected risk. We test the new *risk-value shortfall model* against the existing models in two behavioral experiments. Our results indicate that the proposed model can successfully predict people's preference for options with the higher expected value, lower variance and more positively skewed distribution. Also, we showed the advantage of the risk-value shortfall model over expected utility model.

Keywords: skewness, risk-value model, expected shortfall, risky measurement

Introduction

Risk and uncertainty are an integral part of decision making. Indeed, in decisions ranging from daily activities such as taking a taxi or the bus to investing one life savings' in bonds or stocks one needs to consider if what one stands to gain by taking one option is worth what one can loose. Not surprisingly, the question how much risk humans are willing to take has attracted a lot of research in psychology, economics and cognitive science (e.g. Weber, Shafir, & Blais, 2004; Fishburn, 1977; Lopes, 1984; Holt & Laury, 2002). Frequently it is assumed that people make decisions under risk by weighing the expected gain against the risk, with risk commonly measured by the variance in outcomes (e.g. Markowitz, 1959; Tobler at al., 2007; Weber, Shafir & Blais, 2004; Fishburn, 1977).

However, a substantial amount of literature suggests that also higher statistical moments, such as skewness and kurtosis affect how people make risky decisions (Payne, 1973; Symmonds et al., 2011, Burke & Tobler, 2011). This suggests that using variance as a measure of risk, is insufficient to capture human behavior in decisions under risk (see also Bontempo, Bottom, & Weber, 1997; Luce & Weber, 1986; Weber et al., 2004;). Here, we suggest that peoples' choices can be better explained when using an alternative measure of risk, the expected shortfall. The expected shortfall is a common measure of risk in finance (Acerbi & Tasche, 2002) and, as we will show, it can explain peoples' preferences for variance and skewness in a single measure.

Using the expected shortfall as a measure of risk has several advantages. For one, it captures a choice option's variance and skewness in a single measure and thus represents a more frugal account than incorporating both variance and skewness. Furthermore, it is a psychologically plausible that the expected shortfall underlies decisions under risk, because research suggests that decision makers frequently try to avoid outcomes that fall short of their expectation or "aspiration levels" (see Lopes & Oden, 1999).

In an experimental study, we show that using the expected shortfall as a measure of risk can explain human choices better than traditional models using variance as a measure of risk as well as models that additionally incorporate a preference for skewness such as the mean-variance-skewness model and expected utility theory (von Neumann & Morgenstern, 1953).

The influence of skewness on risky choice

When making choices between risky options, it as assumed that people trade off the expected gain of an option with the risk involved in choosing it. Commonly, the gain of an option is reflected by the expected value and risk by the variance that is the spread of the outcomes (Sharpe, 1964; Weber et al., 2004; Markowitz, 1959). However, the sole use of variance as a measure of risk has been criticized (Weber et al., 2004; Luce & Weber, 1986). In particular, it has been suggested that beside variance people pay attention to higher moments such as the skewness of the outcome distribution (e.g. Symmonds et al., 2011). Skewness refers to asymmetries in the distribution of outcomes, that is the more unevenly the potential outcomes are distributed, the higher the skewness. Statistically skewness of outcomes of option A is defined as,

$$\gamma_A = \frac{E\left[(A - E\left[A\right])^3\right]}{\sigma_A^3} \tag{1}$$

with positive skewness referring to distributions in which low outcomes are more frequent than high outcomes and negative skewness referring to distributions in which high outcomes are more frequent than low outcomes.

For instance, imagine three assets that differ in their prospective returns. The price of Asset A will be 90, 92 or 120 Swiss Francs, the price of asset B will be 81.5, 109.5, 111.5 Swiss Francs and the price of asset C 83.8, 101 and 117.3 Swiss Francs. Although all three assets have the same expected price of 101 and the same variance ($\sigma_A^2 = \sigma_B^2 = \sigma_C^2 = 281$), they differ in their skewness (see Figure 1). Whereas the prices of asset B are evenly distributed, the prices of asset A are positively skewed and the prices of asset C are negatively skewed. How does skewness affect human decision-making?

Prior research suggests that when expected value and variance are kept constant people – and animals – prefer options with a positive skew (i.e. Burke & Tobler, 2011; Chunhachida et al.1997; Symmonds et al., 2011). Furthermore, Garrett and Sobel (1999) argue that a preference for positive skewness can explain why people play lotteries. In addition research in finance and economics suggests that people take skewness into account when selecting portfolios (see Samuelson, 1970; Chiu, 2005; Li, Quin, & Kar, 2010).

Models of Decision Making: Accounting for Skewness Preferences

Various models have been proposed to describe how people make decisions under risk with two types of models dominating the literature: (1) expected utility models (von Neumann & Morgenstern, 1953) and (2) risk-value models (Sarin & Weber, 1993). Both models have been adapted to take preferences for skewness into account. In the following, we shortly describe these models and how they can account for a preference for skewness. Then we introduce the *risk-value shortfall model* as an alternative approach to understand risky decision-making.

Risk-Value Models

Risk-value models assume that risky choices can be understood as a trade-off between the expected gain and the risk associated with an option. In its most classical version, risk is defined as variance of the outcomes of risky options, which defines the range of the possible outcomes. The classic mean-variance model (MV) cannot account for skewness, but it has been recently extended to the mean-variance-skewness model (MVS, see Symmonds et al. 2011; see also Jondeau & Rockinger, 2006; Post, van Vilet & Levy, 2006 for a similar extension of the Capital Asset Pricing Model (CAPM)). The MVS model incorporates preferences for skewness, assuming that the subjective value (*SV*) of a risky option A is a weighted sum of the mean (μ_A)[1], variance (σ_A^2) and skewness (γ_A) of the option,

$$SV(A) = \beta_1\mu_A - \beta_2\sigma_A^2 + \beta_3\gamma_A . \qquad (2)$$

The MVS model successfully accounted for participants' choices in a study by Symmonds et al. (2011) outperforming the classical mean variance model. Furthermore Symmonds et al. (2011) identified separate neural correlates for preferences for skewness and variance suggesting that people are sensitive to both measures.

Expected Utility Models (EU)

Expected utility models assume that preference is driven by the subjectively experienced utility of an option that depends on the options outcomes weighted by their (subjective) probabilities (von Neumann & Morgenstern, 1953; Tversky & Kahneman, 1979). Expected utility theory (von Neumann & Morgenstern, 1953) can account for preference for skewness, if the utility function is appropriately specified.[2] Specifically, the standard one-element utility function of outcomes *a* (von Neumann & Morgenstern, 1953), defined as

$$u(a) = \begin{cases} a^{\alpha} & iff \quad a \geq 0 \\ -(-a)^{\alpha} & iff \quad a < 0 \end{cases} , \qquad (3)$$

can account for choices of higher expected value, lower variance and positive skewness when the free parameter *α* is restricted such that $\alpha \in \Re^{+} \cap (0,1)$. Specifically, when $1 < \alpha < 2$, the

---

[1] Note that throughout the paper, we define the statistical mean as the expected value, $\mu_A = E[A]$.

[2] Golec and Tamarkin (1998) proposed the extension of expected utility to the three element utility function ($u(a)$), where the utility an outcome *x* is a weighted sum of the first three elements of the Taylor's series, $u(x) = b_0 + b_1 a + b_2 a^2 + b_3 a^3$ with positively valued weights b_1 and b_3 and negatively valued b_2. This expansion can account for choices according to the higher expected value, lower variance and positive skew. Accordingly any utility function can for which $u'(a) > 0$, $u''(a) < 0$ and $u'''(a) > 0$, because it would have the same properties as the utility function obtained from the Taylor series expansion (Kroll, Levy, & Markowitz, 1984).

utility function can account for variance and expected value, and when $\alpha \geq 2$, the function can only account for the preference for higher expected value. $\alpha = 0$, in turn, indicates that a decision maker is insensitive to any outcome value.

In sum, classical expected utility models as well as extensions such as cumulative prospect theory can account for positive skewness if the α parameter is specified accordingly. However, even though expected utility theory leads to a preference for options with positively skewed distributions, it does not assume that people actually evaluate the distribution of the outcomes. Instead the skewness preference implicitly results from the curvature of the utility function, which is difficult to reconcile with the findings of Symmonds et al. (2011).

The Risk Shortfall Model (Short)

Alternatively, we propose *risk-value shortfall model*. Similar to the MV and MVS models, the risk-value shortfall model assumes that the subjective value of a choice option can be represented by the expected value ($E[A]$) of the option A and its risk (R).

$$SV[A] = E[A] - \beta \cdot R(A). \quad (4)$$

In contrast, to the MV model, we suggest that risk should not be measured by the variance but by the *expected shortfall* of an option. The expected shortfall reflects the decision maker's expectations about possible losses associated with a choice option. It is a frequently used as a measure of risk in finance for portfolio optimization (Bertsimas, Lauprete, & Samarov, 2004) and in the insurance industry (Embrechts, McNeil, & Straumann, 2002). In finance, it is usually defined as the "average loss in the worst c% of cases", where c a free parameter that can be adjusted depending on the decision maker or the task (Acerbi & Tasche, 2002). Statistically, the expected shortfall ($ES_{c,A}$) of choice option A with I outcomes a_i and their corresponding probabilities p_i is defined as

$$R(A) = ES_{c,A} = \sum_{i=1}^{I} p_i \left[\max\left(c - a_i, 0\right)\right]. \qquad (5)$$

To be able to apply the expected shortfall as a measure of risk in options with only positive outcomes, we here define the expected shortfall as the average expected outcome, which falls below a decision maker's aspiration level. A decision makers' aspiration level is captured by the threshold parameter *c* (Lopes & Oden, 1999). Furthermore, we assume that the aspiration level *c* depends directly on the expected value of the option A, such that

$$c = \delta E\left[A\right], \qquad (6)$$

where δ is a free parameter of the model, such that $\delta \in \left[0,1\right]$.

Accordingly, the value of the expected shortfall is a function of the distance between the outcomes, the threshold *c* and the probability with which the outcomes occurs. To test the psychological plausibility of the risk-value shortfall model, we conducted an experimental study in which we rigorously test the *risk-value shortfall* (Short) model against the *expected utility* theory (EU), and *mean-variance-skewness* model (MVS). For the sake of completeness, we included the *mean-variance* model (MV), even though it does not account for a preference for skewed distributions.

Behavioral Experiment

To test the models we conducted a study following the experimental procedure used by Burke and Tobler (2011). Burke and Tobler (2011) had provided first evidence that skewness influenced peoples' decision between options with three outcomes. In their study, they investigated the influence of expected value, variance and skewness in choices between gambles varying each factor independently. However, due to the scarcity of the behavioral lab experiments on skewness preference[3], and the fact that Burke and Tobler (2011)

[3] Most of the studies on skewness preference either focused on portfolio selection (i.e. Li, Qin, Kar, 2010; Post, van Vilet & Levy, 2005; Jondeau & Rockinger, 2006), or investigating neural correlates of processing skewness (i.e. Symmonds et al. 2011; Burke & Tobler, 2011).

repeatedly used only 5 gambles, we aimed to replicate their results using a wider variety of gambles. Also, we aimed at testing the generalizability of the effect observed by Burke and Tobler (2011) by using gambles with more than three outcomes.

Method

Participants. 24 (N_{female} = 15) persons, aged between 20 and 52 years (M_{age} = 27.4), participated in the study. Most participants were students at the University of Basel. One participant was excluded from the analysis because she chose the gamble with the lower expected value in almost all control trials, suggesting that she did not understand the task. Participation took between 20 and 30 min. The participants received a show-up fee of 10 Swiss Francs and a bonus ranging between 0 and 2 Swiss francs.

Materials. As stimuli we used two blocks of 100 pairs of gambles, the first block contained gambles with three outcomes, the second block gambles with five outcomes. Following Burke and Tobler (2011), we varied the expected mean, the variance, and the skewness such that each pair of gambles differed with respect to only one property, resulting in five sets of 20 pairs of gambles: 1) low mean vs. high mean, 2) low variance vs. high variance, 3) positively skewed vs. negatively skewed, 4) positively skewed vs. not-skewed, 5) negatively-skewed vs. not-skewed. The gambles in the first set had the same variances and not-skewed distributions, whereas the gambles in the second set had the same expected values, not-skewed distributions but different variances. The last three sets of trials contained of gambles with the same expected values and variances but differently skewed outcome distributions.

In our study, the same expected values, variances and skewness distributions means that $E[X]$=35±.5 points, σ^2=450±7 points, γ=0±.05 points. The skewness of the positively skewed gambles ranged between .48 and .70, and of the negatively skewed gambles between

-.70 and -.42. Positively and negatively skewed gambles in group 3 were matched such that their skewness was symmetric across 0 (i.e. -.7 and .7).

One block of trials contained gambles with three outcomes, whereas the second the gambles with five outcomes. In both cases the probabilities of these outcomes were equal, which means in the first case the probabilities were 1/3 and in the second 1/5. The outcome values ranged between 0 and 100 points. Graphically, the values of probabilities were presented as height of bars, whereas the values of outcomes were marked on the x-axis (see Figure 2A). Three-outcome gambles had three bars (Figure 2A), while the five-outcome gambles had five bars (Figure 2B).

Procedure. Participants were randomly assigned to one of two conditions. Participants in condition 1 started the experiment with the block containing 3-outcome gambles and participants assigned to condition 2 started the experiment with the block containing 5-outcome gambles. In each block, the gamble pairs were presented in a random order, which was varied for every participant. Each time, the participants were asked to choose between two gambles, A (presented on the left) and B (presented on the right) by choosing buttons “A” or “B” on the keyboard. In about 50% of the pairs the left gamble was “the better gamble” (i.e. having higher expected value, lower variance, more positively skewed distribution).

Before the start of the experiment the participants were provided with instructions on the computer screen. They were not informed that some gambles differed in their expected values, variances and distributions. They were asked to choose at each time the gamble that they prefer, but their bonus would depend on their performance. At the end of the experiment, one gamble pair form both blocks was randomly chosen and played out. The outcome of the gamble was divided by 5 and paid out as a bonus. Afterwards, the participants were asked to fill out a demographic questionnaire. This experimental design (Figure 2C) was tested earlier in a pilot study.

Results and Dicusssion

Data Analysis. As shown in Figure 3, overall, in both blocks of gambles people preferred the gambles with a higher mean, followed by lower variance and more positively skewed gambles, 3-outcomes: $p < .001$, χ^2 (3) = 53.81, 5-outcomes: $p < .001$, χ^2 (3) = 57.79[4]. Additionally, we compared the distributions of average frequencies of choices in the three skewness conditions against the uniform distribution of choice frequency of .5, with the series of two-sampled Kolmogorov-Smirnov tests, with Bonferroni correction of α. We found significant differences of the frequency distributions in all three skewness conditions. As expected, participants preferred positively skewed options to negatively skewed options and this preference was stronger than the preference of positively skewed to non-skewed options. Participants preferred negatively skewed options to non-skewed options (see Figure 3). However, this preference was not as strong as the preference for the positively skewed options. This last finding is in contrast to our expectations, and contrary to previous findings (see Burke & Tobler, 2011).

Models of Decision Making. We fitted four models: MV, MVS, EU and Short to the data of every participant from both conditions, using the maximum likelihood method. The exact models' specifications are outlined in Appendix A. We constrained the parameters as follows MV: $\beta \in \{0,15\}$, $\theta \in \{0,100\}$, MVS: $\beta_1 \in \{0,15\}$, $\beta_2 \in \{0,15\}$, $\theta \in \{0,100\}$, EU: $\alpha \in \{0,3\}$, $\theta \in \{0,100\}$, Short: $\delta \in \{0,1\}$, $\beta \in \{0,1500\}$, $\theta \in \{0,100\}$. Overall, all models predicted the data better than the baseline model for 96% of the participants (see Table 1). The estimated parameters are consistent across MV, MVS and EU (see Table B1 in Appendix B). This means that when $\beta_{MV} = 0$, then $\beta_{1\,MVS} = 0$, and when $\beta_{2\,MVS} = 0$, then $\alpha_{EU} \geq 1$.

[4] According to Kolmogorov-Smirov test, the data were not normally distributed.

Therefore, if according to one model's parameter estimates, a decision maker is not sensitive to variance and/or skewness, the same conclusion could be drawn from another model's parameter estimates. Short's parameter estimates do not exhibit such regularities, which confirms that Short assumes a different decision process than the three other models.

According to the median Bayesian information criterion (BIC), Short model had the best fit, followed by MV, MVS and EU (see Table 1). According to total BIC, Short model had the best fit, followed by EU, MVS and MS. As expected, MV did not predict skewness preference, while EU, MVS and Short successfully did. These three models have also successfully predicted the strongest preference for gambles with high mean, slightly lower preference for gambles with lower variance and small variations among three skewness conditions. These results are not surprising for EU and MVS. However, in this experiment, we show for the first time that on the qualitative level, expected shortfall can effectively predict people's risk avoidance with respect to mean, variance and skweness.

In addition to this, overall, Short had the best fit on the individual level. According to BIC, it had a better fit than EU for 56% of participants, and than MVS for 39% of the participants. MVS model had a better fit than EU for 52% of the participants. We compared the models based on the individual Bayes factor (Kass & Raftery, 1995), which quantifies how much better is one model over the other while incorporating the complexity of the model, of Short with respect to MVS and EU. According to the median logistic BF there was a substantial evidence in favor of Short and against EU, $Me\left(2\log BF_{Short\_EU}\right)=5.96$ and there was no difference between Short and MVS, $Me\left(2\log BF_{Short\_MVS}\right)=-.13$. Also, MVS had a better fit than EU, $Me\left(2\log BF_{MVS\_EU}\right)=5.32$ (see Kass & Raftery, 1995 for interpretation of BF values). As shown in Figure 5, for the majority of the participants there was a strong and very strong evidence in favor of Short and against EU. The fact that Short performed better than (according to the BIC) or similar to (according to the BF) MVS shows that the more cognitively inspired model can successfully make the same predictions as the purely mathematical model.

General Discussion

In this study, we showed that using the expected shortfall as measure of risk can improve the ability to predict how people make decisions under risk. Overall, Short model predicted participants' responses better than three alternative models that have been proposed in the literature, EU, MV, and MVS. This suggests that the expected shortfall is well suited to capture how people perceive the risk associated with an option when making decisions under risk. A frequent measure of risk in finance (Bertisas et al., 2004), the expected shortfall is based one the idea that when making decisions under risk people compare the number of outcomes that will fall below their level of aspiration. This idea resonates with the research of Lopes and Oden (1999) saying that people try to avoid outcomes that fall below their level of aspiration, which provides further support for the psychological plausibility of the expected shortfall as a measure of risk.

Using expected shortfall as a measure of risk implies that people not only consider the variance but also the skewness of the outcomes' distribution. Indeed, our results suggest that Short outperformed the MV model, because it takes the skewness of the outcome distribution into account. The idea that skewness influences decision making under risk has been widely supported in research on portfolio selection (i.e. Li et al., 2010; Post, van Vilet & Levy, 2005; Jondeau & Rockinger, 2006), but very few experimental test have been provided so far (for exceptions see Burke & Tobler, 2011; Symmonds et al., 2011). By replicating the results obtained by Burke & Tobler (2011) with larger set of gambles, we found that people not only consider the expected value and the variance when making risky choices (Sarin & Weber, 1993), but also the skewness of the outcome distribution. Although the effect of skewness was smaller than the effect of expected value or variance, participants reliably preferred options with positively skewed outcome distributions to options without skew or negatively skewed outcome distributions. These results dovetail with the literature on portfolio selection (see Chiu, 2005). In addition to this, the results provide further evidence for a preference for skewness in experimental fully controlled conditions – an analysis that usually is not possible in portfolio selection because the distribution of market assets is generally skewed (Li et al., 2010).

Unexpectedly, we did not find any evidence that people would prefer gambles without skew to gambles with a negatively skewed outcome distribution. One possible explanation is that the possible bonus that could be obtained was relatively low (up to 2 Swiss francs), and therefore, the negatively skewed gambles would have not been perceived as particularly risky. Alternatively, it is possible, that the range of negative skweness -.70 to -.42 has been too large, meaning that gambles with γ = -.42 has not been perceived as sufficiently risky. Burke and Tobler (2011) used gambles with one negative skewness level, namely γ = -.7. However, there are very few gambles with this exact skenwess level, and expected value and variance as defined in the methods section. Therefore, to diversify the range of gambles, we had to extend the range of skewness.

The preference for skewness was supported by the model comparisons. From the four models in the comparison three – EU, Short and MVS – predicted a preference for positively skewed outcome distributions, in line with the observed data. In contrast, the MV model did not predict a preference for skewness and, accordingly, was outperformed by the other three models in terms of the total BIC. However, it should be noted, that although the majority of participants (57%) showed a preference for positively skewed options, a considerable minority of participants was not sensitive to skewness. For these the MV provided better fit than MVS, but not necessarily better fit than EU and Short. This means, that MVS' complexity is only feasible for participants who are skewness-sensitive, whereas EU and Short provide a very good fit, even when people are not sensitive to skewness.

Short outperformed the competing models EU and MVS based on the total BIC. In addition, when classifying participants according to the BIC, the majority of participants were classified as using Short, suggesting that overall Short was best to describe participants choices. However, based on the BF, there was a comparable evidence for Short and MVS, but very strong evidence in favor of Short. Also, when comparing MVS to EU, the advantage of MVS over EU was slightly lower than the evidence of Short over EU. The good fit of the MVS model, however, is not surprising, because it directly considers the skewness of the choice options and thus will generally predict the same choice as Short in our design. Here, conducting a more focused model comparison is necessary to tease apart the predictions of these two models.

In sum, the results of this study indicate that in general, people prefer choice options with more positively skewed outcome distributions to the options with non- or negatively skewed distributions. This suggests that when evaluating the risk associated with a choice option people not only consider the spread of its outcomes, but also by the distribution of these outcomes. Furthermore, our results suggest that that the perceived risk of the option can be captured by the expected shortfall, which takes variance and skewness of an option's outcomes into account. In line with this idea, using the expected shortfall as a risk measure in the standard risk-value models improved the models fit outperforming the mean-variance model and expected utility, and performed at least as good or better than the mean-variance-skewness model.

References

Acerbi, C., Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance. 26*(7), 1487-1503. doi: 10.1016/S0378-4266(02)00283-2

Bertsimas, D., Lauprete, G.J., Samarov, A. (2004). Shortfall as a risk measure: properties, optimization and applications. *Journal of Economic Dynamics & Control*. *28*(7), 1353-1381. doi: 10.1016/S0165-1889(03)00109-X

Bontempo, R. N., Bottom, W. P., & Weber, E. U. (1997). Cross-cultural differences in risk perception: A model-based approach. *Risk Analysis*, *17*(4), 479-488. doi: 10.1111/j.1539-6924.1997.tb00888.x

Burke, C.J., Tobler, P.N. (2011). Reward skewness coding in the insula independent of probability and loss. *Journal of Neurophysiology. 106*(1), 2415-2422. doi: 10.1152/jn.00471.2011

Chiu, W.H. (2005). Skewness preference, risk aversion, and the precedence relations on stochastic changes. *Management Science. 51*(12), 1816-1828. doi: 10.1287/mnsc.1050.0431

Chunhachida, P., Danpani, K., Hamid, S., Prakash, A.J. (1997). Portfolio selection and skewness: Evidence from international stock markets. *Journal of Banking and Finance, 21*(2), 143-167. doi: 10.1016/S0378-4266(96)00032-5

Embrechts, P., McNeil, A., & Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 176-223.

Fishburn, P.C. (1977). Mean-risk analysis with risk Associated with below-target returns. *The American Economic Review. 67*(2), 116-126. doi: http://www.jstor.org/stable/1807225

Garrett, T.A., Sobel, R. (1999). Gamblers favor skewness, not risk: Further evidence from

United States' lottery games. *Economic Letters. 63*(1), 85-90. doi: 10.1016/j.bbr.2011.03.031

Golec, J., Tamarkin, M. (1998). Bettors love skewness, not risk, at the horse track. *Journal of Political Economy. 106*(1), 205-225. doi: 10.1086/250007

Holt, C.A, & Laury, S.K. (2002). Risk aversion and incentive effects. *The American Economic Review*. *92*(5), 1644-1655. doi: 10.2139/ssrn.893797

Jondeau, E., Rockinger, M. (2006). Optimal portfolio allocation under higher moments. *European Financial Management. 12*(1), 29-55. doi: 10.1111/j.1354-7798.2006.00309.x

Li, X., Qin, Z., Kar, S. (2010). Mean-variance-skewness model for portfolio selection with fuzzy returns. *European Journal of Operational Research. 202*(1), 239-247. doi: 10.1016/j.ejor.2009.05.003

Lopes, L. L. (1984). Bases for preferences among three-outcome bets. *Journal of Experimental Psychology: Human Perception and Performance. 10*(4), 162-169. doi: 10.1037/0096-1523.10.4.465

Lopes, L. L., Oden, G.C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of Mathematical Psychology*. 43(2), 286-313. doi: 10.1006/jmps.1999.1259

Luce, R.D., Weber, E.U. (1986). An axiomatic theory of conjoint, expected risk. *Journal of Mathematical Psychology, 30*(2), 188-205. doi: 10.1016/0022-2496(86)90013-1

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795. doi:10.1.1.19.6821

Kroll, Y., Levy, H., Markowitz, H.M. (1984). Mean-variance versus direct utility maximization. *The Journal of Finance. 39*(1), 47-61. doi: 10.1111/j.1540-6261.1984.tb03859.x

Markowitz, H. (1959). *Portfolio selection: Efficient Diversification of Investment*. New York, John Wiley

von Neumann, J., & Morgenstern, O. (1953) T*heory of Games and Economic Behavior*. Princetor, NJ. Princeton University Press

Payne, J. (1973). Alternative approaches to decision making under risk: Moments versus risk dimensions. Psychological Bulletin. 80(6), 439-453. doi: 10.1037/h0035260

Post, T., van Villet, P., Levy H. (2006). Risk aversion and skewness preference. *Journal of Banking and Finance. 32*(1), 1178-1187. doi: 10.1016/j.jbankfin.2006.02.08

Sarin, R.K., Weber, M. (1993). Risk-value models. *European Journal of Operational Research. 70*(1), 135-149. doi: 10.1016/0377-2217(93)90033-J

Samuelson, P.A. (1970). The fundamental approximation theorem of portfolio analysis in terms of means, variances and higher moments. *The Review of Economic Studies*. 37(4), 537-542. doi: http://www.jstor.org/stable/2296483

Symmonds, M., Wright, N.D., Bach, D.R., Dolan, R.J. (2011). Deconstructing risk: Separable encoding of variance and skewness in the brain. *NeuroImage. 58*(1), 1139-1149. doi: 10.1016/j.neuroimage.2011.06.087

Tobler, P.N., O'Doherty, J.P., Dolan, R., Schultz, W. (2007). Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *Journal of Neurophysiology. 97*(2), 1621-1632. doi: 10.1152/jn.00745.2006

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297-323. doi:10.1007/BF00122574

Weber, E.U., Shafir, S., Blais, A-R. (2004). Predicting Risk-Sensitivity in Humans and Lower Animals: Risk as Variance or Coefficient of Variation. *Psychological Review. 111*(2), 430-445. doi: 10.1037/0033-295X.111.2.430

Table 1

Model fits according to median and total Bayesian Information Criterion of each model with respect to the baseline model.

Model	*Me*(BIC)	Total BIC
Mean-Variance	211.80	4900.5
Mean-Variance-Skewness	212.41	4607.2
Expected Utility	213.04	4589.0
Risk-Value Shortfall	210.81	4360.3

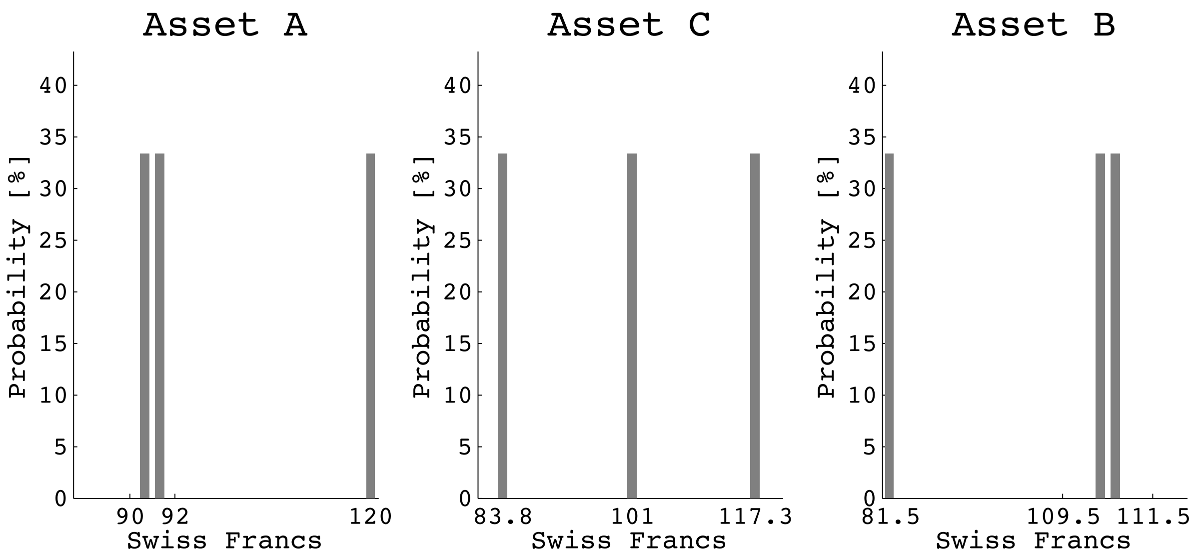


Figure 1. Examples of price distributions of three risky assets whose expected prices and variances are the same, whereas their skewness differs such that $\gamma_A = .70$, $\gamma_B = -.70$, $\gamma_C = 0$.

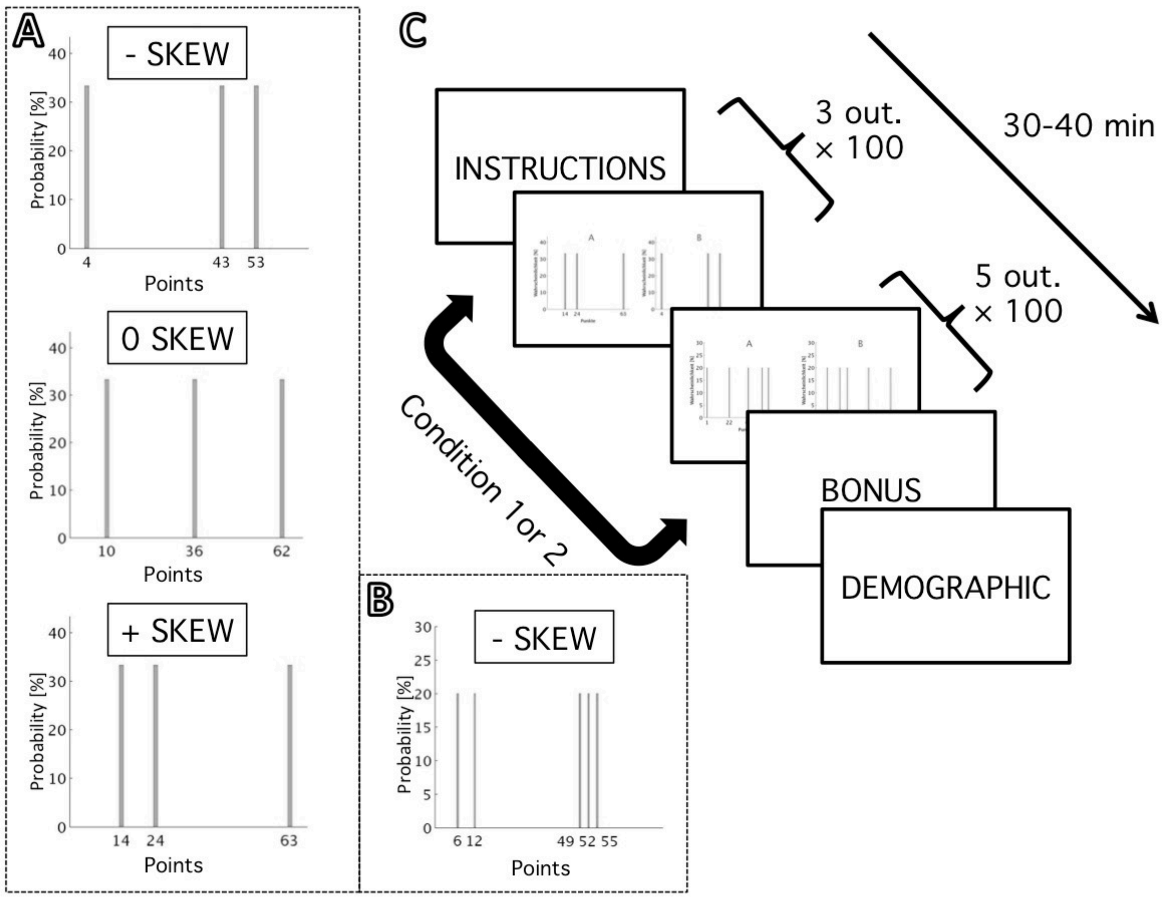


Figure 2. A) Graphical representation of three types of differently skewed three-outcome gambles. Numbers on x-axis correspond to the outcome values, whereas values on y-axis correspond to the probabilities of these outcomes; B) Example of a negatively-skewed five-outcome gamble; C) Experimental design of study 1.

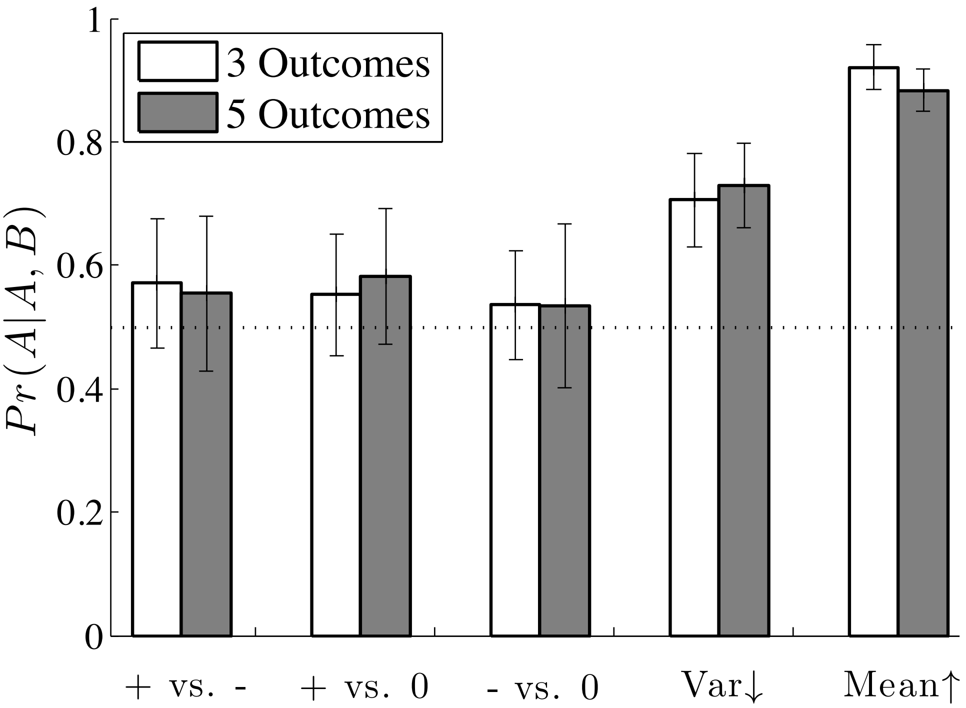


Figure 3. Average frequencies of people's choices of gambles with positively vs. negative (+ vs. -) skewed, positively vs. non-skewed (+ vs. 0), negatively vs. non-skewed (- vs. 0) distributions, lower variance and higher mean. White bars show results for the 3-outcome gambles, whereas the gray gambles correspond to the results for 5-outcome gambles. Error bars correspond to bootstrapped standard error.

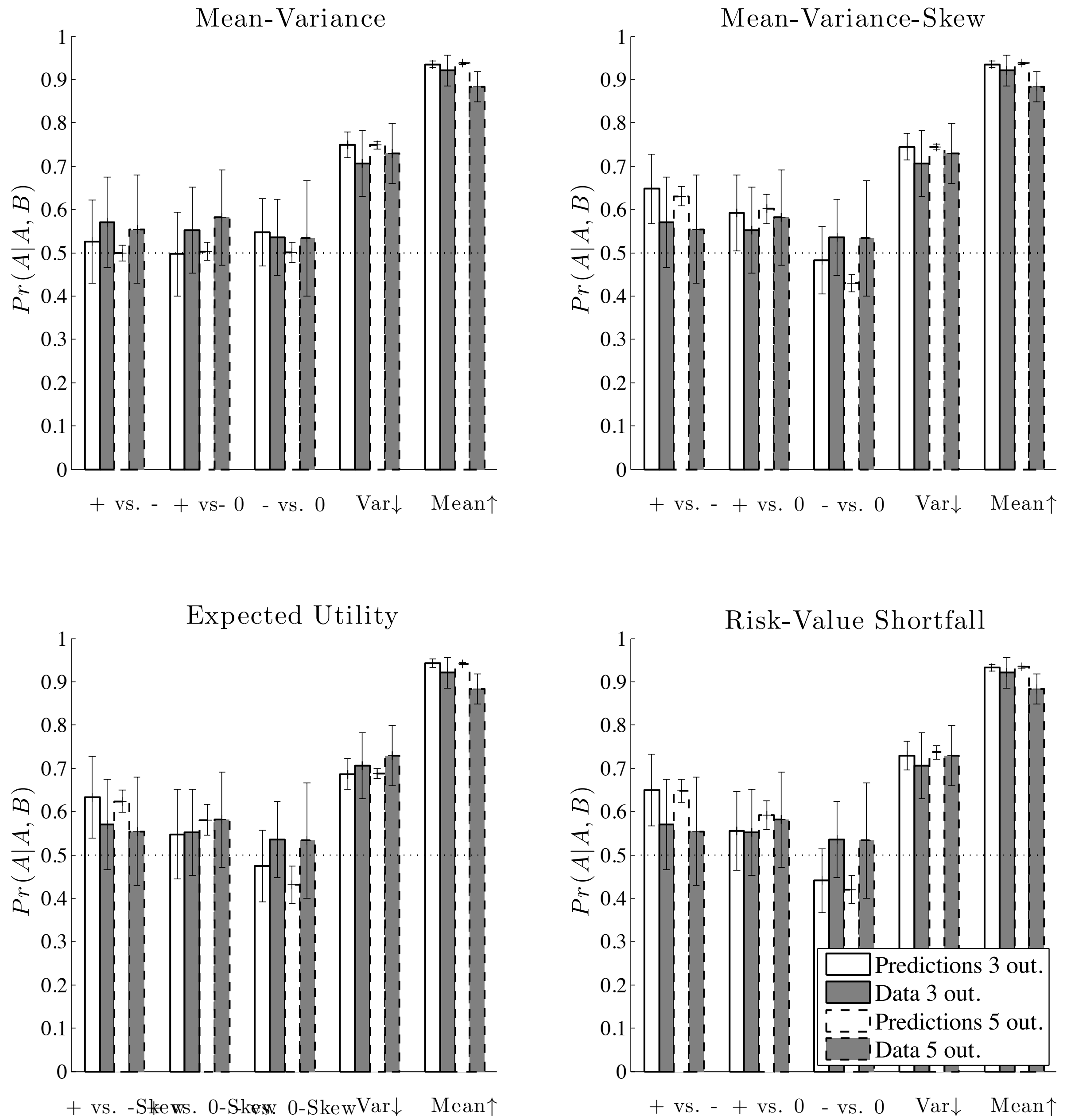


Figure 4. Average predicted probabilities of choosing a gamble with positively vs. negative (+ vs. -) skewed, positively vs. non-skewed (+ vs. 0), negatively vs. non-skewed (- vs. 0) distributions, lower variance and higher mean, of four models: Mean-Variance, Mean-Variance-Skewness, Expected Utility and Risk-Value Shortfall, in two experimental conditions: 3-outcome gambles and 5-outcome gambles. For reference, each cell of the figure includes observed choice frequencies in both conditions, which are the same as in Figure 3. Error bars correspond to bootstrapped standard errors.

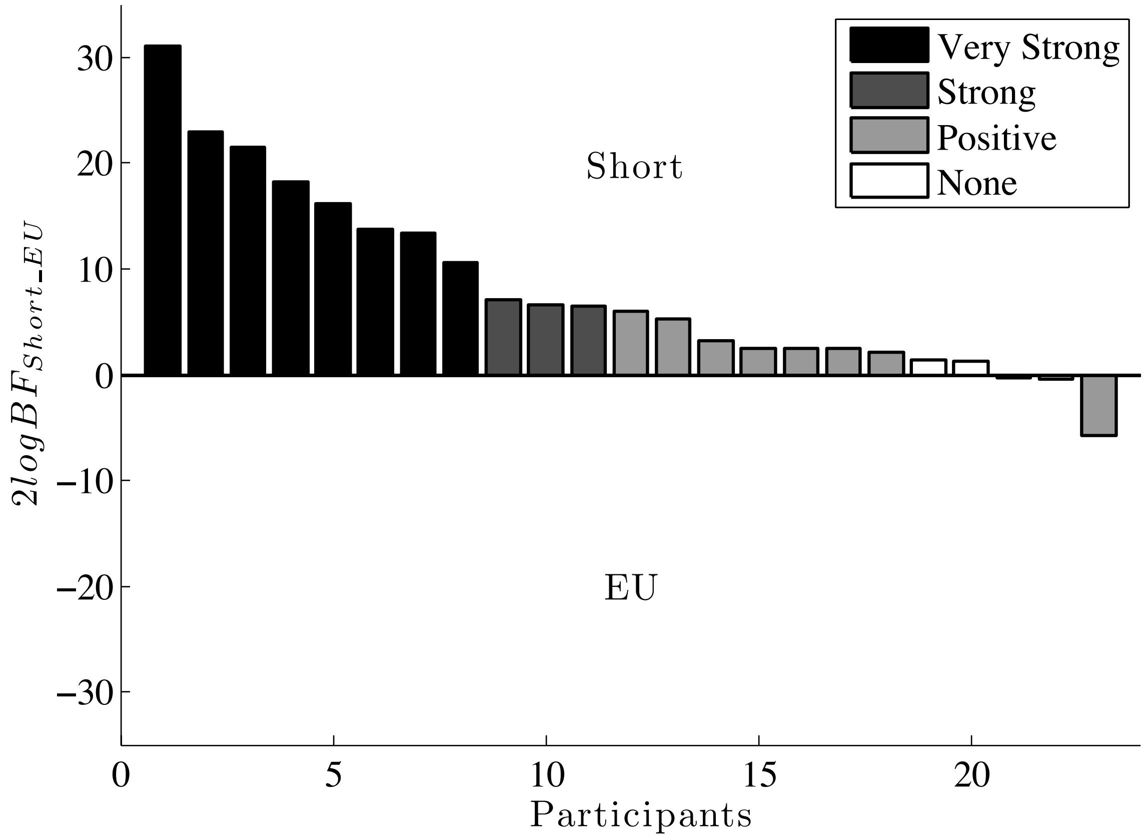


Figure 5. Model comparison of Short and EU, based on logistic Bayes factor. The darker the bar, the stronger the evidence. Bars above 0 indicate evidence in favor of Short, whereas bars pointing downwards indicate evidence in favor of EU. The legend explains the interpretation of the strength of the evidence, as outlined in Kass and Rafter (1995).

Appendix A

Expected Utility Theory

For a gamble A with *I* outcomes a_i, with corresponding probabilities p_i, the utility of each outcome is expressed as in Equation 3, where *α* is a free parameter which defines decision maker's risk attitude. Total expected utility of the gamble equals to

$$U(A) = \sum_{i=1}^{I} p_i u(a_i) . \tag{A1}$$

The probability that a decision maker chooses gamble A over gamble B is defined by the softmax rule:

$$\Pr(A \mid \{A, B\}) = \frac{1}{1 + \exp\left[\theta\left(U(B) - U(A)\right)\right]} . \tag{A2}$$

Mean-Variance Model

The mean-variance model assumes the subjective value of gamble A is expressed as trade-off between expected value and the expected risk

$$SV[A] = E[A] - \beta \cdot \sigma_A^2 , \tag{A3}$$

where *β* is a free parameter which defines decision maker's risk sensitivity. Expected value of gamble A with *I* outcomes a_i, with corresponding probabilities p_i, equals

$$E[A] = \sum_{i=1}^{I} p_i a_i . \tag{A4}$$

The probability of choosing gamble A over gamble B is expressed by the softmax rule

$$\Pr(A \mid \{A, B\}) = \frac{1}{1 + \exp\left[\theta\left(SV[B] - SV[A]\right)\right]} . \tag{A5}$$

Mean-Variance-Skewness Model

Mean-skewness model assumes that the subjective value of gamble A is expressed as a trade-off between expected value and distribution of gambles' outcomes

$$SV[A] = E[A] - \beta_1 \cdot \sigma_A^2 + \beta_2 \cdot \gamma_A, \qquad \text{(A6)}$$

where *E*[*A*] and probability of choosing gamble A over gamble B is expressed the same as in equations A4 and A5 respectively. γ_A denotes skenwess of gamble's outcomes defined as in Equation 1.

Appendix B

Table B1

Estimated parameters and individual model fits measured with Bayesian information criterion in the behavioral experiment.

	MV			MVS				EU			Short			
Ppt.	β	θ	BIC	β_1	β_2	θ	BIC	α	θ	BIC	δ	β	θ	BIC
1	0	0.27	241.2	0	0	0.26	244.4	1.02	0.25	241.2	0.03	1282.	0.26	233.8
2	5	0.81	165.7	0.42	1.13	0.54	156.9	0.63	4.27	169.2	0.76	0.75	0.57	162.6
3	3.62	0.61	169.6	0.09	0.98	0.62	164.2	0.65	4.15	168.6	0.72	0.62	0.68	163.1
4	0.55	0.3	199.3	0.62	0	0.25	234.8	0.54	2.55	213.0	1	0.94	0.29	201.8
5	5	0.52	171.0	0.25	0	0.47	173.4	0.78	1.61	209.0	1	0.71	0.48	183.5
6	0.01	1.4	213.4	0	0	1.45	216.4	1	1.43	213.4	0	0	1.41	207.2
7	1.09	0.4	176.0	0.02	10.32	0.69	25.76	0.31	100	24.02	0.39	2.17	4.39	9.313
8	0.01	1.07	218.9	0	0	1.08	222.1	1.05	0.87	216.5	0	0	1.08	212.8
9	0	3.95	171.3	0	0	3.97	174.5	1.01	3.86	168.8	0	0	3.94	165.1
10	0.42	0.18	241.0	0.01	2.85	0.18	237.2	0.41	2.7	238.6	0.69	1.38	0.15	228.6
11	0	0.09	273.3	0	0	0.1	276.5	1.27	0.03	272.5	0	0	0.08	267.1
12	0.29	0.48	195.1	0.01	3.4	0.58	138.6	0.08	89.61	137.4	0.29	4.99	0.33	126.6
13	0.12	0.04	281.6	0	5.14	0.04	283.7	0.42	0.45	281.2	0.93	0.98	0.04	274.6
14	0	0.29	239.0	0	7.4	0.3	174.1	0.03	83.37	231.4	0.24	4.71	0.19	209.1
15	2.1	0.03	271.4	0.06	13.01	0.03	270.1	0.32	1.23	277.0	0.56	162.4	0	256.7
16	0.26	0.38	210.1	0.01	2.51	0.39	192.8	0.41	5.47	194.0	0.46	1.69	0.31	176.2
17	0.29	0.43	200.3	0.01	4.08	0.48	146.2	0.13	42.83	153.2	0.41	2.93	0.34	126.1
18	0.17	0.5	210.3	0.01	0.27	0.48	212.4	0.75	1.59	210.8	0.57	0.53	0.46	201.4
19	0	0.17	256.5	0	0	0.18	259.6	1.44	0.03	248.7	0	0	0.19	250.3
20	0.08	0.52	222.2	0	0	0.52	225.2	0.92	0.74	225.4	1	0.12	0.53	218.8
21	0.42	0.5	181.3	0.02	0	0.36	185.8	0.84	1.2	217.1	1	0.66	0.51	181.9
22	5	0.36	177.7	0.37	0.52	0.34	178.8	0.5	3.9	195.5	0.89	1.16	0.31	184.9
23	0	1.45	213.0	0	0.27	1.49	212.7	1.01	1.39	212.8	0.13	1.42	1.46	195.3

Title: Neural Underpinnings of Exploitation of Common Goods

Authors*:* Vasily Klucharev, Sandra Andraszewicz, Jörg Rieskamp

Affiliation*:* University of Basel, Department of Psychology, Missionsstrasse 62A, 4055 Basel, Switzerland

Corresponding author: Vasily Klucharev University of Basel, Department of Psychology, Missionsstrasse 62A, 4055 Basel, Switzerland, Phone: +41-61-2670618, Fax: +41-61-2670613, vasily.klucharev@unibas.ch

Keywords: Common Goods; Tragedy of the Commons; Social Competition; Ventral Striatum; Reinforcement Learning; Social Comparison.

Abstract.

Why do people often exhaust unregulated common natural resources but successfully sustain similar private resources? To answer this question the present work combines a neurobiological, economic, and cognitive modeling approach. Using functional magnetic resonance imaging we showed that sharp depletion of a common (shared) and a private resource deactivated the ventral striatum, that is involved in the valuation of outcomes. Across individuals the observed inhibition of the ventral striatum negatively correlated with attempts to preserve the common resource, but the opposite pattern was observed when individuals dealt with their private resource. The results indicate that the basic neural value signals differentially modulate people's behavior in response to the depletion of common versus private resources. The computational modeling of the results suggests that the overharvesting of common resources is facilitated by social comparison. Overall, the results could explain some aspects of people’s tendency to overexploit unregulated common natural resources.

The sustainability of environmental resources is of worldwide concern in the 21st century. Currently the world faces a rapid decline of many natural resources, such as fish, clean air, and primeval forests (Ostrom, 2009). In the present article we explore the neurobiological underpinnings of resource overexploitation. We combine neurobiological, economic and computational approaches to explain why humans treat a resource differently in a competitive social environment as compared to a private environment.

Economic theory predicts the overexploitation of common resources by self-interested people. This claim is illustrated by the "tragedy of the commons" (Hardin, 1968): a dilemma in which multiple individuals, acting independently and rationally, will ultimately deplete a shared limited resource even if it is against their long-term interest. For example, a group of people sharing fishing grounds often realize that they greatly benefit from increasing their own catch. Yet if every person increases his or her own profit this will destroy the fish stock as a whole. This social dilemma is commonly conceptualized as a *common-pool resource* (CPR) situation. In such a situation a natural or urban system generates benefits that can be consumed by individuals who cannot be excluded from consumption (Ostrom, 1990). According to economic theory, open-access CPRs that anyone can enter and/or harvest are likely to be overharvested and destroyed. However, behavioral economics also gives many examples in which people behave fairly and cooperatively contrary to the standard self-interest model (Fehr & Schmidt, 1999): Under some conditions, in particular in two-person interactions, people often show high rates of cooperation. Why, then, is it so difficult even for cooperative people to overlook short-term benefits and sustain CPRs for larger, long-term benefits?

It have shown that overharvesting is particularly high in social groups containing a substantial number of "free riders", that is, people who take benefits without paying any costs (Camerer, 2003). One explanation for a tendency to overharvest CPRs refers to people's social preference for equity and reciprocal cooperation (Falk & Fischbacher, 2006; Fehr & Schmidt, 1999): If others are cooperative, then people act cooperatively, but if others free ride, people retaliate. Accordingly, in a group that contains few free riders, overall average consumption of the CPR will be higher than consumption of its cooperative members. If cooperative members behave reciprocally by choosing the average consumption rate for the future, this will lead to an upward spiral of consumption and to final overexploitation of the CPR (for details see (Fehr & Fischbacher, 2003)). Thus, overexploitation can result even for cooperative people who monitor their own and others' behavior and act reciprocally.

Here we hypothesized that the brain dopaminergic system, a set of brain areas involved in reward and performance monitoring, not only continuously monitors our own outcomes during CPR games but also monitors the outcomes of others. The dopaminergic system has been previously implicated into *social comparison* (Bault, Joffily, Rustichini, & Coricelli, 2011; Dvash, Gilam, Ben-Ze'ev, Hendler, & Shamay-Tsoory, 2010; Fliessbach et al., 2007) – people's strong tendency to compare their own behavior with that of others (Festinger, 1954). We suggest that when dealing with CPRs, the dopaminergic system continually compares personal outcomes with the outcomes of others and facilitates overharvesting in response to free-riding behavior of others causing inequality. But when dealing with private resources, the dopaminergic system monitors deviation from outcomes that maintain long-term resource sustainability. More specifically, we hypothesize that individual overexploitation tendencies have to be depicted in the ventral striatum activity or in the functional connectivity of the ventral striatum with the dorsal prefrontal cortex, known to be involved in control processes that are necessary to achieve long-term harvesting goals (Koechlin & Hyafil, 2007; McClure, Laibson, Loewenstein, & Cohen, 2004). To find a computational explanation of the increasing CPR depletion, we developed a computational model suggesting that the ventral striatum generates a reward prediction-error signal that compares a player’s own outcome with the harvesting behavior of others. This model of social comparison follows the classic idea of people's social preference for equity (Falk & Fischbacher, 2006; Fehr & Schmidt, 1999), with the difference that we assumed that receiving more than the competitors is a positive outcome (see, e.g., Fliessbach et al., 2007 for a similar concept). Thus, we hypothesized that social comparison is encoded in the neural learning signal that facilitates overharvesting of the common natural resources.

Materials and Methods

Participants. Fifty young healthy right-handed students participated in the neuroimaging experiment (aged 18–32 years, mean 23.4 years, 26 females). Subjects were randomly assigned to the social or private condition (with N = 24 for the social and N = 26 for the private condition). None of the subjects reported a history of drug abuse, head trauma, neurological or psychiatric illness. Three participants were rejected from the fMRI analysis due to large head motions exceeding 3 mm; one subject was excluded due to a misunderstanding of the instructions and a high level of error. The study was approved by the local ethics committee of the canton of Basel.

Experiment. Participants were confronted with a resource of fish. In order to avoid any demand effects and suspicion toward the two different (but structurally identical) conditions, we implemented a between-subjects design: Subjects were randomly assigned to the social or private condition of the CPR. Overall they encountered 16 sessions (8 trials per session). In every trial, participants decided between three possible net sizes for fishing one, two, or three fishes. Their task was to collect as much fish as possible and each collected fish led to a monetary payoff (0.25 Swiss Franks per fish). In the social version of the experiment (social condition), two other participants (pre-recorded in a behavioral pre-study) also decided between the three net sizes. In the non-social version of the experiment (private condition), the same number of fishes “migrated” to two neighboring lakes. Importantly, the change of the resources due to the two other pre-recorded participants or the "migration" to the two neighboring lakes was identical in both conditions. Subjects were informed that although the number of fishes in the lake decreases by fishing, it also grows naturally due to proliferation of fish. Therefore, at the end of every trial, the remaining number of fish in the lake was multiplied by 1.5, which gave the total number of fishes for the next trial (with a maximum number of 16 fishes representing the utmost capacity of the lake). In case no fish remained for the next trial, the whole session ended automatically. The instructions clearly explained to the subjects that the amount of fish removed by the players could increase, sustain, or decrease the fish population. For example, the participants were informed that whenever the total number of fishes collected by the three participants was smaller than six, the fish population would increase over the trials. In contrast, whenever the total number of fishes collected by the three participants was larger than six, the fish population would decrease over the trials. If the total number of fishes collected by the three persons was equal to six, the fish population would stay constant over the trials. Thus, the net size of 2 fishes corresponded to a cooperative/sustainable level of harvesting. The experiment started with a short training session. On average, subjects earned 33.3 Swiss Francs (30 SFR – participation fee, 3.3 SFR – monetary payoff) for their participation.

MRI Data Acquisition. Functional MRI was performed with ascending slice acquisition using a T2*-weighted echo-planar imaging sequence (3T Siemens Magnetom Verio whole-body MR unit equipped with a twelve-channel head coil; 40 axial slices; volume repetition time (TR), 2.28 s; echo time (TE), 30 ms; 80° flip angle; slice thickness 3.0 mm; field of view 228 mm; slice matrix 76x76). For structural MRI, we acquired a T1-weighted MP-RAGE sequence (176 sagittal slices; volume TR 2.0 s; TE, 3.37 ms; 8° flip angle; slice matrix 256x256; slice thickness, 1.0 mm; no gap; field of view, 256 mm).

MRI Data Analysis. Image analysis was performed with SPM8 (Wellcome Department of Imaging Neuroscience, London, UK). The first four EPI volumes were discarded to allow for T1 equilibration, and the remaining images were realigned to the first volume. Images were then corrected for differences in slice acquisition time, spatially normalized to the Montreal Neurological Institute (MNI) T1 template, resampled into 3x3x3 mm^3 voxels, and spatially smoothed with a Gaussian kernel of 8 mm full-width at half-maximum. Data were high-pass filtered (cutoff at 1/128 Hz). All five time windows (frames) of the trial were modeled separately in the context of the general linear model as implemented in SPM8. The last trials in each session were excluded from the analysis of interest. Motion parameters were included in the GLM as covariates of no interest.

We constructed separate regressors for different scenarios of resource depletion: The feedbacks on sharp or moderate resource depletion (due to harvesting behavior of others or migration) were modeled as individual hemodynamic responses (2000 ms after trial onset). Based on the ensuing parameter estimates, contrasts of interest were generated. For an additional group analysis, the contrast images were then entered into a second level analysis with subjects as a random effect. To examine regions monitoring perceived fluctuations of the CPR in a separate analysis, one regressor specified for all feedbacks, regardless of specific scenarios of resource depletion, was parametrically modulated by the total amount of fish removed from the lake in each trial (by all parties). In addition, different cognitive models were used to analyze the data: To examine regions generating the reward prediction errors, one regressor specified for all feedback, regardless of specific scenarios of resource depletion, was parametrically modulated by the reward prediction error that was calculated for each trial based on a social or non-social version of the reinforcement learning models (see below, for details). We also calculated a psychophysiological interaction (PPI) analysis (Friston et al., 1997) investigating the functional connectivity of the right ventral striatum. PPI analysis was performed by extracting signal time series from a sphere (5 mm) centered at [9, 5, -5] – the overall group maximum of the right ventral striatum deactivation to the sharp depletion of the resource calculated using a second-level random effects analysis that included all subjects in both conditions.

Social Reinforcement Learning Model. To explain the effect of the social context on harvesting behavior in the CPR task, we constructed a reinforcement learning model (Sutton & Barto, 1998). The model assigns to each choice option a subjective expectation value, which is updated on a trial-by-trial basis. The probability $p_{i,t}$ of choosing an option (net size) i at time t depends on the options' subjective expectations, as specified by a softmax choice rule:

$$p_{i,t} = \frac{\exp[\theta \cdot Q_{i,t-1}]}{\sum_{j=1}^{3} \exp[\theta \cdot Q_{j,t-1}]} \quad (1)$$

where $Q_{i,t\text{-}1}$ is the current subjective expectation for choice option i, θ (θ>0) is the inverse temperature parameter that denotes the stochasticity of the choice process. A larger value for θ implies that the option with the higher expectation value is chosen with a large probability, whereas low values for θ imply random choices. The expectations $Q_{i,t}$ are updated in each trial after the participant makes a decision and obtains the feedback about the two competitors' decisions (social condition) or migration (private condition). Thus, in all trials t, such that $t \in [1,8] \cap \mathrm{Z}$ we calculated the value expectations for each choice option i:

$$Q_{i,t} = Q_{i,t-1} + \alpha(R_{i,t} - Q_{i,t-1}) \qquad (2)$$

where, $R_{i,t}$ is the participants' reinforcement from the current choice and where ($R_{i,t} - Q_{i,t-1}$) represents the prediction error between the participants' expectation and the actual reinforcement from the choice. The parameter α denotes a learning rate ($\alpha \in [0,1]$). Unlike standard reinforcement models (Sutton & Barto, 1998), we assumed that not only the expectation of the chosen option is updated, but also the expectation of the two not chosen options (Camerer & Ho, 1999). Therefore, our model represents a variant of a standard reinforcement model with the difference of updating all options as suggested by fictive updating (Montague, King-Casas, & Cohen, 2006) and recent work in the neuroimaging literature on fictitious prediction errors (Glascher, Hampton, & O'Doherty, 2009; Hampton, Adolphs, Tyszka, & O'Doherty, 2007; Hampton, Bossaerts, & O'Doherty, 2008). For the un-chosen option, the hypothetical payoff given the hypothetical choice was used to determine the fictive prediction error. For the fMRI-analysis, we used the prediction error for the chosen option as a parametric modulator.

The model assumes that when a participant starts fishing in the first trial $t = 1$, she has an a priori expectation about her choice ($Q_{i,t=0}$). To estimate this expectation, we calculated the actual frequencies of choosing net size 1, 2, and 3 in the first trials of all sessions multiplied by 4:

$$Q_{i,t=0} = (n_{i,t=1} / N_{t=1}) \cdot 4 \qquad (3)$$

where $n_{i,t=1}$ is the number of a particular choice (i.e. net size 2) in the first trial in each session, divided by the total number of the sessions $N_{t=1}$. The expected frequencies were multiplied by four to scale the initial expectancies to the real range of rewards that could be obtained in the task (number of fishes: 1, 2, and 3). $Q_{i,t=0}$ is different for every participant, but is constant for each participant and is not estimated as a free parameter when fitting the model. We further suggested that in the social condition people do not only take their personal payoff into account but also compare their payoff with the other players' payoffs. Therefore, the reinforcement of an outcome results from the personal payoff and a social comparison component. According to the social comparison component of our model, the participant received a negative reinforcement if the participant's payoff was lower than the other players' average payoff. When the participant took more than the other players took on average this led to a reward. Therefore, in the *social learning model*, the reinforcement $R_{i,t}$ is a weighted sum of the direct reward from the resource received in a given trial and the social $Comparison_{i,t}$:

$$R_{i,t} = \beta \cdot Payoff_{i,t} + (1-\beta) \cdot Comparison_{i,t} \qquad (4)$$

where $0 \leq \beta \leq 1$ indicates the relative weight given to the personal payoff and the social comparison component. The $Comparison_{i,t}$ is calculated at every trial t for the three net sizes i and is defined as the difference between a person's own payoff ($Payoff_{i,t}$) and the average payoff of the other players ($Others_t$):

$$Comparison_{i,t} = Payoff_{i,t} - \langle Others_t \rangle \qquad (5)$$

Overall, the *social learning model* has three free parameters, the *learning rate α*, the *importance weight* for the individual outcome *β*, and the *sensitivity parameter θ*.

Non-Social Learning Model. We suggest that in the non-social (private) condition, people take their personal payoff into account but are also motivated to sustain the resource in order to maintain a personal payoff in the future. Therefore the reinforcement of an outcome results from the weighted personal payoff and a sustainability component:

$$R_{i,t} = \beta \cdot Payoff_{i,t} + (1-\beta) \cdot Sustain_{i,t} \qquad (6)$$

The $Sustain_{i,t}$ component is determined by the negative value of the absolute value of the difference between the optimal (sustainable) total number of fishes removed from the resource (i.e. *Sustainability* = 6 fishes) and the actual number of fishes taken from the resource by the subject (i.e. $Payoff_{i,t}$) and the total number of fishes that migrated to another lake (i.e. $FishOutflow_{i,t}$):

$$Sustain_{i,t} = -\left| Sustainability - Payoff_{i,t} - FishOutflow_{i,t} \right| \qquad (7)$$

Taking the absolute value of this difference implies that the reinforcement is either zero when the sum of fishes taken from the resource is equal to the *sustainable* number or it leads to a punishment whenever "too much" *or* "too little" is taken from the resource. The rationale behind this punishment is that taking "too little" harms the personal payoff and taking "too much" harms the sustainability of the resource and thereby future payoffs. Thus, according to the sustainability reinforcement component, a participant is penalized for taking too much from the resource if the migration is large. Similarly, the participants are also penalized for taking too little if the migration is small. Importantly, in the social and the private condition of the experiment, the participants were clearly informed in the instructions of the experiment that when the resource decreased by 6 fishes the number of fishes in the lake would stay constant over time. The *non-social learning model* also has three free parameters: the *learning rate* α, the *importance weight* for the individual outcome β, and the *sensitivity parameter* θ.

Evaluation of the Models. We used a maximum-likelihood approach to estimate the three parameters of the two learning models for each participant separately. The evaluation of the models was done by comparing it with a baseline model assuming a random choice between the three choice options (i.e. predicting a choice probability of 1/3) using the Bayesian Information Criterion (BIC, (Schwarz, 1978)). BIC takes the models' complexities (i.e. the number of free parameters, with three free parameters for the two learning models) into account. In the social condition, the *social learning model* was better than the baseline model for 75% of the participants according to BIC and in the private condition, the *non-social learning model* was better than the baseline for 69% of the participants. Paired samples t-tests confirmed a significant difference between the learning models and the baseline model (p=0.0001).

To further examine the empirical validity of the models, we compared the *social learning model* with the *non-social learning model* by following a reverse fitting approach: We fitted the *non-social model* to the participants in the social condition and the *social learning model* to the participants in the private condition. This reversed fitting approach should show that the models are inadequate if they are used in the ”inappropriate“ condition. BIC applied to the reversely fitted models prove the lack of fit. The BIC values for the *social model* fitted to the private condition data was only better than the baseline model for 50% of the participants, and the *non-social learning model* fitted to the social condition was only better than the baseline model for 54% of the participants.

We also tested the two learning models against two competing models. Therefore, we additionally implemented a modified reinforcement learning model (Rescorla and Wagner, 1972) and a modified equity aversion model (Fehr and Schmidt 1999). The reinforcement model only considered the personal payoffs in the task as reinforcement and had no sustainability component, thus it basically is nested within the non-social learning model assuming β = 1. The equity aversion model is identical to the social learning model, but assumes that the comparison component is defined as:

$$Comparison_{i,t} = Payoff_{i,t} - \delta \times \sum_{j=1}^{2} \max\left(Others_{i,j,t} - Payoff_{i,t}, 0\right) + \\ - \gamma \sum_{j=1}^{2} \max\left(Payoff_{i,t} - Others_{i,j,t}, 0\right) \quad (8)$$

We tested the *social* and *non-social learning models* against the two competitive models relying on the BIC. According to this measure, Rescola-Wegner model was better than the baseline model for 62.5% of the participants in the social condition and for only 13% of the participants in the private condition. Fehr-Schmidt model was worse than the baseline model was better for 62.5% of the participants in the social condition and for 37.5% participants in the private condition. Therefore, the social learning model and the non-social learning model performed better than the two competing models. To further express the evidence, we have compared the BIC values of *social* vs. Fehr-Schmidt model and *non-social* vs. Rescola-Wegner model. BIC for the *social model* was better than the Fehr-Schmidt model for 96% of the participants and the *non-social model* was better than the Rescola-Wegner model for 96% participants.

Results

Behavioral results. As expected, subjects depleted the resource of fish significantly faster in the social condition than in the private condition (mean number of trials in the social condition = 6.3 vs. mean number of trials in the private condition = 7.0), $t(1,46)$ = 4.89, p = 0.0001. Furthermore, different styles of harvesting in the two conditions were indicated by a significant interaction of Net Size (one, two, or three fish) × Condition (private, social), $F(2.45)$ = 15.41, p = 0.0001. Subjects used the smallest net size more often in the private condition than in the social condition (Fig. 2), whereas the largest net size was selected more often in the social condition than in the private one. Most importantly, in the social condition after the overexploitation of the fish resource by others (six fish were collected by other players), subjects in return also overexploited the resource in the next trial. In the private condition, in contrast, a similar reduction of the fish stock (six fishes migrated) triggered a trend toward resource preservation. This observation was supported by a significant interaction of Resource Reduction (small, large) × Condition, $F(2.45) = 9.67$, $p = 0.003$ (Fig. 2C).

Neuroimaging results. The sharp depletion of the resource (subtraction of six fish as a result of overexploitation by others or extensive migration) deactivated the ventral striatum more strongly than a moderate change in the resource (subtraction of four or fewer fish) in both conditions (Fig. 3A, Tables S1, $p = 0.001$, uncorrected). However, in the social condition, overexploitation by others evoked stronger deactivation of the ventral striatum than the similar extensive migration (private condition, Fig. 3B). We hypothesized that interindividual variation in the ventral striatum deactivation evoked by the resource depletion would correlate with individual harvesting strategies. Indeed, in the social condition the relative deactivation levels (contrast estimates) evolved by a moderate change in the resource versus a sharp depletion of the resource—observed within peak of the right ventral striatum—negatively correlated with subjects' individual tendency to overcompensate for the resource depletion (Fig. 3C). Thus, the ventral striatum response to the depletion of the resource correlated with opposite behavioral strategies in the two conditions.

To further test the hypothesis that the ventral striatum differently monitors the resource changes in social and private contexts we conducted a more detailed parametric analysis. Using the total number of fish removed from the lake in each trial (by all parties) as the modulation parameter, we found a significant effect of the total resource change on the activity of the ventral striatum: Activity of the ventral striatum negatively correlated with the CPR depletion (total decrease of the CPR, Fig. 4A *middle*, $p = 0.001$, uncorrected). The resource-monitoring modulation of the right ventral striatum activity was significantly stronger in the social condition than in the private condition (Table S2).

As shown in the lower part of Fig. 4, the overexploitation of the CPR was successfully predicted by our social learning model. Using parametric fMRI analyses, we found modulation of the right ventral striatum activity by the reward prediction error signal (encompassing private and social comparison rewards) in the social condition using a more liberal threshold (Fig 4A, *right* and Suppl. Table S2, $p = 0.003$, uncorrected). Similar analysis did not reveal significant modulation ($p < 0.005$) of the ventral striatum activity in the private condition (Suppl. Table S2). These results indicate that the right ventral striatum differentially monitors resources in the social and private conditions. Moreover, the activity of the right ventral striatum is sensitive to the social comparison of the outcomes during CPR depletion. In the social condition, we observed positive task-related functional connectivity (sharp depletion < moderate depletion) between the ventral striatum and the anterior dorsolateral prefrontal cortex (anterior DLPFC): Both decreased activity in response to the overexploitation of the resource by others (Fig. 5, Suppl. Table S3, $p = 0.001$, uncorrected).. In the private condition, anterior DLPFC–ventral striatum connectivity was reduced as a result of a trend toward negative connectivity (Fig. 5, *right*). Interestingly, in the private condition the negative connectivity strength correlated with the tendency to preserve the resource. Thus, the anterior DLPFC could be involved in controlling ventral striatum activity in the context of the private resource, but this control is suppressed during social competition.

Discussion

The results of our study indicate that during the CPR task the ventral striatum encoded opposite harvesting strategies: Relative deactivation of the ventral striatum in response to resource depletion correlated positively with subjects' attempts to preserve their own private resource and correlated negatively with their attempts to preserve the CPR. The ventral striatum receives dopamine projections from the midbrain and is activated by a wide range of rewarding stimuli, from foods, odors, and drugs to beautiful faces (Aharon et al., 2001; Breiter et al., 1997; Gottfried, O'Doherty, & Dolan, 2002; O'Doherty et al., 2004). Activity of the ventral striatum was also associated with social comparison of collected rewards (Fliessbach et al., 2007), voluntarily donations (Harbaugh, Mayr, & Burghart, 2007; Moll et al., 2006), mutual cooperation (Rilling et al., 2002; Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004), and even the punishment of others who have previously behaved unfairly (de Quervain et al., 2004; Singer et al., 2006). A previous study showed that the ventral striatum exhibited more activity when players chose cooperation following a cooperative choice by her partner in the previous round of the iterated Prisoner's Dilemma Game (Rilling et al., 2002). Furthermore, persons with a higher desire for revenge against unfair partners exhibited activation in the nucleus accumbens (Singer et al., 2006). Subjects who made more costly donations to real charitable organizations also exhibited more activity in the striatum (Moll et al., 2006). Overall, our results are consistent with the previous studies indicating the critical role of the ventral striatum in the cooperative behaviour.

To find a computational explanation of the CPR depletion, we developed a computational model suggesting that the ventral striatum generates a reward prediction-error signal where the reward of an outcome is composed of the own monetary reward and a comparison of a person's own outcome with the outcome of others. Our fMRI results indicate that the dopamine system is involved in social comparisons and generates a negative prediction error when a person receives less than the competitors and a positive prediction error when receiving more than the competitors. Thus, ventral striatum activity not only monitors outcomes (resource depletion) but also integrates outcomes into the specific social context. Perhaps the dual nature of the reward-monitoring activity explains our observation that behavioral tendencies underlying competitive depletion of resources are differentially encoded in the activity of the ventral striatum in social and nonsocial contexts. Overall, our results are consistent with the hypothesis that social rewards and social preferences are represented in the ventral striatum similar to primary or monetary rewards (Fehr & Camerer, 2007; Montague & Berns, 2002).

Modeling of the behavioral results further supported the role of social comparisons in overharvesting of CPRs. Perceived depletion of CPRs by others facilitated overharvesting behavior in subsequent trials, particularly by subjects who were more sensitive to social comparison: The individual weights of the model given to the social comparison significantly correlated with the relative increase of harvesting in the trials following CPR depletion (i.e., mean selected net size in the trials following resource depletion by others *minus* mean selected net size in the trials following resource preservation by others), $r = .49$, $p = 0.015$, $n = 24$.

Interestingly, the tendency to preserve private resources also correlated with negative connectivity between the ventral striatum and the anterior DLPFC, which can indicate successful self-regulation of conflicting short-term and long-term goals in the context of private possession. Anterior DLPFC activation has been observed in a variety of response-conflict paradigms (Bench et al., 1993; Carter, Mintun, & Cohen, 1995; van Veen, Cohen, Botvinick, Stenger, & Carter, 2001). Other imaging data have implicated the anterior DLPFC in voluntary decision making under risk (Rao, Korczykowski, Pluta, Hoang, & Detre, 2008) and in second-order control processes, such as integration mechanisms, that are necessary to satisfy more complex or long-term goals (Badre & Wagner, 2004; Braver & Bongiolatti, 2002; Koechlin & Hyafil, 2007; McClure et al., 2004). Our findings indicate a DLPFC influence on the ventral striatum: As it deals with a private resource, the DLPFC can give priority to a more sustainable harvesting strategy that allows for future consumption.

The conclusions of our study need to be limited. Similar to other standard behavioral games participants in our study acted fully anonymously and independently of each other. They were given no opportunity to discuss the situation or to change the institutional rules. However, these opportunities might exist in real-life situations and could also provide a way of avoiding the depletion of the resource (Ostrom, E.,1990). Thus, further studies are clearly needed to investigate strategic aspects of common-pool depletion. Additionally our model of social comparison assumes that receiving more than the competitors is perceived as a positive reward. Although on average this assumption led to a good description of the overall results, there might be an individual difference in social preferences which the model could not account for. Also the follow-up studies will help to examine alternative interpretations of the activity of the ventral striatum observed in our study, e.g. as a neural correlate of the perceived violation of warm glow preferences (Andreoni J., 1990; Harbaugh WT et al., 2007) or of the altruistic norm by others.

For a long time behavioral economics focused on examining factors that favor CPR preservation, including best possible rules, institutions, and communication (Ostrom, 1990). Social psychologists searched for psychological determinants of individual cooperative versus self-interested behavior in commons-dilemma situations (Messick et al., 1983). Our results show that the context of a shared resource versus a private resource (with similar control over the resources in both contexts) modulates neural activity and connectivity of the ventral striatum—a brain area strongly associated with the valuation of outcomes. Overall, the notion of the neurobiological underpinnings of resource overexploitation could help us to develop efficient boundary rules and a better understanding of global commons governance.

Acknowledgments: We thank Markus Klarhoefer, Oliver Schürmann and Stefan Thommen for assistance in the fMRI experiments and Jet Tang for assistance with the computational analysis. This study received financial support from SNSF grant no. 100014-130352 to Vasily Klucharev and Jörg Rieskamp.

References:

Andreoni J. (1990) Impure Altruism and Donations to Public-Goods - a Theory of Warm-Glow Giving. *Economic Journal* 100(401):464-477 (in English).

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron, 32*(3), 537-551.

Badre, D., & Wagner, A. D. (2004). Selection, integration, and conflict monitoring: Assessing the nature and generality of prefrontal cognitive control mechanisms. *Neuron, 41*(3), 473-487.

Bault, N., Joffily, M., Rustichini, A., & Coricelli, G. (2011). Medial prefrontal cortex and striatum mediate the influence of social comparison on the decision process. *Proceedings of the National Academy of Sciences of the United States of America, 108*(38), 16044-16049.

Bench, C. J., Frith, C. D., Grasby, P. M., Friston, K. J., Paulesu, E., Frackowiak, R. S. J., et al. (1993). Investigations of the Functional-Anatomy of Attention Using the Stroop Test. *Neuropsychologia, 31*(9), 907-922.

Braver, T. S., & Bongiolatti, S. R. (2002). The role of frontopolar cortex in subgoal processing during working memory. *Neuroimage, 15*(3), 523-536.

Breiter, H. C., Gollub, R. L., Weisskoff, R. M., Kennedy, D. N., Makris, N., Berke, J. D., et al. (1997). Acute effects of cocaine on human brain activity and emotion. *Neuron, 19*(3), 591-611.

Camerer, C. (2003). *Behavioral Game Theory: Experiments on Strategic Interaction*. New Jersey Princeton University Press.

Camerer, C., & Ho, T. H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica, 67*(4), 827-874.

Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: An (H2O)-O-15 PET study of Stroop task performance. *Neuroimage, 2*(4), 264-272.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science, 305*(5688), 1254-1258.

Dvash, J., Gilam, G., Ben-Ze'ev, A., Hendler, T., & Shamay-Tsoory, S. G. (2010). The Envious Brain: The Neural Basis of Social Comparison. *Human Brain Mapping, 31*(11), 1741-1750.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54*(2), 293-315.

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn Sci, 11*(10), 419-427.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*(6960), 785-791.

Fehr, E., & Schmidt , K. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics, 114*(3), 817–868.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117-140.

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., et al. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science, 318*(5854), 1305-1308.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage, 6*(3), 218-229.

Glascher, J., Hampton, A. N., & O'Doherty, J. P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cereb Cortex, 19*(2), 483-495.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J Neurosci, 22*(24), 10829-10837.

Hampton, A. N., Adolphs, R., Tyszka, M. J., & O'Doherty, J. P. (2007). Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. *Neuron, 55*(4), 545-555.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences of the United States of America, 105*(18), 6741-6746.

Harbaugh, W. T., Mayr, U., & Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science, 316*(5831), 1622-1625.

Hardin, G. (1968). Tragedy of Commons. *Science, 162*(3859), 1243-1248.

Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science, 318*(5850), 594-598.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science, 306*(5695), 503-507.

Messick, D. M., Wilke, H., Brewer, M. B., Kramer, R. M., Zemke, P. E., & Lui, L. (1983). Individual Adaptations and Structural-Change as Solutions to Social Dilemmas. *Journal of Personality and Social Psychology, 44*(2), 294-309.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc Natl Acad Sci U S A, 103*(42), 15623-15628.

Montague, P. R., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron, 36*(2), 265-284.

Montague, P. R., King-Casas, B., & Cohen, J. D. (2006). Imaging Valuation Models in Human Choice. *Annu Rev Neurosci, 29*, 417-448.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304*(5669), 452-454.

Ostrom, E. (1990). *Governing the commons: the evolution of institutions of collective action*. Cambridge: Cambridge University Press.

Ostrom, E. (2009). A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science, 325*(5939), 419-422.

Rao, H., Korczykowski, M., Pluta, J., Hoang, A., & Detre, J. A. (2008). Neural correlates of voluntary and involuntary risk taking in the human brain: An fMRI Study of the Balloon Analog Risk Task (BART). *Neuroimage, 42*(2), 902-910.

Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Classical Conditioning II: Current Research and Theory (Black AH, Prokasy WF, eds), pp 64-99: Appleton-Century-Crofts.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., & Kilts, C. (2002). A neural basis for social cooperation. *Neuron, 35*(2), 395-405.

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *Neuroreport, 15*(16), 2539-2543.

Schwarz G. (1978) Estimating the Dimension of a Model. Ann Statist 6:461-464.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*(7075), 466-469.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An introduction.*: MIT Press.

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *Neuroimage, 14*(6), 1302-1308.

Figures

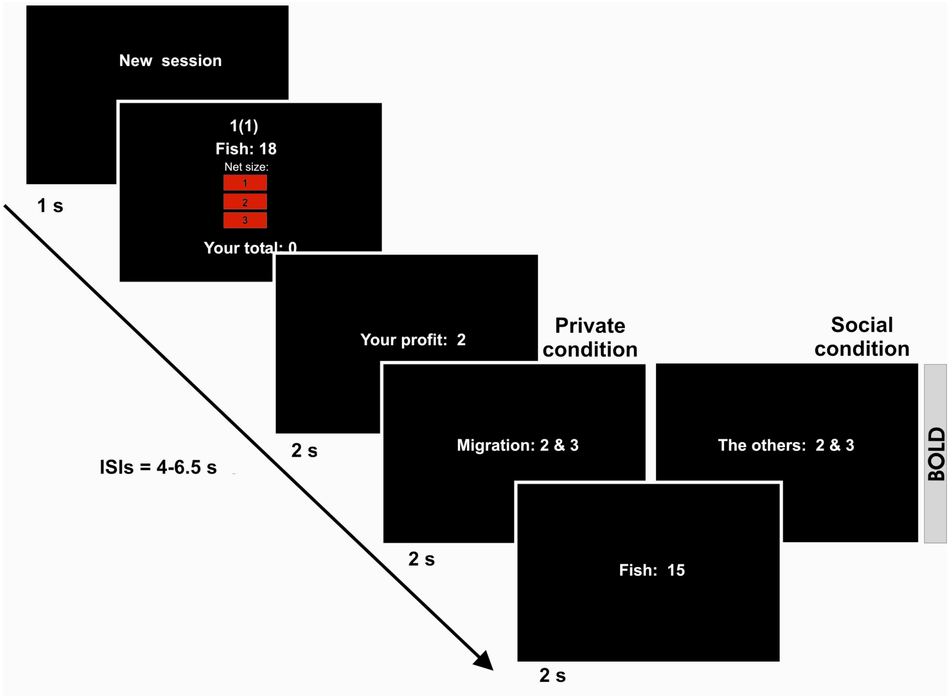


Fig. 1. The private and social versions of the common-pool resource (CPR) task. The sequence of events within a trial is shown. Subjects removed 1, 2, or 3 fish from the CPR and observed follow-up "migration" of the fish into two neighboring lakes (private condition) or "harvesting of fish" by two prerecorded subjects (social condition). At the end of each trial and at the beginning of the next trial subjects were informed about the remaining number of fish in the CPR. ISI: interstimulus interval.

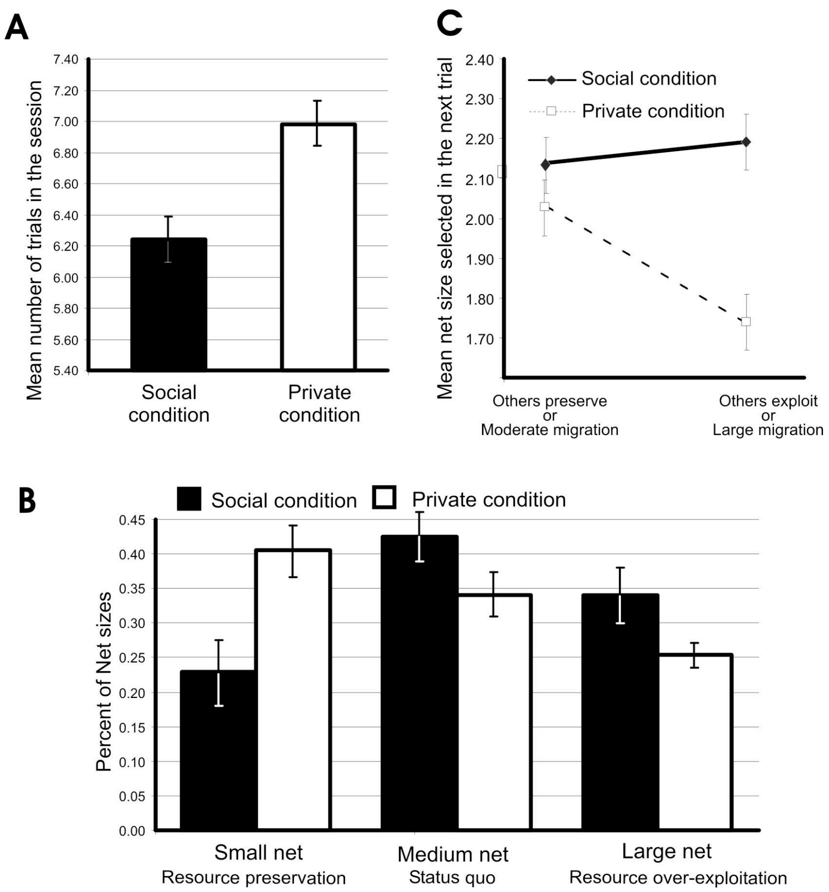


Fig. 2. The effects of experimental conditions on resource depletion in Behavioral Study 2. Subjects used the larger net size and depleted the fish resource faster in the social condition than in the private condition, similar to the main fMRI study. **A**) Mean number of trials per session in two experimental conditions. The graph illustrates faster depletion of the resource in the social condition than in the private one. Each session continued as long as the resource was sustained, with a maximum of 8 trials. **B**) Mean harvest decision. Subjects decided to take one fish more often in the social condition than in the private one. The opposite was true for the largest net size of three fishes. **C**) Mean harvest decision (in next trials) following resource depletion / preservation due to behavior of others or migration. After the depletion of the resource by others (social condition), subjects also depleted the resource in the next trial, whereas in the private condition, identical reduction of the fish stock followed by a trend of resource preservation.

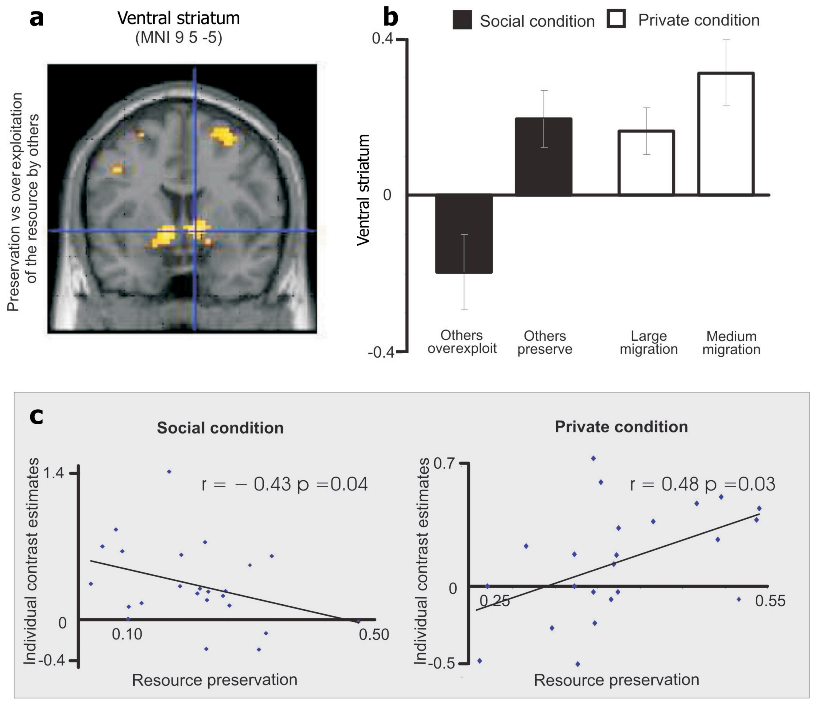


Fig. 3. General effects of resource depletion: neural response to sharp resource depletion (6 fish removed from the resource as a result of migration or overharvesting by others) vs. neural response to resource preservation (2–4 fish removed from the resource). (A) *Z*-maps of deactivations induced by resource depletion in both experimental conditions. (B) The signal change of the hemodynamic response evoked by overexploitation/preservation (social condition, n = 25) or by large/moderate migration (private condition, n = 21). (C) The ventral striatum recruitment (MNI: 9, 5,-5) evoked by perceived resource depletion predicted individual differences in resource preservation. Interestingly, in the social condition (left side) stronger deactivation of the ventral striatum evoked by others' common-pool resource (CPR) overexploitation negatively correlated with subjects' own CPR preservation behavior (proportion of the smallest net size in harvest decisions). In contrast, in the private condition (right side) stronger deactivation of the ventral striatum evoked by extensive migration positively correlated with subjects' own resource preservation behavior.
The data is thresholded at $p < 0.001$, uncorrected.

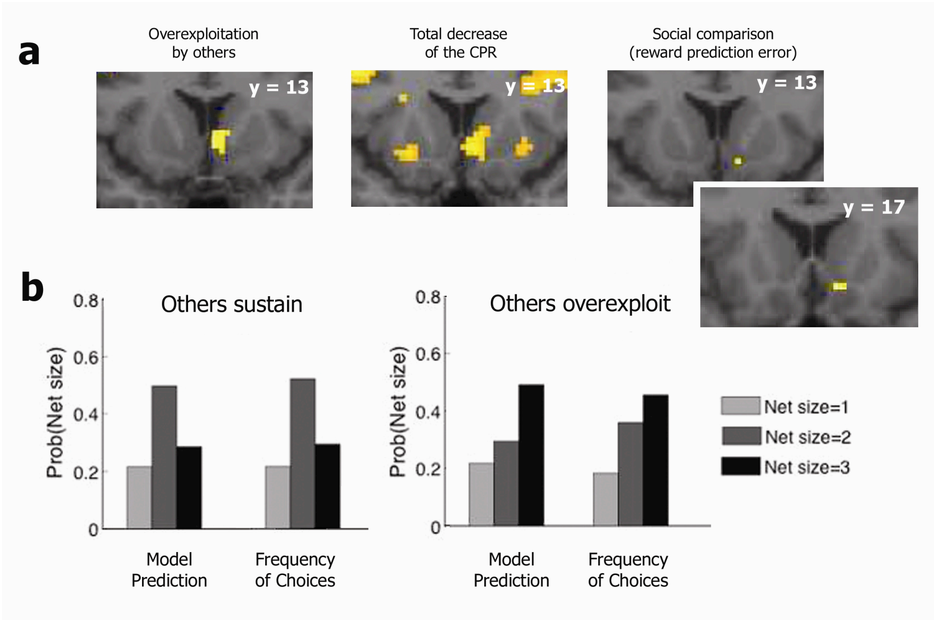


Fig. 4. The computational mechanism of the resource depletion: the right ventral striatum monitors crucial aspects of CPR exploitation.

(**A**) The role of the ventral striatum in CPR monitoring. *Left side*: The right ventral striatum reacted to the sharp depletion of CPR by others. *Center*: Furthermore, activity of the right ventral striatum was involved into in depth monitoring of the overall size of CPR: the activity was parametrically modulated the by the total change of the size of CPR in each trial. *Right side*: Importantly, the right ventral striatum generated the learning signal (reward prediction error signal encompassing *private* and *social comparison* rewards) underlining overexploitation of CPR*.

(**B**) The *social learning model* (incorporating the reward-prediction error signal) correctly predicted the actual behavioral patterns in the social condition. *Left side*: Average probability of choosing *Net size* ∈ {1,2,3} after the CPR depletion by others in the previous trial (six fishes had been removed from the lake by others) accurately predicted the observed high frequency of the overharvesting (selection of the largest *Net size*). *Right side*: the model also accurately predicted the strong tendency for the CPR preservation after the CPR preservation by others in the previous trial (four fishes had been removed from the lake by others). On both sides, model predictions refer to the probability values obtained from the models with fitted parameters, whereas the data is the observed frequency of choices.

* fMRI data for Fig.3A, *Left and Center* is thresholded at $p < 0.001$, for Fig.4A, *Right* at $p < 0.003$, cluster size =3, uncorrected.

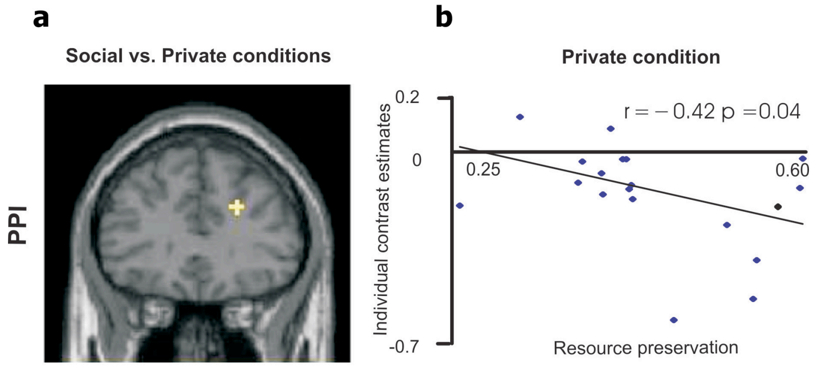


Fig. 5. Stronger functional connectivity between the right ventral striatum and the anterior dorsolateral prefrontal cortex (DLPFC) in the social condition.

(A) Z-maps for psychophysiological interaction (PPI).

(B) A trend toward negative connectivity of the right ventral striatum and the anterior DLPFC was observed in the private condition. The effect negatively correlated with resource preservation behavior.

The data is thresholded at $p < 0.001$, uncorrected.

Supplemental Information

Vasily Klucharev, Sandra Andraszewicz, Jörg Rieskamp

Table of contents

Supplemental Experimental Procedures

I. Behavioral Study 1.

The first behavioral study consisted of a common pool resource (CPR) experiment. The aim of the experiment was to test the CPR paradigm and to collect behavioral data for the follow-up behavioral and imaging studies.

Participants. In the experiment, three persons interacted simultaneously with each other facing the CPR. The participants (N = 24, aged 18–28 years, mean 21.8 years, 9 females) simultaneously participated in a common pool resource (CPR) task (see Fig.1 in the main text). Participants encountered 20 sessions of the CPR task (8 trials per session). The task was performed in large groups (6 people) in separate cubicles to ensure subjects' anonymity.

Experiment. Subjects were confronted with a "common" resource of fish. Subjects were informed that they were participating in a "Fishing Study" project investigating decision making. Participants had to imagine that they were fishing at a lake together with two other fishermen. Their task was to collect as much fish as possible and each collected fish led to a monetary payoff (0.25 Swiss Franks per fish). In every trial, participants decided between three possible net sizes for fishing one, two, or three fishes. Overall, depletion of the resource was caused by own behavior and the behavior of two other anonymous players present in the room. Subjects were informed that although the number of fishes in the lake decreases by fishing, it also grows naturally due to proliferation of fish. Therefore, at the end of every trial, the remaining number of fish in the lake was multiplied by 1.5, which gave the total number of fishes for the next trial (with a maximum number of 20 fishes representing the utmost capacity of the lake). In case no fish remained for the next trial, the whole session ended automatically. The instructions clearly explained to the subjects that the amount of fish removed by the players could increase, sustain, or decrease the fish population. For example, the participants were informed that whenever the total number of fishes collected by the three participants was smaller than six, the fish population would increase over the trials. In contrast, whenever the total number of fishes collected by the three participants was larger than six, the fish population would decrease over the trials. If the total number of fishes collected by the three persons was equal to six, the fish population would stay constant over the trials. Thus, the net size of 2 fishes corresponded to a cooperative/sustainable level of harvesting. The experiment started with a short training session. The CPR was programmed with the software z-Tree (Fischbacher U., 2007). The study design was similar to the social version of the CPR task in the follow-up fMRI study, with the difference that all participants in the lab made decisions at the same time.

Results. Overall, the participants did not follow the game-theoretical prediction of completely self-interested people who would always select the largest net size for all trials in the game. Nevertheless the participants overharvested and depleted the CPR: on average 58.7% (SD = 32.5) of sessions were completed before the 8^{th} trial, which indicated overharvesting behavior (mean number of trials in a session = 7.4). The average selected net size (net size = 2.3) was significantly higher than the "sustainable" size of the net (net size = 2), $t(1,23) = 5.73$, $p = 0.000008$. Two highly competitive subjects (the average net size = 2.6 and 2.7) were selected for the fMRI version of the study and their behavioral results were used in the social and private conditions.

II. Behavioral Study 2.

The goal of the second behavioral study was to examine how people deal with a social as compared to a private resource situation. We used a modified version of the CPR task from Behavioral Study 1. The experiment was identical to the fMRI experimental design, but it was conducted in a behavioral experimental laboratory.

Participants. We invited thirty-seven healthy students to test the CPR task for the follow-up fMRI study. In order to avoid any demand effects and suspicion toward the two different (but structurally identical) conditions, we implemented a between-subjects design: Subjects were randomly assigned to the social or private condition of the CPR task (with N = 19 for the social and N = 18 for the private condition). Overall they encountered 16 sessions (8 trials per session).

Experiment. In every trial, participants decided between three possible net sizes for fishing one, two, or three fishes. In the social version of the experiment (social condition), two other participants (pre-recorded from Study 1) also decided between the three net sizes. In the non-social version of the experiment (private condition), the same number of fishes “migrated” to two neighboring lakes. Importantly, the change of the resources due to the two other pre-recorded participants or the "migration" to the two neighboring lakes was identical in both conditions.

Results. Similar to the fMRI experiment, subjects depleted the resource of fish significantly faster in the social condition than in the private condition (mean number of trials in the social condition = 6.24 vs. 7.00 in the private condition, $t(1,35) = 3.30$, $p = 0.002$, Fig.S1a). The average selected net size was significantly larger in the social condition (2.09) than in the private condition (1.85), $t(1,35) = 2.25$, $p = 0.015$. We observed different styles of behavior in the two conditions as indicated by a significant interaction *net size (one, two, three fishes*) × *condition* (*private, social*): $F(2.34) = 3.99$, $p = 0.028$: Fig. S1 illustrates that subjects more often used the smallest net size in the private condition than in the social condition (Fig.S1b) and the largest net size was selected more often in the social condition than in the private one. Similar to the results in the fMRI study, in the social condition, after the overexploitation of the fish resource by others (6 fishes were collected by other players), subjects then also overexploited the resource in the next trial. However, in the private condition, a similar reduction of the fish stock (6 fishes migrated) led to resource preservation. This observation was supported by a significant interaction *resource reduction (small, large)* x *condition,* $F(1.35) = 7.44$, $p = 0.010$ (Fig.S1c). Overall, the results were later replicated in the behavioral results of the fMRI study reported in the main text, providing independent additional evidence for the observed results.

III. Game-theoretical analysis of the CPR task

What is the game-theoretical solution for the fishing game when assuming only self-interested (i.e. payoff-maximizing) players? In the CPR situation, the solution can be easily determined by backward induction. The task has a final number of trials which are common knowledge to all players. Therefore it is clear that in the very last trial, it is best for everyone to choose the largest net size to maximize payoffs. Given this behavior, it is also rational to choose the largest net size in the second-last trial, and so on. Therefore the game-theoretical solution is to choose the largest net size in all trials of the task.

How should a self-interested player behave in the non-social situation (private condition), in which no other players are involved? Here the solution depends on a person's belief about the amount of fish that migrates to the two other lakes. If a person believes that the migration rate is low, then the person should choose the largest net size all the time. In contrast, if the player believes that the migration rate is high, it can be payoff-maximizing to choose a small net size to sustain the resource to allow for future consumption. However, the optimal behavior will depend on the specific beliefs about the migration rate. When assuming uniform priors of players' beliefs about the migration rate, it can be predicted that the consumption rate should be lower in the private condition than in the social condition, which is in line with the behavioral findings.

More specifically, we determined the optimal behavioral strategy for the game in the private condition given different beliefs about the migration rate. The migration to the first lake is represented by L_t^1, the migration to the second lake by L_t^2, and its sum represents the total migration L_t for trial t. The beliefs about migration can be represented by the probability with which a player believes that the particular migration rate occurs, that is $\Pr(L_t^1)$ and $\Pr(L_t^2)$ (note the migration to each lake is discrete and ranges between 1 and 3 fishes).

We implemented three different assumptions about the players' beliefs: First, we assumed uniform priors, that is all three possible migration rates for each lake were equally likely. Second, we assumed that the players' beliefs about the different migration rates would reflect the average migration observed in the whole task (thus, if a migration of 2 fishes to one lake occurred in 50% of all trials and sessions, the probability would be .50). Third, we assumed that a player would start with an initial belief that every migration rate would be equally likely. After completing the first session, this belief is updated according to the observed migration rates in each trial. To update the belief after the completion of session S we determine:

$$\Pr(L_t^i)_S = \left\lfloor 1 + \sum_{j=1}^{S} f_j(L_t^i) \right\rfloor / (S+1), \tag{1}$$

where L represents the three possible migration rates of 1, 2, or 3 to one of the two lakes i, and $f(.)$ represents an indicator function that takes a value of 1 if the particular migration occurred in the trial t and a value of 0 otherwise.

Given the players' beliefs about the possible migration rates, we determined the optimal strategy for the whole task, specifying the chosen net size for each trial of the task using a dynamic programming approach. For the very last trial (i.e. the eighth trial) it is possible to determine the expected payoff of choosing each of the three possible net sizes. The expected payoff in the eighth trial depends on the chosen net size, the remaining number of fishes, and the probability of the different migrations rates. It can be easily seen that the highest expected payoff will always result from choosing the largest net size in the last trial. From this eighth trial we determine the optimal strategy in the seventh trial. Here, a complete strategy specifies the chosen net size for the seventh trial and the eighth trial. The expected payoff for the strategies in trial seven depends on the payoff in the seventh and the eighth trial. The payoff in the eighth trial depends on the remaining number of fishes, which depends on the chosen net size and migration in the seventh trial. Thus, the chosen net size in the seventh trial does not only influence the immediate payoff but also the possible payoffs in the eighth trial. Following this approach, the optimal strategy can be determined for the sixth trial, where the expected payoffs of all possible strategies depend on the chosen net size in the sixth trial and the chosen net sized in the seventh and eighth trial. Following this backward analysis one can determine the overall best strategy for the whole task starting in the very first trial.
Mathematically, the expected payoff given a player's strategy for the whole task is calculated as the total payoff that can be obtained in the task multiplied by the probability of obtaining this payoff given a particular strategy:

$$EV[Payoff \mid Strategy] = \sum_{payoff} payoff \cdot \Pr(Payoff = payoff \mid Strategy) \qquad (2)$$

The probability of obtaining a specific total payoff depends on the strategy. On the one hand, the strategy defines the net size and affect the payoffs, but it also affects the development of the resource and thereby the size of the resource in subsequent trials .

The results of this analysis are illustrated in Figure S2. When assuming that all migration rates are equally likely ("Belief 1"), then according to the best strategy, one should choose a net size of 1 in the very first trial, increase the net size to 2 in trial two to four and starting from trial five, one should always choose net size 3. When assuming that the players would know the actual migration rates in all trials, they should also choose net size 1 in trial 2, and net size 2 for trials two to four, and always a net size of 3 from trial five onwards. Finally, when assuming equal priors for the first trial that are updated on the observed migration rates, then it is optimal to choose net size 1 for trial 1 and to increase the net size for the following trials with a net size of 3 starting from trial six onwards. Overall, the analysis shows that given a variety of beliefs, the payoff maximizing strategy is not to choose the largest net size at the beginning of the task in the private condition, but to choose the largest net size at the end, starting at the sixth trial at the latest. Thus, according to this analysis, one would expect smaller net sizes in the private as compared to the social condition at the beginning of the task, which is consistent with the experimental findings.

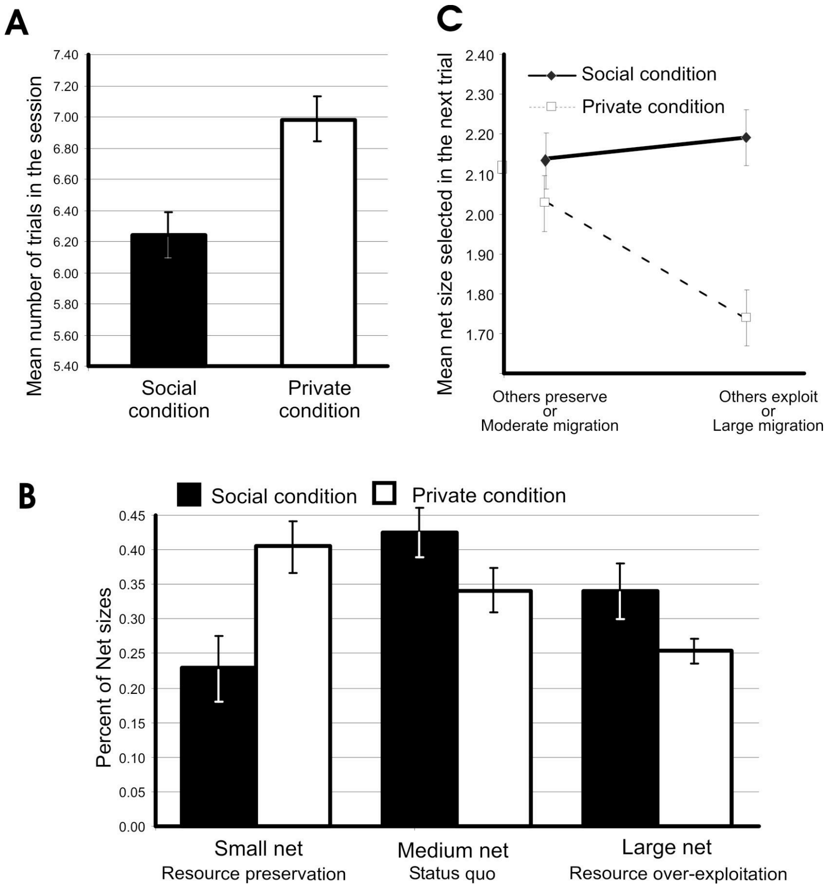


Fig. S1. The effects of experimental conditions on resource depletion in Behavioral Study 2. Subjects used the larger net size and depleted the fish resource faster in the social condition than in the private condition, similar to the main fMRI study. **A**) Mean number of trials per session in two experimental conditions. The graph illustrates faster depletion of the resource in the social condition than in the private one. Each session continued as long as the resource was sustained, with a maximum of 8 trials. **B**) Mean harvest decision. Subjects decided to take one fish more often in the social condition than in the private one. The opposite was true for the largest net size of three fishes. **C**) Mean harvest decision (in next trials) following resource depletion / preservation due to behavior of others or migration. After the depletion of the resource by others (social condition), subjects also depleted the resource in the next trial, whereas in the private condition, identical reduction of the fish stock followed by a trend of resource preservation.

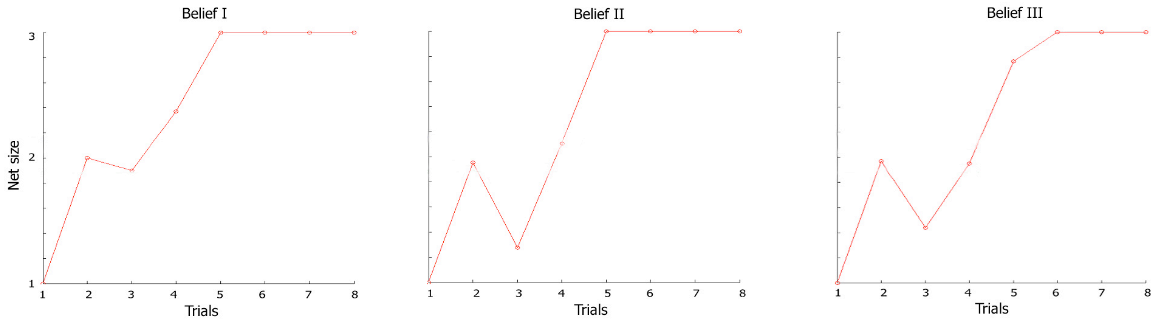


Fig. S2 The optimal behavioral strategy to maximize payoff in the non-social condition. Belief I assuming equal prior beliefs for the three possible migration rates. Belief II assuming beliefs corresponding to the actual migration rates. Belief III assuming equal prior beliefs for the three possible migration rates for the first round of all games and updating of these beliefs for the following rounds.

Supplementary Table S1. Significant activation clusters to sharp resource depletion in both experimental conditions.

Brain Region	x	y	z	No. of Voxels	Z
moderate depletion of the resource < sharp depletion of the resource					
Superior temporal gyrus, BA42	66	-31	22	46	4.45
Postcentral gyrus, BA43	51	-13	19	43	3.83
Superior temporal gyrus, BA38	45	17	-23	11	3.46
Amygdala/Parahippocampal gyrus	27	-7	-17	7	3.44
moderate depletion of the resource > sharp depletion of the resource					
Middle frontal gyrus, BA46	-48	23	31	196	5.22
Cerebellum	42	-67	-41	356	4.74
Cerebellum	-36	-73	-47	175	4.65
Middle frontal gyrus, BA9	48	29	31	522	4.60
Ventral striatum	-12	2	-8	55	4.57
Superior parietal lobule, BA7	-33	-52	49	277	4.41
Ventral striatum	9	5	-5	128	4.39
Inferior parietal lobule, BA40	42	-43	46	258	4.27
Superior frontal gyrus, BA10	-30	59	10	154	4.10
Cuneus, BA17	15	-88	4	117	4.08
Middle frontal gyrus, BA6	30	17	58	94	4.04
Middle frontal gyrus, BA6	-24	17	61	37	3.84
Thalamus	24	-19	19	7	3.57
Middle occipital gyrus, BA18	-24	-85	-5	26	3.56
Middle occipital gyrus, BA18	-21	-91	10	12	3.55
Precuneus, BA7	12	-58	43	14	3.52
Precentral gyrus, BA6	36	-1	28	8	3.51
White matter/ Parahippocampal gyrus, BA19	30	-46	22	6	3.38
Middle temporal gyrus, BA20	57	-43	-11	7	3.28

Local maxima within these clusters are reported together with the number of voxels (No. of Voxels); BA, Brodmann area; x, y, z are MNI coordinates of the local maximum.
Thresholded at $p < 0.001$, cluster size =3.

Supplementary Table S2. Brain regions parametrically modulated by the social and non-social versions of prediction error in the social and private conditions, correspondingly.

Brain Region	x	y	z	No. of Voxels	Z
Social condition					
	negative modulation				
Inferior parietal lobule, BA40	-42	-43	43	102	3.40
Middle frontal gyrus, BA8	-24	8	43	7	3.10
Cerebellum	30	-73	-47	9	3.03
Middle frontal gyrus, BA6	-27	-4	52	6	2.94
Cerebellum	-39	-58	-32	30	2.93
Middle frontal gyrus, BA46	-45	38	19	13	2.91
Ventral striatum	12	17	-11	4	2.75
	positive modulation				
Inferior frontal gyrus, BA47	-33	26	-20	38	4.28
Cingulate gyrus, BA32	-15	20	40	11	3.90
Putamen	-33	-22	-5	17	3.68
Medial frontal gyrus/Anterior cingulate BA10/32	-6	47	13	78	3.55
Cingulate gyrus, BA24	12	-1	37	21	3.52
Superior temporal gyrus, BA42	66	-13	7	19	3.39
Superior temporal gyrus, BA22	-51	-13	-8	35	3.38
Middle temporal gyrus, BA22	-60	-37	4	48	3.28
Superior frontal gyrus, BA6	-9	26	61	22	3.21
Cingulate gyrus, BA24	-6	-10	40	11	3.21
Precentral gyrus, BA6	54	-1	31	10	3.13
Midbrain	0	-31	-5	17	3.06
Insula, BA13	-33	-25	16	12	3.00
Precentral gyrus, BA4	48	-16	55	19	2.99
Fusiform gyrus, BA37	-42	-46	-17	5	2.99
Cuneus, BA18	18	-85	22	8	2.90
Precentral gyrus, BA6	-51	-7	10	4	2.90
Thalamus, BA	3	-13	-2	11	2.88
Superior frontal gyrus, BA9	-15	50	31	19	2.84
Middle temporal gyrus, BA39	-51	-61	7	18	2.81
Private condition					
	negative modulation				
none					
positive modulation					
Insula, BA 13	-39	3	20	103	3.75
Insula, BA 13	-36	-21	10	17	3.74
Cingulate gyrus, BA 24/32	9	7	45	75	3.66
Precentral gyrus, BA 4/6	56	-11	32	59	3.50
Cingulate gyrus, BA 32	-15	21	36	36	3.42
Cingulate gyrus, BA 24	-9	1	42	25	3.31
Cerebellum	-12	-37	-25	9	3.18
Precuneus/Cingulate gyrus BA 7/31	21	-48	42	6	3.11
Thalamus, medial dorsal nucleus	12	-21	10	7	3.09
Inferior frontal gyrus, BA 45	-39	23	5	12	3.07
Medial frontal gyrus, BA 6	-12	-25	52	6	3.06
Superior temporal gyrus, BA 22	53	-10	4	5	3.05
Insula, BA 13	33	-18	18	19	3.04

Superior temporal gyrus, BA 22	-50	-12	4	35	3.01
Cingulate gyrus, BA 31	15	-23	35	25	2.94
Inferior frontal gyrus, BA 45	59	15	22	5	2.87
Middle temporal gyrus, BA 37	53	-56	-1	6	2.84
Middle temporal gyrus, BA 39	-45	-75	30	6	2.75

Thresholded at $p < 0.003$, cluster size =3.

Supplementary Table S3. Stronger functional connectivity between the (deactivated) right ventral striatum and the anterior DLPFC in the social condition.

Brain Region	x	y	z	No. of Voxels	Z
	negative modulation				
none					
	positive modulation				
Cingulate gyrus, BA 13 /white matter	-24	-16	34	28	4.00
Cingulate gyrus, BA 31	18	-46	37	9	3.91
DLPFC:Medial frontal gyrus, BA 9	24	32	28	12	3.82

Thresholded at $p < 0.001$, cluster size =3.

Supplemental References

1. Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. (Translated from English) *Experimental Economics* 10(2):171-178 (in English).

A PRACTICAL TUTORIAL ON BAYESIAN HIERARCHICAL REGRESSION IN MANAGEMENT SCIENCE

Sandra Andraszewicz
University of Basel, Switzerland

Benjamin Scheibehenne
University of Basel, Switzerland

Jörg Rieskamp
University of Basel, Switzerland

Raoul Grasman
University of Amsterdam, The Netherlands

Josine Verhagen
University of Amsterdam, The Netherlands

Eric-Jan Wagenmakers
University of Amsterdam, The Netherlands

Corresponding author: Sandra Andraszewicz, University of Basel Missionsstrasse 62a 4055 Basel, Switzerland

Email: s.andraszewicz@unibas.ch

ABSTRACT

In management science, hierarchical regression models are almost always analyzed using null-hypothesis significance testing (NHST). Here we outline the conceptual and practical advantages of an alternative analysis method: Bayesian hypothesis testing and model selection using the Bayes factor. In contrast to NHST, Bayes factors allow researchers to quantify evidence in favor of the null-hypothesis. Also, Bayes factors do not require adjustment for the intention with which the data were collected. The use of Bayes factors for hierarchical regression is demonstrated with an extended example based on the design of an experiment recently published in the *Journal of Management* (Dierdorff et al., 2012).

Key words: Bayes factor, statistical evidence, optional stopping

A PRACTICAL TUTORIAL ON BAYESIAN HIERARCHICAL REGRESSION IN MANAGEMENT SCIENCE

Across the empirical sciences, one of the most popular and well-known statistical tools is regression analysis: a dependent or criterion variable (e.g., income) is accounted for by a weighted combination of independent or predictor variables (e.g., level of education, age, gender, etc.). In management science, the inclusion of particular predictor variables often amounts to the test of a specific theory or hypothesis, in the sense that statistical support for the inclusion of the predictor variables yields conceptual support for the theory that postulated the importance of those variables.

Almost always, researchers gauge the statistical support for the inclusion of particular predictors by means of the p-value obtained from null-hypothesis significance testing (NHST). Formally, the p-value is defined as the probability of encountering a test statistic at least as extreme as the one that was observed, given that the null-hypothesis is true (Schervish, 1996). Predictors whose weights are significantly different from zero (i.e., $p < .05$) are cause to reject the null-hypothesis, and the lower the p-value, the more compelling the evidence against the null-hypothesis.

Unfortunately, p-values have a number of serious statistical limitations (e.g., Wagenmakers, 2007). In particular, p-values cannot quantify evidence in favor of a null-hypothesis (e.g., Gallistel, 2009; Rouder, Speckman, Sun, Morey, & Iverson, 2009), they overstate the evidence against the null-hypothesis (e.g., Berger & Delampady, 1987; Edwards, Lindman, & Savage, 1963; Johnson, in press; Sellke, Bayarri, & Berger, 2001), and they depend on the sampling plan, that is, they depend on the intention with which the data were collected (e.g., Berger & Wolpert, 1988).

The most prominent alternative to orthodox hypothesis testing using p-values is Bayesian hypothesis testing using so-called Bayes factors (e.g., Jeffreys, 1961; Kass & Raftery, 1995). Bayes factors quantify the support that the data provide for one model versus another; using Bayes factors, researchers can quantify evidence for any hypothesis (including the null), and monitor this evidence as the data come in. In Bayesian inference, the intention with which the data are collected is irrelevant. As will be apparent later, inference using p-values can differ dramatically from inference using Bayes factors. We believe that such differences should be acknowledged rather than ignored.

The main goal of this article is to explain the conceptual foundations and practical complications of model selection and hypothesis testing using Bayes factors. For concreteness, we focus on the case of hierarchical regression analysis as it is commonly conducted in management science. To set the stage, below we first provide an overview of the current analysis method for hierarchical regression. Next, we outline the conceptual basis of Bayesian inference in general and Bayesian hypothesis testing using Bayes factors in particular. We then apply the Bayes factor methodology to a concrete example inspired by the recent work of Dierdorff, Rubin, and Bachrach (2012). The appendix provides R code that allows the reader to carry out Bayes factor hierarchical regression analysis on other data sets.

HIERARCHICAL REGRESSION IN MANAGEMENT SCIENCE: CURRENT STATUS

In a hierarchical regression analysis, predictor variables are added to the regression equation sequentially, either one by one or in batches. The sequence by which the predictors are entered is determined by their hierarchy, which is motivated by theoretical considerations and the structure of the data. Usually, the batch of predictors added in the first step represent nuisance variables that are outside the immediate focus of interest. Such variables may include demographic information such as socioeconomic status, gender, and age. In the next step, the researcher adds a variable of interest (e.g., communication style) and judges the extent to which this variable adds anything over and above the nuisance variables that were added in the first step. At every next step, new predictors can be added to the regression equation, and the order of inclusion usually reflects an increasing level of sophistication of the hypotheses under consideration. For instance, the third step may feature a predictor that quantifies the interaction between communication style and pro-social role expectations. At any step, the statistical support for the hypothesis that postulates the presence of the new predictors is determined by the increase in variance explained as formalized by an F-test (Cohen & Cohen, 1983).

To illustrate the standard hierarchical multiple regression analysis procedure we turn to a classical example from Cohen and Cohen (1983). Length of the stay in a psychiatric hospital was hypothesized to depend on (1) the patient's demographic characteristics such as age or socioeconomic status; (2) the patient's personality; and (3) the hospital the patient was assigned to.

To assess the evidence for each of the above three hypotheses, Cohen and Cohen (1983) conducted a hierarchical regression analysis. In the first step, the regression model featured 9 demographic measures. Table 1 shows that, compared to the null model without predictors, the increase in variance explained is highly significant. Hence, the first hypothesis is supported: demographic characteristics matter for determining length-of-stay.

In the second step, the regression model is expanded to include 9 scales of the Minnesota Multiphasic Inventory (MMPI), a widely used personality test, and a dichotomous variable indicating missingness (number of independent variables $k = 10$). Table 1 shows that, compared to the model that includes the demographic predictors, adding the personality predictors increased the explained variance by only 2%. This increase is not statistically significant, and hence there is no evidence that a patient's personality impacts length-of-stay.

In the final and third step, the regression model is further expanded to include
8 hospitals (number of independent variables k=7, as one of the hospitals serves as the reference group). In total, the third model featured 26 predictors. As shown in Table 1, inclusion of the hospital predictor variables resulted in an 11% increase in the proportion of variance explained. This increase is highly significant, and therefore the data support the hypothesis that length-of-stay is determined partly by the hospital that a patient is assigned to.

Insert Table 1 about here

Although the standard hierarchical regression analysis procedure seems straightforward, the statistical inference ultimately hinges on p-value methodology. This methodology is based on classical, orthodox, or frequentist statistics, in which probability is conceptualized as the proportion of occurrences in the large-sample limit. An alternative statistical paradigm, whose popularity has risen tremendously over the past 20 years (e.g., Poirier, 2006), is Bayesian inference. In Bayesian inference, probability is used to quantify uncertainty or degree-of-belief. As we demonstrate in the next section, the differences between frequentist and Bayesian statistics are considerable, both in theory and in practice.

BAYESIAN INFERENCE

The many aspects of Bayesian inference are explained in detail elsewhere (e.g., Dienes, 2008; Lee & Wagenmakers, in press; Kruschke, 2010; and the articles in this special issue, such as Zyphur & Oswald, in press). Here we explain the essentials in as far as they are required to understand, at a conceptual level, the material covered in later sections.

For concreteness, consider the height advantage of candidates for the US presidency (Stulp, Buunk, Verhulst, & Pollet, 2013). The data from 46 US presidential elections can be analyzed in multiple ways, but here we are concerned with the Pearson correlation between the proportion of the popular vote and the height ratio (i.e., height of the president divided by the height of his closest competitor). Figure 1 shows that taller candidates tend to attract more votes; the sample correlation is .39 and is significantly different from zero (p = .007, two-sided test).

Insert Figure 1 somewhere here

A Bayesian analysis may proceed as follows. The model under consideration assumes that the data come from a bivariate normal distribution, and interest centers on the unknown correlation coefficient ρ. In Bayesian statistics, the uncertainty about ρ before seeing the data is quantified by a probability distribution known as the prior. Here we specify a default prior distribution, one that expresses that we do not have any knowledge about the size of the correlation coefficient beforehand and stipulates that every value of ρ is equally plausible a priori (Jeffreys, 1961); this yields a uniform distribution ranging from -1 to 1, shown in Figure 2 as the dashed line. It is possible to specify different models by changing the prior distribution. For instance, we could have incorporated the knowledge that ρ is expected to be positive and used a uniform prior distribution that ranges only from 0 to 1. We refrain from doing so here because the frequentist analysis is also two-sided, but we note that a complete analysis of this data set requires one to explore whether the statistical conclusions hold across a range of plausible priors.

Insert Figure 2 somewhere here

Next, the prior is combined with the information coming from the data (i.e., the likelihood) and the result is a posterior distribution. This posterior distribution quantifies the uncertainty about ρ after having seen the data. Figure 2 shows that compared to the prior distribution, the posterior distribution assigns relatively little mass to values lower than 0 and higher than .7. Using the posterior distribution, one can quantify how likely it is that ρ falls between, say, .2 and .4; or one can provide a 95% credible interval for ρ. In contrast to the frequentist 95% confidence interval, the Bayesian credible interval has a direct and intuitive interpretation as "the plausibility that ρ is in the specified interval" (Hoekstra, Morey, Rouder, & Wagenmakers, in press).

Bayesian Hypothesis Testing

The posterior distribution allows one to answer the general question "What do we know about the correlation between height and popularity in the US elections, assuming from the outset that such a correlation exists?" From this formulation, it is clear that we cannot use the posterior distribution alone for the purpose of hypothesis testing. As stated by Berger (2006, p. 383): "[...] Bayesians cannot test precise hypotheses using confidence intervals. In classical statistics one frequently sees testing done by forming a confidence region for the parameter, and then rejecting a null value of the parameter if it does not lie in the confidence region. This is simply wrong if done in a Bayesian formulation (and if the null value of the parameter is believable as a hypothesis)."

Hence, when the goal is hypothesis testing, Bayesians need to go beyond the posterior distribution. To answer the question "To what extent do the data support the presence of a correlation?" one needs to compare two models: a null-hypothesis that states the absence of the effect (i.e., H0: $\rho = 0$) and an alternative hypothesis that states its presence. In Bayesian statistics, this alternative hypothesis needs to be specified exactly. In our scenario, the alternative hypothesis is specified as H1: $\rho \sim$ Uniform(-1,1), that is, ρ is distributed uniformly ranging from -1 to 1.

In Bayesian hypothesis testing, hypotheses or models have prior plausibility. Before having seen the data, the relative plausibility of the competing models is known as the prior model odds, that is, p(H1)/p(H0). After having seen the data, the relative plausibility is known as the posterior model odds, that is, p(H1 | D)/p(H0 | D). The change from prior to posterior odds that is brought about by the data is referred to as the Bayes factor, that is, p(D | H1)/p(D | H0). Because of the inherently subjective nature of the prior model odds, the emphasis of Bayesian hypothesis testing is on the amount by which the data shift one's beliefs, that is, on the Bayes factor.

Thus, when the Bayes factor BF_{10} equals 10.5, the data are 10.5 times more likely under H1 than under H0. When the Bayes factor equals $BF_{10} = 0.2$, the data are 5 times more likely under H0 than under H1. Even though the Bayes factor has an unambiguous and continuous scale, it is sometimes useful to summarize the Bayes factor in terms of discrete categories of evidential strength. Jeffreys (1961, his Appendix B) proposed the classification scheme shown in Table 2. We replaced the labels "worth no more than a bare mention" with "anecdotal", "decisive" with "extreme", and "substantial" with "moderate" (Wetzels, van Ravenzwaaij, & Wagenmakers, in press). These labels facilitate scientific communication but should be considered only as an approximate descriptive articulation of different standards of evidence.

Bayes factors represent "the standard Bayesian solution to the hypothesis testing and model selection problems" (Lewis & Raftery, 1997, p. 648) and "the primary tool used in Bayesian inference for hypothesis testing and model selection" (Berger, 2006, p. 378). Nevertheless, Bayes factors come with two important challenges.

Insert Table 2 about here

Challenges for Bayesian Hypothesis Testing

The first challenge for Bayesian hypothesis testing is the specification of sensible prior distributions for the parameters that are subject to test. For Bayesian hypothesis testing, it matters whether we test H0 versus H1: $\rho \sim$ Uniform(-1,1) (the correlation can take on any value), versus H2: $\rho \sim$ Uniform (0,1) (the correlation is positive), or versus, say, H3: $\rho \sim$ Uniform(-0.1, 0.1) (there is a correlation but it is small). The fact that the result depends on the prior specification is not in itself a challenge or a limitation; in fact, it is desirable that different results are obtained for different models: H1 is a relatively flexible model that keeps all options open; H2 is less flexible than H1, because it rules out the possibility that ρ is negative. Finally, H3 is the most parsimonious, least flexible alternative model – it is very similar, in fact, to H0 and therefore a relatively large number of data points will be required before H3 can be discriminated from H0 with much confidence.

Because the Bayesian hypothesis test is relatively sensitive –as it should be– to the prior distributions, the specification of these prior distributions requires considerable care. In the case of the Pearson correlation we may follow Jeffreys (1961) and place a uniform prior on ρ, but this is not feasible for variables with unbounded support, such as the mean of a Normal distribution. Considerable effort has been spent to develop "default" prior distributions, that is, prior distributions that work well across a wide range of substantively different applications. For instance, the default priors we use for linear regression are known as the Jeffreys-Zellner-Siow priors (Jeffreys, 1961; Zellner & Siow, 1980; Liang, Paulo, Molina, Clyde, & Berger, 2008; Rouder & Morey, 2012); as discussed later, these priors fulfill several desiderata and can provide a reference analysis that may, if needed, be fine-tuned using problem-specific information.

The second challenge for Bayesian hypothesis testing is computational: Bayes factors can be relatively difficult to obtain. The Bayes factor is the ratio of marginal likelihoods, for instance BF_{10} = p(D | H1) / P(D | H0), where numerator and denominator indicate the probability of the observed data under the hypothesis at hand. The marginal likelihoods are obtained by integrating or averaging the likelihoods over a model's prior parameter space; this way, all predictions that the model makes are taken into account. Flexible models make many predictions, and if most of these predictions are wrong this drives down the average (Lee & Wagenmakers, in press). This is how Bayes factors implement Occam's razor or the principle of parsimony (e.g., Myung, Forster, & Browne, 2000; Wagenmakers & Waldorp, 2006).

Although integrating the likelihood over the prior distribution is vital to obtain Bayes factors and penalize models for undue complexity, the integration process itself can be analytically infeasible and computationally demanding (e.g., Gamerman & Lopes, 2006). Fortunately, the details of the specific situation may allow Bayes factors to be obtained without conducting the integration process. For instance, consider the set of models for which p-values can be computed; this set features a comparison between a null-hypothesis which is a simplified version of a more complex alternative hypothesis – in the previous example on the US presidents, H1: ρ ~ Uniform(-1,1) can be simplified to H0 by setting ρ equal to zero. For such a comparison between nested models, one can obtain the Bayes factor by the Savage-Dickey density ratio (e.g., Dickey & Lientz, 1970; Wagenmakers et al., 2010).

Figure 2 visualizes the Savage-Dickey density ratio by the dots that indicate the height of the prior and posterior distribution at ρ = 0. Specifically, Figure 2 indicates that, under H1, the prior density at ρ = 0 is higher than the posterior density – in other words, the data have decreased the belief that ρ = 0. The ratio between prior and posterior height equals 6.33, and this ratio equals the Bayes factor. Thus, for nested models one can obtain the Bayes factor without integrating over the prior parameter space; instead, one can consider the prior and posterior distribution for the parameter that is subject to test, and the Bayes factor is given by the ratio of the ordinates.

Advantages of Bayesian Hypothesis Testing

Bayesian hypothesis testing through Bayes factors provide the researcher with several concrete and practical advantages. First and foremost, the Bayes factor quantifies evidence for and against statistical hypotheses. It does not matter whether one of the hypotheses under consideration is a null-hypothesis. Hence, evidence can be quantified in favor of the null-hypothesis, something which is impossible using the p-value (e.g., Gallistel, 2009; Rouder, Speckman, Sun, Morey, & Iverson, 2009).

Related to the previous point, the Bayes factor is inherently comparative: it weighs the support for one model against that of another. This contrasts with the p-value, which is calculated conditional on the null-hypothesis being true; the alternative hypothesis is irrelevant as far as the calculation of the p-value is concerned. Consequently, data that are unlikely under H0 may lead to its rejection, even though these data are just as unlikely under H1 – and are therefore perfectly uninformative. Consequently, p-values are known to overstate the evidence against H0 (e.g., Berger & Delampady, 1987; Edwards, Lindman, & Savage, 1963; Johnson, in press; Sellke, Bayarri, & Berger, 2001). This is also evident from the election example, where a correlation of .39, displayed in Figure 1, yields $p = .007$ and $BF_{10} = 6.33$. Even though in this particular case both numbers support the same conclusion, we believe that the p-value suggests that the evidence is compelling, whereas the Bayes factor leaves room for doubt (for an extensive empirical comparison between p-values and Bayes factors see Wetzels et al., 2011).

An additional advantage is that –in contrast to the p-value– the Bayes factor is not affected by the sampling plan, or the intention with which the data were collected. Consider again the election example and the data shown in Figure 1. We reported that for this correlation, $p = .007$. However, this p-value was computed under a fixed sample size scenario; that is, the p-value was computed under the assumption that an experimenter set out to run 46 elections and then stop. This sampling plan is certainly incorrect, and by extension, so is the p-value. But what is the correct sampling plan? It could be something like “US elections will continue every four years until democracy is replaced with a different system of government or the US ceases to exist”. But even this sampling plan is vague – we only learn that we can expect quite a few elections more.

In order to compute a p-value, one could settle for the fixed sample size scenario and simply not worry about the details of the sampling plan. However, consider the fact that new elections will continue be added to the set. How should such future data be analyzed? One can pretend, after every new election, that the sample size was fixed. However, this myopic perspective induces a multiple comparison problem – every new test has a probability of 5% of falsely rejecting the null-hypothesis, and the myopic perspective therefore fails to control the overall Type I error rate.

For Bayes factors, in contrast, the sampling plan is irrelevant to inference (as dictated by the stopping rule principle; Berger & Wolpert, 1988). This means that researchers can monitor the evidence (i.e., the Bayes factor) as the data come in, and terminate data collection whenever they like, such as when the evidence is deemed sufficiently compelling, or when the researcher has run out of resources. Figure 3 illustrates the process for the election example. The Bayes factor is monitored from the third election onward (the first two elections do not allow the calculation of a Bayes factor, which was therefore set to 1). The evidence in favor of the alternative hypothesis gradually increases until the 46^{th} election, when it stands at a Bayes factor of $BF_{10} = 6.33$. Clearly, new election results can be added and the evidence can be updated, continually and indefinitely, for as long as there are US elections.

Insert Figure 3 somewhere here

BAYESIAN HYPOTHESIS TESTING FOR REGRESSION MODELS

The principles outlined in the previous section also hold for regression models (e.g., Liang et al., 2008; Rouder & Morey, 2012). In regression models, the key question concerns the quantification of statistical evidence for the inclusion of a particular predictor or set of predictors.

Suppose model $\mathcal{M}_X$ includes x predictors, and model $\mathcal{M}_Y$ includes one additional predictor. The evidence for the inclusion of this additional predictor is then given by $BF_{YX} = p(D \mid \mathcal{M}_Y) / p(D \mid \mathcal{M}_X)$. Now suppose a third model, $\mathcal{M}_Z$, again includes one predictor more than $\mathcal{M}_Y$. The evidence for $\mathcal{M}_Z$ over $\mathcal{M}_X$ is $BF_{ZX} = p(D \mid \mathcal{M}_Z) / p(D \mid \mathcal{M}_X)$. Thus, we know the strength of evidence for both M_Y and M_Z vs. the simplest model M_X. Then it is easy to see that the evidence for M_Y vs. M_Z can be obtained by transitivity, as follows: $BF_{ZY} = BF_{ZX} / BF_{YX}$. Thus, all that is required to assess the evidence for and against the inclusion of predictors is the ability to compute the Bayes factor for any specific model against a common null model without predictors; the Bayes factors for different non-null models against each other can then be obtained through transitivity.

The remaining difficulty is to specify suitable priors for the beta regression coefficients. Here we adopt an objective Bayesian perspective and specify priors based on general desiderata instead of on substantive knowledge that is unique to a particular application. In linear regression models, the most popular objective prior specification scheme is inspired by the pioneering work of Harold Jeffreys and Arnold Zellner. This Jeffreys-Zellner-Siow prior specification scheme (Jeffreys, 1961; Zellner & Siow, 1980; Liang et al., 2008; Rouder & Morey, 2012) assigns a multivariate Cauchy distribution to the regression coefficients. The Cauchy distribution is a t-distribution with one degree of freedom. Compared to the Normal distribution, the Cauchy distribution has more mass in the tails.

Detailed mathematical derivation, explanation, and motivation for the Jeffreys-Zellner-Siow prior is provided elsewhere (i.e., Liang et al., 2008; Rouder & Morey, 2012; Wetzels, Grasman, & Wagenmakers, 2012). Here the emphasis is on the conceptual interpretation and practical utility of the Bayes factors associated with the Jeffreys-Zellner-Siow (JZS) specification. In this context, it is important that there exists user-friendly software to obtain the JZS Bayes factors – in particular, we attend the reader to the web-applet from Jeff Rouder (http://pcl.missouri.edu/bf-reg) and the BayesFactor package in R. The appendix includes the R code that we used for the analysis of the examples in this article.

Before turning to a more detailed example related to management science, consider again the prototypical data on the length of hospital stays presented by Cohen and Cohen (1983; see Table 1). The null model, $\mathcal{M}_0$, contained no predictors and functions as the common null model. Hypothesis 1 was implemented as model $\mathcal{M}_1$ that contains 9 predictors (i.e., 9 beta coefficients, which are the additional free parameters of $\mathcal{M}_1$ compared to $\mathcal{M}_0$). As shown in Table 1, the JZS Bayes factor indicates that the data are 10.34×10^{13} times more likely to have occurred under model $\mathcal{M}_1$ than under model $\mathcal{M}_0$. This is extreme evidence in favor of $\mathcal{M}_1$ over model $\mathcal{M}_0$.

Hypothesis 2 was implemented as model $\mathcal{M}_2$ containing 19 predictors, 10 predictors more than model $\mathcal{M}_1$. Again, the JZS Bayes factor indicates that the data are much more likely under $\mathcal{M}_2$ than under model $\mathcal{M}_0$, that is, $BF_{20} = 19.30 \times 10^7$. As explained above, we can now determine the Bayes factor for $\mathcal{M}_2$ versus $\mathcal{M}_1$ by transitivity, and conclude that $BF_{21} = 1.89 \times 10^{-6}$, extreme evidence against the inclusion of the 10 additional predictors. Note that this is evidence in favor of a null-hypothesis that postulates the absence of predictor effects.

Finally, hypothesis 3 was implemented as model $\mathcal{M}_3$ containing 26 predictors, 7 predictors more than $\mathcal{M}_2$. Repeating the previous steps (i.e., computing the JZS Bayes factor against the common null model, and obtaining the desired Bayes factors through transitivity) shows that there is extreme evidence in favor of $\mathcal{M}_3$, both when compared against $\mathcal{M}_1$ (i.e., $BF_{31} = 54.41 \times 10^3$) and when compared against $\mathcal{M}_2$ (i.e., $BF_{32} = 29.14 \times 10^9$).

The analysis of this simple, prototypical example highlights that Bayes factors can quantify evidence in favor of either the simpler model or the more complex model. The strength of evidence is quantified on a continuous scale, providing a more informative assessment than whether the p-value is smaller than .05, .01, or .001.

The remaining section provides a more realistic and detailed example, inspired by an article recently published in the *Journal of Management* (i.e., Dierdorff et al., 2012).

AN EXAMPLE FROM THE *JOURNAL OF MANAGEMENT*

The goal of this section is to underscore the advantages of JZS Bayes factor hypothesis testing for hierarchical regression when applied to a practical analysis problem in management science. We will outline two different ways in which Bayes factors allow researchers to assess the importance of predictors: covariance testing and model comparison (Rouder & Morey, 2012).

For concreteness, our points are illustrated using a design from Dierdorff et al. (2012). Unfortunately, the authors declined our request to use their data for this article, and we therefore chose to make our points using simulated data, generated to yield summary statistics as similar as possible to those that were reported for the original data. These simulated data form the basis of our analysis; the file with simulated data can be found online[1] so that the interested reader can confirm the analysis. Because the data are simulated, no substantive conclusions can be attached to the results. Instead, our aim is to illustrate the JZS Bayes factor procedure using an example of realistic complexity.

Theoretical Background of the Dierdorff et al. Study

The study of Dierdorff et al. (2012) focused on citizenship, a concept defined as the set of “cooperative, helpful behaviors extending beyond job requirements” (Dierdorff et al., 2012, p. 573). Citizenship is affected both by work context and by role expectations, that is, the “beliefs about what is required for successful role performance” (Dierdorff et al., 2012, p. 575).

Based on an extensive literature review and detailed reasoning process, Dierdorff et al. (2012) proposed the following five hypotheses about the effects of work context and role expectations on citizenship:

“Hypothesis 1: Prosocial role expectations are positively related to citizenship behavior.” (Dierdorff et al., 2012, p. 577)

“Hypothesis 2: The relationship between role expectations and citizenship is stronger in more interdependent contexts.” (Dierdorff et al., 2012, p. 579)

“Hypothesis 3: The relationship between role expectations and citizenship is stronger in more socially supportive contexts.” (Dierdorff et al., 2012, p. 580)

“Hypothesis 4: The relationship between role expectations and citizenship is stronger in more autonomous contexts.” (Dierdorff et al., 2012, p. 581)

“Hypothesis 5: The relationship between role expectations and citizenship is weaker in more ambiguous contexts.” (Dierdorff et al., 2012, p. 581)

In the Dierdorff et al. study, these hypotheses were tested using data from two sources: (1) self-report surveys filled out by 198 full-time employees; (2) a performance evaluation form completed by the employee's immediate supervisor.

Frequentist Analysis

As mentioned above, we used the information reported in the Dierdorff study to create a simulated data set that was as similar as possible to the original. All of the following analyses were conducted on the simulated data set. In this section, we discuss the frequentist analysis plan as followed by Dierdorff and colleagues. Table 3 summarizes the main findings (cf. Table 2 in Dierdorff et al., 2012).

As is customary in hierarchical regression, the predictors of interest were added in steps. In the first step, the control variable “task-specific performance” was included as a predictor (i.e., Model 1), and this yields $R^2 = .37$. In the second step, the variable “role expectations” was added (i.e., Model 2), allowing a test of Hypothesis 1. As expected, Hypothesis 1 was confirmed: inclusion of “role expectation” increases R^2 from .37 to .57; in addition, the beta coefficient equals .42 ($p < .001$). In the third step, all remaining variables were added simultaneously (i.e., Model 3). The assessment of Hypothesis 2-5 then proceeds by inference on the beta-coefficients for the specific predictors from Model 3.

In particular, frequentist inference suggests that the data do not support Hypothesis 2 (β = -.04, $p > .05$), but they do support Hypotheses 3, 4, and 5 (β = -.04, $p < .05$; β = .07, $p < .05$; β = -.17, $p < .001$, respectively).[2]

Insert Table 3 about here

Bayesian Analysis

From R^2, the number of predictors, and the sample size one can compute the JZS Bayes factors for Models 1, 2, and 3 against the null model (see appendix for R code using the BayesFactor package); as before, the other Bayes factors can then be obtained by transitivity. Consistent with the frequentist analysis, the JZS Bayes factors indicated decisive support for Model 3 over Model 2 (i.e., $BF_{32} = 2.10\times10^8$), Model 2 over Model 1 (i.e., $BF_{21} = 3.04\times10^{14}$), and Model 1 over the null model ($BF_{10} = 3.43\times10^{18}$).

Unfortunately, Model 3 comprises eight additional predictors, four of which connected to specific hypotheses. To evaluate the evidence that the data provide for the inclusion of particular predictors and the associated hypotheses we now turn to a more detailed analysis (see also Rouder & Morey, 2012).

Bayesian Method 1: Covariance testing. This approach is most similar to the NHST approach that is currently popular in management science and other social sciences (Faraway, 2002). With covariance testing, the researcher assesses the importance of specific predictors or covariates by eliminating them from the full model that includes all predictors. This method, applied to the simulated Dierdorff data, is illustrated in Table 4. As before, Model 3 is the full model that contains all predictors. To test each of the four hypotheses (i.e., Dierdorff's Hypothesis 2-5), four matching regression models were created by excluding a single predictor of interest from the full model.

Using the same tools as in the previous sections, we then computed the Bayes factors for each of the simpler regression models against the full model. When the JZS Bayes factor $BF_{nf} > 1$ this signifies evidence in favor of the simpler model representing the null hypothesis of no effect, and against the hypothesis under investigation.

As Table 4 shows, covariance testing indicates moderate evidence against Hypothesis 2 (i.e., $BF_{2f} = 7.16$) and in favor of the simpler regression model that lacks the relevant predictor. This illustrates the point that Bayes factors can quantify support in favor of a null hypothesis. For Hypotheses 3 and 4, the evidence provided by the data is anecdotal and does not warrant any conclusions (i.e., $BF_{3f} = 1.93$ and $BF_{4f} = 0.87$, respectively). This illustrates the point that p-values overestimate the evidence against the null-hypothesis; in the corresponding frequentist analysis, the beta coefficients corresponding to Hypothesis 3 and 4 both yielded $p < .05$, prompting researchers and readers to reject the null-hypothesis.

Finally, the data provide extreme support in favor of Hypothesis 5 ($BF_{5f} = 1,630,000$).

This illustrates the point that the Bayes factor provides a more informative measure of evidence than the p-value cut-off, “$p < .001$”.

Covariate testing is straightforward and conceptually similar to standard testing procedures. However, covariance testing may fail in the presence of collinearity (Rouder & Morey, 2012). Specifically, assume that two predictors (e.g., a person's weight and height) are highly correlated. Leaving only one of the two predictors (e.g., weight or height) out of the full model will do little harm, as the other predictor is able to take over and accommodate the data. Based on covariance testing, one may therefore conclude that the highly correlated predictors are irrelevant; this reasoning, however, ignores the possibility that the fit may worsen dramatically when both predictors are left out of the model at the same time.

Insert Table 4 about here

Bayesian method 2: Model comparison. In contrast to covariance testing, model comparison represents a more elaborate and complete approach to the variable selection problem. In the Dierdorff design, there are four crucial predictors, each of which is associated with a specific hypothesis. By including or excluding each of these four predictors independently, one can create 15 different regression models, listed as the first column in Table 5.

The full model $\mathcal{M}_f$ includes all four hypotheses and is equivalent to Model 3 in the original study. All other models are simpler and include either 3, 2, or 1 hypotheses. The right-most column of Table 5 provides for each of these 14 models, their Bayes factors against the full model. Whenever the Bayes factor BF_{nf} is greater than 1, the data provide evidence in favor of the simpler model over the full model.

Insert Table 5 about here

As is evident from Table 5, the data provide strong support in favor of Hypothesis 5: all models that outperform the full model (M_4, M_9, M_{10}, M_{14}) feature the predictor that represents Hypothesis 5. In addition, the data show little support for Hypothesis 2, neither in isolation (M_{11}) nor in combination with other predictors.

The evidence for Hypothesis 3 and 4 is mixed; These hypotheses do not fare well alone (M_{12} and M_{13}), or together (M_8), but added to the predictor for Hypothesis 5 they perform reasonably well, both separately and together (M_4, M_9, M_{10}). The crucial element appears to be the omission of the predictor for Hypothesis 2.

Overall, these conclusions are similar to those obtained with covariance analysis: the data provide support for Hypothesis 5 and against Hypothesis 2, whereas the support for Hypothesis 3 and 4 is ambiguous. A more precise quantification of evidence using model comparison makes use of model averaging (e.g., Hoeting, Madigan, Raftery, & Volinsky, 1999; Liang et al., 2008); in model averaging, one computes the overall inclusion probability for each predictor as the sum of relevant posterior model probabilities – for instance, the inclusion probability for the predictor corresponding to Hypothesis 3 is the sum of posterior model probabilities for M_f, M_1, M_2, M_4, M_5, M_8, M_9, and M_{12}.

BAYES FACTORS VERSUS ABSOLUTE GOODNESS OF FIT

The Bayes factor is inherently comparative: it assesses the support that the data provide for one model versus another. This is useful and informative, but it can also be misleading: even though a specific model may outperform another in terms of the Bayes factor, both models may provide a poor account of the data, invalidating the inference. Thus, before drawing conclusions it is important to assess absolute goodness-of-fit and confirm that the best model is also a good model.

This important issue is highlighted in Anscombe's quartet (Anscombe, 1973), shown here as Figure 4. Each panel shows a different data set, carefully constructed so that the variables have the same means, variances, and linear regression coefficient. For each data set, the Bayes factor is 23, indicating strong support for the presence of a linear association. A casual glance at Figure 4, however, convinces one that the statistical models and associated inference are valid only for panel A.

Model misfit can be assessed in several ways. Anscombe's quartet suggests that inspecting data by eye can often be highly informative. In general, one can inspect structure in the residuals and assess the impact of individual data points by successively leaving them out of the analysis. Such methods for assessing absolute model fit can be carried out within both the frequentist and the Bayesian paradigm.

Insert Figure 4 somewhere here

CONCLUDING COMMENTS

Using Bayes factor hypothesis testing, researchers may monitor evidence as the data come in; they may quantify support in favor of the null-hypothesis; and they may prevent themselves from prematurely rejecting the null-hypothesis. To the best of our knowledge, the combination of these benefits is unique to the Bayesian statistical paradigm.

The JZS Bayes factor regression analysis is relatively easy to carry out. Researchers can construct their regression models and apply either the covariance testing or model comparison method described in the previous sections. Scripts and data used for our analysis are available online for download.

One may argue that in many situations the data will pass the “interocular traumatic test” (i.e., when the pattern in the data is so evident that the conclusion hits you straight between the eyes; Edwards et al., 1963), and the results will be clear no matter what statistical paradigm is being used. Luckily this is true; however, some data fail the interocular traumatic test and the results may indeed depend on the statistical paradigm that is used. In such cases, it seems counterproductive and potentially misleading to base one's statistical inference on frequentist methods alone.

REFERENCES

Anscombe, F. J. 1973. Graphs in statistical analysis. *The American Statistician,* 27: 17-21.

Berger, J. O. 2006. Bayes factors. In S. Kotz, N. Balakrishnan, C. Read, B. Vidakovic, & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences, vol. 1* (2nd ed.) (pp. 378-386). Hoboken, NJ: Wiley.

Berger, J. O., & Delampady, M. (1987). Testing precise hypotheses. *Statistical Science,* 2: 317-352.

Berger, J. O. and Wolpert, R. L. 1988. *The likelihood principle* (2nd ed.). Harward (CA): Institute of Mathematical Statistics

Cohen, J., & Cohen, P. 1983. *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed). Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.

Dickey, J. M., & Lientz, B. P. 1979. The weighted likelihood ratio, sharp Hypotheses about chances, the order of a Markov Chain. T*he Annals of Mathematical Statistics*, 41: 214-226.

Dienes, Z. 2008. *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*, New York: Palgrave MacMillan.

Dierdorff, E. C., Rubin, R. S., & Bachrach, D. G. 2012. Role expectations as antecedents of citizenship and the moderating effects of work context. *Journal of Management,* 38: 573-598.

Edwards, W., Lindman, H., & Savage, L. J. 1963. Bayesian statistical inference for psychological research. *Psychological Review,* 70: 193-242.

Faraway, J. J. 2002. *Practical Regression and Anova Using R*. Available from http://www.maths.bath.ac.uk/~jjf23/book/

Gallistel, C. R. 2009. The importance of proving the null. *Psychological Review*, 116: 439-453.

Gamerman, D., & Lopes, H. F. 2006. *Markov Chain Monte Carlo: Stochastic simulation for Bayesian Inference*. Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A. 2010. Bayesian statistics then and now. *Statistical Science,* 25: 162-165.

Hoekstra, R. and Morey, R. D. and Rouder, J. N., &Wagenmakers, E.-J. in press. Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*.

Hoeting, J. A. and Madigan, D. and Raftery, A. E. , &Volinsky, C. T. 1999. Bayesian model averaging: A tutorial. *Statistical Science*. 14: 382-417.

Jeffreys, H. 1961. Theory of probability (3rd ed.). New York: Oxford University Press.

Johnson, V. E. in press. Revised standards for statistical evidence. *Proceedings of the National Academy of Sciences of the United States of America.*

Kass, R. E., & Raftery, A. E. 1995. Bayes factors. *Journal of the American Statistical Association*, 90: 773-795.

Kruschke, J. K. 2010. *Doing Bayesian data analysis: A tutorial introduction with R and BUGS*. Burlington, MA: Academic Press.

Lee, M. D. and Wagenmakers, E.-J. in press. B*ayesian modeling for cognitive science: A practical course*. Cambridge University Press.

Lewis, S. M., & Raftery, A. E. 1997. Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association,* 92: 648-655.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. 2008. Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association,* 103: 410-423.

Meng, X.-L. 1994. Predictive p-values. *The Annals of Statistics,* 22: 1142-1160.

Myung, I. J., Forster, M. R., & Browne, M. W. 2000. Model selection [Special Issue]. *Journal of Mathematical Psychology*, 44: 1-2.

Poirier, D. J. 2006. The growth of Bayesian methods in statistics and economics since 1970. *Bayesian Analysis*, 1: 969-980.

Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, 47: 877-903.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. 2009. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16: 225-237.

Schervish, M. J. 1996. P values: What they are and what they are not. *The American Statistician*, 50: 203-206.

Sellke, T., Bayarri, M. J., & Berger, J. O. 2001. Calibration of p values for testing precise null hypotheses. *The American Statistician,* 55: 62-71.

Stulp, G., Buunk, A. P., Verhulst, S., & Pollet, T. V. 2013. Tall claims? Sense and nonsense about the importance of height of US Presidents. *The Leadership Quarterly*, 24: 159-171.

Wagenmakers, E.-J. 2007. A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review,* 14: 779-804.

Wagenmakers, E.J., Lodewyckx, T., Kuriyal, H., & Grasman, R. 2010. Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology,* 60: 158-189.

Wagenmakers, E.-J., & Waldorp, L. 2006. Model selection: Theoretical developments and applications [Special Issue]. *Journal of Mathematical Psychology*, 50: 2.

Wetzels, R., Grasman, R. P. P. P. , & Wagenmakers, E.-J. 2012. A default Bayesian hypothesis test for ANOVA designs. *The American Statistician*, 66: 104-111.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J. , & Wagenmakers, E.-J. 2011. Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, 6: 291-298.

Wetzels, R., Wagenmakers, E.J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin Review,* 19: 1057-1064.

Wetzels, R., van Ravenzwaaij, D., & Wagenmakers, E.-J. in press. Bayesian analysis. In R. Cautin, & S. Lilienfeld (Eds.), *The Encyclopedia of Clinical Psychology*. Wiley-Blackwell.

Zellner, A., & Siow, A. 1980. Posterior odds ratios for selected regression hypotheses In Bernardo, J. M., DeGroot, M. H., Lindley, D. V. ,& Smith, A. F. M *Bayesian Statistics.* Valencia: University Press

Zyphur, M. J., & Oswald, F. L. in press. Bayesian estimation and inference: A user's guide. *Journal of Management*.

FOOTNOTES

[1] The link to the simulated data is: https://drive.google.com/folderview?id=0B5vDt-Crky_DeHkyYndTS2tMVU0&usp=sharing

[2] In what follows we deliberately ignore the complication that, for the simulated data set, the beta coefficient corresponding to Hypothesis 3 (social support × role expectations) does not have the correct sign – in the original data, the beta coefficient was estimated to be +.11 instead of -.04. This qualitative mismatch reveals that, despite considerable effort, we were unable to generate simulated data that matched the original data exactly.

TABLES

Table 1

Hierarchical Regression for the Classic Psychiatric Hospital Stay Example from Cohen and Cohen (1983). Bayes Factors (BF) Are Discussed in a Later Section.

Model/ Hypothesi	Nº Predictors	R^2	ΔR^2	BF_{n0}	BF_{n1}	BF_{n2}
1	9	.20	.20***	10.34×10^{13}		
2	19	.22	.02	19.30×10^{7}	1.89×10^{-6}	
3	26	.33	.11***	56.25×10^{17}	54.41×10^{3}	29.14×10^{9}

N = 500
*p<.05; ** p<.01; ***p<.001

Table 2

Evidence Categories for the Bayes Factor BF_{12} (Jeffreys, 1961).

Bayes factor BF_{12}			Interpretation
	>	100	Extreme evidence for M_1
30	-	100	Very Strong evidence for M_1
10	-	30	Strong evidence for M_1
3	-	10	Moderate evidence for M_1
1	-	3	Anecdotal evidence for M_1
	1		No evidence
1/3	-	1	Anecdotal evidence for M_2
1/10	-	1/3	Moderate evidence for M_2
1/30	-	1/10	Strong evidence for M_2
1/100	-	1/30	Very Strong evidence for M_2
	<	1/100	Extreme evidence for M_2

Table 3

Hierarchical Regression Results for Simulated Data Based on the Study of Dierdorff et al. (2012).

Predictors	Beta parameters		
	Model 1	Model 2	Model 3
Task-specific performance	.61***		
Role expectations		.42***	
Interdependence			.19**
Social support			.23***
Ambiguity			.17**
Autonomy			.04
Interdependence × Role expectations			-.04
Social support × Role expectations			-.04*
Ambiguity × Role expectations			-.17***
Autonomy × Role expectations			.07*
R^2	.37	.57	.71
ΔR^2	.37***	.20***	.14***
BF_{n0}	3.43×10^{18}	1.04×10^{33}	2.18×10^{43}
BF_{n1}		3.04×10^{14}	6.36×10^{22}
BF_{n2}			2.10×10^{8}

$N = 198$
*$p<.05$; **$p<.01$; ***$p<.001$

Table 4

Covariance Testing Results for Simulated Data Based on the Study of Dierdorff et al. (2012)

		Hypotheses			
Predictors	Full Model	2	3	5	4
Task-specific performance	+	+	+	+	+
Role expectations	+	+	+	+	+
Interdependence	+	+	+	+	+
Social support	+	+	+	+	+
Ambiguity	+	+	+	+	+
Autonomy	+	+	+	+	+
Interdependence × Role expectations	+	-	+	+	+
Social support × Role expectations	+	+	-	+	+
Ambiguity × Role expectations	+	+	+	-	+
Autonomy × Role expectations	+	+	+	+	-
R^2	.6883	.6879	.6835	.6325	.6808
ΔR^2	.6883	-.0004	-.0048	-.0558***	-.00075*
BF_{nf}	1	7.16	1.93	1.63×10^{-6}	0.87

N = 198
**p*<.05; ****p*<.001

Table 5

Model Comparison Results for Simulated Data Based on the Study of Dierdorff et al. (2012)

Model	Hypotheses	R^2	BF_{nf}
M_f	$H_2+H_3+H_4+H_5$	.6883	1
M_1	$H_2+H_3+H_4$	.6325	1.63×10^{-6}
M_2	$H_2+H_3+H_5$	.6808	0.87
M_3	$H_2+H_4+H_5$	.6835	1.93
M_4	$H_3+H_4+H_5$	.6879	7.16
M_5	H_2+H_3	.6310	8.75×10^{-6}
M_6	H_2+H_4	.6213	7.81×10^{-7}
M_7	H_2+H_5	.6760	1.81
M_8	H_3+H_4	.6328	1.39×10^{-5}
M_9	H_3+H_5	.6810	7.81
M_{10}	H_4+H_5	.6821	10.81
M_{11}	H_2	.6201	4.68×10^{-6}
M_{12}	H_3	.6315	8.35×10^{-5}
M_{13}	H_4	.6202	4.80×10^{-6}
M_{14}	H_5	.6754	13.55

M_f = full model
H_2 = Hypothesis 2
H_3 = Hypothesis 3
H_4 = Hypothesis 4
H_5 = Hypothesis 5

FIGURES

Figure 1

Correlation between the Proportion of the Popular Vote and the Height Ratio between a US President and his Closest Competitor.

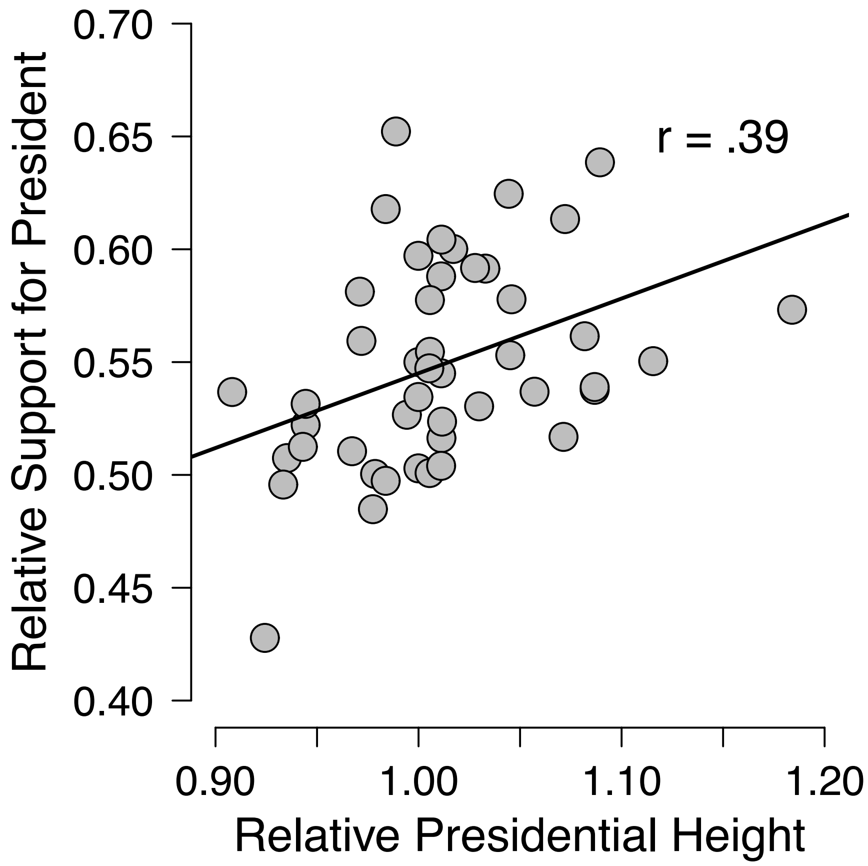


Figure 2

Prior and Posterior Distribution for the Correlation between the Proportion of the Popular Vote and the Height Ratio between a US President and his Closest Competitor.

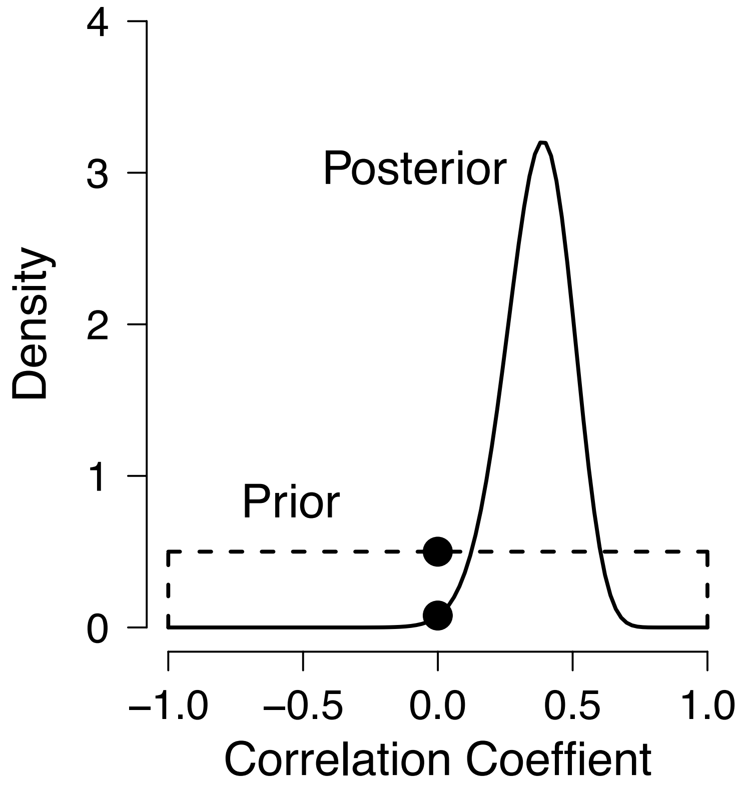


Figure 3

Sequential Analysis of the Evidence for and against a Height-Effect in the US Presidential Elections.

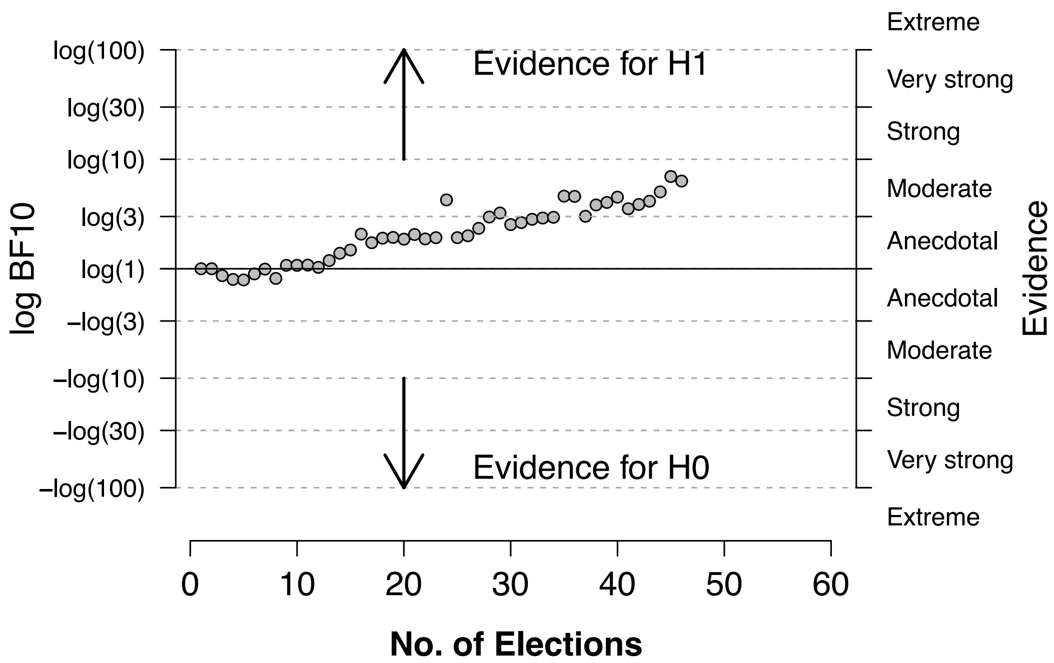
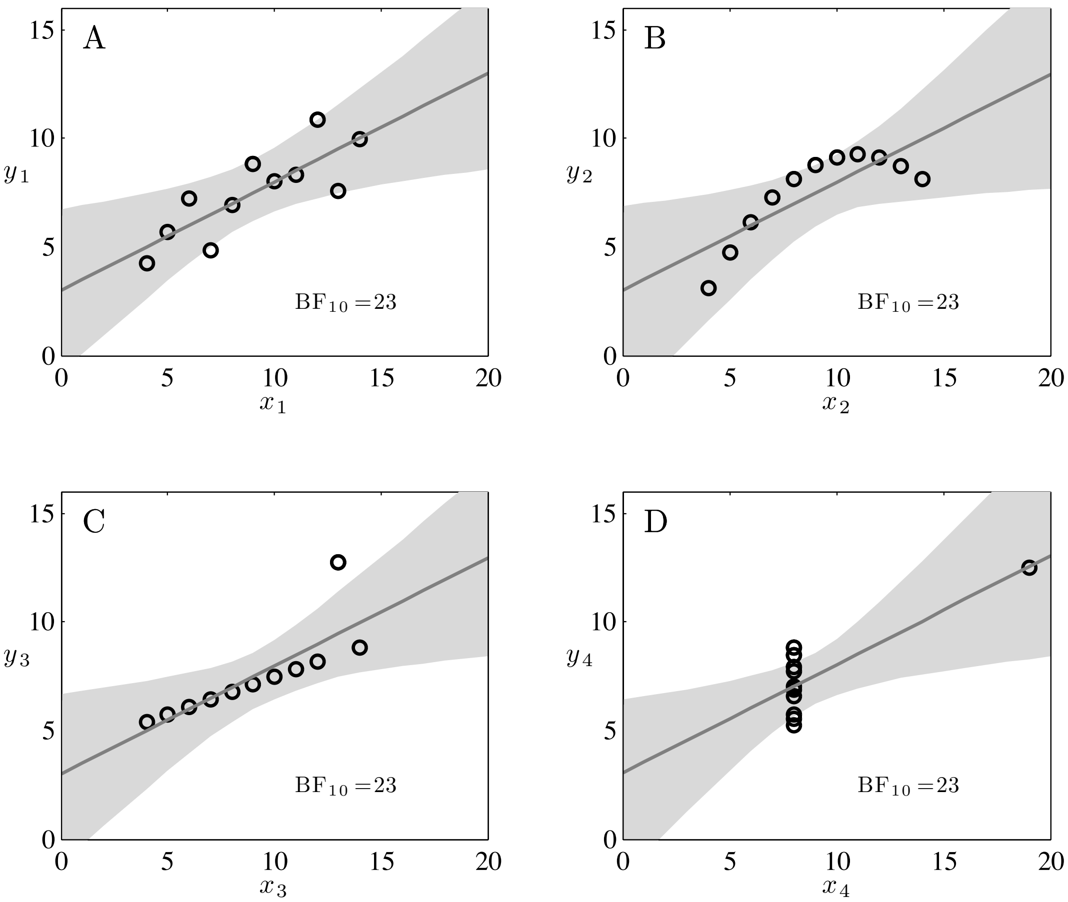


Figure 4

Anscombe's Quartet. Gray Regions Indicate the 95% Bayesian Posterior Predictive Interval

(Gelman, 2010; Meng, 1994).



APPENDIX

Below we provide functions implemented in R programming language which can be used to 1) compute a *p*-value given the R^2 change, number of predictors and sample size (function "R2.to.p"), 2) compute a Bayes factor from the R^2 given the number of predictors and sample size (function "R2.to.bf"), and 3) generate a dataset given means, standard deviations, correlations and regression coefficients of the original data.

In order to use these functions you need to have the R programming language (version 2.15.2 or higher) installed on your computer (http://cran.r-project.org/). Make sure that you install R-version which is not older than 2.15.2. You should install from the following packages, which are not part of the standard R-version: BayesFactor, caTools, coda, lattice, mvtnorm, pbapply, bitops.

When you already have R and all necessary packages installed on your computer, you are ready to use the functions that we provide. Copy the text and save it in a file with .R extension. Alternatively, the files can be downloaded from https://drive.google.com/folderview?id=0B5vDt-Crky_DeHkyYndTS2tMVU0&usp=sharing. Before calling the functions, you have to "source" the file, in which you saved the functions. Type in the R-console source("path-to-your-file"). Then, you can simply call the functions from the console, by typing for example:

```
>  p <- R2.to.p(500, )
```

Obtaining P-value from the R^2 Change

```
R2.to.p <- function(n, nPred1, nPred2, r21, r22) {
  # Calculator of p-values for the R^2 change
  # Author: Sandra Andraszewicz, University of Basel
  # Last update: 09-11-2012
  # Args:
  #   n: Sample size
  #   nPred1: Number of predictors in Model 1
  #   nPred2: Number of predictors in Model 2
  #   r21: R^2 of Model 1
  #   r22: R^2 of Model 2
  # Returns:
  #   p: significance level of the regression test
  # The function prints out the input R^2 and the corresponding p-value

  deltaR2 <- r22 - r21 #Change in R^2 from Model 1 to Model 2

  #Conduct F-test for the R^2 change and obtain p-value
  fchange <- ((n-nPred2-1) * (r22-r21)) / ((nPred2 - nPred1) * (1-r22)) #F-statistic
  df1 <- nPred2 - nPred1 #Degrees of freedom related to the nominator
  df2 <- n-nPred2 - 1 #Degrees of freedom related to the denominator
  p <- pf(fchange, df1, df2, lower.tail=FALSE) #p-value from F-distribution

  #Display the results
  print("The R^2 change equals:")
  print(deltaR2)
  print("The p-value of the R^2 change is:")
  print(p)

  return(p)
}
```

Obtaining Bayes Factor from R^2

```
R2.to.bf <- function (n, nPred, r2) {
  # Computes Bayes factor of Models M to the Null Model and to all other models
  # from R^2 obtained from hierarchical regression analysis
  # The output can be provided for all models conducted for the same sample
  # Author: Sandra Andraszewicz, University of Basel
  # Last update: 06-11-2012
  # Args:
  #   n: Sample size
  #   nPred: an array containing mount of predictors
  #   r2: an array containing R^2 values
  # Returns:
  #   output: a matrix containing combinations of all Bayes factors
  #           first two columns correspond to the models, e.g. 2 1
  #           is a Bayes factor of model 2 to model 1
  # Additionally, the function displays the output in a descriptive form
  # in the console
  # Example:
  #   Hierarchical regression analysis was conduced for 500 participants,
  #   where the predictors are entered in 3 steps:
  #   Step 1) 9 predictors, R^2=.20
  #   Step 2) 19 predictors, R^2=.22
  #   Step 3) 26 predictors, R^2=.26
  #   To obtain Bayes factors type in the console:
  #   R2.to.bf(500, c(9, 19, 26), c(.20, .22, .26))

  #Load neccessary packages
  library(BayesFactor)
  library(caTools)
  library(coda)
  library(lattice)
  library(mvtnorm)
  library(pbapply)
  library(bitops)

  nModels <- length(r2) #Aomunt of models

  #Initialise output Bayes Factors comparing Model M to the Null Model
  bfm0 <- c(rep(0, nModels))

  #Obtain Bayes Factors comparing Models M to the Null Model
  for(i in seq(1:nModels)) {
    bfm0[i] <- linearReg.Quad(n, nPred[i], r2[i], logbf <- FALSE)
  }

  #Obtain Bayes factors comparing Models M with each other
  myCombs <- combs(seq(1:nModels), 2) #Create all combinations of BF
  bfmm <- c(rep(0, nrow(myCombs))) #Initialise vector containing BF among other models
  bfmmnames <- c(rep("", nrow(myCombs))) #Initialise vector containing names of BFs
  for(i in seq(1:nrow(myCombs))){
    bfmm[i] <- bfm0[myCombs[i, 2]] / bfm0[myCombs[i, 1]]
    bfmmnames[i] <- paste("BF", paste(as.character(myCombs[i, 2]),
      as.character(myCombs[i, 1]), collapse=NULL), collapse=NULL)
  }

  #Display all results in the console
  for(i in seq(1:nModels)){
    print(c(paste("BF", paste(as.character(i),as.character(0), collapse=NULL),
      collapse=NULL), bfm0[i]))
  }
  #Display the results in the console
  for(i in seq(1:nrow(myCombs))){
    print(c(bfmmnames[i], bfmm[i]))
  }
  output <- cbind(myCombs, bfmm)
  return(output)
}
```

Generating Data from a Regression Model

```
regrconstr.datgen = function(n, R, means, sds, model, beta) {
	# This function generates multivariate normal data such that the predictors
	# have the same means, sd's and correlations as given by 'means', 'sds' & 'R',
	# respectively, and will give the regression coefficients 'beta' if the
	# linear regression model 'model' is fit with lm. The model must be specified
	# in terms of the variable names V1, V2, V3, etc. where the order in the
	# means and sds vector and correlation matrix R is used.
	#
	# Example usage generating 20 observations:
	#  regrconstr.datgen(20, matrix(.5,3,3)+diag(.5,3), rep(1,3), 1:3,
	#     V3 ~ V1*V2, c(0, .5, .5, .25)) # don't forget the intercept coefficient!
	#
	M = means;
	SD = sds;
	N = n;

	# generate multivariate normal data with the same correlations, sd's and means
	m = nrow(R)
	C = diag(SD) %*% R %*% diag(SD)
	Xc = scale(matrix(rnorm(N*m),N),TRUE,FALSE)
	Xc = Xc %*% solve(chol(cov(Xc))) %*% chol(C)
	X = as.data.frame(t(t(Xc)+M))

	# adjust the simulated dependent variable such that the covariance remains as
	# specified and the model coefficient estimates are exactly as specified
	fit = lm(model, data = X, x = TRUE, y = TRUE)
	x = fit$x # matrix of used predictors
	y = fit$y # dependent variable
	# The covariance- and beta coefficient-constraints boils down to y lying in a
	# subspace of R^N:
	# Projections are computed from qr decompositions of the Model matrix of
	# predictors and from
	# the matrix of variables.
	U = qr.Q(fit$qr)
	D = qr.R(fit$qr)
	Y1 = U %*% solve(t(D)) %*% t(D) %*% D %*% beta
	IPu = (diag(nrow(U)) - U %*% t(U))
	Y = Y1 + IPu %*% y
	X[,as.character(model[[2]])] = Y
	X
}
```