Salience Effects in Economic Choice

Inauguraldissertation
zur Erlangung des akademischen Grades
doctor rerum politicarum

eingereicht an der

Wirtschaftswissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

von

Mats Köster, M.Sc.
geb. am 27. März 1991 in Düsseldorf.

Rektorin der Heinrich-Heine-Universität Düsseldorf:
Univ.-Prof. Dr. Anja Steinbeck

Dekan der Wirtschaftswissenschaftlichen Fakultät:
Univ.-Prof. Dr. Stefan Süß

Gutachter:
1. Univ.-Prof. Paul Heidhues, PhD
2. Univ.-Prof. Dr. Hans-Theo Normann

Acknowledgements

Writing this dissertation would not have been possible without the support and feedback of way too many people to list all of them here. So, while this section is highly incomplete, I am deeply grateful to anyone who has contributed to this work in one or the other way.

First and foremost, I would like to thank Paul Heidhues, who was and still is the best advisor I could have wished for. Throughout the last four and a half years, he has been extremely generous with his time, and I learned a lot about economics and — maybe more importantly — asking the "right" questions from the many discussions we have had.

Secondly, I want to thank Hans-Theo Normann, who has agreed to evaluate this thesis, for all his support and feedback throughout the years (already before graduate school).

A special thanks goes to my co-author and friend Markus Dertwinkel-Kalt. At least topic-wise, he had by far the biggest impact on who I am as a researcher today, and the main ideas spelled out in this dissertation we developed together. I learned so much from working with him, and he always took my ideas seriously — even back when I was "just" a Master student —, which is something that I will never forget or take for granted.

I am also grateful to Philipp Strack for hosting me at UC Berkeley (and for his feedback on my research) as well as to Andrei Shleifer for inviting me to spend a week at Harvard. This one week, but also Andrei's generous support over the last year had an enormous impact on Chapter 3 of this dissertation, which eventually became my job-market paper.

Moreover, I would like to thank my other co-authors Katrin Gödker, Botond Kőszegi, Florian Peiseler, and Matthias Sutter not only for working with me, but also for all the insightful and fun discussions we have had during this time. The same is true for all of my colleagues at Düsseldorf who had to listen to a talk about "salience effects" at least once a year; in particular, David Heine and Frank Schlütter who had to hear these talks much more often. As mentioned before, many more people have improved the research in this thesis through their thoughtful comments, and I am deeply grateful to all of you.

Finally, I would like to extend a special thanks to my family and friends, who have supported me throughout the way. Without doing injustice to the others, I have to highlight my father Thomas — who suggested that I should enroll in Economics back when I had no idea what else to do — and my wife Annika — who has supported me through all the ups and downs of trying to do research, which I will never take for granted.

Contents

Introduction

It is well-known that our attention is limited and sometimes (perhaps subconsciously) influenced by salient cues that stand out in the choice context. But the economic implications of salience-driven (or non-directed) attention — in contrast to directed attention that is chosen to balance the costs and benefits of being attentive — have only recently been modeled formally (e.g. Bordalo, Gennaioli and Shleifer, 2012, 2013b; Kőszegi and Szeidl, 2013). These economic models formalize (mostly experimental) evidence from the psychology literature on so-called *contrast effects*: attention is directed to the attribute(s) of an option along which it differs most from the alternative options (e.g. Simonson and Tversky, 1992; Tversky and Simonson, 1993; Schkade and Kahneman, 1998). This dissertation works out and tests the behavioral implications of such contrast effects for consumer behavior (with a particular focus on choice under risk) and the functioning of markets.

To understand the testable implications of salience-driven attention for behavior in various domains, we need a coherent model to start with. Chapter 1, therefore, axiomatizes a generalization of *salience theory* that nests its (somewhat detached) applications to choice under risk (Bordalo *et al.*, 2012) and consumer choice (Bordalo *et al.*, 2013b) as special cases. Despite allowing for a completely flexible interaction of the different attributes of an option (e.g., the quality and price of a product), this general salience model makes clear-cut predictions on behavior that are consistent with experimental evidence. More importantly, this model provides a basis for my research on salience effects.

Chapter 2, published in the *Journal of the European Economic Association*, goes on to investigate the implications of salience-driven attention for choice under risk. We start from an empirical regularity in behavior across a variety of settings: whether people seek or avoid risks on gambling, insurance, asset, or labor markets seems to crucially depend on the skewness of the underlying probability distribution. More specifically, people typically seek positively skewed risks and avoid negatively skewed risks. We show that salience theory of choice under risk (Bordalo *et al.*, 2012) can explain this preference for positive skewness, because according to the contrast effect unlikely, but outstanding payoffs attract attention. In contrast to alternative models, however, salience theory predicts that choices under risk not only depend on the absolute skewness of the available options, but also on how skewed these options appear to be relative to each other. We exploit this fact to derive novel, experimentally testable predictions that are unique to the salience model and that we find support for in two laboratory experiments. We thereby argue that skewness preferences — typically attributed to cumulative prospect theory (Tversky and Kahneman, 1992) — are more naturally accommodated by salience theory.

Building on these insights regarding "simple" choices under risk, Chapter 3 studies the interaction between salience effects and the complexity of a decision problem. Intuitively, through channeling attention to a subset of the relevant information (e.g. a certain state of the world), salience reduces the "dimensionality" of a decision problem. We thus hypothesize that people behave more consistently across differently complex problems if there is a common salient cue that guides their attention and, consequently, behavior. Combining the experimental design delineated in Chapter 2 with ideas from Eyster and Weizsäcker (2016), we test and confirm this hypothesis in the context of choice under risk: while revealed attitudes toward skewed risks — which have extreme and salient outcomes — are consistent across differently complex problems, revealed attitudes toward symmetric risks — where such a salient cue is missing — vary significantly with complexity. We provide suggestive evidence that these findings are driven by the extreme outcomes of skewed risks attracting a subject's attention and guiding choices. To rationalize our experimental results, we propose a variant of Bordalo *et al.*'s (2012) salience theory.

Chapter 4, published in the *European Economic Review*, investigates the implications of salience-driven attention in a market context where firms optimally adjust their contract design in response to consumers being susceptible to contrast effects. Precisely, we try to make sense of the observation that supply contracts (e.g., for electricity, telephony, or banking services) often include many small regular payments made by consumers and a single, large bonus that is paid to consumers at some point during the contractual period. Since bonus payments create inefficiencies such as sending out checks and redeeming them, the use of bonuses is puzzling from the perspective of the classical model. We offer an explanation for the frequent occurrence of bonus contracts based on the contrast effect: a large bonus is particularly salient and attracts a consumer's attention. Specifically, we show that a monopolist offers a bonus contract only for low-value goods, while firms standing in competition always offer bonus contracts, independent of the consumers' valuation for the product. Thus, competition may not eliminate, but exacerbate inefficiencies arising from contracting with agents who are susceptible to salience effects. Common contract schemes on markets for electricity, telephony, and bank accounts mirror our findings.

I conclude by briefly summarizing the key take-aways from my dissertation and my related research as well as the additional questions that it raises. Working on these additional questions will hopefully keep me busy for a couple of years to come.

Chapter 1

Multivariate Salience Theory of Choice under Risk

1.1 Introduction

Ample evidence from psychology, but also economics suggests that the choice context often affects behavior in a systematic way. To explain the context-dependency of observed behavior, Bordalo, Gennaioli, and Shleifer have recently developed what is called *salience theory*, and they have subsequently applied different versions of the model to study choices between univariate risks (Bordalo *et al.*, 2012) as well as choices between deterministic options with multiple attributes (Bordalo *et al.*, 2013b). Specifically, Bordalo *et al.* (2012) assume that people choose a monetary risk X over a monetary risk Y if and only if

$$\sum_{\text{supp } p} p(x, y)(x - y)\sigma(x, y) > 0, \tag{1.1}$$

where p denotes the joint distribution of these two random variables and $\sigma : \mathbb{R}^2 \to \mathbb{R}_+$ is a so-called salience function that distorts how much weight the decision maker puts on the different states of the world. For the choice between the two deterministic options $\mathbf{A} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{B} = (b_1, b_2, \ldots, b_n)$, where an option is characterized by the value it takes in each of $n \in \mathbb{N}$ distinct attributes (e.g. the price and quality of a product), Bordalo *et al.* (2013b) assume that the agent chooses option $\mathbf{A}$ if and only if

$$\sum_{i=1}^{n}(a_i - b_i)\sigma(a_i, b_i) > 0. \tag{1.2}$$

While in Eq. (1.1) salience distorts the weights assigned to the objectively feasible states of the world, it applies to a set of exogenously specified (but subjective) and additively separable attributes in Eq. (1.2). But does there exist a coherent salience model that can be applied to both situations as well as combinations thereof?

Extending the axiomatization of the decision criterion in Eq. (1.1), as proposed by Lanzani (2020), we develop a general salience model that allows us to study the binary choice between multivariate random variables in one coherent framework. As in Ellis and Masatlioglu (2020), who axiomatize a version of the decision criterion in Eq. (1.2), our main representation theorem deals with the case of just two attributes. More precisely, we adopt techniques on additive conjoint measurement from Fishburn (1990a,b, 1991) to extend Lanzani's axiomatization of univariate salience theory of choice under risk to the choice between bivariate random variables. We validate the choice of our axioms by the following result: the axioms are necessary and sufficient for the model to nest as special cases, the decision criteria in Eq. (1.1) and Eq. (1.2) as well as variants thereof that have been discussed in the literature. Extending the model to the choice between random variables with more than two attributes is straightforward, but it requires some technical adjustments (without much economic substance) that we discuss in the Appendix.

Despite allowing for a completely flexible interaction between the different attributes of an option, our model makes testable predictions: the model predicts subjects to avoid a positive correlation across the two attributes of non-negative bivariate random variables, which is consistent with experimental evidence by Ebert and van de Kuilen (2016) in the context of intertemporal choice. The model further predicts that correlation preferences flip for non-positive bivariate random variables, which is reminiscent of the reflection of risk preferences on the negative domain (e.g., Kahneman and Tversky, 1979).

1.2 A Motivating Example

Consider an agent who chooses between two n-dimensional, real-valued random variables $\mathbf{X} = (X_i)_{i=1,\dots,n}$ and $\mathbf{Y} = (Y_i)_{i=1,\dots,n}$ with a joint distribution p that has finite support.

The univariate case. Let $n = 1$. According to salience theory of choice under risk (Bordalo *et al.*, 2012), when evaluating and comparing the options, the agent distorts the objective probability of a given realization $(x, y) \in \mathbb{R}^2$ by how salient this state of the world is. The salience of a state is measured by a *salience function* defined as follows.

Definition 1. *A symmetric and continuous function $\sigma : \mathbb{R}^2 \to \mathbb{R}_+$ is a* salience function *if and only if it satisfies the following three properties:*

1. Ordering. *Let $x \geq y$. Then for any $\epsilon, \epsilon' \geq 0$ with $\epsilon + \epsilon' > 0$,*

$$\sigma(x + \epsilon, y - \epsilon') > \sigma(x, y).$$

2. Diminishing sensitivity. *Let $x > y \geq 0$. Then for any $\epsilon > 0$,*

$$\sigma(x+\epsilon, y+\epsilon) < \sigma(x,y).$$

3. Weak reflection. *For any $x, y, x', y' \geq 0$ with $|x-y| = |x'-y'|$, we have*

$$\sigma(x,y) < \sigma(x',y') \quad \textit{if and only if} \quad \sigma(-x,-y) < \sigma(-x',-y').$$

We say that a given realization $(x,y) \in \mathbb{R}^2$ is the *more salient* the larger its salience value $\sigma(x,y)$ is. The ordering property implies that a pair of outcomes receives a larger weight in making a decision the more these outcomes differ. Ordering thus captures the well-known *contrast effect* (e.g., Schkade and Kahneman, 1998), whereby people focus their attention on those states of the world in which the attainable outcomes differ a lot. Diminishing sensitivity reflects *Weber's law* of perception, and it implies that the salience of a pair of outcomes decreases if these outcomes uniformly increase in absolute terms. Hence, diminishing sensitivity can be described as a *level effect* according to which a given contrast in outcomes is more salient at a lower outcome level. Finally, by the weak reflection property, diminishing sensitivity (with respect to zero) reflects from the domain of positive outcomes to the domain of negative outcomes.[1]

We follow Bordalo *et al.* (2012) in assuming that a *salient thinker* evaluates monetary outcomes via a linear value function, $u(x) = x$, and aims to maximize her salience-weighted utility, which is defined as follows: the *salience-weighted utility* of the random variable X evaluated in the choice set $\mathcal{C} = \{X, Y\}$ is given by

$$U^s(X|\mathcal{C}) = \sum_{\text{supp } p} x \cdot \frac{p(x,y) \cdot \sigma(x,y)}{\sum_{\text{supp } p} p(s,t) \cdot \sigma(s,t)}. \tag{1.3}$$

A multivariate example. Extending salience theory to n-dimensional random variables entails certain degrees of freedom. As an illustration, let us assume that the salient thinker first reduces the dimensionality of the problem from n to $m \leq n$ dimensions, by additively aggregating non-separable dimensions. More specifically, a salient thinker represents the n-dimensional random variable $\mathbf{X}$ as an m-dimensional variable $\mathbf{X}^r$, where m denotes the number of non-separable dimensions. We denote, for any dimension $i \in \{1, \ldots, n\}$, the corresponding set of non-separable dimensions as

$$D_i := \{j \in \{1, \ldots, n\} \setminus \{i\} : j \text{ and } i \text{ are additively integrable}\},$$

[1] Bordalo *et al.* (2012) impose a stronger version of the reflection property: for any $x, y, x', y' \geq 0$, we have $\sigma(x,y) < \sigma(x',y')$ if and only if $\sigma(-x,-y) < \sigma(-x',-y')$. This stronger notion is, however, not identified from choice data (see, for instance, Lanzani, 2020).

and iteratively define $X_i^r := X_i + \sum_{j \in D_i} X_j$ for any $i \in \{1, \ldots, n\}$ so that $i \notin \bigcup_{k=1,\ldots,i-1} D_k$.

Suppose, for instance, $\mathbf{X} = (X_1, X_2, X_3)$ is a product with a base price X_1, a surcharge X_2, and a quality X_3. A salient thinker who just cares about the total price she pays, represents the product as $\mathbf{X}^r = (X_1 + X_2, X_3)$ and assigns it a salience-weighted utility of

$$U^s(\mathbf{X}|\mathcal{C}) = \sum_{\text{supp } p} p(\mathbf{x}, \mathbf{y}) \cdot \frac{(x_1 + x_2)\sigma(x_1 + x_2, y_1 + y_2) + x_3 \sigma(x_3, y_3)}{\sum_{\text{supp } p} p(\mathbf{s}, \mathbf{t}) \cdot \big(\sigma(s_1 + s_2, t_1 + t_2) + \sigma(s_3, t_3)\big)}.$$

The following definition generalizes this example.

Definition 2. *The* salience-weighted utility *of the n-dimensional random variable $\mathbf{X}$, with an m-dimensional representation $\mathbf{X}^r$, evaluated in the choice set $\mathcal{C} = \{\mathbf{X}, \mathbf{Y}\}$ equals*

$$U^s(\mathbf{X}|\mathcal{C}) = \sum_{\text{supp } p} p(\mathbf{x}, \mathbf{y}) \cdot \sum_{i=1}^{m} x_i^r \cdot \frac{\sigma(x_i^r, y_i^r)}{\sum_{\text{supp } p} p(\mathbf{s}, \mathbf{t}) \cdot \sum_{j=1}^{n} \sigma(s_j^r, t_j^r)}, \tag{1.4}$$

where $\sigma : \mathbb{R}^2 \to \mathbb{R}_+$ is a salience function.

In particular: (1) if $m = n$, then we say that the salience-weighted utility admits an additively separable *representation;*[2] *and (2) if $m = 1$, then we say that the salience-weighted utility admits an* additively integrable *representation.*

The following examples illustrate that our model nests the salience theories of choice under risk (Bordalo *et al.*, 2012) and consumer choice (Bordalo *et al.*, 2013b) as well as Dertwinkel-Kalt and Köster's extension to portfolio selection problems as special cases.

Example 1 (Bordalo *et al.*, 2012: Salience Theory of Choice under Risk). For $n = 1$, Definition 2 collapses to Equation (1.3), which gives the continuous version of Bordalo *et al.*'s (2012) salience theory of choice under risk (see Dertwinkel-Kalt and Köster, 2020b).

Example 2 (Bordalo *et al.*, 2013b: Salience and Consumer Choice). Suppose that $\mathbf{X} = \mathbf{x} \in \mathbb{R}^n$ and $\mathbf{Y} = \mathbf{y} \in \mathbb{R}^n$ are deterministic (e.g. products with n features such as price, quality, etc.), and that the salience-weighted utility admits an additively separable representation. Then, the model reduces to a continuous version of Bordalo *et al.* (2013b):

$$U^s(\mathbf{x}|\mathcal{C}) = \sum_{i=1}^{n} x_i \cdot \frac{\sigma(x_i, y_i)}{\sum_{j=1}^{n} \sigma(x_j, y_j)}.$$

More generally, our model also resembles that of Bordalo *et al.* (2013b), when replacing the n-dimensional options $\mathbf{x}$ and $\mathbf{y}$ by their m-dimensional representations $\mathbf{x}^r$ and $\mathbf{y}^r$.

[2] To be precise, due to the normalization factor, the salience-weighted utility cannot be additively separable. But, for any $\mathcal{C} = \{\mathbf{X}, \mathbf{Y}\}$, a salient thinker (weakly) prefers $\mathbf{X}$ over $\mathbf{Y}$ if and only if $\sum_{\text{supp } p} p(\mathbf{x}, \mathbf{y}) \cdot \sum_{i=1}^{m} (x_i^r - y_i^r) \cdot \sigma(x_i^r, y_i^r) \geq 0$, whereby the left-hand side represents the same preferences and is additively separable in the n dimensions if and only if $m = n$.

Example 3 (Dertwinkel-Kalt and Köster, 2020c: Salience and Portfolio Selection). If the salience-weighted utility admits an additively integrable representation, we obtain the model proposed in Dertwinkel-Kalt and Köster (2020c) to study portfolio selection:

$$U^s(\mathbf{X}|\mathcal{C}) = \sum_{\text{supp } p} p(\mathbf{x},\mathbf{y}) \cdot \underbrace{\left(\sum_{i=1}^{n} x_i\right)}_{\text{portfolio return}} \cdot \frac{\sigma(\sum_{i=1}^n x_i, \sum_{i=1}^n y_i)}{\sum_{\text{supp } p} p(\mathbf{s},\mathbf{t}) \cdot \sigma(\sum_{i=1}^n s_i, \sum_{i=1}^n t_i)}.$$

Here, a salient thinker evaluates different portfolios based on the aggregate portfolio returns instead of separately looking at the returns of the n underlying assets. This model resembles salience theory of choice under risk when replacing the n-dimensional random variables $\mathbf{X}$ and $\mathbf{Y}$ by the one-dimensional variables $\tilde{X} = \sum_{i=1}^n X_i$ and $\tilde{Y} = \sum_{i=1}^n Y_i$.

1.3 A General Characterization for Bivariate Risks

We propose a general framework to study salience effects in the choice between two bivariate random variables. We build on the univariate axiomatization by Lanzani (2020), which itself relies on the seminal work by Fishburn (1989, 1991) on conjoint measurement.

1.3.1 Setup and Notation

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$. Building on Lanzani (2020), we define preferences over the joint distribution of the two available options; here, two-dimensional, real-valued random variables. Let $\Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ be the set of all probability measures over $\mathbb{R}^2 \times \mathbb{R}^2$ with finite support. We introduce a *preference set* $P \subseteq \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ with the following interpretation: For any joint distribution $p \in \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ the agent has to decide whether she wants to be paid according to the first two or according to the last two components of any realized vector. We then assume that $p \in P$ if and only if the agent (weakly) prefers to be be paid according to first two components of a vector. For any given distribution $p \in \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$, we define its *conjugate distribution* as

$$\overline{p}(\mathbf{x},\mathbf{y}) = p(\mathbf{y},\mathbf{x}).$$

We say that $\mathbf{x} \geq \mathbf{y}$ if and only if $x_i \geq y_i$ for any $i \in \{1,2\}$ and $\mathbf{x} > \mathbf{y}$ if, in addition, $x_i > y_i$ for some $i \in \{1,2\}$. Denote as $\mathbf{e}_i$ the i-th unit vector in $\mathbb{R}^2$. We say that a function $f : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is *additively separable* if there exists some function $g : \mathbb{R}^2 \to \mathbb{R}$ such that $f(\mathbf{x},\mathbf{y}) = g(x_1, y_1) + g(x_2, y_2)$.[3] Finally, let $\delta_{(\mathbf{x},\mathbf{y})}$ be the *Dirac measure* on $\mathbb{R}^2 \times \mathbb{R}^2$.

[3] Requiring that there exists a single function $g : \mathbb{R}^2 \to \mathbb{R}$ that applies to both dimensions is stricter than "simple" additive separability. But, as we will see in the proof of our representation theorem, this stricter notion of additive separability is sensible in our context, because both dimensions can take

1.3.2 General Axioms

The first set of axioms in Lanzani (2020), which are not specific to salience theory, straightforwardly extend to the case of two-dimensional random variables. First, we assume that the preferences over the joint distributions of the two random variables are complete.

Axiom 1 (Completeness)**.** *For any* $p \in \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$*, it holds that* $p \notin P \implies \overline{p} \in P$*.*

Given that the preferences are complete, we can define the *strict* preference set as $\hat{P} = \{p \in P : \overline{p} \notin P\}$. This allows us to impose a (strong) version of independence.

Axiom 2 (Strong Independence)**.** *For any* $p, q \in P$ *and any* $\lambda \in (0, 1)$*, it holds that*

$$\lambda p + (1 - \lambda)q \in P,$$

and, whenever $p \in \hat{P}$*,*

$$\lambda p + (1 - \lambda)q \in \hat{P}.$$

We further assume that the preferences over joint distributions satisfy three increasingly more demanding versions of continuity: Part (a) of Axiom 3 is a weak requirement ruling out the case where one of the marginals is "infinitely preferred" to the other. Parts (b) and (c) are stronger requirements that ensure continuity of the representing function.

Axiom 3 (Continuity)**.**

(a) Archimedean Continuity: *for any* $p \in \hat{P}$ *and* $q \notin P$*, there exist* $\alpha, \beta \in (0, 1)$ *so that*

$$\alpha p + (1 - \alpha)q \in \hat{P} \quad \text{and} \quad \beta p + (1 - \beta)q \notin P.$$

(b) Continuity in Outcomes: *Let* $(\mathbf{x}_k)_{k\in\mathbb{N}}$ *be a sequence that converges to* $\mathbf{x}$ *as* $k \to \infty$*. Then, for any* $\alpha \in [0, 1]$*,* $\mathbf{y} \in \mathbb{R}^2$*, and* $p \in \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$*,*

$$\alpha\delta_{(\mathbf{x}_k,\mathbf{y})} + (1 - \alpha)p \in P \ \ \forall k \in \mathbb{N} \implies \alpha\delta_{(\mathbf{x},\mathbf{y})} + (1 - \alpha)p \in P,$$

and

$$\alpha\delta_{(\mathbf{y},\mathbf{x}_k)} + (1 - \alpha)p \in P \ \ \forall k \in \mathbb{N} \implies \alpha\delta_{(\mathbf{y},\mathbf{x})} + (1 - \alpha)p \in P.$$

(c) Continuity at Identity: *Let* $(\mathbf{x}_k)_{k\in\mathbb{N}}$ *be a sequence that converges from above to* $\mathbf{x}$ *as* $k \to \infty$*, and let* $(\mathbf{y}_k)_{k\in\mathbb{N}}$ *be a sequence that converges from below to* $\mathbf{x}$ *as* $k \to \infty$*.*

values in $\mathbb{R}$. This would be different, for instance, if the feasible values in dimension i would be given by $X_i \subset \mathbb{R}$ with $X_1 \cap X_2 = \emptyset$. In this case, a sensible notion of additive separability would require functions $g_1, g_2 : \mathbb{R}^2 \to \mathbb{R}$ such that $f(\mathbf{x}, \mathbf{y}) = g_1(x_1, y_1) + g_2(x_2, y_2)$.

For any $\mathbf{z} \in \mathbb{R}^2$ *and* $\mathbf{w} \in \mathbb{R}^2_{>0}$, *there exists some* $K \in \mathbb{N}$ *such that for any* $k \geq K$,

$$\Big((\mathbf{x}, \mathbf{x}_k), 1 - ||\mathbf{x}_k - \mathbf{x}||_2; (\mathbf{z} + \mathbf{w}, \mathbf{z}), ||\mathbf{x}_k - \mathbf{x}||_2\Big) \in P$$

and

$$\Big((\mathbf{y}_k, \mathbf{x}), 1 - ||\mathbf{y}_k - \mathbf{x}||_2; (\mathbf{z} + \mathbf{w}, \mathbf{z}), ||\mathbf{y}_k - \mathbf{x}||_2\Big) \in P.$$

Finally, we assume that the preferences over joint distributions are monotonic with respect to state-wise improvements, which is not to be confused, however, with monotonicity with respect to first-order stochastic dominance in the marginals. As we show in Dertwinkel-Kalt and Köster (2020b), salience theory of choice under risk satisfies monotonicity with respect to first-order stochastic dominance for independent random variables, but not for correlated ones (see also Lanzani, 2020, for a similar discussion).

Axiom 4 (Monotonicity)**.** *For* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^2$ *and* $p \in \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$, *if* $\mathbf{x} > \mathbf{y}$ *and* $\alpha \in (0, 1)$,

$$\alpha\delta_{(\mathbf{y},\mathbf{z})} + (1 - \alpha)p \in P \implies \alpha\delta_{(\mathbf{x},\mathbf{z})} + (1 - \alpha)p \in \hat{P}.$$

The "general" Axioms 1 – 4 allow us to state a first representation result, which follows immediately from Theorem 1 in combination with Remarks 2 and 3 in Lanzani (2020).

Lemma 1. *The preference set* $P \subseteq \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ *satisfies Axioms 1, 2, 3 (a) and (b), and 4 if and only if there exists a continuous and skew-symmetric function* $\Psi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ *that is strictly increasing in its first two arguments and strictly decreasing in its last two arguments such that*

$$p \in P \iff \sum_{(\mathbf{x},\mathbf{y}) \in \text{supp } p} p(\mathbf{x}, \mathbf{y})\Psi(\mathbf{x}, \mathbf{y}) \geq 0. \tag{1.5}$$

Moreover, the function Ψ *is unique up to a positive linear transformation.*

Theorem 1 in Lanzani (2020) deals with finite distributions over the space $X \times X$, where X is an arbitrary set of prizes. This includes, in particular, sets of two-dimensional, real-valued vectors, so that his theorem directly applies to our setup. In the proof of Remarks 2 and 3 in Lanzani (2020) — here, he assumes that X is an interval on the real line — we have to simply replace the scalars by two-dimensional vectors to obtain the monotonicity and continuity properties of the representing function. The only difference compared to his result is the fact that Ψ is increasing in its first two arguments (not only the first one) and decreasing in its last two arguments (not only the last one).

1.3.3 Salience-Specific Axioms

In order to capture the additional structure of salience theory, we follow Lanzani (2020) and translate the properties of the salience function — i.e., ordering, diminishing sensitivity, and weak reflection — into behaviorally testable axioms. While Lanzani's axioms of ordering and reflexivity straightforwardly extend from univariate to bivariate random variables, we have to extend his diminishing sensitivity axiom substantially.

The Ordering Axiom. Following Lanzani (2020), we define ordering as follows.

Axiom 5 (Ordering)**.** *For any* $\mathbf{x} > \mathbf{y}$ *and* $\alpha, \beta \in [0, 1]$ *with* $\beta > \alpha$ *and* $\beta < 1$ *or* $\alpha > 0$*,*

$$p = \left((\mathbf{x}, \mathbf{y}), \frac{\beta - \alpha}{1 + \beta - \alpha}; (\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}, \beta\mathbf{x} + (1 - \beta)\mathbf{y}), \frac{1}{1 + \beta - \alpha}\right) \in \hat{P}.$$

Each marginal of the joint distribution p in Axiom 5 corresponds to a binary lottery in each of the two dimensions. Notice that the marginals have the same expected value in each of the two dimensions. Moreover, since for any $\mathbf{x} > \mathbf{y}$ and $\alpha, \beta \in (0, 1)$ with $\beta > \alpha$, $\mathbf{x} > \beta\mathbf{x} + (1 - \beta)\mathbf{y} > \alpha\mathbf{x} + (1 - \alpha)\mathbf{y} > \mathbf{y}$, the contrast in outcomes (along both dimensions) is larger in the first state, $(\mathbf{x}, \mathbf{y})$, than in the second state, $(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}, \beta\mathbf{x} + (1 - \beta)\mathbf{y})$. Importantly, since the contrast in outcomes is larger in the first state along *each* dimension, it is still larger when the two dimensions interact with each other. Hence, according to the contrast effect, the larger contrast in the first state — where the first marginal performs better than the second one — attracts the decision maker's attention, which then yields the strict preference for the first marginal distribution. This is captured by the axiom.

The Diminishing Sensitivity Axiom. The main difficulty in extending the axiomatization by Lanzani (2020) to two-dimensional random variables is the proper definition of diminishing sensitivity. On the one hand, we need to account for the fact that with two-dimensional random variables there can be uniform shifts in ouctomes along some, but not all dimensions. On the other hand, we have to take into account how the different dimensions of an option interact. We first illustrate the latter, more serious issue.

Example 4. Suppose that the preference set P can be represented by the salience-weighted utility in Definition 2. In addition, consider the joint distribution given by

$$p = \left(\Big((x_1, x_2), (y_1, y_2)\Big), \frac{1}{2}; \Big((y_1 + \epsilon, y_2), (x_1 + \epsilon, x_2)\Big), \frac{1}{2}\right),$$

where $x_1 = y_1 \geq 0$, $x_2 > y_2 \geq 0$ and $\epsilon > 0$. If the salience-weighted utility admits an additively separable representation, then the difference in salience-weighted utility of the two options (corresponding to the two marginals) is proportional to

$$\underbrace{(x_1 - y_1)}_{=\,0} \cdot \Big(\sigma(x_1, y_1) - \sigma(y_1 + \epsilon, x_1 + \epsilon)\Big) + (x_2 - y_2) \cdot \underbrace{\Big(\sigma(x_2, y_2) - \sigma(y_2, x_2)\Big)}_{=\,0 \text{ due to symmetry}} = 0,$$

which in turn implies that $p \in P \setminus \hat{P}$. If, in contrast, the salience-weighted utility admits an additively integrable representation, then the difference in salience-weighted utility of the two options is proportional to

$$(x_2 - y_2) \cdot \underbrace{\Big(\sigma(x_1 + x_2, y_1 + y_2) - \sigma(y_1 + y_2 + \epsilon, x_1 + x_2 + \epsilon)\Big)}_{>\,0 \text{ due to diminishing sensitivity of the salience function}} > 0,$$

and, therefore, $p \in \hat{P}$. In consequence, a uniform shift in outcomes, along some dimension in which the available options perform equally well, can have very different implications for behavior depending on how the two dimensions of an option interact.

To address the potential interdependence of the two dimensions, we build on results by Fishburn (1991) on the additive separability of a skew-symmetric representation of non-transitive preferences. For any $m \in \mathbb{N}$, we introduce the set

$$\begin{aligned} E_m := \Big\{ (\delta_{(\mathbf{x}^1, \mathbf{y}^1)}, \ldots, \delta_{(\mathbf{x}^m, \mathbf{y}^m)}) : \delta_{(\mathbf{x}^j, \mathbf{y}^j)} \in \mathbb{R}^2 \times \mathbb{R}^2 \text{ for any } j \leq m, \text{ and } \forall i \text{ and } \mathbf{a} \in \mathbb{R}^2, \\ |\{j : (x_i^j, y_i^j) = (a_1, a_2)\}| = |\{j : (x_i^j, y_i^j) = (a_2, a_1)\}| \Big\}. \end{aligned}$$

In words, $(\delta_{(\mathbf{x}^1, \mathbf{y}^1)}, \ldots, \delta_{(\mathbf{x}^m, \mathbf{y}^m)}) \in E_m$ if and only if the i-th component of $\delta_{(\mathbf{x}^j, \mathbf{y}^j)}$ is exactly matched by its inverse. The usefulness of this set becomes clearer through the following observation: if Axioms 1 – 4 hold and the representation Ψ is additively separable, then $(\delta_{(\mathbf{x}^1, \mathbf{y}^1)}, \ldots, \delta_{(\mathbf{x}^m, \mathbf{y}^m)}) \in E_m$ implies $(\delta_{(\mathbf{x}^1, \mathbf{y}^1)}, \ldots, \delta_{(\mathbf{x}^m, \mathbf{y}^m)}) \in P \setminus \hat{P}$. In particular, if the representation Ψ is additively separable, then the following condition has to hold.

Condition 1 (Dimension Independence)**.** *For any $m \geq 1$ and any $\delta_{(\mathbf{x}^j, \mathbf{y}^j)} \in \mathbb{R}^2 \times \mathbb{R}^2$, whenever $(\delta_{(\mathbf{x}^1, \mathbf{y}^1)}, \ldots, \delta_{(\mathbf{x}^m, \mathbf{y}^m)}) \in E_m$ and $\delta_{(\mathbf{x}^j, \mathbf{y}^j)} \in P$ for any $j < m$, then $\delta_{(\mathbf{x}^m, \mathbf{y}^m)} \notin \hat{P}$.*

Following Fishburn (1991), we next define a (auxiliary) binary relation $\succ_0$ (and $\sim_0$) on $\mathbb{R}^2$ as follows: for any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, let

$$\begin{aligned} \mathbf{a} \sim_0 \mathbf{b} \quad \text{and} \quad \mathbf{b} \sim_0 \mathbf{a} \quad &\text{if} \quad \delta_{((a_1, b_2),(a_2, b_1))} \in P \setminus \hat{P}, \\ \mathbf{a} \succ_0 \mathbf{b} \quad &\text{if} \quad \delta_{((a_1, b_2),(a_2, b_1))} \in \hat{P}, \\ \mathbf{b} \succ_0 \mathbf{a} \quad &\text{if} \quad \delta_{((a_2, b_1),(a_1, b_2))} \in \hat{P}. \end{aligned}$$

Denote as $\succeq_0$ the union of $\succ_0$ and $\sim_0$, and define its modified transitive closure on $\mathbb{R}^2$ as $\succ_*$, such that for any $\mathbf{c}, \mathbf{d} \in \mathbb{R}^2$, it holds that $\mathbf{c} \succ_* \mathbf{d}$ if there exists an $m \geq 2$ and

$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m \in \mathbb{R}^2$ with $\mathbf{c} = \mathbf{r}_1, \mathbf{d} = \mathbf{r}_m, \mathbf{r}_k \succeq_0 \mathbf{r}_{k+1}$ for $k = 1, \ldots m-1$, and $\mathbf{r}_k \succ_0 \mathbf{r}_{k+1}$ for some k. We obtain the following lemma. All missing proofs can be found in Appendix A.

Lemma 2. *If Condition 1 holds, then $\succ_*$ on $\mathbb{R}^2$ is a weak order.*[4]

Since, under Condition 1, $\succ_*$ on $\mathbb{R}^2$ is a weak order, its symmetric complement $\sim_*$ is an equivalence relation. In general, for a weak order $\succ'$ on Z, define the set of equivalence classes as $\mathbb{R}^2/\sim'$. We then say that, for $A, B \in \mathbb{R}^2/\sim'$, $A \succ' B$ if $a \succ' b$ for some (and hence all) $a \in A$ and $b \in B$. A subset W of $\mathbb{R}^2/\sim'$ is called *$\succ'$-order dense in $\mathbb{R}^2/\sim'$* if whenever $A \succ' B$ and $A, B \notin W$, then there exists $C \in W$ such that $A \succ' C$ and $C \succ' B$.

Condition 2 (Denseness)**.** *If $\succ_*$ on $\mathbb{R}^2$ is a weak order, then $\mathbb{R}^2/\sim_*$ includes a countable subset that is $\succ_*$-order dense in $\mathbb{R}^2/\sim_*$.*

Using Conditions 1 and 2, we can finally define diminishing sensitivity as follows:

Axiom 6 (Diminishing Sensitivity)**.** *For any $\mathbf{x} > \mathbf{y} \geq 0$ and any $\epsilon > 0$, it holds that*

$$p = \left((\mathbf{x}, \mathbf{y}), \frac{1}{2}; (\mathbf{y} + \mathbf{e}_i\epsilon, \mathbf{x} + \mathbf{e}_i\epsilon), \frac{1}{2} \right) \in P,$$

and, whenever $x_i > y_i$, then $p \in \hat{P}$. For any $\mathbf{x} > \mathbf{y} \geq 0$ with $x_i = y_i$ and any $\epsilon > 0$,

$$p = \left((\mathbf{x}, \mathbf{y}), \frac{1}{2}; (\mathbf{y} + \mathbf{e}_i\epsilon, \mathbf{x} + \mathbf{e}_i\epsilon), \frac{1}{2} \right) \in P \setminus \hat{P}$$

if and only if Condition 1 and Condition 2 are satisfied.

As long as there is a positive contrast in outcomes, uniformly increasing the payoff level along some or all dimensions in a given state makes this state less salient. This is captured by the first part of the axiom. But, if the two dimensions are additively separable *and* the uniform shift occurs in a state with zero contrast in outcomes (in the relevant dimension), then it has no effect on behavior. This part of the axiom is motivated by the fact that salience theory describes a diminished sensitivity to contrasts in outcomes.

The Reflexivity Axiom. Reflexivity ensures that diminishing sensitivity reflects to the domain of negative outcomes and it straightforwardly extends to the bivariate case.

Axiom 7 (Reflexivity)**.** *For any $\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z} \geq 0$ with $w_i - x_i = z_i - y_i$ for $i \in \{1, 2\}$,*

$$\left((\mathbf{w}, \mathbf{x})\frac{1}{2}; (\mathbf{y}, \mathbf{z}), \frac{1}{2} \right) \in \hat{P} \iff \left((-\mathbf{x}, -\mathbf{w})\frac{1}{2}; (-\mathbf{z}, -\mathbf{y}), \frac{1}{2} \right) \in \hat{P}.$$

[4]A binary relation on a set Z is a weak order if and only if it is asymmetric and negatively transitive.

1.3.4 The Representation Theorem

In this subsection, we characterize the skew-symmetric function $\Psi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ that represents not only the general Axioms 1 – 4, but also the salience-specific Axioms 5 – 7.

Definition 3. *A preference set $P \subseteq \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ admits a* salience representation *if there exists a continuous and skew-symmetric function $\Psi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ that is strictly increasing in its first two and strictly decreasing in its last two arguments, such that*

$$p \in P \quad \Longleftrightarrow \quad \sum_{(\mathbf{x},\mathbf{y}) \in \text{supp } p} p(\mathbf{x}, \mathbf{y})\Psi(\mathbf{x}, \mathbf{y}) \geq 0, \tag{1.6}$$

and if this function Ψ further satisfies the following three properties:

(a) For any $\mathbf{x} > \mathbf{y}$ and $\alpha, \beta \in [0, 1]$ with $\beta > \alpha$ and either $\beta < 1$ or $\alpha > 0$, we have

$$(\beta - \alpha)\Psi(\mathbf{x}, \mathbf{y}) > \Psi(\beta\mathbf{x} + (1 - \beta)\mathbf{y}, \alpha\mathbf{x} + (1 - \alpha)\mathbf{y}).$$

(b) For any $\mathbf{x} > \mathbf{y} \geq \mathbf{0}$ and any $\epsilon > 0$, we have

$$\Psi(\mathbf{x} + \mathbf{e}_i\epsilon, \mathbf{y} + \mathbf{e}_i\epsilon) \leq \Psi(\mathbf{x}, \mathbf{y}),$$

holding with equality if and only if $x_i = y_i$ and Ψ is additively separable.

(c) For any $\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z} \geq 0$ with $w_i - x_i = z_i - y_i$ for all $i \in \{1, 2\}$, we have

$$\Psi(\mathbf{w}, \mathbf{x}) > \Psi(\mathbf{z}, \mathbf{y}) \quad \textit{if and only if} \quad \Psi(-\mathbf{w}, -\mathbf{x}) < \Psi(-\mathbf{z}, -\mathbf{y}).$$

Our main theorem extends the representation of salience theory of choice under risk derived in Lanzani (2020) to the choice between *bivariate* random variables by establishing that Axioms 1 – 7 are necessary and sufficient for the existence of a salience representation.

Theorem 1. *The preference set $P \subseteq \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ admits a salience representation if and only if it satisfies the Axioms 1, 2, 3 (a) and (b), 4 as well as 5 – 7.*

1.3.5 Validation of the Axioms

Using the technical Axiom 3 (c), we validate our extension of Lanzani's axioms by proving the following result: Impose the salience-weighted utility introduced in Definition 2 for $n = 2$. Then, the corresponding preference set P satisfies Axioms 1 – 7 if and only if σ is a salience function. In this sense, Axioms 1 – 7 are necessary and sufficient for nesting existing versions of salience theory as special cases.

Proposition 1. *Let $n = 2$. The preference set $P \subseteq \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ implied by the salience-weighted utility introduced in Definition 2 satisfies the Axioms 1 – 7 if and only if σ is symmetric, continuous, and satisfies ordering, diminishing sensitivity, and weak reflection.*

One remark is in order regarding the role of the technical Axiom 3 (c). Without Axiom 3 (c), Axiom 5 implies only a weaker version of ordering: namely, for any $x > y$ and any $\epsilon, \epsilon' > 0$ with $\epsilon + \epsilon' > 0$, we have $\sigma(x + \epsilon, y - \epsilon') > \sigma(x, y)$. Since $(x - x)\sigma(x, x) = 0$ irrespective of the value of $\sigma(x, x)$, the axioms used in Theorem 1 are not sufficient for the salience function to be continuous at $(x, x) \in \mathbb{R}^2$. In other words, without imposing the additional structure of Axiom 3 (c), the value $\sigma(x, x)$ is not identified from choice data.

1.3.6 The Structure of the Salience Representation

We conclude this section by proposing simple tests to elicit the structure of the salience representation; in particular, to elicit how the different dimensions of an option interact.

Definition 4. *The salience representation $\Psi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is said to be*

(a) additively separable *if there exists some $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that*

$$\sum_{(\mathbf{x},\mathbf{y})\in \text{supp } p} p(\mathbf{x}, \mathbf{y})\Psi(\mathbf{x}, \mathbf{y}) = \sum_{(\mathbf{x},\mathbf{y})\in \text{supp } p} p(\mathbf{x}, \mathbf{y})[\phi(x_1, y_1) + \phi(x_2, y_2)]; \tag{1.7}$$

(b) additively integrable *if there exists a function $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that*

$$\sum_{(\mathbf{x},\mathbf{y})\in \text{supp } p} p(\mathbf{x}, \mathbf{y})\Psi(\mathbf{x}, \mathbf{y}) = \sum_{(\mathbf{x},\mathbf{y})\in \text{supp } p} p(\mathbf{x}, \mathbf{y})\phi\left(x_1 + x_2, y_1 + y_2\right). \tag{1.8}$$

Buildung on the seminal ideas in Richard (1975) as well as some recent advances by Ebert and van de Kuilen (2016), we propose a test of additive separability by eliciting preferences over the correlation between the two dimensions of a bivariate random variable.

Definition 5 (Ebert and van de Kuilen, 2016)**.** *For the univariate random variables X_1 and X_2, and the constants $k, c \in \mathbb{R}_{>0}$, we define the following two compound lotteries $\mathbf{L} = \big((X_1 - k, X_2), 1/2; (X_1, X_2 - c), 1/2\big)$ and $\mathbf{L}' = \big((X_1, X_2), 1/2; (X_1 - k, X_2 - c), 1/2\big)$.*

(a) An agent is correlation averse *if she prefers* $\mathbf{L}$ *over* $\mathbf{L}'$ *for any such pair of lotteries.*

(b) An agent is correlation neutral *if she is indifferent between any such lotteries.*

(c) An agent is correlation seeking *if she prefers* $\mathbf{L}'$ *over* $\mathbf{L}$ *for any such pair of lotteries.*

An agent is said to be correlation averse if she avoids positive correlation across the two dimensions of an option. If, in contrast, an agent seeks such positive correlation across dimensions, she is said to be correlation seeking. Finally, an agent who does not care about the correlation structure across dimensions is said to be correlation neutral.

Due to diminishing sensitivity, a salient thinker is correlation neutral if and only if her behavior can be represented by an additively separable function Ψ. Otherwise, the salient thinker has strict preferences over the correlation across the two dimensions.

Theorem 2. *The salience representation Ψ is additively separable if and only if it implies correlation neutral behavior.*

A second, trivial observation is that, when assuming additively inegrable dimensions, only the joint distribution over the sum of outcomes (across dimensions) matters.

Remark 1. *By construction, if the salience representation is additively integrable, a change in the joint distribution of $\mathbf{X} = (X_1, X_2)$ and $\mathbf{Y} = (Y_1, Y_2)$ that does not affect the joint distribution of $\tilde{X} = X_1 + X_2$ and $\tilde{Y} = Y_1 + Y_2$ does not affect choices. In particular, a salient thinker with an additively integrable salience representation prefers $\mathbf{X}$ over $\mathbf{Y}$ if and only if she prefers $\tilde{X}$ over $\tilde{Y}$.*

1.4 Reflection of Correlation Preferences

We apply our insights from the previous section to study a salient thinker's correlation preferences under the assumption that the preference set P admits a salience representation that is *not* additively separable. Our main result in this section is the following:

Proposition 2. *If the salience representation is not additively separable, a salient thinker is neither correlation averse nor correlation neutral or correlation seeking. In particular:*

(a) If the two compound lotteries are non-negative, a salient thinker chooses $\mathbf{L}$ over $\mathbf{L}'$, which reveals correlation aversion.

(b) If the two compound lotteries are non-positive, a salient thinker chooses $\mathbf{L}'$ over $\mathbf{L}$, which reveals correlation seeking.

Proposition 2 is reminiscent of the difference in a salient thinker's attitudes toward non-negative and non-positive univariate risks (e.g., Dertwinkel-Kalt and Köster, 2020b). In line with the predictions of salience theory, for univariate risks, most people are risk averse on the positive domain, but risk seeking on the negative domain. To draw the connection between this result and Proposition 2, denote $\mathbf{L} = (L_1, L_2)$ and $\mathbf{L}' = (L'_1, L'_2)$. Since $Cov(L_1, L_2) < Cov(L'_1, L'_2)$, it follows immediately that $Var(L_1 + L_2) < Var(L'_1 + L'_2)$. In this sense, a correlation averse (seeking) choice reflects a risk averse (seeking) choice; that is, a choice of the option with a lower (higher) variance in outcomes. Hence, correlation aversion on the positive domain reflects a form of risk aversion on the positive domain, while correlation seekingness on the negative domain reflects a form of risk seekingness on the negative domain. This intuition is robust to allowing for $k, c \in \mathbb{R}$, taking into account that, for k and c having opposite signs, choosing $\mathbf{L}'$ over $\mathbf{L}$ reveals correlation aversion.[5]

Corollary 1. *Consider the choice between the compound lotteries* $\mathbf{L}$ *and* $\mathbf{L}'$ *as introduced in Definition 5, with the one exception that* $k, c \in \mathbb{R}$*. Then, the following statements hold:*

(a) If k *and* c *have the same signs, then the predictions derived in Proposition 2 hold.*

(b) If k *and* c *have opposite signs, then the predictions derived in Proposition 2 flip.*

In sum, if the two compound lotteries are non-negative (non-positive), a salient thinker is correlation averse (seeking).

Proposition 2 (a) is consistent with experimental evidence by Ebert and van de Kuilen (2016) on how subjects make intertemporal trade-offs. As predicted, a majority of subjects avoids a positive correlation in the sense of Definition 5 among a non-negative sum received today and a non-negative sum received three weeks later. To the best of our knowledge, Proposition 2 (b) has not been tested yet. But Ebert and van de Kuilen (2016) provide further evidence that is consistent with the more refined predictions in Corollary 1.

1.5 Concluding Remarks

We propose a general version of salience theory that allows us to study the binary choice between two bivariate random variables that nests its applications to choice under risk (Bordalo *et al.*, 2012) and consumer choice (Bordalo *et al.*, 2013b) as special cases. Extending the model from bivariate to multivariate random variables is, in principle, straightforward. Axioms 1 – 4 as well as Ordering (Axiom 5) and Reflexivity (Axiom 7) apply to bivariate as well as to multivariate random variables. But, with more than just two distinct dimensions, we need to slightly adjust Diminishing Sensitivity (Axiom 6) to capture the potential interdependence of the different dimensions. We discuss this technical issue, with basically no economic substance, in Appendix B.

Our model is not restricted to the extreme cases of additively separable (a key assumption in Bordalo *et al.*, 2013b) or additively integrable dimensions, but allows for an arbitrary interdependence of the dimensions. As a consequence, our general framework offers tools to test whether the two dimensions (e.g., quality and price) of a bivariate random variable are additively separable. This partially closes a gap in the model proposed by Bordalo *et al.* (2013b) for the choice between deterministic options, which takes as a primitive a set of exogenously given attributes and assumes that these attributes affect

[5] It is easily verified that $Cov(L_1, L_2) = Cov(X_1, X_2)$ and $Cov(L_1', L_2') = Cov(X_1, X_2) + 3/8kc$, which implies that $Cov(L_1, L_2) - Cov(L_1', L_2') = -3/8kc$ and thereby yields the claim.

behavior in an additively separable way. Our test of additive separability relies on the idea that an agent might have preferences over the correlation between the different dimensions of an option (e.g., Richard, 1975), which would be inconsistent with the assumption that the dimensions are additively separable. The novelty of our paper compared to existing studies — such as Richard (1975) or Ebert and van de Kuilen (2016), who also emphasize the elicitation of correlation preferences as way to test for additive separability — is that we explicitly characterize the relationship between correlation preferences and additive separability for a certain class of non-transitive preference relations.[6]

Appendix A: Proofs

A.0: Preliminary Results on Additive Separability

In this section, we prove some preliminary results on additive separability. Lemma 2 (stated in the main text) and Lemma 4 are special cases of Lemma 4 and Theorem B in Fishburn (1991). Since we use a different notation and since parts of the proofs simplify due to our specialized setup, however, we provide our own proofs here.

To begin with, we prove Lemma 2 which says that, *if Condition 1 holds, then $\succ_*$ on $\mathbb{R}^2$ (as defined in the main text) is a weak order, and thus $\sim_*$ is an equivalence relation.*

Proof of Lemma 2. $\succ_*$ is irreflexive: Suppose, for the sake of contradiction, that $\succ_*$ is not irreflexive; that is, there exists some $\mathbf{a} \in \mathbb{R}^2$ such that $\mathbf{a} \succ_* \mathbf{a}$. This implies, in particular, that there exist some vectors $\mathbf{r}_1 = \mathbf{a}, \mathbf{r}_2, \ldots, \mathbf{r}_m = \mathbf{a} \in \mathbb{R}^2$ with $\mathbf{r}_k \succeq_0 \mathbf{r}_{k+1}$ for all k and $\mathbf{r}_k \succ_0 \mathbf{r}_{k+1}$ for at least some k.

Let $m \geq 2$ and m even, as well as $\mathbf{a} \succeq_0 \mathbf{r}_2 \succeq_0 \ldots \succeq_0 \mathbf{r}_{m-1} \succeq_0 \mathbf{a}$ with at least one $\succ_0$ somewhere. Then, we can construct $\delta_{((a_1,r_{22}),(a_2,r_{21}))}$, $\delta_{((r_{32},r_{21}),(r_{31},r_{22}))}$, $\delta_{((r_{31},r_{42}),(r_{32},r_{41}))}$, $\ldots$, $\delta_{((a_2,r_{(m-1)1}),(a_1,r_{(m-1)2}))}$ as well as $\delta_{((r_{22},a_1),(r_{21},a_2))}$, $\delta_{((r_{21},r_{32}),(r_{22},r_{31}))}$, $\ldots$, $\delta_{((r_{(m-1)1},a_2),(r_{(m-1)2},a_1))}$ all in P, and at least one of them in $\hat{P}$. But

$$\left(\delta_{((a_1,r_{22}),(a_2,r_{21}))}, .., \delta_{((a_2,r_{(m-1)1}),(a_1,r_{(m-1)2}))}, \delta_{((r_{22},a_1),(r_{21},a_2))}, .., \delta_{((r_{(m-1)1},a_2),(r_{(m-1)2},a_1))}\right) \in E_{2(m-1)},$$

which contradicts Condition 1.

Let $m \geq 3$ and m odd, as well as $\mathbf{a} \succeq_0 \mathbf{r}_2 \succeq_0 \ldots \succeq_0 \mathbf{r}_{m-1} \succeq_0 \mathbf{a}$ with at least one $\succ_0$ somewhere. Then, we can construct $\delta_{((a_1,r_{22}),(a_2,r_{21}))}$, $\delta_{((r_{32},r_{21}),(r_{31},r_{22}))}$, $\delta_{((r_{31},r_{42}),(r_{32},r_{41}))}$,

[6] Richard (1975) characterizes correlation preferences only for expected utility theory (via the sign of the utility function's cross-derivatives). Ebert and van de Kuilen (2016) allow for violations of expected utility theory, but do not provide a formal analysis of the class of non-transitive preferences that we study in this paper.

..., $\delta_{((a_2,r_{(m-1)1}),(a_1,r_{(m-1)2}))}$ all in P, and at least one of them in $\hat{P}$. But

$$\left(\delta_{((a_1,r_{22}),(a_2,r_{21}))}, \delta_{((r_{32},r_{21}),(r_{31},r_{22}))}, \delta_{((r_{31},r_{42}),(r_{32},r_{41}))}, \ldots, \delta_{((a_2,r_{(m-1)1}),(a_1,r_{(m-1)2}))}\right) \in E_{m-1},$$

which contradicts Condition 1. Hence, $\succ_*$ is irreflexive.

$\succ_*$ is asymmetric: Given that $\succ_*$ is irreflexive, it is sufficient to show that $\succ_*$ is transitive. Suppose $\mathbf{a} \succ_* \mathbf{b}$ and $\mathbf{b} \succ_* \mathbf{c}$. The former implies that there exist some $\mathbf{q}_1 = \mathbf{a}, \mathbf{q}_2, \ldots, \mathbf{q}_m = \mathbf{b} \in \mathbb{R}^2$ with $\mathbf{q}_k \succeq_0 \mathbf{q}_{k+1}$ for all k and $\mathbf{q}_k \succ_0 \mathbf{q}_{k+1}$ for some k, while the latter implies that there exist $\mathbf{r}_1 = \mathbf{b}, \mathbf{r}_2, \ldots, \mathbf{r}_{m'} = \mathbf{c} \in \mathbb{R}^2$ with $\mathbf{r}_k \succeq_0 \mathbf{r}_{k+1}$ for all k and $\mathbf{r}_k \succ_0 \mathbf{r}_{k+1}$ for some k. Together this implies that there exist $\mathbf{s}_1 = \mathbf{a}, \mathbf{s}_2 = \mathbf{q}_2, \ldots, \mathbf{s}_m = \mathbf{b}, \mathbf{s}_{m+1} = \mathbf{r}_1, \ldots, \mathbf{s}_{m+m'} = \mathbf{c} \in \mathbb{R}^2$ with $\mathbf{s}_k \succeq_0 \mathbf{s}_{k+1}$ for all k and $\mathbf{s}_k \succ_0 \mathbf{s}_{k+1}$ for some k. But this implies $\mathbf{a} \succ_* \mathbf{c}$, which was to be proven.

$\succ_*$ is negatively transitive: Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ with $\mathbf{a} \succ_* \mathbf{b}$. It is sufficient to show that for any $\mathbf{r} \in \mathbb{R}^2$ either $\mathbf{a} \succ_* \mathbf{r}$ or $\mathbf{r} \succ_* \mathbf{b}$. Since $\mathbf{a} \succ_* \mathbf{b}$, there exists some $\mathbf{s} \in \mathbb{R}^2$ such that $\mathbf{a} \succeq_0 \ldots \succeq_0 \mathbf{s} \succeq_0 \ldots \succeq_0 \mathbf{b}$ with $\succ_0$ somewhere. Without loss of generality, assume that $\succ_0$ occurs between $\mathbf{a}$ and $\mathbf{s}$. If $\mathbf{s} \succeq_0 \mathbf{r}$, then $\mathbf{a} \succ_* \mathbf{r}$. If $\mathbf{r} \succ_0 \mathbf{s}$, then $\mathbf{r} \succ_* \mathbf{b}$. □

In order to prove Lemma 4 below — which is the main result of this section — we will use the following basic result on the representation of a weak order on an uncountable set, which can be found, for instance, in Fishburn (1970, Theorem 3.1) or Krantz, Luce, Suppes and Tversky (1971, Theorem 2 in Chapter 2).

Lemma 3. *There is a function ϕ on Z such that, for all $\mathbf{a}, \mathbf{b} \in Z$,*

$$\mathbf{a} \succ' \mathbf{b} \quad \Longleftrightarrow \quad \phi(\mathbf{a}) > \phi(\mathbf{b}),$$

if and only if $\succ'$ on Z is a weak order whose quotient set $Z/\sim'$ includes a countable subset that is $\succ'$-order dense in $Z/\sim'$.

Lemma 4 (a special case of Theorem B in Fishburn, 1991) finally links Conditions 1 and 2 to the additive separability of a skew-symmetric representation as in Lemma 1.

Lemma 4. *Let the preference set $P \subseteq \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ satisfy Axioms 1 – 4. Its skew-symmetric representation Ψ is additively separable if and only if Conditions 1 and 2 hold.*

Proof. Since Axioms 1 – 4 hold, by Lemma 1, $\mathbf{a} \succ_0 \mathbf{b}$ if and only if $\Psi((a_1, b_2), (a_2, b_1)) > 0$. We have to show that *there exists some $\phi : \mathbb{R}^2 \to \mathbb{R}$ such that $\Psi((a_1, b_2), (a_2, b_1)) = \phi(a_1, a_2) + \phi(b_2, b_1)$ if and only if Condition 1 and Condition 2 are satisfied.*

(If) Suppose Conditions 1 and 2 hold. By Lemma 2, $\succ_*$ on $\mathbb{R}^2/\sim_*$ is a weak order.

Denote as $D := \{(x, x) \in \mathbb{R}^2\}$ and let $\mathbf{d} = (d, d) \in D$. By Condition 1, for any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ and any $\mathbf{d} \in D$, we obtain (i) $(a_1, a_2) \succ_* (d, d) \Leftrightarrow (d, d) \succ_* (a_2, a_1)$ as well as (ii) $(a_1, a_2) \succ_* (b_1, b_2) \Leftrightarrow (b_2, b_1) \succ_* (a_2, a_1)$ and (iii) $(a_1, a_2) \sim_* (b_1, b_2) \Leftrightarrow (b_2, b_1) \sim_* (a_2, a_1)$. We only prove (i) here, since the other two cases are analogous. To save on tedious notation and without loss of generality, we can consider the case where $(a_1, a_2) \succ_0 (r_1, r_2) \succeq_0 (d, d)$ for some $\mathbf{r} \in \mathbb{R}^2$. First, $(a_1, a_2) \succ_0 (r_1, r_2)$ implies that $\delta_{((a_1,r_2),(a_2,r_1))} \in \hat{P}$. Since $\left(\delta_{((a_1,r_2),(a_2,r_1))}, \delta_{((a_2,r_1),(a_1,r_2))}\right) \in E_2$, Condition 1 implies $\delta_{((a_2,r_1),(a_1,r_2))} \notin P$. But this implies that $(r_2, r_1) \succ_0 (a_2, a_1)$ has to hold. By the same type of argument, $(r_1, r_2) \succeq_0 (d, d)$ implies $(d, d) \succeq_0 (r_2, r_1)$. In conclusion, we have $(a_1, a_2) \succ_* (d, d) \Rightarrow (d, d) \succ_* (a_2, a_1)$. The other direction works in the same way.

The preceding considerations imply that D is an equivalence class in $\mathbb{R}^2/\sim_*$. Let $(d, d), (e, e) \in D$. For the sake of a contradiction, assume $(d, d) \succ_* (e, e)$. By (i), it has to hold that $(e, e) \succ_* (d, d)$. But $\succ_*$ is a weak order and thus asymmetric; a contradiction.

Now, for some $\mathbf{d} \in D$, define

$$\mathcal{A} := \{\mathbf{a} \in \mathbb{R}^2 : \mathbf{a} \succ_* \mathbf{d}\} \quad \text{and} \quad \mathcal{B} := \{\mathbf{b} \in \mathbb{R}^2 : \mathbf{d} \succ_* \mathbf{b}\}.$$

The observations in the preceding paragraph imply that $\mathcal{B}$ is the inverse of $\mathcal{A}$, and if $A \in \mathbb{R}^2/\sim_*$ and $A \neq D$, then $A \subseteq \mathcal{A}$ if and only if $A^{-1} \subseteq \mathcal{B}$. Finally, any $C \in (\mathbb{R}^2/\sim_*)\backslash D$ has to be either in $\mathcal{A}$ or in $\mathcal{B}$.

By Condition 2 and Lemma 3, there exists some functional ϕ on $\mathcal{A} \cup D$ such that for any $\mathbf{a}, \mathbf{b} \in \mathcal{A} \cup D$ and any $\mathbf{d} \in D$, $\mathbf{a} \succ_* \mathbf{b} \iff \phi(\mathbf{a}) > \phi(\mathbf{b})$ and $\phi(\mathbf{d}) = 0$. Next, we extend ϕ to $\mathcal{B}$ as follows: for any $(b_1, b_2) \in \mathcal{B}$, let $\phi(b_1, b_2) = -\phi(b_2, b_1)$. (Here, it is important to keep in mind that $(b_2, b_1) \in \mathcal{A}$.) Since $\mathcal{A} \cup D \cup \mathcal{B} = \mathbb{R}^2$, the preceding observations imply that ϕ preserves $\succ_*$ on $\mathbb{R}^2$.

Finally, we observe that, for any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$,

$$\begin{aligned} \delta_{((a_1,b_1),(a_2,b_2))} \in \hat{P} &\Rightarrow (a_1, a_2) \succ_0 (b_2, b_1) \Rightarrow (a_1, a_2) \succ_* (b_2, b_1) \\ &\Rightarrow \phi(a_1, a_2) > \phi(b_2, b_1) \Rightarrow \phi(a_1, a_2) + \phi(b_1, b_2) > 0 \end{aligned}$$

as well as

$$\begin{aligned} \delta_{((a_1,b_1),(a_2,b_2))} \in P \setminus \hat{P} &\Rightarrow (a_1, a_2) \sim_0 (b_2, b_1) \stackrel{(\star)}{\Rightarrow} (a_1, a_2) \sim_* (b_2, b_1) \\ &\Rightarrow \phi(a_1, a_2) = \phi(b_2, b_1) \Rightarrow \phi(a_1, a_2) + \phi(b_1, b_2) = 0. \end{aligned}$$

The implication $(\star)$ can be verified as follows: Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ with $\mathbf{a} \sim_0 \mathbf{b}$. For the sake of a contradiction, assume $\mathbf{a} \succ_* \mathbf{b}$. This implies that there exist $\mathbf{r}_1 = \mathbf{a}, \mathbf{r}_2, \ldots, \mathbf{r}_m = \mathbf{b} \in \mathbb{R}^2$ with $\mathbf{r}_k \succeq_0 \mathbf{r}_{k+1}$ for all k and $\mathbf{r}_k \succ_0 \mathbf{r}_{k+1}$ for some k. Define $\mathbf{q}_1 = \mathbf{r}_1$ for any $k \leq m$ and $\mathbf{q}_{m+1} = \mathbf{a}$. Since $\mathbf{b} \succeq_0 \mathbf{a}$, we have that $\mathbf{q}_k \succeq_0 \mathbf{q}_{k+1}$ for all k and $\mathbf{q}_k \succ_0 \mathbf{q}_{k+1}$ for some k. Hence, $\mathbf{a} \succ_* \mathbf{a}$; a contradiction, since $\succ_*$ on $\mathbb{R}^2/\sim_*$ is irreflexive.

In sum, under Conditions 1 and 2, the preference set P can be represented by an additively separable function $\Psi((a_1,b_2),(a_2,b_1)) = \phi(a_1,a_2) + \phi(b_2,b_1)$.

(Only if) Suppose that Ψ is additively separable. Then, skew-symmetry implies that, for any $(\delta_{(\mathbf{x}^1,\mathbf{y}^1)},\ldots,\delta_{(\mathbf{x}^m,\mathbf{y}^m)}) \in E_m$, we have $\sum_{j=1}^m \phi(x_1^j,y_1^j) + \phi(x_2^j,y_2^j) = 0$. Now suppose that $\delta_{(\mathbf{x}^1,\mathbf{y}^1)},\ldots\delta_{(\mathbf{x}^{m-1},\mathbf{y}^{m-1})} \in P$, which implies $\sum_{j=1}^{m-1} \phi(x_1^j,y_1^j) + \phi(x_2^j,y_2^j) \geq 0$. Then, we must have $\phi(x_1^m,y_1^m) + \phi(x_2^m,y_2^m) \leq 0$ and thus $\delta_{(\mathbf{x}^m,\mathbf{y}^m)} \notin \hat{P}$. Hence, Condition 1 holds.

We define the binary relation $\succ'$ on $\mathbb{R}^2$ as follows: for any $\mathbf{a},\mathbf{b} \in \mathbb{R}^2$, let $\mathbf{a} \succ' \mathbf{b}$ if and only if $\phi(a_1,a_2) > \phi(b_1,b_2)$. Since Ψ is additively separable by assumption, we have $\mathbf{a} \succ' \mathbf{b}$ if and only if $\mathbf{a} \succ_0 \mathbf{b}$. By Lemma 3, it will be sufficient to show that *in this case* $\mathbf{a} \succ_* \mathbf{b}$ *if and only if* $\mathbf{a} \succ' \mathbf{b}$. (If) $\mathbf{a} \succ' \mathbf{b}$ implies $\mathbf{a} \succ_0 \mathbf{b}$ and therefore $\mathbf{a} \succ_* \mathbf{b}$. (Only if) If $\mathbf{a} \succ_* \mathbf{b}$, then there exist $\mathbf{r}_1 = \mathbf{a}, \mathbf{r}_2, \ldots, \mathbf{r}_m = \mathbf{b} \in \mathbb{R}^2$ with $\mathbf{r}_k \succeq_0 \mathbf{r}_{k+1}$ for all k and $\mathbf{r}_k \succ_0 \mathbf{r}_{k+1}$ for some k. Given that Ψ is additively separable, this implies $\phi(a_1,a_2) \geq \phi(r_{11},r_{12}) \geq \ldots \geq \phi(r_{(m-1)1},r_{(m-1)2}) \geq \phi(b_1,b_2)$ with at least one inequality being strict. As a consequence, we have $\phi(a_1,a_2) > \phi(b_1,b_2)$ and therefore $\mathbf{a} \succ' \mathbf{b}$. Given that $\mathbf{a} \succ_* \mathbf{b}$ if and only if $\mathbf{a} \succ' \mathbf{b}$, we also know that $\mathbf{a} \succ_* \mathbf{b}$ if and only if $\phi(a_1,a_2) > \phi(b_1,b_2)$. But then Lemma 3 implies that Condition 2 has to hold. □

A.1: General Framework

Proof of Theorem 1. By Lemma 1, the preference set $P \subseteq \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ satisfies Axioms 1, 2, 3 (a) and (b), and 4 if and only if there exists a continuous and skew-symmetric function $\Psi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ that is strictly increasing in its first two arguments and strictly decreasing in its last two arguments, such that

$$p \in P \quad \Longleftrightarrow \sum_{(\mathbf{x},\mathbf{y})\in \mathrm{supp}\ p} p(\mathbf{x},\mathbf{y})\Psi(\mathbf{x},\mathbf{y}) \geq 0. \tag{1.9}$$

To complete the proof, we establish the following three lemmata

Lemma 5. *Suppose that the Axioms 1, 2, 3 (a) and (b), and 4 are satisfied. Then, the preference set $P \subseteq \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ satisfies Ordering if and only if Ψ satisfies Property (a).*

Proof. Let $\mathbf{x} > \mathbf{y}$ and $\alpha, \beta \in [0,1]$ with $\beta > \alpha$ and either $\beta < 1$ or $\alpha > 0$. Then, we have

$$p = \left((\mathbf{x},\mathbf{y}), \frac{\beta-\alpha}{1+\beta-\alpha}; (\alpha\mathbf{x}+(1-\alpha)\mathbf{y}, \beta\mathbf{x}+(1-\beta)\mathbf{y}), \frac{1}{1+\beta-\alpha} \right) \in \hat{P}$$

if and only if

$$\frac{(\beta-\alpha)}{1+\beta-\alpha}\Psi(\mathbf{x},\mathbf{y}) > \frac{1}{1+\beta-\alpha}\Psi(\beta\mathbf{x}+(1-\beta)\mathbf{y}, \alpha\mathbf{x}+(1-\alpha)\mathbf{y}),$$

which yields the statement. □

Lemma 6. *Suppose that the Axioms 1, 2, 3 (a) and (b), 4 and 5 are satisfied. Then, $P \subseteq \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ satisfies Diminishing Sensitivity if and only if Ψ satisfies Property (b).*

Proof. (If) Let Ψ satisfy Property (b). Then, for any $\mathbf{x} > \mathbf{y} \geq 0$ with $x_i > y_i$ and $\epsilon > 0$,

$$\Psi(\mathbf{x}, \mathbf{y}) + \Psi(\mathbf{y} + \mathbf{e}_i\epsilon, \mathbf{x} + \mathbf{e}_i\epsilon) > 0,$$

which in turn implies that

$$p = \left((\mathbf{x}, \mathbf{y}), \frac{1}{2}; (\mathbf{y} + \mathbf{e}_i\epsilon, \mathbf{x} + \mathbf{e}_i\epsilon), \frac{1}{2}\right) \in \hat{P}.$$

If $x_i = y_i$, then by Property (b),

$$\Psi(\mathbf{x}, \mathbf{y}) + \Psi(\mathbf{y} + \mathbf{e}_i\epsilon, \mathbf{x} + \mathbf{e}_i\epsilon) \geq 0,$$

holding with equality if and only if Ψ is additively separable. By Lemma 4, this is indeed the case if and only if Condition 1 and Condition 2 are satisfied, which was to be proven.

(Only if) Suppose that the preference set P satisfies Diminishing Sensitivity. Then, for any $\mathbf{x} > \mathbf{y} \geq \mathbf{0}$ with $x_i > y_i$ and any $\epsilon > 0$, we have

$$p = \left((\mathbf{x}, \mathbf{y}), \frac{1}{2}; (\mathbf{y} + \mathbf{e}_i\epsilon, \mathbf{x} + \mathbf{e}_i\epsilon), \frac{1}{2}\right) \in \hat{P},$$

which implies $\Psi(\mathbf{x}, \mathbf{y}) + \Psi(\mathbf{y} + \mathbf{e}_i\epsilon, \mathbf{x} + \mathbf{e}_i\epsilon) > 0$. Hence, the first Part of Property (b) follows by skew-symmetry. If $x_i = y_i$, then, by Diminishing Sensitivity,

$$p = \left((\mathbf{x}, \mathbf{y}), \frac{1}{2}; (\mathbf{y} + \mathbf{e}_i\epsilon, \mathbf{x} + \mathbf{e}_i\epsilon), \frac{1}{2}\right) \in P \setminus \hat{P}$$

if and only if Condition 1 and Condition 2 are satisfied, which is, by Lemma 4, equivalent to Ψ being additively separable. This proves the claim. □

Lemma 7. *Suppose that the Axioms 1, 2, 3 (a) and (b), 4, 5, and 6 are satisfied. Then, $P \subseteq \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ satisfies Reflexivity if and only if Ψ satisfies Property (c).*

Proof. Let $\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z} \geq 0$ with $|w_i - x_i| = |z_i - y_i|$ for any $i \in \{1, 2\}$. Then,

$$\left((\mathbf{w}, \mathbf{x})\frac{1}{2}; (\mathbf{y}, \mathbf{z}), \frac{1}{2}\right) \in \hat{P} \quad \Longleftrightarrow \quad \left((-\mathbf{x}, -\mathbf{w})\frac{1}{2}; (-\mathbf{z}, -\mathbf{y}), \frac{1}{2}\right) \in \hat{P}$$

if and only if

$$\Psi(\mathbf{w}, \mathbf{x}) + \Psi(\mathbf{y}, \mathbf{z}) > 0 \quad \Longleftrightarrow \quad \Psi(-\mathbf{x}, -\mathbf{w}) + \Psi(-\mathbf{z}, -\mathbf{y}) > 0.$$

By skew-symmetry, this statement is equivalent to

$$\Psi(\mathbf{w},\mathbf{x}) > \Psi(\mathbf{z},\mathbf{y}) \quad \Longleftrightarrow \quad \Psi(-\mathbf{w},-\mathbf{x}) < \Psi(-\mathbf{z},-\mathbf{y}),$$

which was to be proven. □

We have established that the preference set $P \subseteq \Delta(\mathbb{R}^2 \times \mathbb{R}^2)$ admits a salience representation if and only if Axioms 1, 2, 3 (a) and (b), 4 as well as 5 – 7 are satisfied. □

A.2: Validation of the Axioms

Proof of Proposition 1. Without loss of generality, we can assume that the two dimensions are additively integrable. We then have to show that *the preference set P implied by*

$$\Psi(\mathbf{x},\mathbf{y}) := \Big((x_1+x_2)-(y_1+y_2)\Big)\sigma\left(x_1+x_2, y_1+y_2\right)$$

satisfies Axioms 1 – 7 if and only if $\sigma : \mathbb{R}\times\mathbb{R} \to \mathbb{R}_+$ is symmetric, continuous, and satisfies ordering, diminishing sensitivity, and weak reflection.

(If) This direction is straightfoward and the proof is therefore omitted.

(Only if) W.l.o.g., we can treat the two options as one-dimensional random variables. We will use the properties of the representing function derived in Theorem 1.

"Skew-symmetry $\Rightarrow \sigma$ is symmetric": Trivial.

"Continuity $\Rightarrow \sigma$ is continuous at $(x,y) \in \mathbb{R}^2$ with $x \neq y$": Since the representation Ψ is continuous, for any $(x,y) \in \mathbb{R}^2$ and any sequence $(z_n)_{n\in\mathbb{N}}$ that converges to zero,

$$\lim_{n\to\infty}(x+z_n-y)\sigma(x+z_n,y)-(x-y)\sigma(x,y)=0.$$

This implies in particular that

$$(x-y)\lim_{n\to\infty}\sigma(x+z_n,y)-\sigma(x,y)=0.$$

Hence, for any $x \neq y$ and any such sequence $(z_n)_{n\in\mathbb{N}}$, we must have

$$\lim_{n\to\infty}\sigma(x+z_n,y)-\sigma(x,y)=0,$$

which implies that the function σ is continuous at any $(x,y) \in \mathbb{R}^2$ with $x \neq y$.

"Continuity at Identity $\Rightarrow \sigma$ is continuous at $(x,x) \in \mathbb{R}^2$": Take two sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ with x_n converging to $x \in \mathbb{R}$ from above and y_n converging to the same x from below. Let n be large enough such that $\max\{x_n-x, x-y_n\} < 1$. By Axiom 3 (c), there exists some $m \in \mathbb{N}$, such that for any $n \geq m$ and any $z \in \mathbb{R}$ as well as any $w > 0$,

$$(x - x_n)\sigma(x, x_n)(1 - (x_n - x)) + w\sigma(z + w, z)(x_n - x) \geq 0$$

and

$$(y_n - x)\sigma(y_n, x)(1 - (x - y_n)) + w\sigma(z + w, z)(x - y_n) \geq 0.$$

Re-arranging these two inequalities gives

$$(x_n - x)\Big(w\sigma(z + w, z) - \sigma(x, x_n)(1 - (x_n - x))\Big) \geq 0$$

and

$$(x - y_n)\Big(w\sigma(z + w, z) - \sigma(y_n, x)(1 - (x - y_n))\Big) \geq 0,$$

which in turn requires that, for large enough n,

$$w\sigma(z + w, z) \geq \sigma(x_n, x)(1 - (x_n - x))$$

and

$$w\sigma(z + w, z) \geq \sigma(y_n, x)(1 - (x - y_n)),$$

where we further use the symmetry of the salience function in the first inequality. Since these inequalities have to hold for all $z \in \mathbb{R}$ and all $w > 0$, it has to be true that

$$\lim_{n\to\infty} \sigma(x_n, x) = 0 = \lim_{n\to\infty} \sigma(y_n, x).$$

Since the right- and left-limit coincide, we conclude that σ is continuous at $(x, x) \in \mathbb{R}^2$.

"(a) $\Rightarrow$ σ satisfies ordering": Suppose, for the sake of a contradiction, that (a) holds, but that σ does not satisfy ordering. Hence, there exist $x, y, w, z \in \mathbb{R}$ with $x \geq z > w \geq y$ and $[w, z] \subset [y, x]$ such that $\sigma(x, y) \leq \sigma(w, z)$. Define $\alpha := \frac{w-y}{x-y}$ and $\beta := \frac{z-y}{x-y}$, and notice that $\alpha, \beta \in [0, 1]$ with $\beta > \alpha$ and either $\beta < 1$ or $\alpha > 0$. Now Property (a) implies that

$$\begin{aligned}(\beta - \alpha)(x - y)\sigma(x, y) &> (\beta - \alpha)(x - y)\sigma(\alpha x + (1 - \alpha)y, \beta x + (1 - \beta)y)\\ &= (\beta - \alpha)(x - y)\sigma(w, z),\end{aligned}$$

which holds if and only if $\sigma(x, y) > \sigma(w, z)$; a contradiction.

It remains to be shown that (a) implies $\sigma(x, y) > \sigma(w, w)$ for any $x \geq w \geq y$ with at least one inequality being strict. Suppose, for the sake of a contradiction, that there exist $x \geq w \geq y$, with at least one inequality being strict, such that $\sigma(x, y) \leq \sigma(w, w)$. W.l.o.g., let $x > w > y$. Then, by the preceding argument, there exist $x' \in (w, x)$ and $y' \in (y, w)$ such that $\sigma(x', y') < \sigma(w, w)$. Since we have already seen that σ is continuous, there exists some $\epsilon > 0$ with $w + \epsilon < x'$ such that $\sigma(w + \epsilon, w) > \sigma(x', y')$; a contradiction.

"(b) $\Rightarrow$ σ satisfies diminishing sensitivity": By (b), for any $x > y \geq 0$ and any $\epsilon > 0$,

$$(x - y)\sigma(x, y) > (x - y)\sigma(x + \epsilon, y + \epsilon),$$

which holds if and only if $\sigma(x, y) > \sigma(x + \epsilon, y + \epsilon)$.

"(c) $\Rightarrow$ σ satisfies weak reflection": By (d), for any $w, x, y, z \geq 0$ with $w-x = z-y > 0$,

$$(w - x)\sigma(w, x) > (z - y)\sigma(z, y) \quad \iff \quad (w - x)\sigma(-w, -x) < (z - y)\sigma(-z, -y),$$

which indeed holds if and only if $\sigma(w, x) > \sigma(z, y) \iff \sigma(-w, -x) < \sigma(-z, -y)$. □

A.3: The Structure of the Salience Representation

Proof of Theorem 2. (If) Without loss of generality, we can consider the choice between the lotteries $\mathbf{L} = [(x, 0), 1/2; (0, x), 1/2]$ and $\mathbf{L}' = [(0, 0), 1/2; (x, x), 1/2]$ for some $x > 0$. A salient thinker is indifferent between $\mathbf{L}$ and $\mathbf{L}'$ if and only if

$$\begin{aligned} 0 = \Psi((x, 0), (0, 0)) + \Psi((0, x), (x, x)) &= \Psi((x, 0), (0, 0)) - \Psi((x, x), (0, x)) \\ &\leq \Psi((x, x), (0, x)) - \Psi((x, x), (0, x)) = 0, \end{aligned}$$

whereby the second equality holds by skew-symmetry and the weak inequality holds by Property (b). The claim then follows from the fact that the inequality is strict if and only if the function Ψ is not additively separable. (Only if) Trivial. □

A.4: Reflection of Correlation Preferences

Proof of Proposition 2. We can fully parameterize the joint distribution of the compound lotteries $\mathbf{L}$ and $\mathbf{L}'$ via a single parameter $\eta \in [0, 1]$ as follows:

Probability	$\frac{1}{2}\eta$	$\frac{1}{2}(1-\eta)$	$\frac{1}{2}(1-\eta)$	$\frac{1}{2}\eta$
$\mathbf{L}$	$(X_1 - k, X_2)$	$(X_1 - k, X_2)$	$(X_1, X_2 - c)$	$(X_1, X_2 - c)$
$\mathbf{L}'$	(X_1, X_2)	$(X_1 - k, X_2 - c)$	(X_1, X_2)	$(X_1 - k, X_2 - c)$

Table 1.1: *Joint distribution of the compound lotteries* $\mathbf{L}$ *and* $\mathbf{L}'$.

Let $X_1 + X_2 > k + c$ with $k, c > 0$. A salient thinker chooses $\mathbf{L}$ over $\mathbf{L}'$ if and only if

$$\sum_{\text{supp } p} p(x_1, x_2)\eta \Big[\Psi(x_1 - k, x_2, x_1, x_2) + \Psi(x_1, x_2 - c, x_1 - k, x_2 - c)\Big] + \sum_{\text{supp } p} p(x_1, x_2)(1-\eta)\Big[\Psi(x_1 - k, x_2, x_1 - k, x_2 - c) + \Psi(x_1, x_2 - c, x_1, x_2)\Big] > 0.$$

By skew-symmetry, this inequality holds if and only if

$$\sum_{\text{supp } p} p(x_1, x_2)\eta \Big[\Psi(x_1, x_2 - c, x_1 - k, x_2 - c) - \Psi(x_1, x_2, x_1 - k, x_2)\Big] + \sum_{\text{supp } p} p(x_1, x_2)(1-\eta)\Big[\Psi(x_1 - k, x_2, x_1 - k, x_2 - c) - \Psi(x_1, x_2, x_1, x_2 - c)\Big] > 0.$$

Since $k, c > 0$, by Property (b), both brackets are positive if and only if the salience representation is not additively separable. The proof of Part (b) is analogous. □

Proof of Corollary 1. Let $X_1 + X_2 > k + c$ with $k, c < 0$. By similar arguments as in the proof of Proposition 2, a salient thinker chooses $\mathbf{L}$ over $\mathbf{L}'$ if and only if

$$\sum_{\text{supp } p} p(x_1, x_2)\eta \Big[\Psi(x_1 - k, x_2, x_1, x_2) - \Psi(x_1 - k, x_2 - c, x_1, x_2 - c)\Big] + \sum_{\text{supp } p} p(x_1, x_2)(1-\eta)\Big[\Psi(x_1, x_2 - c, x_1, x_2) - \Psi(x_1 - k, x_2 - c, x_1 - k, x_2)\Big] > 0.$$

Since $k, c < 0$, by Property (b), both brackets are positive if and only if the salience representation is not additively separable. The proof of Part (b) is analogous. □

Appendix B: Diminishing Sensitivity with Three or More Dimensions

As discussed in Fishburn (1991), the Conditions 1 and 2, as introduced in Section 3, are necessary and sufficient for the additive separability of the representing function if and only if $n = 2$. If $n \geq 3$, Fishburn (1990a,b) provides the following necessary and sufficient conditions for the representation being additively separable, that is, for $\Psi(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \phi_i(x_i, y_i)$ to hold:[7]

[7] To be precise, Fishburn (1990a,b) adds one more axiom that imposes structure and ensures that all dimensions are relevant: Since $\mathbb{R}^n$ endowed with the base of open balls (i.e., the product topology) is a connected and separable topological space, the first part of Axiom 1 in Fishburn (1990a,b) is automatically satisfied in our setup. The second part of his Axiom 1 says that for any $i \in \{1, \dots, n\}$ there exist $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ with $x_i \neq y_i$ and $x_j = y_j$ for any $j \neq i$ such that $\delta_{(\mathbf{x},\mathbf{y})} \in \hat{P}$.

Condition A.1 (Dimension Independence). *For any* $\delta_{(\mathbf{x}^j,\mathbf{y}^j)} \in \mathbb{R}^n \times \mathbb{R}^n$, *whenever* $(\delta_{(\mathbf{x}^1,\mathbf{y}^1)}, \delta_{(\mathbf{x}^2,\mathbf{y}^2)}, \delta_{(\mathbf{x}^3,\mathbf{y}^3)}, \delta_{(\mathbf{x}^4,\mathbf{y}^4)}) \in E_4$ *and* $\delta_{(\mathbf{x}^j,\mathbf{y}^j)} \in P$ *for any* $j \in \{1,2,3\}$, $\delta_{(\mathbf{x}^4,\mathbf{y}^4)} \notin \hat{P}$.

Condition A.2 (Continuity). *For any* $i \in \{1,\dots,n\}$ *and any* $\mathbf{w} \in \mathbb{R}^{n-1}$, *the set* $\{\mathbf{a} \in \mathbb{R}^2 : \delta_{(\mathbf{x},\mathbf{y})} \in \hat{P}$ *for* $(\mathbf{x}_{-i},\mathbf{y}_{-i}) = \mathbf{w}$ *and* $(x_i,y_i) = \mathbf{a}\}$ *lies in the product topology of* $\mathbb{R}^n \times \mathbb{R}^n$.

Using these conditions, we could state the following version of diminishing sensitivity.

Axiom 8 (Diminishing Sensitivity, $n \geq 3$). *For any* $\mathbf{x} > \mathbf{y} \geq 0$ *and any* $\epsilon > 0$,

$$p = \left((\mathbf{x},\mathbf{y}), \frac{1}{2}; (\mathbf{y}+\mathbf{e}_i\epsilon, \mathbf{x}+\mathbf{e}_i\epsilon), \frac{1}{2}\right) \in P,$$

and, whenever $x_i > y_i$, *then* $p \in \hat{P}$. *For any* $\mathbf{x} > \mathbf{y} \geq 0$ *with* $x_i = y_i$ *and any* $\epsilon > 0$,

$$p = \left((\mathbf{x},\mathbf{y}), \frac{1}{2}; (\mathbf{y}+\mathbf{e}_i\epsilon, \mathbf{x}+\mathbf{e}_i\epsilon), \frac{1}{2}\right) \in P \setminus \hat{P}$$

if Condition A.1 and Condition A.2 are satisfied.

The main difference compared to the diminishing sensitvity axiom for $n = 2$ is that in the last line it reads “if” instead of “if and only if”, which makes Axiom 8 obviously less useful. The following example illustrates that this is indeed a limitation.

Example 5. Let $n = 3$. Consider an agent who aims to maximize the salience-weighted utility introduced in Definition 2, whereby she integrates Dimensions 1 and 2, but separates Dimension 3. This agent chooses the safe option (x_1, x_2, x_3) over the safe option (y_1, y_2, y_3) if and only if

$$\big((x_1+x_2)-(y_1+y_2)\big)\sigma(x_1+x_2, y_1+y_2) + (x_3-y_3)\sigma(x_3,y_3) > 0.$$

Hence, for any $\epsilon > 0$ and any $z \in \mathbb{R}$, we have

$$p = \big((x_1,x_2,z,y_1,y_2,z), 1/2; (y_1,y_2,z+\epsilon,x_1,x_2,z+\epsilon), 1/2\big) \in P \setminus \hat{P},$$

although the salience representation does not satisfy Conditions A.1 and A.2.

Fixing this limitation seems to be a purely technical issue without much economic substance. In this sense our model straightforwardly extends to higher dimensions by simply replacing Condition A.1 and A.2 by the appropriate technical conditions.

Chapter 2

Salience and Skewness Preferences

Co-authored by Markus Dertwinkel-Kalt

2.1 Introduction

Most puzzles in choice under risk can be attributed to the *skewness* (i.e., the third standardized moment) of the underlying probability distribution. First, due to monotonicity of preferences most people like, *ceteris paribus*, risks with a higher expected value (i.e., a higher first moment). Second, due to intrinsic risk aversion most people dislike, *ceteris paribus*, risks with a higher variance (i.e, a higher second moment). But the conventional wisdom that *in general* people prefer risks with a higher expected value and/or a lower variance can be overturned by preferences over the skewness of a risk. Many individuals, for instance, overpay for insurance with low deductibles against left-skewed risks that yield a rather large loss with a small probability (e.g., Sydnor, 2010; Barseghyan, Molinari, O'Donoghue and Teitelbaum, 2013). But at the same time, these individuals often seek right-skewed risks such as casino gambles that realize a large gain with a tiny probability (e.g., Mao, 1970; Kahneman and Tversky, 1979; Golec and Tamarkin, 1998; Garrett and Sobel, 1999; Forrest, Simmons and Chesters, 2002). Also the famous Allais paradoxes are at the heart a manifestation of the tendency to avoid negatively skewed risks. The fact that people seek right-skewed and avoid left-skewed risks is often referred to as *skewness preferences*.

A compelling explanation for skewness preferences is still missing. Since Bernoulli (1738) expected utility theory (EUT) has been the predominant model of choice under risk. Given the assumption of a concave utility function,[1] however, EUT cannot explain

[1]This assumption is necessary to explain why people avoid symmetric mean-preserving spreads over

why people behave risk-averse in some and risk-seeking in other situations. As Kahneman and Tversky (1979) have pointed out, "Choices among risky prospects exhibit several pervasive effects that are inconsistent with the basic tenets of [expected] utility theory." In response, cumulative prospect theory (CPT; Tversky and Kahneman, 1992) has proposed a non-linear probability weighting, which allows us to rationalize that people dislike some, but not all risks. Since a CPT agent overweights the probabilities of extreme events by assumption, she exhibits a preference for right- and an aversion toward left-skewed risks. Indeed, the prevalence of people who reveal a preference for positive skewness was a main reason for why CPT became the gold standard for descriptive modeling of choice under risk. Since CPT assumes that the value of an option is context-independent, however, it implies that only the *absolute* skewness of a lottery matters, but not how skewed a lottery appears to be *relative* to alternative options. We will provide experimental evidence that documents the importance of relative skewness for choice under risk, suggesting that neither EUT nor CPT offer a satisfactory explanation for skewness preferences.

As an illustration of relative skewness, let us consider the problem of whether to invest in some right-skewed asset or not. Depending on its correlation with the market (that is typically left-skewed), such a right-skewed asset appears to be more or less skewed relative to the market portfolio. If the right-skewed asset is, for instance, negatively correlated with the market, then it yields extreme, above-average returns exactly when the market performs badly, which happens to be the case only with a rather small probability. As a consequence, the asset's distribution of excess returns (relative to the market) is right-skewed, which we interpret as the asset being skewed relative to the market. If the asset is, however, positively correlated with the market, then its distribution of excess returns can be left-skewed. In this case we would argue that the market is skewed relative to this asset, which suggests that a right-skewed asset is not necessarily also skewed in relative terms. Based on this idea we propose a notion of *relative* skewness, and we experimentally show that subjects decide on the grounds of the options' *relative* rather than *absolute* skewness. Referring to the preceding example this means that people have a stronger preference for the right-skewed asset the more negatively correlated it is with the market, and in our experiment we can rule out that this correlation-dependence of preferences is driven by reasons of diversification. As in CPT an asset's value only depends on its marginal distribution, CPT cannot account for the role of relative skewness in choice under risk.

Salience theory of choice under risk (Bordalo *et al.*, 2012) provides an intuitive account for why people like skewness, both in absolute and relative terms. Here, attention is automatically directed toward states of the world (i.e., in binary choices, toward pairs of outcomes) that stand out in the choice context while less salient states tend to be neglected. Similar to CPT, salience theory incorporates non-linear probability weighting,

positive outcomes.

but the distortion of a state's probability weight is determined by the contrast in the corresponding outcomes. Probabilities of outstanding states are inflated, while probabilities of less salient states are underweighted. In a typical lottery game, for instance, the large jackpot differs by much from the rather low price of the lottery ticket, thereby attracting a great deal of attention. As a consequence, the agent overweights the probability of winning the salient jackpot and behaves as if she was risk-seeking. In contrast, an agent typically demands insurance against unlikely, but potentially large losses. Compared to the rather small insurance premium the large loss stands out, its probability is inflated, and the agent behaves as if she was risk-averse. In summary, the salience mechanism yields both a preference for right- and an aversion toward left-skewed risks. Moreover, since in salience theory the value of an option depends on the joint distribution of all available alternatives, salience can account for the role of relative skewness in choice under risk.

Our contribution in this paper is threefold. First, we show that the salience model predicts the behavior that the literature has referred to as skewness preferences; that is, whether a salient thinker opts for a risky instead of a safe option depends in a systematic way on the skewness of the risk at hand. To be more precise, we study a salient thinker's preferences over the skewness of binary lotteries buildung on the methodology developed in Ebert (2015), and we find that a salient thinker is more likely to choose a binary risk over its expected value if it is ceteris paribus (i.e., for a given expected value and a given variance) skewed further to the right. By varying a lottery's skewness independent of its expected value and variance, we are able to identify a salient thinker's preferences over the skewness of a risk, which extends the analysis of salience-dependent risk attitudes provided in Bordalo *et al.* (2012). We also single out the channel—namely, the *contrast effect*—through which salience theory predicts skewness preferences. The contrast effect means that when comparing a risky and a safe option, a certain outcome of the risky option receives the more attention the more it differs from the safe option's payoff.

Second, when controlling for the variance of the available lotteries, we show that not only a lottery's absolute skewness is important, but it also matters how skewed this lottery is relative to the alternative option. To formalize this fact, we propose a novel measure of *relative skewness*: we say that a lottery L_x is skewed relative to a lottery L_y if and only if the difference in outcomes, $L_x - L_y$, is right-skewed. This measure of relative skeweness is closely aligned to the idea of the contrast effect. By the contrast effect, a salient thinker's behavior depends on the difference in attainable outcomes in the different states of the world, and it can thereby vary with changes in a lottery's relative skewness. In particular, since the correlation of the two lotteries determines the set of feasible states of the world and thereby also the lotteries' relative skewness, the lotteries' correlation can affect a salient thinker's behavior even if it does not convey any relevant information. More specifically, we delineate under which conditions a salient thinker prefers a right-over a left-skewed binary lottery with the same expected value and the *same* variance, and we formulate novel, experimentally testable predictions that are based on our measure of relative skewness. As we argue in Appendix A, the concept of relative skewness is not only useful in the context of our experiment, but also helps us to better understand the seemingly unrelated Allais paradoxes. As it turns out, not only the common-consequence Allais paradox, but also prominent versions of the common-ratio Allais paradox (see, e.g., Kahneman and Tversky, 1979; O'Donoghue and Sprenger, 2018) can be explained by a preference for relative skewness.

Third, we conducted two laboratory experiments in order to test for our predictions. These experiments allow us to precisely disentangle the salience-based explanation for skewness preferences from alternative approaches such as CPT. As predicted by salience, we find that the more skewed the risky option is, the more likely it is that subjects will choose a risky option over the safe option that pays its expected value. Also in line with the salience model, this preference for positive skewness becomes stronger for lotteries with a larger expected value. In a second experiment we control for the options' variance, but vary both the lotteries' absolute skewness and their correlation structure, which enables us to identify the effect of relative skewness on choice under risk. Our experimental results confirm that relative skewness indeed plays an important role, which is consistent with salience, but not with CPT. Hence, given the empirical relevance of the lotteries' relative skewness, we argue that skewness preferences typically attributed to CPT are more naturally accommodated by the salience model.

Skewness preferences are not only relevant for insurance and gambling decisions, but they also have important implications for other economic and financial decision situations. On asset markets, for instance, investors pay a premium for positive skewness (Boyer, Mitton and Vorkink, 2010; Bali, Cakici and Whitelaw, 2011; Conrad, Dittmar and Ghysels, 2013). This may help us to understand the well-known growth puzzle (Fama and French, 1992) according to which value stocks, that are underpriced relative to financial indicators, yield higher average returns than (overpriced) growth stocks. Bordalo, Gennaioli and Shleifer (2013a) suggest that this discrepancy arises as value stocks are typically left-skewed while growth stocks are usually right-skewed. Along these lines, Barberis (2013) argues that firms conducting an initial public offering (IPO) have a lower average return on their stocks than comparable firms that did not conduct an IPO, because stocks for which an IPO is conducted typically yield a right-skewed distribution of returns and are thus overpriced. Accordingly, Green and Hwang (2012) find that the more skewed the distribution of expected returns is, the lower the long-term average return of an IPO stock is. Chen, Hong and Stein (2001) even argue that managers strategically disclose information to create positive skewness in the distribution of stock returns. Skewness preferences also play an important role for portfolio selection (Chunhachinda, Dandapani, Hamid and

Prakash, 1997; Prakash, Chang and Pactwa, 2003; Mitton and Vorkink, 2007), and allow us to understand the prevalent use of technical analysis for asset trades, even though it is futile in light of the efficient market hypothesis (Ebert and Hilpert, 2019). Finally, a preference for positive skewness also matters in labor economics. Hartog and Vijverberg (2007) and Berkhout, Hartog and Webbink (2010) argue that workers accept a lower expected wage if the distribution of wages in a cluster (i.e., education-occupation combination) is right-skewed. Grove, Jetter and Papps (2018) investigate factors that induce junior tennis players to pursue the risky career of a professional tennis player. Using longitudinal data they show that junior tennis players are attracted to the profession by highly right-skewed earnings distributions. In line with this evidence, Choi, Lou and Mukherjee (2016) observe that the number of college students majoring in a certain field is the higher the more right-skewed the distribution of stock returns of potential employers is.

We proceed as follows. In Section 2.2, we present the salience model and derive first results. Section 2.3 introduces measures of absolute and relative skewness, which we use in Section 2.4 to delineate our main theoretical results on skewness preferences. We present two experiments on salience and skewness preferences in Section 2.5, and provide the experimental results in Section 2.6. Section 2.7 discusses alternative accounts for skewness preferences. Section 2.8 concludes. All proofs are relegated to the Appendix.

2.2 A Continuous Version of Salience Theory of Choice under Risk

2.2.1 Model

Suppose the choice set $\mathcal{C}$ contains exactly two real-valued lotteries, L_x and L_y. The corresponding space of states of the world $S \subseteq \mathbb{R}^2$ is determined by the joint distribution of these lotteries. We denote the joint cumulative distribution function as F. If some lottery is degenerate, we call it a *safe* option. With slight abuse of notation, we refer to $L_z \in \{L_x, L_y\}$ both as a lottery and the corresponding random variable.

According to salience theory of choice under risk (Bordalo *et al.*, 2012, henceforth: BGS), a decision-maker evaluates a lottery by assigning a subjective probability to each state of the world $s \in S$ that depends on the state's objective probability and on its *salience*. The salience of a state $s \in S$ is determined by a so-called *salience function* that is defined as follows.

Definition 1 (Salience Function)**.** *A symmetric, bounded, continuous, and almost everywhere continuously differentiable function* $\sigma : \mathbb{R}^2 \to \mathbb{R}_+$ *is a* salience function *if and only if it satisfies the following three properties:*

1. Ordering. *Let* $\mu = \mathrm{sgn}(x - y)$*. Then for any* $\epsilon, \epsilon' \geq 0$ *with* $\epsilon + \epsilon' > 0$*,*

$$\sigma(x + \mu\,\epsilon, y - \mu\,\epsilon') > \sigma(x, y).$$

2. Diminishing sensitivity. *Let* $x, y \geq 0$*. Then for any* $\epsilon > 0$*,*

$$\sigma(x + \epsilon, y + \epsilon) < \sigma(x, y).$$

3. Reflection. *For any* $x, y, x', y' \geq 0$*, we have*

$$\sigma(x, y) < \sigma(x', y') \quad \textit{if and only if} \quad \sigma(-x, -y) < \sigma(-x', -y').$$

We say that a given state of the world $(x, y) \in S$ is the *more salient* the larger its salience value $\sigma(x, y)$ is. The ordering property implies that a state is the more salient the more the lotteries' payoffs in this state differ. Thus, ordering captures the well-known *contrast effect* (e.g., Schkade and Kahneman, 1998), whereby decision makers focus their attention on those states of the world where the attainable outcomes differ a lot.[2] Diminishing sensitivity reflects *Weber's law* of perception and it implies that the salience of a state decreases if the outcomes in this state uniformly increase in absolute terms. Hence, diminishing sensitivity can be described as a *level effect* according to which a given contrast in outcomes is more salient for lower outcome levels. Finally, by the reflection property, diminishing sensitivity (with respect to zero) reflects to the negative domain. Instead of the rather abstract terminology of ordering and diminishing sensitivity, we mainly use the more intuitive notions of contrast and level effects.

Since the relative importance of the contrast and the level effect may vary with the payoff level, we need to impose more structure on the salience function to derive certain comparative statics with respect to a lottery's expected value (namely, Proposition 4).

Definition 2 (Decreasing Level Effect)**.** *Suppose that* $x, y, z \in \mathbb{R}$ *with* $x + y, x + z \geq 0$*. For a given salience function* σ*, denote*

$$\varepsilon_\sigma(x, y, z) := -\frac{\frac{d}{dx}\sigma(x + y, x + z)}{\sigma(x + y, x + z)}.$$

The salience function σ *satisfies a* decreasing level effect *if and only if both* $\varepsilon_\sigma(x, y, z)$ *and* $\varepsilon_\sigma(-x, -y, -z)$ *strictly decrease in* y *and* z*.*

The decreasing level effect states that the decrease in salience due to a uniform increase in payoffs (which is due to the level effect) is smaller for states with larger outcomes in

[2] In Appendix B.1 we provide a novel, equivalent definition of the ordering property that is based on the partial derivatives of the salience function (see Lemma 4).

absolute terms. In general, the decreasing level effect implies that the contrast effect is the more important relative to the level effect the larger the payoff level is. The property of a decreasing level effect is a stronger notion of what BGS define as *convexity*,[3] and it turns out that the commonly used salience functions satisfy also this stricter property.[4]

We follow BGS in assuming that the salient thinker evaluates monetary outcomes via a linear value function, $u(x) = x$. Using the concept of a salience function, we can define the *salience-weighted utility* of a lottery L_x that is evaluated in the choice set $\mathcal{C}$ as follows.

Definition 3 (Salience-Weighted Utility). *The* salience-weighted utility *of a lottery L_x evaluated in the choice set $\mathcal{C} = \{L_x, L_y\}$ is given by*

$$U^s(L_x|\mathcal{C}) = \int_{\mathbb{R}^2} x \cdot \frac{\sigma(x,y)}{\int_{\mathbb{R}^2} \sigma(s,t)\ dF(s,t)}\ dF(x,y),$$

where $\sigma : \mathbb{R}^2 \to \mathbb{R}_+$ is a salience function that is bounded away from zero.

This gives a more general version of the continuous salience model proposed by BGS. Note that the denominator of the integrand normalizes salience-weighted probabilities so that they sum up to one. Thus, we obtain $U^s(c|\mathcal{C}) = c$ for any safe option $c \in \mathbb{R}$ and any choice set $\mathcal{C}$. In words, the normalization factor ensures that a salient thinker's valuation for any safe option is undistorted, irrespective of the composition of the choice set.

Example 1 (Binary Lotteries). *Suppose that $L_x = (x_1, p; x_2, 1-p)$ and $L_y = (y_1, q; y_2, 1-q)$. Here, the lottery L_x realizes x_1 with probability p and x_2 with probability $1-p$, while the lottery L_y realizes y_1 with probability q and y_2 with probability $1-q$. Depending on the correlation structure of the lotteries, the state space comprises two states (under perfect correlation) to four states (under independence). Denote the probability that the state $s_{ij} := (x_i, y_j)$ is realized by π_{ij}. The salience-weighted utility of lottery L_x is given by*

$$U^s(L_x|\mathcal{C}) = \sum_{s_{ij} \in S} x_i \cdot \frac{\pi_{ij}\ \sigma(x_i, y_j)}{\sum_{s_{kl} \in S} \pi_{kl}\ \sigma(x_k, y_l)},$$

and if the lottery L_y is a safe option, that is, if $L_y = (y, 1)$, the preceding formula further simplifies to

$$U^s(L_x|\mathcal{C}) = \frac{px_1\sigma(x_1,y) + (1-p)x_2\sigma(x_2,y)}{p\sigma(x_1,y) + (1-p)\sigma(x_2,y)}.$$

[3] To arrive at a definition equivalent to BGS's notion of convexity we have to restrict Definition 2 to the case of $y = z$. While convexity is *not* a sufficient condition for Lemma 1 in BGS to hold, the decreasing level effect is.

[4] In Appendix C, we verify that this property holds for basically all salience functions used in the literature.

2.2.2 Certainty Equivalents and the Role of First-Order Stochastic Dominance

To meaningfully discuss risk attitudes under salience theory, we first verify that a salient thinker's certainty equivalent, and thus also her risk premium, to a lottery are well-defined.

Definition 4 (Certainty Equivalent)**.** *Consider any two lotteries L and L' with finite expected values.*

(a) Suppose that an agent faces some choice set $\mathcal{C} = \{L, c\}$ where c represents a safe option. We say that the safe option c is a certainty equivalent *to the lottery L if and only if $U^s(L|\mathcal{C}) = c$.*

(b) Let c be the unique certainty equivalent to L and c' be the unique certainty equivalent to L'. Let L' first-order stochastically dominate L. The certainty equivalent is monotonic *if and only if $c' > c$.*

Proposition 1. *A salient thinker's certainty equivalent to any lottery L is unique and monotonic.*

As an illustration consider a binary lottery $L = (x_1, p; x_2, 1-p)$ with outcomes $x_2 > x_1$. Since the lottery's salience-weighted utility is a convex combination of its payoffs, at least one certainty equivalent to L exists and any certainty equivalent lies between x_1 and x_2. Now suppose that the agent chooses between the lottery L and a safe option $c \in [x_1, x_2]$, and denote as $\phi(x_i, c) := \sigma(x_i, c)(x_i - c)$ the salience-weighted difference between one of the lottery's outcomes and the safe option. We conclude that any certainty equivalent c to lottery L has to solve $p\phi(x_1, c) + (1-p)\phi(x_2, c) = 0$. In addition, any increase in the safe option's payoff does not only make the safe option more attractive, but also makes the lottery less attractive: the lottery's lower payoff becomes more salient while its higher payoff becomes less salient. Consequently, for any fixed lottery $L = (x_1, p; x_2, 1-p)$, the expected salience-weighted difference, $p\phi(x_1, c)+(1-p)\phi(x_2, c)$, is a monotonic function of c, which proves uniqueness of the certainty equivalent. Monotonicity follows immediately from the fact that $\partial\phi(x_i, c)/\partial x_i > 0$ almost everywhere, which implies that, for a fixed safe option c, first-order stochastic dominance with respect to outcomes x_i translates into first-order stochastic dominance with respect to salience-weighted differences $\phi(x_i, c)$.[5]

Given that the certainty equivalent is well-defined, we can define a salient thinker's *risk premium* r for a lottery L as the difference in the lottery's expected value $\mathbb{E}[L]$ and its certainty equivalent $c(L)$: $r(L) := \mathbb{E}[L] - c(L)$. According to Proposition 1, a salient

[5]Notice that in the rank-based salience model which BGS have analyzed, the certainty equivalent is not monotonic and may not even exist (Kontek, 2016). The fact that the certainty equivalent is not monotonic in the rank-based model highlights that Proposition 1 is not a trivial result.

thinker's risk premium for a lottery L is well-defined and monotonic with respect to first-order stochastic dominance. Indeed, we can show that, more generally, a salient thinker's preferences over any pair of independent lotteries are monotonic with respect to first-order stochastic dominance.

Proposition 2. *Let $\mathcal{C} = \{L_x, L_y\}$ and suppose that the lotteries L_x and L_y are stochastically independent. Then, if the lottery L_x first-order stochastically dominates the lottery L_y, we have $U^s(L_x|\mathcal{C}) > U^s(L_y|\mathcal{C})$.*

2.3 Measures of Absolute and Relative Skewness

In this section, we introduce concepts to measure a lottery's absolute and relative skewness. We thereby build on Ebert (2015) who has shown that for binary lotteries, various popular measures of absolute skewness coincide so that absolute skewness is unambiguously defined in this case. To assess relative skewness, we introduce a novel measure that proves useful in our analysis of skewness preferences under salience theory. We also discuss the relationship between absolute and relative skewness. Whenever we simply speak of skewness, we refer to absolute skewness.

2.3.1 Absolute Skewness is Well-Defined for Binary Risks

Often the absolute skewness of a lottery L is associated with its third standardized central moment

$$S(L) := \mathbb{E}\left[\left(\frac{L - \mathbb{E}[L]}{\sqrt{Var(L)}}\right)^3\right]. \tag{2.1}$$

Other notions of absolute skewness refer to "long and lean" tails of the risk's probability distribution. In general, these different notions of absolute skewness are not equivalent. For instance, Ebert (2013) delineates an example of a distribution that has a third moment of zero (which is usually interpreted as the distribution being symmetric), but that is clearly left-skewed when judged by its tails. That said, the analysis of skewness effects on choice under risk is not unambiguous for lotteries with a general distribution.

Importantly, Ebert (2015) shows that for binary risks all conventional notions of skewness are equivalent (see Proposition 2 in his paper), so that the skewness of a binary risk is well-defined.

Definition 5 (Absolute Skewness). *Consider binary lotteries L_x and L_y. We say that the lottery L_x is* more skewed *than the lottery L_y if and only if $S(L_x) > S(L_y)$. A lottery*

L_x is called right-skewed *(or, equivalently, positively skewed) if $S > 0$,* left-skewed *(or, equivalently, negatively skewed) if $S < 0$, and* symmetric *otherwise.*

The main difficulty in identifying the effect of skewness on risk attitudes is that in general variance and skewness are not independent. As an illustration, let us consider a stylized horse race bet $L = (1/p, p; 0, 1-p)$ for some $p \in (0,1)$. If we fix the lottery's lower payoff to zero, then any increase in the probability p implies (i) an increase in the lottery's skewness and (ii) a decrease in the lottery's variance. More generally, as long as we keep a lottery's outcome(s) fixed, any shift in choice that is attributed to a change in skewness can be likewise attributed to a change in variance. Indeed most empirical studies on skewness effects do not properly disentangle preferences over variance from preferences over skewness. Ebert (2015) argues, for instance, that inferring skewness preferences at the horse track from the study by Golec and Tamarkin (1998) might be misleading: increasing the skewness of a horse race bet, while holding its expected value and variance constant, does not yield a new horse race bet, but a lottery with very different properties. Conversely, Ebert concludes that "a choice between two horse-race bets is never a choice between different levels of skewness only."

In order to disentangle preferences for variance and skewness, we need to vary a lottery's skewness for a *fixed* expected value and a *fixed* variance. For that, we rely on the characterization of binary lotteries in terms of their first three moments as provided in Ebert (2015).

Lemma 1 (Ebert's Moment Characterization of Binary Risks)**.** *For any constants $E \in \mathbb{R}$, $V \in \mathbb{R}_+$ and $S \in \mathbb{R}$, there exists exactly one binary lottery $L = (x_1, p; x_2, 1-p)$ with $x_2 > x_1$ such that $\mathbb{E}[L] = E$, $Var(L) = V$ and $S(L) = S$. Its parameters are given by*

$$x_1 = E - \sqrt{\frac{V(1-p)}{p}}, \; x_2 = E + \sqrt{\frac{Vp}{1-p}}, \quad \text{and } p = \frac{1}{2} + \frac{S}{2\sqrt{4+S^2}}. \tag{2.2}$$

We denote the binary lottery with expected value E, variance V, and skewness S as $L(E, V, S)$.

Besides allowing us to investigate how skewness affects choice under risk, the preceding lemma also speaks to the potential richness of predictions based on binary lotteries. An immediate implication of Ebert's Moment Characterization is that any probability distribution can be approximated by a binary lottery up to its first three moments. Thus, when restricting our analysis to binary risks we still allow for three-moment approximations of arbitrary risks. Studying preferences over binary lotteries is also of intrinsic interest since many insurance and gambling applications can be modeled as a binary lottery. Finally, a good understanding of choices among binary lotteries proves useful in other domains such as in the analysis of dynamic gambling behavior (e.g., Ebert and Strack, 2015, 2018; Dertwinkel-Kalt, Frey and Köster, 2020a).

2.3.2 Relative Skewness and Mao Pairs

One way to define the relative skewness of two lotteries L_x and L_y is via the third standardized central moment of their difference $\Delta_{xy} := L_x - L_y$.

Definition 6 (Relative Skewness)**.** *We say that the* lottery L_x is skewed relative to the lottery L_y *if and only if* $S(\Delta_{xy}) > 0$ *holds.*

As we verify in Appendix A.1, the relative skewness of binary lotteries L_x and L_y—when defined this way—depends on (i) the difference in third moments (or *skewness*), $S(L_x) - S(L_y)$, and (ii) the difference in third cross-moments (or *coskewness*), $Cos(L_y, L_x) - Cos(L_x, L_y)$. Since the coskewness of two lotteries is determined by the lotteries' joint distribution, our measure of relative skewness varies with the state space. Thereby, relative skewness captures how skewed the distribution of lottery L_x appears to be in comparison to the distribution of lottery L_y.

At this point it might be useful to highlight some important properties of the proposed measure of relative skewness, which suggest that it is not an arbitrary choice. First, if we compare a lottery to a safe option, then our measure of relative skewness boils down to the lottery's absolute skewness. Hence, a lottery is skewed relative to a safe option if and only if it is right-skewed, which allows us to use the same skewness measure to study (1) the choice between a binary lottery and a safe option and (2) the choice between two binary lotteries. Second, when comparing two lotteries with equal variance, whether some lottery is skewed relative to the other is fully determined by the differences in third moments (i.e., skewness) and third cross-moments (i.e., coskewness). In particular, if these lotteries are stochastically independent, then a lottery L_x is skewed relative to a lottery L_y if and only if it is more skewed in absolute terms.

Next, we illustrate the concept of relative skewness using a class of binary lotteries that was introduced by Ebert and Wiesen (2011, Definition 2). This class of binary lotteries contains the lotteries used in Mao's (1970) seminal survey on skewness preferences of company managers.

Definition 7 (Mao Pair)**.** *Let* $S \in \mathbb{R}_+$*. The lotteries* $L(E, V, S)$ *and* $L(E, V, -S)$ *denote a* Mao pair*.*

The lotteries of a so-called *Mao pair* have the same expected value and the same variance, but differ in their direction of skewness, which makes these lotteries ideal for eliciting preferences over the skewness of a risk. Since the lotteries of a Mao pair only differ in the direction of skewness, the difference between the left-skewed lottery's lower outcome and the right-skewed lottery's lower outcome is the same as the difference between the higher outcomes of the two lotteries. As a consequence, depending on the correlation structure, the difference in the outcomes of the lotteries of a Mao pair takes at most three different values and often only two, which will prove useful in illustrating the lotteries' relative skewness (see Figure 2.2).

As an example of a Mao pair consider the lotteries depicted in Figure 2.1. Both lotteries have an expected value of $E = 108$ and a variance of $V = 1296$. The lottery L_x pays €120 Euro with 90% probability and €0 with 10% probability; it is left-skewed. The lottery L_y pays €96 with 90% probability and €216 with 10% probability; it is right-skewed.

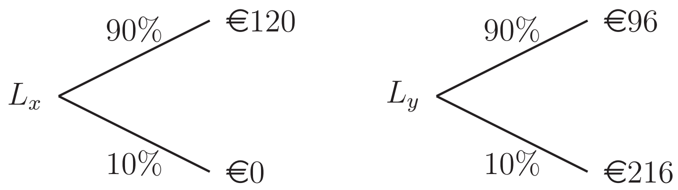


Figure 2.1: *An example of a Mao pair.*

If L_x and L_y are perfectly negatively correlated, only the states $(120, 96)$ and $(0, 216)$ can occur. If the lotteries are independent, all four payoff combinations $(120, 96)$, $(0, 216)$, $(120, 216)$, and $(0, 96)$ can occur. If L_x and L_y are as positively correlated as possible only the three states $(120, 216)$, $(0, 96)$ and $(120, 96)$ are realized with a positive probability.

In general, a Mao pair's joint distribution can be parameterized by one parameter $\eta \in [0, 1]$. Let us denote the outcomes of the left-skewed lottery $L(E, V, -S)$ as $x_1 = x_1(E, V, -S)$ and $x_2 = x_2(E, V, -S)$, respectively, and the outcomes of the right-skewed lottery $L(E, V, S)$ as $y_1 = y_1(E, V, S)$ and $y_2 = y_2(E, V, S)$, respectively. In addition, let $p = p(-S) \in (0, 1/2)$ be the probability of the left-skewed lottery's lower payoff, which is, by construction, identical to the probability of the right-skewed lottery's higher payoff. Table 2.1 depicts the joint distribution of a Mao pair where the parameter $\eta \in [0, 1]$ pins down the correlation structure. The correlation of the two lotteries monotonically increases in η, with $\eta = 0$ corresponding to the perfectly negative correlation and $\eta = 1$ corresponding to the maximal positive correlation.[6]

Since the lotteries of a Mao pair are uniquely characterized by their first three moments and the correlation structure is identified by η, we denote a Mao pair by $M(E, V, S, \eta)$. In addition, let $\Delta(V, S, \eta) := L(E, V, -S) - L(E, V, S)$ denote the difference in outcomes for the lotteries of a Mao pair, which is by Lemma 1 independent of the expected value.

Now we turn to the relative skewness of Mao pairs and illustrate how it depends on the absolute skewness and the correlation of the lotteries. Figure 2.2 depicts the

[6] Lemma 3 in Appendix A.1 summarizes this and further properties of Mao pairs.

Table 2.1: *Joint distribution of the lotteries of a Mao pair.*

Probability	ηp	$p - \eta p$	$1 - p - \eta p$	ηp
$L(E, V, -S)$	x_1	x_1	x_2	x_2
$L(E, V, S)$	y_1	y_2	y_1	y_2

distribution of $\Delta = \Delta(V, S, \eta)$, for the Mao pair introduced above (top row) and a Mao pair with the same variance but a lower absolute skewness (bottom row), both under the perfectly negative correlation (left column) and the maximal positive correlation (right column). For starters, consider the Mao pair with a lower absolute skewness depicted in the bottom row. We observe that the distribution of the difference in outcomes is right-skewed under the perfectly negative correlation (bottom left), but left-skewed under the maximal positive correlation (bottom right). Thus, by Definition 6, the left-skewed lottery is skewed relative to the right-skewed lottery under the maximal positive correlation while the opposite is true under the perfectly negative correlation. For the Mao pair with a higher absolute skewness depicted in the top row, however, the right-skewed lottery is skewed relative to the left-skewed one, irrespective of the correlation structure. More formally, as we show in Lemma 3 in the Appendix, the left-skewed lottery is skewed relative to the right-skewed lottery if and only if their correlation is sufficiently positive—that is, if and only if $\eta > (2/3)(1 + S/\sqrt{4 + S^2})$—, which can only be the case for Mao pairs that are sufficiently symmetric in absolute terms, that is, only for Mao pairs with $S < (2/3)\sqrt{3} \approx 1.15$.

2.4 Salience and the Role of Skewness in Choice under Risk

In this section, we first show that whether a salient thinker appears to be risk seeking or risk averse depends on the skewness of the risk at hand. Put differently, for a fixed expected value and a fixed variance, a binary lottery is chosen over its expected value if and only if it is sufficiently skewed. Subsequently, we delineate that a salient thinker's choice among two binary lotteries with the same expected value and the same variance is determined by their relative skewness.

2.4.1 Skewness-Dependent Risk Attitudes

Suppose that an agent chooses between the binary lottery $L(E, V, S)$ and the safe option that pays the lottery's expected value; that is, the choice set is $\mathcal{C} = \{L(E, V, S), E\}$. The

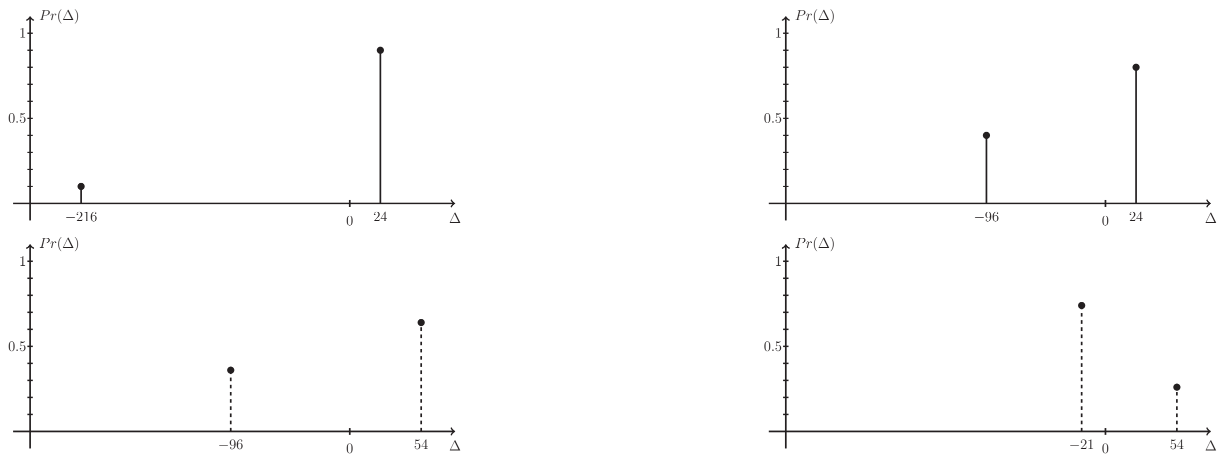


Figure 2.2: *The figures in the top row depict the probability mass functions of $\Delta(V, S, \eta)$, for any Mao pair with $V = 1296$ and $S = 2.7$, under the perfectly negative correlation ($\eta = 0$, top left) and the maximal positive correlation ($\eta = 1$, top right). As both $\Delta(1296, 2.7, 0)$ and $\Delta(1296, 2.7, 1)$ are left-skewed, the lottery $L(E, 1296, 2.7)$ is also skewed relative to the lottery $L(E, 1296, -2.7)$ for both $\eta = 0$ and $\eta = 1$. The figures in the bottom row depict the probability mass functions of $\Delta(V, S, \eta)$, for any Mao pair with $V = 1296$ and $S = 0.6$, under the perfectly negative correlation ($\eta = 0$, bottom left) and the maximal positive correlation ($\eta = 1$, bottom right). As $\Delta(1296, 0.6, 0)$ is left-skewed, but $\Delta(1296, 0.6, 1)$ is right-skewed, for $\eta = 0$ the lottery $L(E, 1296, 0.6)$ is skewed relative to the lottery $L(E, 1296, -0.6)$, while for $\eta = 1$ the lottery $L(E, 1296, -0.6)$ is skewed relative to the lottery $L(E, 1296, 0.6)$.*

salient thinker's risk premium for the lottery $L(E, V, S)$ is given by

$$r(E, V, S) = \sqrt{\frac{V}{4 + S^2}} \cdot \left(\frac{\sigma(x_1, E) - \sigma(x_2, E)}{p\sigma(x_1, E) + (1 - p)\sigma(x_2, E)} \right), \tag{2.3}$$

where the outcomes $x_k = x_k(E, V, S)$, $k \in \{1, 2\}$, and the probability $p = p(S)$ are defined in Eq. (3.1). To break ties we assume that the agent chooses the lottery if and only if the risk premium is strictly negative. Then, an immediate implication of Eq. (2.3) is that a salient thinker chooses the lottery $L(E, V, S)$ over the safe option E if and only if the lottery's higher payoff is salient, which turns out to be the case if and only if the lottery is sufficiently skewed.

Proposition 3. *For any expected value E and any variance V, there exists a unique skewness value $\hat{S} = \hat{S}(E, V) \in \mathbb{R}$ such that $r(E, V, \hat{S}) = 0$. The salient thinker strictly prefers the binary lottery $L(E, V, S)$ over its expected value E if and only if $S > \hat{S}$.*

By Eq. (3.1), an increase in the lottery's skewness is equivalent to an increase in both of the lottery's payoffs and in the probability that the lower payoff is realized. Since the lottery's expected value is fixed, the difference between the lower payoff and the expected value decreases in the lottery's skewness, while the difference between the expected value and the higher payoff increases in the lottery's skewness. Thus, by the contrast effect, the salience of the lottery's lower (higher) payoff monotonically decreases (increases) in the lottery's skewness. We conclude that a salient thinker is the more likely to take up a binary risk the more skewed this risk is. The statement then follows from the fact that the lottery's lower (higher) payoff converges to the expected value as the lottery's skewness approaches (minus) infinity.

A useful corollary to Proposition 3 is the *fourfold pattern of risk attitudes* (Tversky and Kahneman, 1992): people are risk-averse (risk-seeking) over gambles with non-negative payoffs and a likely (unlikely) upside, and risk-seeking (risk-averse) over gambles with non-positive payoffs and a likely (unlikely) downside. Put differently, people avoid symmetric lotteries with non-negative payoffs, but seek symmetric risks with non-positive payoffs.

Corollary 1 (Fourfold Pattern of Risk Attitudes)**.** *Let $\hat{S} \in \mathbb{R}$ be the cutoff value derived in Proposition 3.*

(a) For any binary lottery with non-negative payoffs it holds that $\hat{S} > 0$.

(b) For any binary lottery with non-positive payoffs it holds that $\hat{S} < 0$.

Consider a symmetric binary lottery. Due to symmetry, Eq. (3.1) gives $p = 1/2$, which implies that the difference between the low payoff and the expected value equals the difference between the high payoff and the expected value. If the lottery's payoffs are non-negative, diminishing sensitivity implies that the lower payoff is salient. Hence, a salient thinker avoids any symmetric binary risk with non-negative payoffs. Likewise, diminishing sensitivity implies that the higher payoff is salient if the lottery's payoffs are non-positive, so that a salient thinker seeks symmetric binary risks with non-positive payoffs. In any case, the salient thinker chooses a binary lottery over its expected value if it is sufficiently skewed. A lottery with non-negative payoffs has to be right-skewed in order to be selected, while a lottery with non-positive payoffs can be attractive even if it is left-skewed.[7]

A second straightforward implication of Eq. (2.3) is that the risk premium converges to zero as the lottery's skewness becomes arbitrarily large in absolute terms. This result is driven by the fact that the salience function is bounded from above and also bounded away from zero.

Corollary 2. *For any expected value E and any variance V, $\lim_{S\to\pm\infty} r(E,V,S) = 0$. In particular, for any binary lottery with a fixed expected value E, the certainty equivalent*

$$c_E : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R},\ \ (V,S) \mapsto c_E(V,S) := c(L(E,V,S))$$

is bounded.

[7] BGS derive a similar result for the rank-based salience model.

We conclude from Corollary 2 that a salient thinker's certainty equivalent to the lottery $L(E, V, S)$—which corresponds to her willingness-to-pay for this lottery—is a non-monotonic function of its skewness. In particular, for any fixed expected value, a salient thinker's certainty equivalent $c_E(V, S)$—as a function of the lottery's variance V and skewness S—is bounded from above. While increasing a lottery's skewness boosts the salience of the lottery's higher payoff, it also lowers the probability that this higher payoff will occur. Hence, as salience effects are bounded, increasing a lottery's skewness beyond some threshold decreases a salient thinker's certainty equivalent and therefore her willingness-to-pay for a given lottery.

So far, our results are driven by *either* the contrast *or* the level effect. Next, we study the interaction of both effects which requires us to impose more structure on the salience function. More precisely, we assume that the salience function satisfies the decreasing level effect, which allows us to derive comparative statics with respect to the lottery's expected value.

Proposition 4. *If the salience function satisfies the decreasing level effect, then, for any $\epsilon > 0$ such that $L(E, V, \hat{S}(E, V))$ and $L(E + \epsilon, V, \hat{S}(E + \epsilon, V))$ have either both positive or both negative payoffs, it holds that*

$$\hat{S}(E + \epsilon, V) < \hat{S}(E, V).$$

Under the assumption of the decreasing level effect positive skewness becomes even more desirable at larger payoff levels. Since the contrast effect becomes relatively more important if the attainable outcome(s) increase, the larger contrast on a right-skewed lottery's upside becomes more salient at higher payoff levels. Consequently, a uniform increase in the payoff level—that is, an increase in the lottery's expected value from E to $E+\epsilon$—lowers the minimum level of skewness that renders a lottery with a given variance attractive. This prediction is easily testable in the lab where we can vary the payoff level across choices (see Experiment 1 in Section 2.5).

2.4.2 Not Absolute, But Relative Skewness Shapes Choice Under Risk

Next, we try to isolate the effect of skewness on choice under risk by fixing not only the expected value but also the variance across options. Formally, we study a salient thinker's choice from the set $\mathcal{C} = \{L(E, V, S_x), L(E, V, S_y)\}$ with $S_y > S_x = -S_y$, which makes the two lotteries a Mao pair.[8] In particular, we want to understand the role that relative

[8]While Proposition 5 (a) includes the case of independent Mao pairs, we analyze in Appendix A.2 the choice among two independent binary lotteries in general.

skewness plays in predicting a salient thinker's choice among these two lotteries. For that, recall from Section 2.3 that the parameter $\eta \in [0,1]$ pins down the joint distribution of a Mao pair, and that the right-skewed lottery of this Mao pair is skewed relative to the left-skewed one if and only if $\eta < (2/3)(1 + S/\sqrt{4+S^2})$. The following proposition shows that a salient thinker indeed exhibits a preference for relative rather than absolute skewness, which is reflected in the fact that she chooses the left-skewed lottery if and only if η exceeds a certain threshold (Part (a) of Proposition 5).

Proposition 5. *For any Mao pair, there exist some $\check{\eta}(S) \in (0,1]$ and $\check{S} \in \mathbb{R}_+$ so that the following statements hold:*

(a) A salient thinker prefers $L(E,V,S)$ to $L(E,V,-S)$ if and only if $\eta \leq \check{\eta}(S)$.

(b) For $\eta = 0$, a salient thinker always prefers $L(E,V,S)$ to $L(E,V,-S)$.

(c) For $\eta = 1$, a salient thinker prefers $L(E,V,S)$ to $L(E,V,-S)$ if and only if $S \geq \check{S}$.

If the right-skewed lottery of a given Mao pair is also skewed relative to the left-skewed lottery, which happens to be the case only if η is sufficiently small, then the distribution of $L(E,V,S) - L(E,V,-S)$ is right-skewed *or*, in other words, the contrast in outcomes is largest in those states of the world in which the right-skewed lottery yields the higher payoff. In particular, in the case of the perfectly negative correlation (i.e., $\eta = 0$), the high payoff of the right-skewed lottery always occurs simultaneous to the left-skewed lottery's low payoff, thereby attracting a great deal of attention and rendering the right-skewed lottery attractive to a salient thinker (Part (b) of Proposition 5). As an illustration consider the example of Figure 2.1: under the perfectly negative correlation the two states of the world are $(120, 96)$ and $(0, 216)$, where the latter is more salient according to the contrast effect, so that a salient thinker chooses the right-skewed lottery, thereby revealing a preference for relative skewness. Under a sufficiently positive correlation (i.e., for larger values of η), however, the left-skewed lottery could become skewed relative to the right-skewed lottery. In this case, there would be a larger contrast in those states of the world where the left-skewed lottery outperforms the right-skewed one, which would, due to the contrast effect, render the left-skewed lottery attractive to a salient thinker. But, as we have seen in Section 2.3, only if the lotteries are sufficiently symmetric the left-skewed lottery of a given Mao pair could be skewed relative to the right-skewed lottery (at least under the maximal positive correlation). For the Mao pair presented in Figure 2.1, for instance, the right-skewed lottery is also skewed in relative terms, irrespective of the correlation structure (see the top row of Figure 2.2). This is different for the Mao pair with a lower absolute skewness depicted in the bottom row of Figure 2.2. Here, the lotteries' relative skewness changes as we move from the perfectly negative to the maximal positive correlation. Altogether, the contrast effect implies that even under the maximal positive correlation (i.e., for $\eta = 1$) a salient thinker chooses the left-skewed lottery if and only if the lotteries are sufficiently symmetric (Part (c) of Proposition 5), which again reflects a preference for relative skewness.

Whether relative skewness affects choice under risk is easily testable in a lab experiment, where we can manipulate the relative skewness of a Mao pair via the lotteries' absolute skewness and their correlation (Experiment 2 in Section 2.5). Such an experiment allows us to distinguish our salience-based explanation for skewness preferences from alternative approaches (e.g., CPT) which suggest that only absolute and *not* relative skewness matters (see Section 2.7).

2.5 Two Experiments on Salience and Skewness

Our analytical results give rise to novel predictions that we have tested for in two laboratory experiments. While Experiment 1 tests for skewness-dependent risk attitudes, Experiment 2 investigates the role of relative skewness in choices among the lotteries of a Mao pair.

2.5.1 Design and Predictions

We ensure incentive compatibility by relying only on lotteries that have non-negative payoffs. In order to improve power, we implement both experiments using a within-subjects design. Detailed instructions on the experiments can be found in Appendix D.

Experiment 1. In our first experiment, subjects choose repeatedly between a binary lottery $L(E, V, S)$ and the safe option paying its expected value E, where the lottery's skewness S is gradually increased. We repeat all these choices with the only difference being that all payoffs (the safe payoff E and the lottery's payoffs) are uniformly increased by some amount $\epsilon > 0$.

The choices between E and $L(E, V, S)$ as well as those between $E+\epsilon$ and $L(E+\epsilon, V, S)$ allow us to test for the prediction arising from Proposition 3, saying that for each subject there exists a certain threshold value $\hat{S}$ so that this subject prefers a binary lottery with a fixed variance over its expected value if and only if the lottery's skewness exceeds $\hat{S}$. Since the threshold value $\hat{S}$ depends on the curvature of the salience function, it may vary across subjects. Thus, Proposition 3 implies that the share of subjects choosing the binary lottery $L(E, V, S)$ over its expected value E monotonically increases in the lottery's skewness S.

Prediction 1. *For any E and any V, the share of subjects choosing $L(E, V, S)$ over E increases in S.*

Comparing choices from the set $\{E, L(E, V, S)\}$ with those from $\{E+\epsilon, L(E+\epsilon, V, S)\}$ allows us to test for Proposition 4 whereby skewness preferences become stronger at larger payoff levels. Accordingly, for each subject, the skewness level $\hat{S}$ that makes this subject indifferent between the safe option E and the lottery $L(E, V, S)$ should decrease in E. Given that subjects are heterogeneous with respect to their threshold values $\hat{S}$, we predict that the share of risk takers increases in the lottery's expected value.

Prediction 2. *For any V and any S, the share of subjects choosing $L(E, V, S)$ over E increases in E.*

Experiment 2. In the second experiment, subjects choose repeatedly between two lotteries that form a Mao pair. On the one hand, for a fixed expected value E, variance V, and skewness S, we vary the lotteries' correlation structure as captured by η. On the other hand, for a fixed expected value E, variance V, and correlation structure η, we vary the lotteries' absolute skewness S. Both manipulations affect the relative skewness of the two lotteries. Given that subjects are heterogeneous with respect to their salience functions, Proposition 5 implies that the share of subjects choosing the right-skewed lottery (a) weakly decreases when we move from the perfectly negative correlation to the maximal positive correlation and (b) that this decrease is larger for more symmetric Mao pairs.

Prediction 3. *Consider two Mao pairs $M(E, V, S', \eta)$ and $M(E, V, S'', \eta)$ with $S' < S''$.*

(a) For each of the Mao pairs the share of subjects choosing the right-skewed lottery is weakly larger for $\eta = 0$ (i.e., the perfectly negative correlation) than for $\eta = 1$ (i.e., the maximal positive correlation).

(b) The correlation effect described in (a) is larger for the more symmetric Mao pair $M(E, V, S', \eta)$.

As we have already discussed in the previous section, for sufficiently skewed Mao pairs, the salience model does not predict a shift in choice in response to a change in correlation. In line with the intuition provided by Figure 2.2, we expect to observe a shift from the right-skewed towards the left-skewed lottery when moving from the perfectly negative to the maximal positive correlation only if under the maximal positive correlation the left-skewed lottery is skewed relative to the right-skewed lottery, which is indeed the case if and only if $S < (2/3)\sqrt{3}$.

2.5.2 Implementation

For both experiments students were invited to our laboratory via ORSEE (Greiner, 2015) and both experiments were implemented with z-Tree (Fischbacher, 2007). In each experiment subjects made multiple choices and, in order to warrant incentive compatibility, only one of these decisions was randomly drawn to be payoff-relevant (Azrieli, Chambers and Healy, 2018). For the payoff-relevant decision the outcome of the chosen lottery was determined by a computer simulation.

Experiment 1. Each subject made twelve choices between a lottery and a safe option that paid the lottery's expected value. The order of all decisions was randomized at the subject level. We used experimental currency units (ECU) and a conversion ratio of 2 ECU : 1 Euro. The lotteries that we used in the experiment are listed in Table 2.2. All twelve lotteries have the same variance of $V = 225$. In addition, the lotteries one to six and the lotteries seven to twelve also have the same expected value and differ only with respect to their skewness.

Table 2.2: *Lotteries used in Experiment 1.*

Lottery	Exp. Value	Skewness
(37.5, 80%; 0, 20%)	30	-1.5
(41.25, 64%; 10, 36%)	30	-0.6
(45, 50%; 15, 50%)	30	0
(60, 20%; 22.5, 80%)	30	1.5
(75, 10%; 25, 90%)	30	2.7
(135, 2%; 27.85, 98%)	30	6.9
(57.5, 80%; 20, 20%)	50	-1.5
(61.25, 64%; 30, 36%)	50	-0.6
(65, 50%; 35, 50%)	50	0
(80, 20%; 42.5, 80%)	50	1.5
(95, 10%; 45, 90%)	50	2.7
(155, 2%; 47.85, 98%)	50	6.9

We ran three sessions ($n = 62$) in January 2018 at the DICE experimental laboratory. The experiment lasted, on average, 25 minutes with average earnings of €21. Due to the lotteries' skewness and the corresponding payoff profiles, subjects earned up to €47.

Experiment 2. Each subject made twelve choices between two lotteries that form a Mao pair. Again, we randomized the order of all decisions at the subject level. Experimental earnings were converted at a ratio of 4 ECU : 1 Euro. We used the six Mao pairs listed in Table 2.3, where the absolute skewness is chosen in a way that *only* for the more symmetric Mao pairs (with $S = 0.6$) the left-skewed lottery becomes skewed relative to the right-skewed lottery when moving from negative to positive correlation, but not for the more skewed Mao pairs (with $S = 2.7$). For each Mao pair, we implemented the correlation structures defined by $\eta = 0$ (i.e., perfectly negative correlation) and by $\eta = 1$ (i.e., maximal positive correlation), which gives us in total six paired choices (i.e., two choices for each Mao pair) per subject.

Table 2.3: *Mao pairs used in Experiment 2.*

Left-skewed Lottery	Right-skewed Lottery	Variance	Abs. Skewness	Rel. Skewness	
				$\eta = 0$	$\eta = 1$
(120, 90%; 0, 10%)	(96, 90%; 216, 10%)	1296	± 2.7	-2.7	-1.5
(135, 64%; 60, 36%)	(81, 64%; 156, 36%)	1296	± 0.6	-0.6	1.0
(40, 90%; 0, 10%)	(32, 90%; 72, 10%)	144	± 2.7	-2.7	-1.5
(45, 64%; 20, 36%)	(27, 64%; 52, 36%)	144	± 0.6	-0.6	1.0
(80, 90%; 0, 10%)	(64, 90%; 144, 10%)	576	± 2.7	-2.7	-1.5
(90, 64%; 40, 36%)	(54, 64%; 104, 36%)	576	± 0.6	-0.6	1.0

The initial experiment consisted of three sessions ($n = 79$) that were conducted in February and March 2018 at the DICE experimental laboratory. On the request of a referee, we ran a replication with four sessions ($n = 113$) in November 2018. A power analysis based on the results of the initial experiment implied that a replication study requires a sample size of $n = 110$.[9] We ran the replication also at the DICE experimental laboratory, but excluded subjects that participated already in the initial experiment. We pre-registered the replication study in the AEA RCT Registry with the identifying number AEARCTR-0003545.[10] The experiment lasted 25 minutes on average, subjects earned up to €54, and average earnings were €18.

2.5.3 Related Experimental Literature

To the best of our knowledge none of Predictions 1 to 3 has been tested yet.

Experiment 1. Previous experiments that studied choices between a safe and a risky option changed the variance and skewness of the risky option simultaneously and/or

[9] Using a paired t-test with clustered standard errors (and a cluster size of three symmetric Mao pairs), this sample size allows us to detect an effect size of 13.5 percentage points, which is the effect size observed in our initial study for Mao pairs with $S = 0.6$ (see Table D.2 in Appendix D), at a significance level of 5% with a power of 80%. Based on the initial experiment we assumed an intra-cluster correlation of 0.48.

[10] `https://www.socialscienceregistry.org/trials/3545/history/37113.AEARCTRegistry`.

included more than one lottery into the choice set, both of which make it impossible to test for Predictions 1 and 2. The large experimental literature on probability weighting (e.g., Gonzalez and Wu, 1999; Bruhin, Fehr-Duda and Epper, 2010) as well as the experiments on salience by BGS and Frydman and Mormann (2018) document violations of EUT that might be attributed to a preference for skewness, but do not identify the causal effect of skewness on risk attitudes. Also recent experiments that want to study skewness preferences do not offer tests of Predictions 1 and 2. Brocas, Carrillo, Giga and Zapatero (2016), for instance, try to identify the determinants of skewness-dependent risk attitudes by comparing investments in an asset with normally distributed and thus symmetric returns to investments in a less profitable asset with right-skewed binary returns. Since the return distributions are chosen in a way such that the right-skewed asset also has a lower variance than the symmetric one, however, investments in the right-skewed asset—that Brocas *et al.* (2016) observe for a majority of subjects—could be driven by both a lower variance or a higher skewness. The experiment by Grossman and Eckel (2015) controls for the options' variance, but does also not precisely test for decisions between a safe and a risky option. In their experiment subjects first choose from a menu of six options, one of which is safe. Subsequently, subjects are offered a second set of (non-binary) lotteries that differ only in terms of third moments, that is, for each lottery in the first choice set there is a lottery with the same expected value and the same variance but a larger third moment in the second choice set. Subjects can then switch from their initial choice to any lottery in the second choice set and a majority of subjects do so at least once. Grossman and Eckel (2015) interpret switching towards a lottery with a larger third moment as a skewness-seeking choice. In contrast to the existing literature, our first experiment studies the choice between a safe option and a binary lottery in isolation while properly controlling for variance. This allows us to identify the causal effect of skewness on risk attitudes and provides a test of Predictions 1 and 2.

Experiment 2. Previous experiments that investigated the role of skewness in choices between two risky options used non-binary lotteries and/or did not specify the correlation structure. Building on seminal work by Eeckhoudt and Schlesinger (2006), a series of experiments analyzed whether preferences over risky options satisfy *prudence* (see Trautmann and van de Kuilen, 2018, for a survey). Under EUT prudence is equivalent to a positive third derivative of the utility function and it implies—for a fixed expected value and a fixed variance—a preference for positive skewness. While these studies document that a majority of subjects prefer lotteries with larger third moments, they do neither use binary lotteries (which comes with the caveat that skewness is not well-defined) nor do they manipulate the lotteries' correlation structure. More closely related is the experiment by Ebert (2015) who investigated preferences over binary lotteries with the same expected value and variance, but different levels of skewness, without describing the exact state space. In line with the experiments on prudence, Ebert (2015) finds a preference for right- over left-skewed risks, but he also observes that more absolute skewness is not attractive per se, meaning that subjects do not necessarily choose the more skewed option if both lotteries are right-skewed. As we discuss in Appendix A.2, this finding is consistent with salience theory under the assumption that subjects perceived the lotteries as being independent. Our second experiment directly builds on Ebert and Wiesen (2011) who formally introduced Mao pairs and experimentally investigated choices between these lotteries. We extend their experimental setup by comparing choices under different correlation structures.[11] This allows us to identify the causal effect of relative skewness on choice under risk and provides a test of Prediction 3.

2.6 Experimental Results

2.6.1 Experiment 1: Skewness-dependent Risk Attitudes

In line with Prediction 1, the share of risk-takers strictly increases in the lottery's skewness, S, both for lotteries with an expected value of $E = 30$ (see the hatched bars in Figure 2.3) and for those with an expected value of $E = 50$ (see the dotted bars in Figure 2.3). Moreover, for all skewness levels the share of risk takers strictly increases in the lottery's expected value, E, which is in line with Prediction 2 (compare the hatched and dotted bars in Figure 2.3).

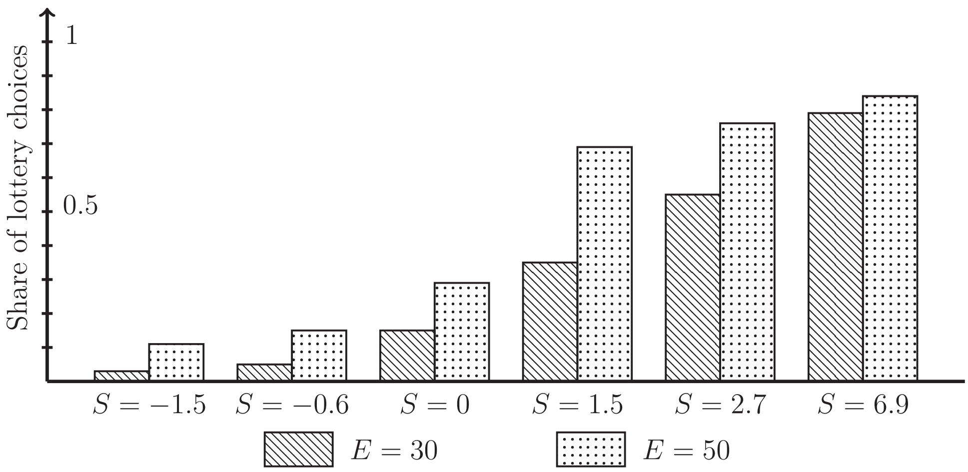


Figure 2.3: *The figure depicts the share of lottery choices depending on the lottery's skewness for a low and a high expected value. The skewness values are presented in ascending order, but not in a proper scale.*

[11]Relatedly, a series of papers tests salience theory against CPT using correlated versions of the Allais paradoxes (e.g., BGS; Frydman and Mormann, 2018; Bruhin, Manai and Santos-Pinto, 2019).

In order to formally test for Predictions 1 and 2, we regress a binary indicator of whether a subject chooses the lottery over its expected value on the lottery's skewness and a dummy indicating whether the lottery's expected value is high or low (see Table 2.4). Since each subject made twelve choices during the experiment, we cluster the standard errors at the subject level. The regressions confirm what Figure 2.3 suggests: the share of risk takers *significantly* increases with both a lottery's skewness and its expected value.

Table 2.4: *Main regressions for Experiment 1.*

Parameter	**(1)**	**(2)**
Constant	0.247***	0.175***
	(0.022)	(0.027)
Skewness	0.097***	0.097***
	(0.008)	(0.008)
High Expected Value	-	0.145***
	-	(0.025)
# Subjects	62	62
# Choices	744	744

Notes: *The table presents the results of OLS regressions of a dummy indicating the choice between the lottery and the safe option (where a value of one indicates the choice of the lottery and a value of zero indicates the choice of the safe option) on the lottery's skewness and a dummy indicating whether its expected value is high or low. All standard errors are clustered at the subject level and provided in parenthesis. Significance level: ***: 1%.*

Returning to Figure 2.3, we further observe that for both payoff levels there is a particularly big shift in the share of risk takers when moving from the symmetric (i.e., $S = 0$) to the least right-skewed (i.e., $S = 1.5$) lottery. This finding is in line with the salience-based explanation of skewness preferences: the contrast on a binary lottery's upside is larger than the contrast on its downside if and only if the lottery is right-skewed, so that—due to the level effect—a binary lottery with non-negative payoffs becomes attractive to a salient thinker only if it is right-skewed. Moreover, we observe that for a low expected value the least right-skewed lottery—$(60, 0.2; 22.5, 0.8)$—is still chosen only by a minority, while the lottery that is least right-skewed in the set of lotteries with a larger expected value—$(80, 0.2; 42.5, 0.8)$—is selected by a large majority. The salience model can also explain this big shift in the share of risk takers: due to the level effect the lower payoff of lottery $(60, 0.2; 22.5, 0.8)$ is slightly more salient than the higher payoff of this lottery although there is a larger contrast on the lottery's upside, while the decreasing level suggests that the upside payoff becomes salient for the lottery $(80, 0.2; 42.5, 0.8)$. A detailed overview of the results is provided in Table D.1 in Appendix D.

Finally, we scrutinize whether the choice patterns predicted by Propositions 3 and 4 also hold at the individual level. We observe that 63% of all subjects (39 out of 62 subjects) have a unique switching point consistent with Proposition 3, both for the set of lotteries with a low and a high expected value, which is a much larger share than we would expect under random choice. Among the remaining 23 subjects there are 18 subjects who have a unique switching point in line with Proposition 3 for exactly one set of lotteries (i.e., either for the lotteries with low or for those with a high expected value), and only one subject has a unique switching point in the opposite direction (i.e., choosing the lottery if and only if it is sufficiently left-skewed) for both sets of lotteries.[12] Moreover, we observe that 89% of those subjects who reveal a unique switching point for both sets of lotteries also chose the lottery weakly more often (and 51% of these subjects chose it strictly more often) in the case of a high expected value, which is consistent with Proposition 4.

In summary, the results of our first experiment confirm the importance of *absolute* skewness for risk attitudes as well as our salience-based explanation of skewness preferences.

2.6.2 Experiment 2: A Preference for Relative Skewness

First, when pooling all choices for the more symmetric Mao pairs (i.e., those with $S = 0.6$), we observe that the right-skewed lottery is chosen much more often under the perfectly negative correlation than under the maximal positive correlation. Second, when pooling all choices for the more skewed Mao pairs (i.e., those with $S = 2.7$), we do not find much of a difference in the frequency with which the right-skewed lottery is chosen under the perfectly negative and the maximal positive correlation, respectively. Both findings are consistent with Prediction 3 and, as illustrated in Figure 2.4, do not only hold for our initial study, but are successfully replicated in a properly powered replication. Table D.2 in Appendix D provides a detailed overview of the results. Moreover, using the combined data for the initial study and the replication,[13] we show in Appendix D that the results

[12] A more formal treatment of skewness effects at the individual level is provided in Figure D.5 in Appendix D, where we plot the point estimates for the coefficient on the lottery's skewness in individual-level versions of Regression Model (1) in Table 2.4. While in line with Proposition 3 the majority of point estimates is positive, the confidence intervals have to be interpreted with caution due to the small number of observations per subject. The results remain basically the same when using Regression Model (2) from Table 2.4.

[13] Since we determined the necessary sample size for the replication in order to detect the same effect size for the symmetric Mao pairs as observed in our initial study under the assumption that each subject makes three paired choices, we do not have enough power to test our hypothesis for each symmetric Mao

are robust across the different Mao pairs (see Figure D.6).

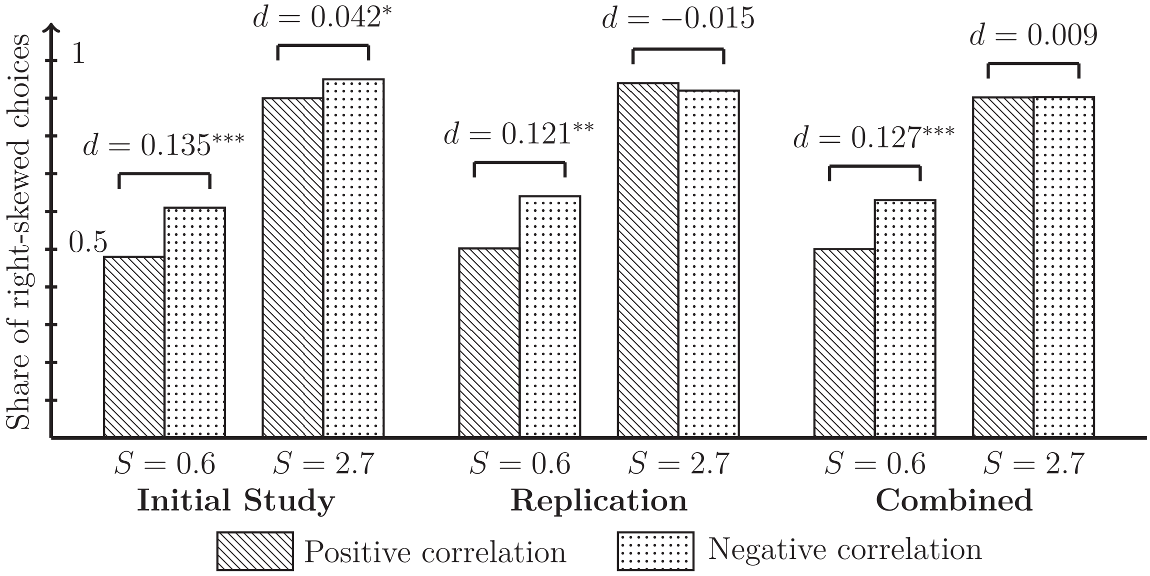


Figure 2.4: *The figure illustrates the share of choices of the right-skewed lottery under positive and negative correlation. We present results separately for the initial study and the replication as well as the combined results for both studies. We further report results of paired t-tests with standard errors being clustered at the subject level (see Table 2.5 for the full regression). Significance level: *: 10%, **: 5%, ***: 1%.*

Based on our power calculation—for which we used a paired t-test with standard errors being clustered at the subject level—we construct the dependent variable for a linear regression model as follows. First, we create for each choice (i.e., twelve per subject) a dummy variable that takes a value of one if the right-skewed lottery is chosen and a value of zero otherwise. Second, using these dummies, we construct for each paired-choice (i.e., six per subject) a variable that indicates the *shift in choice* due to a change in correlation and can take three values: a value of *one* if a subject switches from the right-skewed lottery under the perfectly negative correlation to the left-skewed lottery under the maximal positive correlation, a value of *minus one* if a subject switches in the opposite direction, and a value of *zero* if a subject does not change her choice. Notice that the mean shift in choice corresponds to the difference in the shares of subjects switching in either direction. Then, Prediction 3 (a) implies that the average shift in choice is positive (i.e., there are more subjects who switch from the right-skewed lottery under the perfectly negative correlation to the left-skewed lottery under the maximal positive correlation than subjects who switch in the opposite direction), and Prediction 3 (b) suggests that the average shift in choice is strictly larger for the more symmetric Mao pairs (with $S = 0.6$).

Now in order to formally test for Predictions 3 (a) and (b), we regress the shift in

pair separately when using only the data from the replication.

choice as constructed above on a constant—which indicates the mean shift in choice for symmetric Mao pairs and therefore tests for Prediction 3 (a)—and a dummy *Skewed* indicating whether a given Mao pair is more symmetric (i.e., $Skewed = 0$ if $S = 0.6$) or more skewed (i.e., $Skewed = 1$ $S = 2.7$)—which yields the difference in means between symmetric and skewed Mao pairs and therefore tests for Prediction 3 (b). This regression (as presented in Table 2.5) computes paired t-tests of the average shift in choice being different from zero, and it also allows us to account for the fact that each subject makes multiple choices by clustering the standard errors at the subject level. In line with Prediction 3 (a), we observe that for the symmetric Mao pairs the share of choices of the right-skewed lottery *significantly* decreases by around 12 to 13 percentage points when moving from the perfectly negative to the maximal positive correlation (i.e., the constant takes a value between 0.121 and 0.135, and it is always significantly different from zero). And, in line with Prediction 3 (b), we also find that the shift in choice is *significantly* smaller for the more skewed Mao pairs (i.e., the coefficient of the dummy takes a value between -0.093 and -0.136, and it is always (weakly) significantly different from zero). As illustrated in Figure 2.4, for the more skewed Mao pairs there is at most a weakly significant shift in choice, that vanishes once we take all the data into account.

Table 2.5: Main regressions for Experiment 2.

Parameter	**Initial Study**	**Replication**	**Combined**
Constant	0.135***	0.121**	0.127***
	(0.046)	(0.047)	(0.034)
Skewed	-0.093*	-0.136**	-0.118***
	(0.054)	(0.053)	(0.038)
# Subjects	79	113	192
# Paired Choices	474	678	1,152

Notes: *The table presents the results of OLS regressions of the shift in choice (i.e., the difference in choices under the perfectly negative correlation and maximal positive correlation) on a constant (which corresponds to the mean shift in choice for the more symmetric Mao pairs and therefore tests for Part (a) of Prediction 3) and a dummy indicating skewed Mao pairs (which gives the difference in means between the more symmetric and the more skewed Mao pairs and therefore tests for Part (b) of Prediction 3). The first column uses only data from the initial study, the second column uses only data from the replication, and the third column combines both datasets. All standard errors are clustered at the subject level and provided in parenthesis. Significance level: *: 10%, **: 5%, ***: 1%.*

The preceding results provide a test of the qualitative salience predictions derived in this paper, but the data reveals further interesting patterns that are consistent with salience theory and a preference for relative skewness. While for the more skewed Mao pairs the right-skewed lottery is chosen in more than 90% of the cases, both under the perfectly negative and the maximal positive correlation, for the more symmetric Mao pairs even under the perfectly negative correlation the right-skewed lottery is chosen in only around 60% of the cases (see Table D.2 in Appendix D). Since the difference in salience-weighted utility between the right-skewed and the left-skewed lottery under the perfectly negative correlation is much smaller for the more symmetric Mao pairs, this finding is easily reconciled with our salience model given an appropriate assumption on the decision noise. A more intuitive way to grasp this result, however, is in terms of the lotteries' relative skewness. Recall that the relative skewness of two lotteries depends on both their absolute skewness and their correlation structure. As illustrated in Table 2.3, the left-skewed lottery is *less* skewed relative to the right-skewed lottery for the more skewed Mao pairs under the maximal positive correlation than it is for the more symmetric Mao pairs under the perfectly negative correlation. Along these lines, Figure 2.5 shows that the right-skewed lottery is chosen less often the more skewed the left-skewed lottery is in relative terms. Also a linear regression confirms that the average probability of choosing the right-skewed lottery significantly decreases by around 12.6 to 14 percentage points in the relative skewness of the left-skewed lottery (see Table D.3 in Appendix D). Altogether, these results reveal a strong preference for relative skewness.

Finally, using the combined data from the initial study and the replication, we analyze whether the choice patterns predicted by Proposition 5 also hold at the individual level. First, we observe that 60% of all subjects (i.e., 115 out of 192 subjects) either switched from the right-skewed lottery under the perfectly negative correlation toward the left-skewed lottery under the maximal positive correlation or did not change their behavior for *all* six Mao pairs. This behavior is fully in line with Proposition 5 and a share of 60% is much larger than the share predicted by random choice. In addition, we find that 77% of all subjects (i.e., 147 out of 192 subjects) switched weakly more often in the direction predicted by salience theory. In order to study preferences for relative skewness more directly we ran individual-level regressions of the dummy indicating the choice between the right-skewed and the left-skewed lottery of a given Mao pair on the left-skewed lottery's relative skewness and Figure D.7 in Appendix D depicts the point estimates. While in line with Proposition 5 the majority of point estimates is negative, the confidence intervals have to be interpreted with caution due to the small number of observations per subject.

In summary, the results of our second experiment confirm the importance of *relative* skewness for choice under risk and support our salience-based explanation of skewness preferences. Moreover, if some subjects noticed that each decision problem occurred

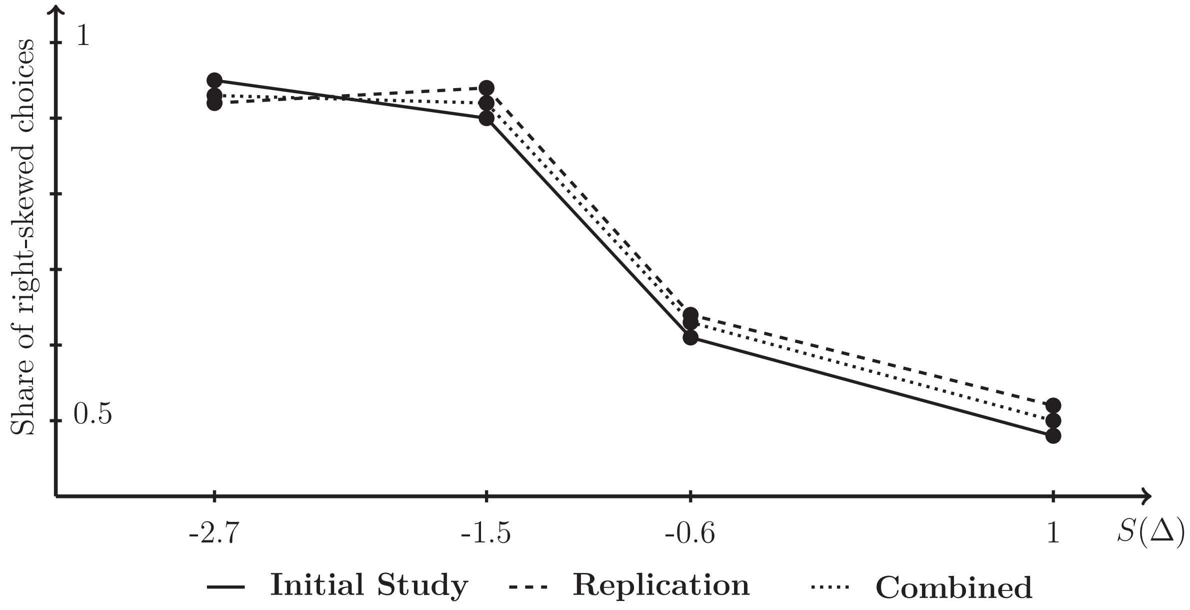


Figure 2.5: *The figure illustrates the share of choices of the right-skewed lottery as a function of the left-skewed lottery's relative skewness $S(\Delta)$, where $\Delta = L(E, V, -S) - L(E, V, S)$, separately for the initial study and the replication as well as for the combined data. We observe that the share of choices of the right-skewed lottery declines in the left-skewed lottery's relative skewness.*

twice in the experiment (with the only difference being that payoff-irrelevant correlations are modified) and therefore tried to choose in a consistent manner, we might actually underestimate the effect of relative skewness on choices. In this sense, due to the within-subjects design, our estimates of the average correlation effect on choice might constitute a lower bound.

2.7 Alternative Explanations of Skewness Preferences

2.7.1 Expected Utility Theory

In order to explain the fact that most people are risk-averse with respect to symmetric mean-preserving spreads over positive outcomes, EUT needs to assume that the utility function is strictly concave (Bernoulli, 1738). Under this assumption, however, EUT cannot account for skewness-dependent risk attitudes as elicited in Experiment 1. More specifically, it cannot explain why—depending on the skewness of a risk—people like variance in some, but dislike variance in other decision situations. While EUT can in principle explain why people prefer right-skewed over left-skewed lotteries with the same expected value and variance (e.g., Menezes, Geiss and Tressler, 1980; Ebert, 2015), it cannot account for the correlation effects that we detected in Experiment 2. According to EUT a subject's choice between two lotteries should be independent of the correlation structure. In summary, EUT may explain a preference for absolute skewness when the variance is fixed across options, but it cannot account either for the fact that a lottery's skewness affects risk attitudes or for the role of relative skewness in choice under risk.

2.7.2 Cumulative Prospect Theory

In response to the weaknesses of EUT as a descriptive model of choice under risk, Kahneman and Tversky proposed *prospect theory* (Kahneman and Tversky, 1979) and later developed it into *cumulative prospect theory* (Tversky and Kahneman, 1992). CPT fundamentally deviates from EUT in two directions by assuming that people (1) are loss-averse with respect to some reference point and (2) weight probabilities according to a non-linear function. In order to guarantee that preferences over lotteries satisfy first-order stochastic dominance, CPT assumes that outcomes are ranked before probability weights are computed, which in turn implies that the size of the outcomes affects probability weighting in an ordinal way.[14]

CPT can explain why decision makers seek positively skewed and avoid negatively skewed risks. More specifically, given an S-shaped value function with a sufficiently positive third derivative and an inverse S-shaped weighting function that overweights not only small but also moderate probabilities, CPT can account for both results of Experiment 1. The non-linear probability weighting function gives the effect of a lottery's skewness on risk attitudes while the curvature of the value function yields the effect of the expected value on risk attitudes.

In contrast to the salience model, however, in CPT the value of a given lottery is independent of the choice context in the sense that it only depends on the lottery's marginal probability distribution. As a consequence, CPT cannot explain the role that the correlation structure plays for the choice between two lotteries and, therefore, it cannot account for the results on relative skewness in Experiment 2. In other words, CPT predicts that only the lotteries' absolute skewness and *not* their relative skewness matters for choice under risk.

Finally, there are further theoretical arguments that favor the salience-based over the CPT-based explanation of skewness preferences. Unlike the salience model, CPT inherently produces puzzling predictions on the strength of skewness effects. Rieger and Wang (2006) and Azevedo and Gottlieb (2012) show that, according to CPT, a firm selling sufficiently right-skewed lotteries with a fixed expected value can earn arbitrarily large profit margins in expectation. This result is based on the fact that increasing a lottery's upside payoff and simultaneously reducing the corresponding probability increases a CPT agent's willingness-to-pay for this lottery, namely due to the overweighting of small prob-

[14] This rank-dependence was first proposed by Quiggin (1982).

abilities. This prediction arises for "virtually all functional forms that have been proposed in the literature" (Azevedo and Gottlieb, 2012, p. 1294). Since, by Corollary 2, a salient thinker's certainty equivalent to any lottery with a fixed expected value is bounded, this puzzling prediction does not arise in the salience model.[15] Also unrealistic predictions of CPT on dynamic gambling and investment behavior (Ebert and Strack, 2015, 2018) can be ruled out in the salience model (Dertwinkel-Kalt *et al.*, 2020a).

2.7.3 Regret

Also in regret theory (Bell, 1982; Loomes and Sugden, 1982, 1987; Diecidue and Somasundaram, 2017; Gollier, 2020) context matters for valuations. Given a binary choice set, regret theory assumes that the unchosen alternative directly affects the decision maker's utility via a regret/rejoice term—added to a standard utility function—which captures the regret (rejoice) a decision maker anticipates to feel when the unchosen option yields a higher (lower) payoff in the realized state of the world. The disutility from regret (the utility from rejoice) is defined to be monotonically increasing in the difference in outcomes in a given state, so that regret models share the mathematical properties of the contrast effect. Lanzani (2020) shows that, for binary choices, salience theory is a special case of generalized regret theory by Loomes and Sugden (1987), which implies that based on choice data alone salience and regret are indistinguishable in our experiment. The psychology behind regret theory is, however, very different from the psychology behind salience theory. To affect behavior, regret has to be anticipated, which necessitates that the subject expects to receive information on the counterfactual outcome, as summarized in a review by Zeelenberg (1999, p.103): "When [post-decisional] feedback is present people anticipate possible regret, but when it is absent regret does not play a significant role in the decision process." Thus, additional information on the choice process might allow us to disentangle salience and regret theory also for binary choices.

Along these lines, even though we cannot formally exclude it, we do not regard regret theory as a plausible driver of our experimental results. Consider, for instance, Experiment 1 where subjects had to choose between a safe and a risky option. Here, conditional on choosing the safe option, subjects should not expect to learn the counterfactual outcome. In fact, if a subject chooses the safe option, the lottery will not even be played out, so that there is *no* counterfactual outcome. When choosing the lottery, in contrast, the subject knows the counterfactual outcome; namely, the value of the safe option.[16] As argued by Zeelenberg (1999), regret-averse subjects should therefore choose the safe

[15] It is straightforward to show that the certainty equivalent to a lottery with an expected value E is bounded by a function that is affine in E.

[16] In the instructions we write: "If you have chosen [the safe option] in this task you will receive the according sum. If you have chosen [the lottery] your payoff will be determined through the simulation of the turn of a wheel of fortune. Your payoff will be paid in cash at the end of the experiment."

option in order to avoid any feelings of regret. In this sense, models based solely on regret aversion (e.g. Strack and Viefers, forthcoming) cannot plausibly account for our results. Regret models that also include rejoice can, in principle, explain why subjects reveal skewness preferences in our first experiment. But despite the massive literature on regret, we are not aware of any psychological evidence substantiating rejoice.

In other experiments where salience effects have been documented, observing choices within a given context suffices to distinguish between salience and regret theory, because more than two options are presented and/or the consequences of choosing a certain option are deterministic. For instance, regret theory cannot account for various effects that decoys have. State-wise dominated decoys yield the well-established effect of asymmetric dominance (as introduced by Huber, Payne and Puto, 1982), which salience theory, but not regret theory can account for (see Dertwinkel-Kalt and Köster, 2017). Also the effect of non-available decoys, so-called phantom decoys (e.g., Soltani, De Martino and Camerer, 2012) can be explained by salience, but not by regret (see Frydman and Mormann, 2018). Moreover, regret theory makes predictions only for choices involving risk, that is, regret cannot give an alternative explanation for salience effects in risk-free domains as delineated in Bordalo *et al.* (2013b) and experimentally supported by Dertwinkel-Kalt, Lange, Köhler and Wenzel (2017).

2.7.4 Context-Dependent Attention Models

We are not the first to draw the connection between salience and skewness preferences. Bordalo *et al.* (2013a) have already pointed out that the salience model can explain why individuals like right-skewed and dislike left-skewed assets, but they have not disentangled a salient thinker's preferences for variance and for skewness, respectively.

Not only the salience model, but also other approaches that model context-sensitive behavior and build on the contrast effect can account for many of our findings. For instance, our results on skewness effects carry over to the closely related *focusing model* (Kőszegi and Szeidl, 2013) if it is applied to choice under risk. In this case, a focusing function—i.e., the pendant to the salience function satisfying only the contrast and not the level effect—determines which states an agent's attention is directed to. As the level effect can be included in the value function (Kőszegi and Szeidl, 2013, Section III.D), the focusing model can be closely aligned to the salience model in which case it shares salience theory's predictions for choice under risk. Namely, this version of the focusing model is mathematically equivalent to the generalized regret theory (Loomes and Sugden, 1987).

Since our results are driven by the contrast effect, the model of relative thinking by Bushong, Rabin and Schwartzstein (forthcoming), which builds on the setup by Kőszegi and Szeidl (2013), but assumes a reverse contrast effect (i.e., the attention a state attracts decreases in the range of attainable payoffs in this state), cannot account for skewness preferences.

2.7.5 Optimal Expectations

According to the model on optimal expectations proposed by Brunnermeier and Parker (2005), an agent receives utility not only from her actions, but also from her beliefs over the likelihood of favorable future outcomes. Therefore, an agent intentionally inflates the "perceived likelihood" of upside events in order to enhance the pleasure from expecting these events. As a consequence, a model of optimal expectations predicts an excessive demand for right-skewed lotteries. But this model yields weaker predictions on skewness preferences than our salience-based approach (see, e.g., Proposition 2 in Brunnermeier and Parker, 2005). First, Brunnermeier and Parker explain a preference for sufficiently right-skewed risks, but they do not obtain precise predictions on the demand for less skewed or left-skewed assets. Second, utility from pleasant expectations can be obtained only *before* an event is realized. Thus, it seems plausible that optimal expectations matter only when there is a considerable amount of time between an investment decision and the event realization. Our salience-based approach instead explains skewness preferences irrespective of whether the realization of outcomes is delayed or not. More importantly, also a model of optimal expectations implies that only the absolute skewness of a lottery matters, but not how skewed it is relative to alternative options. Thus, Brunnermeier and Parker (2005) cannot account for the results of our second experiment.

2.8 Conclusion

Preferences over the skewness of the underlying probability distribution are a robust observation not only in humans, but also in animals (Strait and Hayden, 2013; Genest, Stauffer and Schultz, 2016). Choices on, for example, gambling, insurance, asset, and labor markets are crucially affected by skewness preferences. As a consequence, it is important to understand the mechanism driving skewness effects. In this paper, we have identified the contrast effect as a plausible driver of skewness preferences. Accordingly, when comparing a risky and a safe option, an outcome of the risky option attracts the more attention the more it differs from the safe option's payoff. Thereby, the contrast effect induces a focus on the large, but unlikely upside of right-skewed risks, and a focus on the large potential loss in the case of left-skewed risks. Alongside our theoretical results, we offer a novel set of experimental predictions that we have found support for in two laboratory experiments.

Our two experiments show that (1) when choosing between a risky and safe option subjects exhibit a preference for positive skewness that becomes more pronounced as the payoff level increases and that (2) the choice between two lotteries crucially depends on their correlation structure, even though this correlation is irrelevant for subjects' earnings. As Bordalo *et al.* (2013a) have already pointed out, the results of our first experiment might allow us to better understand, for instance, the countercyclical relationship between the aggregated stock market returns and the current economic situation. Proposition 4 suggests that we should observe more risk-seeking behavior at higher payoff levels, as for higher payoffs a salient thinker focuses even more on the large upside of a right-skewed risk and less on its downside. We confirm this prediction in our first experiment, which can also explain why in the aggregate stocks are often overvalued in boom times, but undervalued in the times of a bust (e.g., Campbell and Shiller, 1988; Guiso, Sapienza and Zingales, 2018). In our second experiment we manipulated the correlation structure of the available lotteries, as this allows us to test salience theory against alternative models of skewness preferences where valuations are context-independent, such as cumulative prospect theory. Since in reality the correlation structure often carries important information and therefore *should* affect behavior, however, it is harder to come up with practical applications of our second experiment. Interestingly, while we show that making the correlation structure explicit affects behavior even if it should not, a recently growing literature has detected that in (rather complex) choice situations where correlation indeed matters subjects tend to neglect it, thereby forming incorrect conditional probabilities when making inferences (e.g., Levy and Razin, 2015; Enke and Zimmermann, 2019). One conclusion that might be drawn from our experimental results is that helping people to overcome *correlation neglect* and to thereby learn the objective probabilities of the states of the world does not necessarily improve people's decisions as salience effects can distort choices even (more) if objective probabilities are known.

A potential limitation of our study is that we restrict our analysis of skewness preferences to the class of binary risks. But despite the fact that skewness is not well-defined for lotteries with a general distribution, the basic insights derived in this paper should carry over to a broader class of distribution functions. Consider for instance a symmetric distribution with continuous and bounded support on the positive real numbers (e.g., a truncated normal distribution). When choosing between this symmetric risk and a safe option paying its expected value, a salient thinker goes for the safe option due to the level effect. Now extend the support of the distribution's right tail and shift some probability mass there, which skews the distribution to the right. When compared to the risk's expected value, outcomes in the right tail attract a salient thinker's attention (due to the contrast effect) and render the risk attractive. Conversely, skewing such a symmetric distribution to the left by extending the support of its left tail makes the risk less attractive to the salient thinker, due to the outcomes in the left tail attracting an overproportionate amount of attention. While the intuition why salience can explain skewness preferences, thus, appears to be quite robust, general results as those derived in this paper cannot be expected to hold, since for continuous distributions the salience predictions hinge on the precise definition of skewness and on the exact curvature of the salience function over the entire realm.

To sum up, our paper adds to the theoretical (Bordalo *et al.*, 2012, 2013a) and experimental literature on salience effects in choice under risk (e.g., Dertwinkel-Kalt and Köster, 2017; Frydman and Mormann, 2018). Using both theoretical arguments and experimental data, we argue that skewness preferences typically attributed to cumulative prospect theory are more naturally accommodated by salience theory.

Appendix A: Relative Skewness and Choice under Risk

A.1: Relative Skewness of Binary Lotteries and Mao Pairs

In this subsection, we derive some properties of our measure of relative skewness. Before we can state these properties, however, we have to introduce the concept of *coskewness*.

Definition 8 (Coskewness)**.** *The* coskewness of a lottery L_x relative to a lottery L_y *is given by*

$$Cos(L_x, L_y) := \frac{\mathbb{E}[(L_x - \mathbb{E}[L_x])(L_y - \mathbb{E}[L_y])^2]}{\sqrt{Var(L_x)}Var(L_y)}.$$

The coskewness of two lotteries refers to their third (non-trivial) cross-moments and it is often used in the asset-pricing literature. Using the concept of coskewness, we can make the following statements on the relative skewness of two binary lotteries L_x and L_y. Recall that $\Delta_{xy} = L_x - L_y$.

Lemma 2. *Consider any two binary lotteries $L_x = L(E_x, V_x, S_x)$ and $L_y = L(E_y, V_y, S_y)$. Then , it holds that*

$$S(\Delta_{xy}) = \frac{S_x\sqrt{V_x}^3 - S_y\sqrt{V_y}^3 + 3Cos(L_x, L_y)\sqrt{V_x}V_y - 3Cos(L_y, L_x)V_x\sqrt{V_y}}{\sqrt{V_{xy}}^3},$$

where V_{xy} gives the variance of Δ_{xy}. In addition, the following statements hold:

(a) If both lotteries have the same variance, then we obtain

$$S(\Delta_{xy}) = \left(\frac{V_x}{V_{xy}}\right)^{3/2} \Big[S_x - S_y + 3Cos(L_x, L_y) - 3Cos(L_y, L_x)\Big].$$

(b) *If the two lotteries are stochastically independent, then we obtain*

$$S(\Delta_{xy}) = \frac{S_x\sqrt{V_x}^3 - S_y\sqrt{V_y}^3}{\sqrt{V_x + V_y}^3},$$

which simplifies to $S(\Delta_{xy}) = (S_x - S_y)/(2\sqrt{2})$ *in case of equal variances.*

(c) *If lottery* L_y *is degenerate, then we obtain* $S(\Delta_{xy}) = S_x$.

Proof. Straightforward calculations yield

$$\begin{aligned}
\mathbb{E}\Big[(\Delta_{xy} - \mathbb{E}[\Delta_{xy}])^3\Big] &= \mathbb{E}\Big[\Big([L_x - E_x] - [L_y - E_y]\Big)^3\Big] \\
&= \mathbb{E}\Big[(L_x - E_x)^3\Big] - 3\mathbb{E}\Big[(L_x - E_x)^2(L_y - E_y)\Big] \\
&\qquad + 3\mathbb{E}\Big[(L_x - E_x)(L_y - E_y)^2\Big] - \mathbb{E}\Big[(L_y - E_y)^3\Big],
\end{aligned}$$

which gives the above formula for $S(\Delta_{xy})$. Now Parts (a) and (c) directly follow, so we only need to prove Part (b). Suppose that L_x and L_y are stochastically independent. Then, we obtain

$$\begin{aligned}
\mathbb{E}\Big[(L_x - E_x)(L_y - E_y)^2\Big] &= \mathbb{E}\Big[L_x L_y^2\Big] - E_x\mathbb{E}\Big[L_y^2\Big] \\
&= \mathbb{E}\Big[L_x\Big]\mathbb{E}\Big[L_y^2\Big] - E_x\mathbb{E}\Big[L_y^2\Big] \\
&= 0,
\end{aligned}$$

where the first equality holds by $Cov(L_x, L_y) = 0$ and the second equality holds by $Cov(L_x, L_y^2) = 0$. □

Next, we restrict attention to those binary lotteries that form a Mao pair (see Definition 7). The following lemma summarizes the properties of Mao pairs that we have already discussed in the main text and extends the list by some further useful properties. In particular, the left-skewed lottery of a Mao pair is skewed relative to the right-skewed lottery only if the lotteries are positively correlated and not too skewed in absolute terms.

Lemma 3. *For any Mao pair* $M(E, V, S, \eta)$ *the following statements hold:*

(a) *The covariance of* $L(E, V, -S)$ *and* $L(E, V, S)$ *strictly increases in* η.

(b) $Cov\Big(L(E, V, -S), L(E, V, S)\Big) > 0$ *if and only if* $\eta > \Big(1 + S/\sqrt{4 + S^2}\Big)/2$.

(c) *The coskewness of* $L(E, V, -S)$ *relative to* $L(E, V, S)$ *strictly increases in* η.

(d) $Cos\Big(L(E, V, S), L(E, V, -S)\Big) = -Cos\Big(L(E, V, -S), L(E, V, S)\Big)$.

(e) *Cos*$\Big(L(E,V,-S), L(E,V,S)\Big) > 0$ *if and only if* $\eta > \Big(1 + S/\sqrt{4+S^2}\Big)/2$.

(f) *The third standardized central moment of* $\Delta = L(E,V,-S) - L(E,V,S)$ *strictly increases in* η.

(g) *The lottery* $L(E,V,-S)$ *is skewed relative to the lottery* $L(E,V,S)$ *if and only if* $\eta > (2/3)\Big(1 + S/\sqrt{4+S^2}\Big)$. *In particular, for* $\eta = 1$, $L(E,V,-S)$ *is skewed relative to* $L(E,V,S)$ *if and only if* $S < (2/3)\sqrt{3} \approx 1.15$.

Proof. The proof is straightforward and therefore omitted. □

A.2: Independent Binary Lotteries

Suppose the choice set is given by $\mathcal{C} = \{L(E,V,S_x), L(E,V,S_y)\}$ with $S_y > S_x$. In addition, let the lotteries be independent, in which case the relative skewness of the two lotteries is fully determined by the difference in their absolute skewness (see Lemma 2 in Appendix A.1). In this sense, the choice between two independent binary lotteries with the same expected value and the same variance is structurally similar to the choice between a binary lottery and a safe option, as also in the latter case the relative skewness is given by the difference in absolute skewness, that is, simply by the skewness of the binary lottery (see Lemma 2).

The first result in this subsection relates a salient thinker's choice between two independent binary lotteries with the same expected value and the same variance to her choice between a binary lottery and the safe option paying its expected value. Proposition 6 shows, in particular, that a salient thinker does not necessarily choose the lottery that is more skewed in absolute terms if either both lotteries are right-skewed or both are left-skewed. This prediction is in line with experimental evidence by Ebert (2015).

Proposition 6. *Let* $\hat{S} \in \mathbb{R}$ *be the cutoff value derived in Proposition 3. Then, if the two binary lotteries are independent, the following statements hold:*

(a) $\lim_{S_y \to \infty} \Big[U^s\Big(L(E,V,S_y)|\mathcal{C}\Big) - U^s\Big(L(E,V,S_x)|\mathcal{C}\Big)\Big] > 0 \iff S_x < \hat{S}$.

(b) $\lim_{S_x \to -\infty} \Big[U^s\Big(L(E,V,S_y)|\mathcal{C}\Big) - U^s\Big(L(E,V,S_x)|\mathcal{C}\Big)\Big] > 0 \iff S_y > \hat{S}$.

Proof. Denote as π_{ij} the probability of state $s_{ij} = (x_i, p_j)$. First, we define

$$\lambda := \frac{\sqrt{p_x(1-p_x)p_y(1-p_y)}}{\sum_{s_{ij} \in S} \pi_{ij}\sigma(x_i, y_j)}.$$

In addition, let $p_z = p_z(S_z)$ be the marginal probability with which the lower payoff of the lottery $L_z := L(E,V,S_z)$ is realized. For the sake of brevity, we further write

$U^s(L_z) = U^s(L_z|\mathcal{C})$. Then, by Eq. (3.1), we obtain

$$\frac{U^s(L_x) - E}{\lambda\sqrt{V}} = \sqrt{\frac{p_y}{1-p_y}}\Big(\sigma(x_2, y_1) - \sigma(x_1,y_1)\Big) + \sqrt{\frac{1-p_y}{p_y}}\Big(\sigma(x_2, y_2) - \sigma(x_1, y_2)\Big)$$

as well as

$$\frac{U^s(L_y) - E}{\lambda\sqrt{V}} = \sqrt{\frac{p_x}{1-p_x}}\Big(\sigma(x_1, y_2) - \sigma(x_1,y_1)\Big) + \sqrt{\frac{1-p_x}{p_x}}\Big(\sigma(x_2, y_2) - \sigma(x_2, y_1)\Big).$$

Let sgn : $\mathbb{R} \to \{-1, 0, 1\}$ be the signum-function. Taking the difference of the above expressions yields

$$\begin{aligned}\operatorname{sgn}\Big(U^s(L_y) - U^s(L_x)\Big) = \operatorname{sgn}\Bigg(&\Big[\sigma(x_1, y_1) - \sigma(x_2, y_1)\Big]\sqrt{p_x}\sqrt{p_y} \\ &+ \Big[\sigma(x_1, y_2) - \sigma(x_1, y_1)\Big]p_x\sqrt{\frac{1-p_y}{1-p_x}} \\ &+ \Big[\sigma(x_1, y_2) - \sigma(x_2, y_2)\Big](1-p_y)\sqrt{\frac{p_x}{p_y}} \\ &+ \Big[\sigma(x_2, y_2) - \sigma(x_2, y_1)\Big]\sqrt{1-p_x}\sqrt{1-p_y}\Bigg).\end{aligned}$$

Now, as S_y approaches infinity, we obtain

$$\begin{aligned}&\operatorname{sgn}\Bigg(\lim_{S_y\to\infty}\Big[U^s(L_y) - U^s(L_x)\Big]\Bigg) \\ &\qquad = \operatorname{sgn}\Bigg(\sqrt{p_x(1-p_x)}\sigma(x_1, E) - \sqrt{p_x(1-p_x)}\sigma(x_2, E)\Bigg),\end{aligned}$$

which exceeds zero if and only if $S_x < \hat{S}$. This yields part (a). Analogously,

$$\begin{aligned}&\operatorname{sgn}\Bigg(\lim_{S_x\to-\infty}\Big[U^s(L_y) - U^s(L_x)\Big]\Bigg) \\ &\qquad = \operatorname{sgn}\Bigg(\sqrt{p_y(1-p_y)}\sigma(E, y_2) - \sqrt{p_y(1-p_y)}\sigma(E, y_1)\Bigg),\end{aligned}$$

which exceeds zero if and only if $S_y > \hat{S}$. This yields part (b). □

Proposition 6 suggests that under the assumption of independence a salient thinker prefers right- over left-skewed binary lotteries with the same expected value and the same variance. This prediction is consistent with evidence from the lab (Ebert and Wiesen, 2011; Ebert, 2015). More precisely, Proposition 6 (a) states that a salient thinker chooses an extremely right-skewed lottery $L(E, V, S_y)$ over the less skewed alternative $L(E, V, S_x)$ if and only if lottery $L(E, V, S_x)$ is *sufficiently* less skewed in the sense that $S_x < \hat{S}$, where the threshold value $\hat{S}$ is the same as the one derived in Proposition 3. This result follows from the fact that, by Corollary 2, a salient thinker's valuation of $L(E, V, S_y)$ approaches the lottery's expected value as S_y approaches infinity, so that in this limit case the choice between the two lotteries $L(E, V, S_y)$ and $L(E, V, S_x)$ is basically the same as the choice between the safe option E and the lottery $L(E, V, S_x)$. Analogously, Proposition 6 (b) implies that a salient thinker does *not* choose an extremely left-skewed lottery as long as the alternative option is sufficiently more skewed.

In summary, the preceding proposition implies that a lottery's absolute skewness matters, but it also suggests that more absolute skewness is not per se attractive to a salient thinker. By Proposition 6 (a) an extremely right-skewed option is *not* chosen over a less but still sufficiently skewed alternative, while by Proposition 6 (b) an extremely left-skewed lottery can be attractive relative to a *slightly* more skewed alternative. These predictions link back to Corollary 2, according to which the certainty equivalent to a binary lottery with a fixed expected value—when interpreted as a function of the lottery's skewness—is bounded. Notably, Ebert (2015) finds in his lab experiment that many subjects choose the less skewed of two right-skewed binary lotteries with the same expected value and variance. In this sense, Proposition 6 (a) is consistent with the experimental evidence by Ebert (2015) under the assumption that the subjects in his experiment perceived the lotteries as being independent.

Finally, by observing that a salient thinker's behavior is driven by the difference in absolute skewness, we can formulate the above result also in terms of relative skewness: Proposition 6 basically says that there exists some threshold value $\hat{\Delta} \in \mathbb{R}$ such that the salient thinker chooses $L(E, V, S_y)$ over $L(E, V, S_x)$ if its relative skewness, $S(\Delta_{yx})$, exceeds the threshold $\hat{\Delta}$. It is in this sense that Proposition 6 points towards a salient thinker's preference for relative skewness.

If we assume that one of the lotteries is symmetric, we further obtain the following corollary, which relates to the fourfold pattern of risk attitudes.

Corollary 3. *Suppose that the two lotteries are stochastically independent. Then, the following statements hold:*

(a) Let $S_x = 0$ and suppose that lottery $L(E, V, S_x)$ has non-negative payoffs. Then, there exists some $S' \in \mathbb{R}$ such that for any $S_y > S'$ the salient thinker chooses $L(E, V, S_y)$.

(b) Let $S_y = 0$ and suppose that lottery $L(E, V, S_y)$ has non-positive payoffs. Then,

there exists some $S'' \in \mathbb{R}$ such that for any $S_x < S''$ the salient thinker chooses $L(E, V, S_y)$.

Recall from Corollary 1 (a) that a binary lottery $L(E, V, S_x)$ with non-negative payoffs is chosen over its expected value if and only if $S_x > \hat{S} > 0$ holds. Then, it follows immediately from Proposition 6 (a) that a salient thinker chooses the more skewed alternative $L(E, V, S_y)$ if its absolute skewness exceeds a certain threshold (Part (a) of Corollary 3). Using the same line of argumentation, Corollary 3 (b) follows directly from Corollary 1 (b) and Proposition 6 (b). Again, both parts of Corollary 3 suggest that a salient thinker opts for the lottery that is more skewed in absolute terms if it is also sufficiently skewed in relative terms.

A.3: A Preference for Relative Skewness can Explain the Allais Paradoxes

Not only choices on gambling, insurance, asset, and labor markets are crucially affected by skewness preferences, but also long-standing puzzles in choice under risk such as the *Allais paradoxes* can be attributed to skewness preferences or, more precisely, a preference for relative skewness.

Common-Consequence Allais Paradox. In order to highlight the role that relative skewness plays in predicting whether subjects exhibit the common-consequence Allais paradox, we build on a recent experiment by Frydman and Mormann (2018). Suppose a subject chooses between the lotteries $L^1(z) = (25, 0.33; 0, 0.01; z, 0.66)$ and $L^2(z) = (24, 0.34; z, 0.66)$ where $z \in \{0, 24\}$. The common finding in the literature is that without making the state space explicit a majority of subjects choose $L^2(24)$ over $L^1(24)$, but $L^2(0)$ over $L^1(0)$. This preference reversal constitutes a puzzle from a classical economics point of view as according to EUT the common consequence z should not affect behavior. It can, however, be easily rationalized by CPT or salience theory. In addition, Frydman and Mormann (2018) have demonstrated that the emergence of the Allais paradox crucially depends on the lotteries' correlation structure. Can the idea of relative skewness explain this phenomenon?

In order to answer this question, let us start by considering the case in which the common consequence is given by $z = 24$. In this case, subjects choose between the safe option $L^2(24)$ and the left-skewed lottery $L^1(24) = (25, 0.33; 0, 0.01; 24, 0.66)$. According to the contrast effect, the left-skewed lottery's lowest payoff attracts a great deal of attention, which results in a preference for the safe option. Now suppose that the common consequence is given by $z = 0$. In this case, in order to derive predictions on a salient thinker's behavior, we need to state what the state space is. Adopting the notation by

Frydman and Mormann (2018), we parameterize the lotteries' joint distribution via some $\beta \in [\frac{1}{2}, 1]$. The last row in Table 2.6 depicts the difference in the outcomes of the lotteries, which determines their relative skewness.

Table 2.6: *Joint distribution common-consequence Allais paradox lotteries.*

Probability	$0.33(2\beta - 1)$	$0.67 - 0.66\beta$	0.66β	$0.66(1-\beta)$
$L^1(0)$	25	0	0	25
$L^2(0)$	24	24	0	0
$L^1(0) - L^2(0)$	1	-24	0	25

Frydman and Mormann (2018) verify in their Appendix that (i) the correlation between $L^1(0)$ and $L^2(0)$ strictly increases in $\beta \in [\frac{1}{2}, 1]$ and that (ii) a salient thinker's preference for $L^1(0)$ over $L^2(0)$ is diminished as the lotteries become more positively correlated. Moreover, as illustrated in Figure 2.6, lottery $L^1(0)$ is the more skewed relative to lottery $L^2(0)$ the smaller β is, where we denote $S(\beta) := S\big(L^1(0) - L^2(0)\big)$.[17] In this sense, salience theory again predicts a preference for relative skewness and it suggests that subjects are the more likely to exhibit the Allais paradox the more skewed lottery $L^1(0)$ is relative to lottery $L^2(0)$.

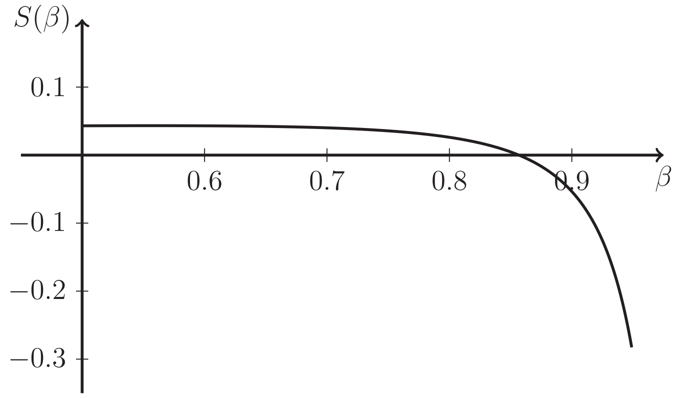


Figure 2.6: *The figure illustrates how skewed $L^1(0)$ is relative to $L^2(0)$ depending on β.*

In their expriments, Frydman and Mormann (2018) vary the lotteries' correlation across three conditions—*no correlation* for $\beta = 0.67$, *intermediate correlation* for $\beta = 0.98$, and *maximal correlation* for $\beta = 1$—and they observe that, in line with salience, the share of subjects exhibiting the Allais paradox strictly decreases in β. Since we obtain $S(0.67) > S(0.98) > S(1)$, these findings can be understood as subjects revealing a preference for relative skewness.

[17] To be precise, $S(\beta)$ monotonically decreases in $\beta \in [0.56, 1]$, with $S(\beta) > 0$ if and only if $\beta < 0.86$.

Common-Ratio Allais Paradox. Finally, we demonstrate that also prominent versions of the common-ratio Allais paradox proposed by Kahneman and Tversky (1979) or O'Donoghue and Sprenger (2018) can be attributed to skewness preferences. We borrow the lotteries introduced in Problems 3 and 4 of Kahneman and Tversky (1979), but divide all payoffs by 1,000 for the sake of comparability with the common-consequence example. Let $L^3(q) = (40, 0.8q; 0, 1 - 0.8q)$ and $L^4(q) = (30, q; 0, 1 - q)$ for $q \in \{0.25, 1\}$. Kahneman and Tversky (1979) observe that around 80% of the subjects choose $L^4(1)$ over $L^3(1)$, which can be explained by an aversion toward left-skewed risks (see Proposition 3), while around 65% of the subjects choose $L^3(0.25)$ over $L^4(0.25)$. This preference reversal contradicts EUT, but is in line with Proposition 6 given that the subjects perceived the lotteries as being independent. In order to say more on the role of relative skewness, Table 2.7 parameterizes the joint distribution for $q = 0.25$ via $\gamma \in [0.6875, 0.9375]$.

Table 2.7: *Joint distribution of common-ratio Allais paradox lotteries.*

Probability	$0.8(1-\gamma) - 0.05$	$0.25 - 0.8(1-\gamma)$	$0.8(1-\gamma)$	0.8γ
$L^3(0.25)$	40	40	0	0
$L^4(0.25)$	0	30	30	0
$L^3(0.25) - L^4(0.25)$	40	10	-30	0

Note that for any salience function the states of the world can be unambiguously ranked according to salience: $\sigma(40, 0) > \sigma(30, 0) > \sigma(40, 30)$, where the first inequality follows by ordering and the second one by ordering and diminishing sensitivity. Then, it is easy to check that either the salient thinker always prefers $L^3(0.25)$ or she always prefers $L^4(0.25)$ or there exists some $\hat{\gamma} \in (0.6875, 0.9375]$ so that she prefers $L^3(0.25)$ over $L^4(0.25)$ if and only if $\gamma < \hat{\gamma}$. In addition, as illustrated in Figure 2.7, lottery $L^3(0.25)$ is the more skewed relative to lottery $L^4(0.25)$ the smaller γ is,[18] where we denote $S(\gamma) := S\Big(L^3(0.25) - L^4(0.25)\Big)$. As before, salience theory suggests that subjects are the more likely to exhibit the Allais paradox the more skewed $L^3(0.25)$ is relative to $L^4(0.25)$. In this sense, also the common-ratio Allais paradox can be understood as a manifestation of skewness preferences.

[18] To be precise, $S(\gamma)$ monotonically decreases in $\gamma \in [0.7257, 0.9375]$, with $S(\gamma) > 0$ if and only if $\gamma < 0.89$.

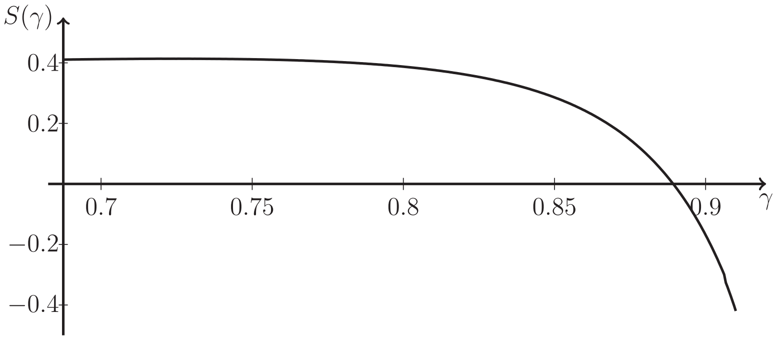


Figure 2.7: *The figure illustrates how skewed $L^3(0.25)$ is relative to $L^4(0.25)$ depending on γ.*

Appendix B: Proofs

B.1: Auxiliary Results

In this subsection, we derive auxiliary results on the salience-weighted utility that we will apply in the proofs of our main results. In a first step, we characterize the ordering property via properties of the partial derivatives of the salience function (Lemma 4). In a second step, we argue that the salience-weighted utility of any lottery with a finite expected value is bounded (Lemma 5). In a third step, we formally introduce the notion of first-order stochastic dominance and we state an implication for positive monotone transformations of lotteries that can be ordered in terms of first-order stochastic dominance (Lemma 6). While Lemmata 5 and 6 are only relevant for the proof of Proposition 1, we will use Lemma 4 also in the proof of Proposition 4.

The first result in this subsection pins down the sign of the partial derivatives of the salience function (whenever these partial derivatives exist). In particular, we show that, on any dense subset of $\mathbb{R}$, the partial derivatives of a salience function are different from zero.

Lemma 4. *Without loss of generality assume $x \geq y$. Consider a symmetric, continuous, and almost everywhere (a.e.) continuously differentiable function $\sigma : \mathbb{R}^2 \to \mathbb{R}_+$. Denote as $N_x \subset \mathbb{R}$ the set on which the partial derivative $\partial\sigma(x,y)/\partial x$ does not exist and as $N_y \subset \mathbb{R}$ the set on which the partial derivative $\partial\sigma(x,y)/\partial y$ does not exist. Then, the following two statements are equivalent:*

i) The function σ satisfies ordering.

ii) $\partial\sigma(x,y)/\partial x \geq 0$ for all $(x,y) \in \{\mathbb{R} \setminus N_x\} \times \mathbb{R}$ and $\partial\sigma(x,y)/\partial y \leq 0$ for all $(x,y) \in \mathbb{R} \times \{\mathbb{R} \setminus N_y\}$. Moreover, $\partial\sigma(x,y)/\partial x > 0 > \partial\sigma(x,y)/\partial y$ on any dense subset of $\mathbb{R}$.

Proof. i) $\Rightarrow$ ii): For any $\epsilon > 0$ ordering implies $\sigma(x+\epsilon, y) - \sigma(x,y) > 0$. Thus, whenever the partial derivative $\partial\sigma(x,y)/\partial x$ exists, then $\partial\sigma(x,y)/\partial x \geq 0$ follows immediately from its definition. By the Mean Value Theorem, there exists some $\xi \in [x, x+\epsilon]$ such that

$$\frac{\partial}{\partial x}\sigma(x,y)\bigg|_{x=\xi} = \frac{\sigma(x+\epsilon,y)-\sigma(x,y)}{\epsilon} > 0.$$

Since N_x has measure zero by assumption, we obtain $\partial\sigma(x,y)/\partial x > 0$ on any dense subset of $\mathbb{R}$. The statement then follows by symmetry.

ii) $\Rightarrow$ i): For any $\epsilon, \epsilon' \geq 0$ with $\epsilon + \epsilon' > 0$ it holds that

$$\begin{aligned}&\sigma(x+\epsilon, y-\epsilon') - \sigma(x,y)\\ &\qquad = \int_{[x,x+\epsilon]\setminus N_x} \frac{\partial}{\partial z}\sigma(z, y-\epsilon')dz - \int_{[y-\epsilon',y]\setminus N_y} \frac{\partial}{\partial z}\sigma(x,z)dz > 0\end{aligned}$$

since N_x and N_y have measure zero and either $[x, x+\epsilon] \setminus N_x$ or $[y-\epsilon', y] \setminus N_y$ or both are non-empty and include dense subsets of $\mathbb{R}$. Hence, σ satisfies ordering, which was to be proven. □

The second result in this subsection replaces the linear value function assumed in the main text by any strictly increasing value function $u(\cdot)$, and it states that the salience-weighted utility of any lottery with a finite expected utility (in particular, those with a finite expected value) is bounded.

Lemma 5. *Fix a choice set $\mathcal{C} := \{L_x, L_y\}$. There exist constants $\underline{K}, \overline{K} \in \mathbb{R}$ so that $U^s(L_z|\mathcal{C}) \in [\underline{K}, \overline{K}]$. In particular, the bounds on the salience-weighted utility of lottery L_z are independent of lottery L_{-z}.*

Proof. Since the expected utility of lottery $L_z \in \mathcal{C}$ is finite, the statement follows from the fact that the salience function is bounded (away from zero). □

Before we state the last result of this subsection, we formally introduce the notion of first-order stochastic dominance: we say that *a lottery L_x first-order stochastically dominates the lottery L_y (i.e., L_x fosd L_y) if and only if $F_x(z) \leq F_y(z)$ for any $z \in \mathbb{R}$ and $F_x(z) < F_y(z)$ for some $z \in \mathbb{R}$.* Lemma 6 states that not only has a first-order stochastic dominant lottery a higher expected value than the corresponding dominated lottery, but that this inequality survives under any positive monotone transformation of the lotteries.

Lemma 6. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a strictly increasing function with $\phi'(\cdot) > 0$ almost everywhere. Then, it holds:*

$$L_x \;\; fosd \;\; L_y \quad \Longrightarrow \quad \int_{\mathbb{R}} \phi(x)\; dF_x > \int_{\mathbb{R}} \phi(y)\; dF_y$$

Proof. The proof is straightforward and therefore omitted. □

B.2: Main Results

Proof of Proposition 1. We prove a slightly more general version of this proposition. Indeed, the statement holds not only for a linear value function, but for any strictly increasing value function $u(\cdot)$ with $u'(\cdot) > 0$ such that the expected utility, $U(L) = \int_{\mathbb{R}} u(x)\, dF(x)$, is finite.

To begin with, notice that a salient thinker with a value function $u(\cdot)$ strictly prefers the lottery L to the safe option c if and only if

$$u(c) < \int_{\mathbb{R}} u(x) \cdot \frac{\sigma(u(x), u(c))}{\int_{\mathbb{R}} \sigma(u(s), u(c))\, dF(s)}\, dF(x) =: h(c).$$

The proof now proceeds in four steps. In a first step, we use the Intermediate Value Theorem and Lemma 5 to argue that $u(c) = h(c)$ has at least one solution, which implies that a certainty equivalent exists. In a second step, we show that for any c that solves $u(c) = h(c)$, there exists some $\epsilon > 0$ such that $u(c') > h(c')$ for any $c' \in (c, c+\epsilon)$ and $u(c'') < h(c'')$ for any $c'' \in (c-\epsilon, c)$. This implies that the certainty equivalent is *locally* unique. In a third step, we take two lotteries L_x and L_y where $L_x\ fosd\ L_y$ and show—by the use of Lemma 6—that the smallest certainty equivalent to the lottery L_x strictly exceeds the largest certainty equivalent to the lottery L_y. In a fourth step, we combine the results from the previous steps to prove that the certainty equivalent is (globally) unique. Monotonicity then follows immediately from the third step.

1. STEP: Since both $u(\cdot)$ and $h(\cdot)$ are continuous, the Intermediate Value Theorem states that at least one solution to $u(c) = h(c)$ exists if it is true that

$$\lim_{c\to-\infty} u(c) < \lim_{c\to-\infty} h(c) \quad \text{and} \quad \lim_{c\to\infty} u(c) > \lim_{c\to\infty} h(c). \tag{2.4}$$

In order to verify that Eq. (2.4) is indeed fulfilled, we distinguish three cases.

First, suppose that $\lim_{c\to-\infty} u(c) = -\infty$ and $\lim_{c\to\infty} u(c) = \infty$. By Lemma 5, there exist constants $\underline{K}, \overline{K} \in \mathbb{R}$ such that $h(c) \in [\underline{K}, \overline{K}]$ for any $c \in \mathbb{R}$. Hence, Eq. (2.4) holds.

Second, let both $\lim_{c\to-\infty} u(c)$ and $\lim_{c\to\infty} u(c)$ be finite. Since $u'(\cdot) > 0$,

$$\begin{aligned}
\lim_{z\to-\infty} u(z) &= \int_{\mathbb{R}} \frac{\sigma(u(x), u(c))}{\int_{\mathbb{R}} \sigma(u(s), u(c))\, dF(s)} \cdot \lim_{z\to-\infty} u(z)\, dF(x) \\
&< \int_{\mathbb{R}} \frac{\sigma(u(x), u(c))}{\int_{\mathbb{R}} \sigma(u(s), u(c))\, dF(s)} \cdot u(x)\, dF(x) \\
&= h(c),
\end{aligned}$$

for any $c \in \mathbb{R}$. This implies, in particular, that $\lim_{c\to-\infty} u(c) < \lim_{c\to-\infty} h(c)$ has to hold. By the same type of argument, we obtain $\lim_{z\to\infty} u(z) > h(c)$ for any $c \in \mathbb{R}$, which, in turn, implies that $\lim_{c\to\infty} u(c) > \lim_{c\to\infty} h(c)$ has to hold. Hence, Eq. (2.4) is satisfied.

Third, suppose that *either* $\lim_{c\to-\infty} u(c) = -\infty$ and $\lim_{c\to\infty} u(c) \in \mathbb{R}$ *or* $\lim_{c\to-\infty} u(c) \in \mathbb{R}$ and $\lim_{c\to\infty} u(c) = \infty$. Here, combining the arguments used in the first and second case yields the claim. Altogether, we conclude that at least one certainty equivalent exists.

2. STEP: Next, consider some certainty equivalent c. We show that there exists some $\epsilon > 0$ such that $u(c') > h(c')$ for any $c' \in (c, c+\epsilon)$ and $u(c'') < h(c'')$ for any $c'' \in (c-\epsilon, c)$. Since we have $u'(c) > 0$ by assumption, it is sufficient to verify that $h'(c) \le 0$ holds.

Denote as $N_c \subset \mathbb{R}$ the set on which the partial derivative $\partial\sigma(u(x), u(c))/\partial c$ does not exist. Since N_c has measure zero, we know that $h'(c)$ exists. Moreover, as σ is bounded and as $|u|$ is integrable, the Dominated Convergence Theorem implies that we can reverse the order of differentiation and integration. Simple re-arrangements show that $h'(c) \le 0$ holds if and only if

$$\underbrace{\frac{\int_{\mathbb{R}} u(x)\sigma(u(x), u(c))\ dF(x)}{\int_{\mathbb{R}} \sigma(u(s), u(c))\ dF(s)}}_{=h(c)} \left(\int_{\mathbb{R}\setminus N_c} \frac{\partial}{\partial c}\sigma(u(s), u(c))\ dF(s) \right)$$
$$\ge \int_{\mathbb{R}\setminus N_c} u(x) \frac{\partial}{\partial c}\sigma(u(x), u(c))\ dF(x).$$

Let $\underline{X} := \{x \in \mathbb{R} : u(x) \le u(c)\} \setminus N_c$ and $\overline{X} := \{x \in \mathbb{R} : u(x) > u(c)\} \setminus N_c$. Since c is a certainty equivalent by assumption, we have $h(c) = u(c)$ and the preceding inequality is equivalent to

$$\int_{\underline{X}} \underbrace{(u(c) - u(x))}_{\ge 0} \underbrace{\frac{\partial}{\partial c}\sigma(u(x), u(c))}_{\ge 0}\ dF(x)$$
$$+ \int_{\overline{X}} \underbrace{(u(c) - u(x))}_{<0} \underbrace{\frac{\partial}{\partial c}\sigma(u(x), u(c))}_{\le 0}\ dF(x) \ge 0,$$

where the sign of $\partial\sigma(u(x), u(c))/\partial c = u'(c) \cdot \partial\sigma(u(x), u(c))/\partial u(c)$ follows from Lemma 4. As a consequence, we have $h'(c) \le 0$, which implies that any certainty equivalent c is locally unique.

3. STEP: Consider two lotteries L_x and L_y where L_x $fosd$ L_y. Suppose that $c_z \in \mathbb{R}$ denotes some certainty equivalent to lottery $L_z \in \{L_x, L_y\}$. Then any such certainty equivalent c_z solves

$$\int_{\mathbb{R}} \phi(s, c_z)\ dF_z(s) = 0, \tag{2.5}$$

where $\phi(s, c) := \Big(u(s) - u(c)\Big)\sigma(u(s), u(c))$ and F_z is the cumulative distribution function of L_z.

Since ordering (by Lemma 4) implies that

$$\frac{\partial}{\partial s}\sigma(u(s),u(c)) \geq 0 \ \ \text{(a.e.)} \qquad \text{if and only if} \qquad u(s) \geq u(c),$$

$\partial\phi(s,c)/\partial s > 0$ almost everywhere. Analogously, $\partial\phi(s,c)/\partial c < 0$ almost everywhere.

Now, for the sake of a contradiction, let $c_x \leq c_y$. This assumption implies

$$0 = \int_{\mathbb{R}} \phi(s,c_x)\ dF_x(s) \geq \int_{\mathbb{R}} \phi(s,c_y)\ dF_x(s) > \int_{\mathbb{R}} \phi(s,c_y)\ dF_y(s) = 0$$

where the equalities follow from (2.5), the weak inequality follows from $\partial\phi(s,c)/\partial c < 0$ (a.e.) and $c_x \leq c_y$, and the strict inequality holds by $\partial\phi(s,c)/\partial s > 0$ (a.e.) and Lemma 6; a contradiction.

Hence, we conclude that the smallest certainty equivalent to the lottery L_x is larger than the largest certainty equivalent to the lottery L_y.

4. STEP: Consider some lottery L with a cumulative distribution function F. For the sake of a contradiction, suppose that at least two certainty equivalents to L exist; that is, there exist some $c_1, c_2 \in \mathbb{R}$, $c_1 < c_2$, such that $u(c_1) = h(c_1)$ and $u(c_2) = h(c_2)$.

Now consider a sequence of lotteries $(L_n)_{n\in\mathbb{N}}$ with cumulative distribution functions $(F_n)_{n\in\mathbb{N}}$ such that $L\ fosd\ L_n$ for any $n \in \mathbb{N}$ and F_n converges to F pointwise. By the second step, for each $k \in \{1,2\}$ there is some $\epsilon_k > 0$ such that $u(z) > h(z)$ for any $z \in (c_k, c_k+\epsilon_k)$ and $u(z) < h(z)$ for any $z \in (c_k-\epsilon_k, c_k)$. In addition, it is straightforward to see that

$$\int_{\mathbb{R}} \frac{u(x)\sigma(u(x),u(c))}{\int_{\mathbb{R}}\sigma(u(s),u(c))\ dF_n(s)}\ dF_n(x) \overset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} \frac{u(x)\sigma(u(x),u(c))}{\int_{\mathbb{R}}\sigma(u(s),u(c))\ dF(s)}\ dF(x).$$

Together, these two observations imply that there exists some $n' \in \mathbb{N}$ such that for any $n \geq n'$ also L_n has at least two certainty equivalents c_1^n and c_2^n which smoothly converge to c_1 and c_2, respectively, when n approaches infinity. This implies that there exists some $n'' \geq n'$ such that for any $n \geq n''$ we have $c_2^n > c_1$. But this yields a contradiction to the fact that the smallest certainty equivalent to L is larger than the largest certainty equivalent to L_n (see 3. STEP). Hence, the certainty equivalent is also globally unique. $\square$

Proof of Proposition 2. Let $\mathcal{C} = \{L_x, L_y\}$ and suppose that L_x and L_y are stochastically independent. Again, we prove a slightly more general version, allowing for any increasing value function $u(\cdot)$ with $u'(\cdot) > 0$ and finite expected utility, $U(L_z) = \int_{\mathbb{R}} u(z)\ dF(z) < \infty$. Denote the marginal cumulative distribution functions as F_x and F_y, respectively. And, as before, we define $\phi(x,y) := (u(x)-u(y))\sigma(u(x),u(y))$. Now suppose that L_x first-order stochastically dominates L_y and denote $\Phi(x,L_y) := \int_{\mathbb{R}} \phi(x,y)\ dF_y(y)$. As, for any $y \in \mathbb{R}$, $\partial\phi(x,y)/\partial x > 0$ (a.e.), we also have $\partial\Phi(x,L_y)/\partial x > 0$. In addition, it is easy to check that, for any lottery L_y, we obtain $\int_{\mathbb{R}} \Phi(s, L_y)\, dF_y(s) = 0$. We thus conclude that

$$\int_{\mathbb{R}^2} \phi(x, y)\, dF(x, y) = \int_{\mathbb{R}} \Phi(s, L_y)\, dF_x(s) > \int_{\mathbb{R}} \Phi(s, L_y)\, dF_y(s) = 0,$$

where the first equality holds by independence, and the inequality holds by $\partial\Phi(x, L_y)/\partial x > 0$ and Lemma 6. By definition, $U^s(L_x|\mathcal{C}) > U^s(L_y|\mathcal{C})$ if and only if $\int_{\mathbb{R}^2} \phi(x, y)\, dF(x, y) > 0$, which yields the claim. □

Proof of Proposition 3. Consider a binary lottery L with expected value E and variance V. Using the characterization of binary risks in Eq. (3.1), we observe that the lottery's lower payoff becomes more likely if the lottery's skewness increases. Formally, we have

$$\frac{\partial p}{\partial S} = 2 \cdot (S^2 + 4)^{-3/2} > 0.$$

Using (3.1) again, we conclude that—for a fixed expected value and a fixed variance—both the lower payoff $x_1 = x_1(E, V, S)$ and the higher payoff $x_2 = x_2(E, V, S)$ increase in the lottery's skewness S. Therefore, the difference between the lower (higher) payoff and the expected value monotonically decreases (increases) in the lottery's skewness S:

$$\frac{\partial(E - x_1)}{\partial S} < 0 \quad \text{and} \quad \frac{\partial(x_2 - E)}{\partial S} > 0.$$

Since the expected value E is fixed, an increase in the contrast $|x_k - E|$ is equivalent to an increase in the salience of state (x_k, E) due to the ordering property. Hence, the lower payoff's salience decreases in S, while the higher payoff's salience increases in S.

Since $\lim_{S\to\infty} x_2(E, V, S) = \infty > E$, we obtain

$$\lim_{S\to\infty} \sigma(x_2, E) > \sigma(E, E) = \lim_{S\to\infty} \sigma(x_1, E)$$

by ordering and continuity of the salience function. Continuity of the salience function further implies that there exists some $\hat{S} < \infty$ such that for any $S > \hat{S}$ the lottery's higher payoff is salient. As we have seen that the salience of both outcomes is monotonic in the lottery's skewness S, we conclude that the salient thinker chooses the risky option if and only if $S > \hat{S}$. Finally, the fact that $\lim_{S\to-\infty} \sigma(x_1, E) > \sigma(E, E) = \lim_{S\to-\infty} \sigma(x_2, E)$ together with monotonicity ensure that there is a unique skewness value $\hat{S} \in \mathbb{R}$ such that $r(E, V, \hat{S}) = 0$. □

Proof of Proposition 4. Let either $x_1 = x_1(E, V, S) \geq 0$ or $x_2 = x_2(E, V, S) \leq 0$. By definition, the threshold value $\hat{S} = \hat{S}(E, V)$ solves $r = r(E, V, S) = 0$. We proceed in two steps. First, we determine the sign of $\partial\hat{S}(E, V)/\partial E$ given that this partial derivative exists. Second, we argue that $\partial\hat{S}(E, V)/\partial E$ exists for any $E \in \mathbb{R}\backslash N$ where N has measure zero. This suffices to prove our claim.

1. STEP: If $\partial\hat{S}(E,V)/\partial E$ exists, then the Implicit Function Theorem yields

$$\frac{\partial}{\partial E}\hat{S}(E,V) = -\left.\frac{\frac{\partial}{\partial E}r(E,V,S)}{\frac{\partial}{\partial S}r(E,V,S)}\right|_{S=\hat{S}}.$$

For any expected value E, variance V, and skewness S, we have

$$\frac{\partial}{\partial E}r(E,V,S) = \sqrt{Vp(1-p)}\cdot\frac{d}{dE}\left(\frac{\sigma(x_1,E)-\sigma(x_2,E)}{p\sigma(x_1,E)+(1-p)\sigma(x_2,E)}\right),$$

so that the sign of the preceding derivative is equal to the sign of

$$\begin{aligned}&\frac{d}{dE}\left(\frac{\sigma(x_1,E)-\sigma(x_2,E)}{p\sigma(x_1,E)+(1-p)\sigma(x_2,E)}\right)\\ &\qquad\qquad = \frac{\sigma(x_2,E)\frac{d}{dE}\sigma(x_1,E)-\sigma(x_1,E)\frac{d}{dE}\sigma(x_2,E)}{\left(p\sigma(x_1,E)+(1-p)\sigma(x_2,E)\right)^2}.\end{aligned}$$

Thus, by definition of the decreasing level effect, we obtain $\partial r(E,V,S)/\partial E < 0$.

In addition, for any expected value E, and any variance V, we have

$$\begin{aligned}\left.\frac{\partial}{\partial S}r(E,V,S)\right|_{S=\hat{S}} &= \sqrt{V}\underbrace{\left(\frac{\hat{\sigma}_1-\hat{\sigma}_2}{p\hat{\sigma}_1+(1-p)\hat{\sigma}_2}\right)}_{=0\text{ by definition of }\hat{S}}\cdot\left.\frac{\partial}{\partial S}\left(\sqrt{p(1-p)}\right)\right|_{S=\hat{S}}\\ &\qquad + \sqrt{Vp(1-p)}\cdot\left.\frac{\partial}{\partial S}\left(\frac{\sigma_1-\sigma_2}{p\sigma_1+(1-p)\sigma_2}\right)\right|_{S=\hat{S}}\end{aligned}$$

where $\sigma_k := \sigma(x_k(E,V,S),E)$ and $\hat{\sigma}_k := \sigma(x_k(E,V,\hat{S}),E)$ for $k\in\{1,2\}$. Since $\hat{\sigma}_1 = \hat{\sigma}_2$ by the definition of $\hat{S}$ and since $\partial\hat{S}(E,V)/\partial E$ exists by assumption, it is easy to verify that $\partial r(E,V,S)/\partial S|_{S=\hat{S}}$ has the same sign as

$$\left.\frac{\partial}{\partial S}\Big(\sigma(x_1(E,V,S),E)-\sigma(x_2(E,V,S),E)\Big)\right|_{S=\hat{S}}.$$

As $\partial\hat{S}(E,V)/\partial E$ exists by assumption, it follows, by Lemma 4 and Eq. (3.1),

$$\left.\frac{\partial}{\partial S}\Big(\sigma(x_1(E,V,S),E)-\sigma(x_2(E,V,S),E)\Big)\right|_{S=\hat{S}} < 0.$$

This implies that $\partial\hat{S}(E,V)/\partial E < 0$, whenever this partial derivative exists.

2. STEP: Notice that

$$\left.\frac{\partial}{\partial S}\Big(\sigma(x_1(E,V,S),E)-\sigma(x_2(E,V,S),E)\Big)\right|_{S=\hat{S}} = 0$$

holds if and only if $\partial\sigma(x_k, E)/\partial x_k = 0$ for $k \in \{1, 2\}$. For a fixed variance V and a fixed skewness S, we define

$$N := \Big\{E \in \mathbb{R} : \frac{\partial}{\partial x_k}\sigma(x_k(E, V, S), E) = 0,\ k \in \{1, 2\},$$
$$\text{or } \frac{d}{dE}\sigma(x_k(E, V, S), E) \text{ does not exist}\Big\}.$$

Hence for any expected value $E \in \mathbb{R} \setminus N$ the partial derivative $\partial\hat{S}(E, V)/\partial E$ exists. By Lemma 4, the set N has measure zero, which implies that for any $\epsilon > 0$ it holds that

$$\hat{S}(E + \epsilon, V) - \hat{S}(E, V) = \int_{[E,E+\epsilon]\setminus N} \frac{\partial}{\partial x}\hat{S}(x, V)\ dx < 0.$$

This completes the proof. □

Proof of Proposition 5. Let $L_y := L(E, V, S)$ and $L_x := L(E, V, -S)$, and denote as $p = p(-S)$ the probability with which the left-skewed lottery's lower payoff is realized. In addition, denote as $\text{sgn} : \mathbb{R} \to \{-1, 0, 1\}$ the signum-function. For the sake of brevity, we again write $U^s(L_z) = U^s(L_z|\mathcal{C})$.

We prove Parts (a), (b), and (c) successively.

PART (a): Since

$$\begin{aligned}
&\text{sgn}\Big(U^s(L_y) - U^s(L_x)\Big)\\
&\quad = \text{sgn}\Bigg((1-p)\Big[2\sigma(y_2, x_1) + \eta\Big(\sigma(y_1, x_1) - 2\sigma(y_2, x_1) + \sigma(y_2, x_2)\Big)\Big]\\
&\qquad\qquad - \Big[2(1-p)\sigma(y_1, x_2) + \eta\Big(\sigma(y_1, x_1) - 2\sigma(y_2, x_1) + \sigma(y_2, x_2)\Big)\Big]\Bigg),
\end{aligned}$$

we conclude that $U^s(L_y) - U^s(L_x) > 0$ holds if and only if

$$\eta < \frac{2(1-p)[\sigma(y_2, x_1) - \sigma(y_1, x_2)]}{2[(1-p)\sigma(y_2, x_1) - p\sigma(y_1, x_2)] - (1-2p)[\sigma(y_1, x_1) + \sigma(y_2, x_2)]} =: \tilde{\eta}(S).$$

where $x_k = x_k(E, V, S)$, $y_k = y_k(E, V, S)$, $k \in \{1, 2\}$, and $p = p(-S)$ are defined in Eq. (3.1). Now we define $\check{\eta}(S) := \min\Big\{1, \tilde{\eta}(S)\Big\}$. Then, it is straightforward to check that $\check{\eta}(S) < 1$ if and only if

$$\sigma(y_1, x_1) + \sigma(y_2, x_2) - 2\sigma(y_1, x_2) < 0. \tag{2.6}$$

As $\lim_{S\to\infty} x_1 = -\infty$, as $\lim_{S\to\infty} x_2 = E = \lim_{S\to\infty} y_1$, and as $\lim_{S\to\infty} y_2 = \infty$, the ordering property implies that Inequality (2.6) does *not* hold in the limit of $S \to \infty$.

Analogously, as $\lim_{S\to 0} x_1 = E-\sqrt{V} = \lim_{S\to 0} y_1$ and as $\lim_{S\to 0} x_2 = E+\sqrt{V} = \lim_{S\to 0} y_2$, ordering implies that Inequality (2.6) holds in the limit of $S \to 0$.

PART (b): Since we have

$$\begin{aligned}\operatorname{sgn}\Big(U^s(L_y) - U^s(L_x)\Big) &= \operatorname{sgn}\Big(p(y_2 - x_1)\sigma(y_2, x_1) + (1-p)(y_1 - x_2)\sigma(y_1, x_2)\Big)\\ &= \operatorname{sgn}\Big(p(y_2 - x_1)\Big(\sigma(y_2, x_1) - \sigma(y_1, x_2)\Big)\Big),\end{aligned}$$

where the second equality follows from the fact that $\mathbb{E}[L_x] = E = \mathbb{E}[L_y]$, and since $y_2 > x_2 > y_1 > x_1$ by Eq. (3.1), the statement follows from ordering.

PART (c): Notice that y_1 and y_2 monotonically increase in S while x_1 and x_2 monotonically decrease in S. Thus, since $y_2 > x_2 > y_1 > x_1$, ordering implies that $\sigma(y_1, x_1)$ and $\sigma(y_2, x_2)$ monotonically increase in S while $\sigma(y_1, x_2)$ monotonically decreases in S. Together these observations imply that the left-hand side of (2.6) strictly increases in S. The statement then follows from $\lim_{S\to\infty} \check{\eta}(S) = 1$ and $\lim_{S\to 0} \check{\eta}(S) < 1$ (see Part (a) of the proof). $\square$

Appendix C: Decreasing Level Effect

In this section, we show that a wide class of salience functions satisfies the decreasing level effect, as introduced in Definition 2. Consider, for instance, the following class of salience functions:

$$\sigma_\theta(x, y) := \frac{(x-y)^{2n}}{(|x| + |y| + \theta)^{2n}}, \; n \in \mathbb{N}, \; \theta > 0.$$

For the sake of the argument, let $y, z \in \mathbb{R}$ and $x \geq \max\{-y, -z, 0\}$. Then,

$$\frac{\frac{d}{dx}\sigma_\theta(x+y, x+z)^n}{\sigma_\theta(x+y, x+z)^n} = -\frac{4n}{x+y+x+z+\theta} = \frac{\frac{d}{dx}\sigma_\theta(-x-y, -x-z)^n}{\sigma_\theta(-x-y, -x-z)^n},$$

which implies that, for any $n \in \mathbb{N}$, the salience function $\sigma_\theta(x, y)$ indeed satisfies the decreasing level effect. It is easy to check that also positive transformations of the salience function $\sigma_\theta(x, y)$ satisfy this property.

Similar calculations show that also the salience function

$$\sigma(x, y) = \frac{|x-y|}{|x| + |y| + \theta},$$

which was proposed by Bordalo *et al.* (2012), satisfies the decreasing level effect, and so do the weighting functions proposed by Bordalo, Gennaioli and Shleifer (2016b) or Thakral and Tô (forthcoming).

Appendix D: Experiments on Salience and Skewness

D.1: Instructions

Information about the experiment

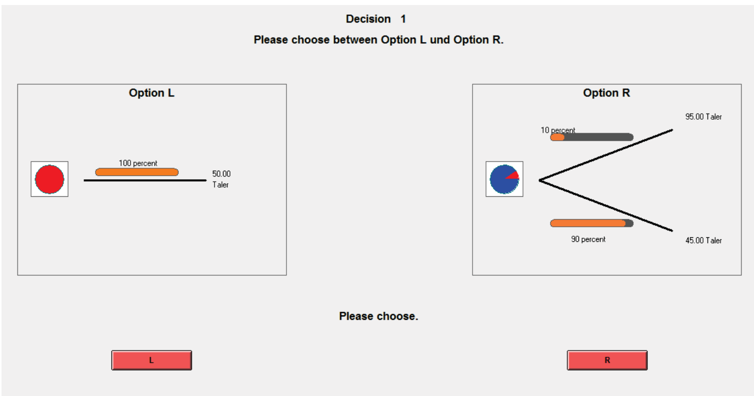
Welcome to this experimental study. Please do not talk to other participants or use your mobile from now on and throughout the entire experiment. Please read the following instructions carefully. For the successful completion of the experiment it is important that you have fully understood the instructions. Should you have any questions at any point in time please raise your hand. An experimenter will then answer your questions at your seat.

In this experiment you can earn an experimental currency (Taler) which will be converted into Euro at the end of the experiment. The conversion rate is

1 Euro = 2 Taler.

Altogether you will make 12 decisions. These decisions only concern your personal preferences, there are no right or wrong answers. You choose between two choice options that are denoted (L) and (R). Option (L) always denotes a safe option that gives you a certain payoff with 100% probability. Option (R) gives you a payoff that depends on a turn of the wheel of fortune with 100 fields that is simulated by your computer. In the following we show you some examples. Please study them carefully.

Example 1:

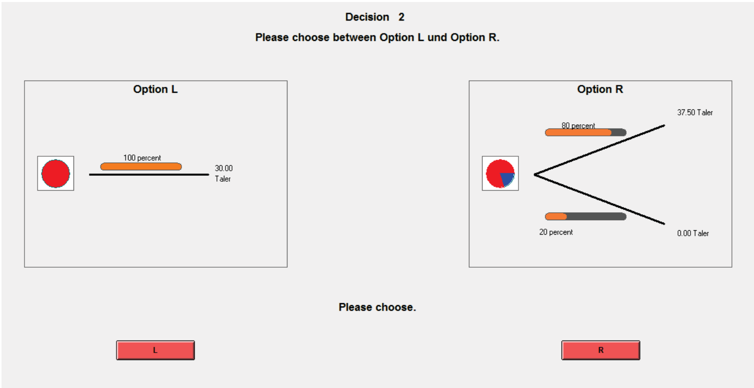


Option (L): You obtain 50 Taler.

Option (R): You obtain 45 Taler with 90% probability (that is, if the wheel of fortune stops on fields 1-90) and 95 Taler with 10% probability (that is, if the wheel of fortune stops on fields 91-100).

Figure 2.8: Instructions for Experiment 1, translated into English (first part).

Example 2:



Option (L): You obtain 30 Taler.

Option (R): You obtain 0 Taler with 20% probability (that is, if the wheel of fortune stops on fields 1-20) and 37.5 Taler with 80% probability (that is, if the wheel of fortune stops on fields 21-100).

Payoffs:

At the end of the experiment the computer will choose one of your 12 choice tasks randomly. If you have chosen (L) in this task you will receive the according sum. If you have chosen (R) your payoff will be determined through the simulation of the turn of a wheel of fortune. Your payoff will be paid in cash at the end of the experiment.

Please look carefully at each of the 12 choice tasks. Between tasks the payoff probabilities and the corresponding payoffs change.

Figure 2.9: Instructions for Experiment 1, translated into English (second part).

Information about the experiment

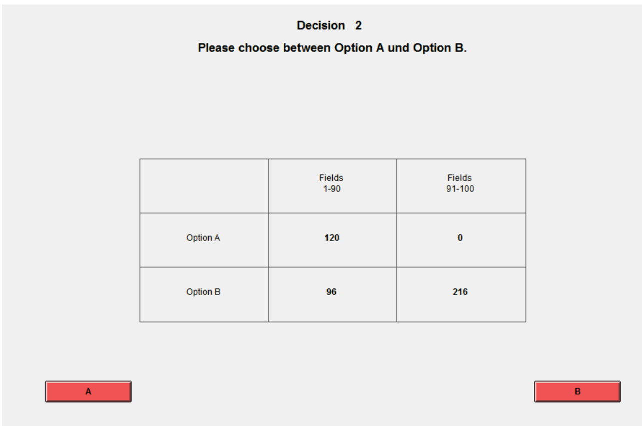
Welcome to this experimental study. Please do not talk to other participants or use your mobile from now on and throughout the entire experiment. Please read the following instructions carefully. For the successful completion of the experiment it is important that you have fully understood the instructions. Should you have any questions at any point in time please raise your hand. An experimenter will then answer your questions at your seat.

In this experiment you can earn an experimental currency (Taler) which will be converted into Euro at the end of the experiment. The conversion rate is

1 Euro = 4 Taler.

Altogether you will make 12 decisions. These decisions only concern your personal preferences, there are no right or wrong answers. You choose between two choice options that are denoted A and B. The payoffs of these options depend on a turn of the wheel of fortune with 100 fields that is simulated by your computer. The probability of being hit is the same for all fields. In the following we show you some examples. Please study them carefully.

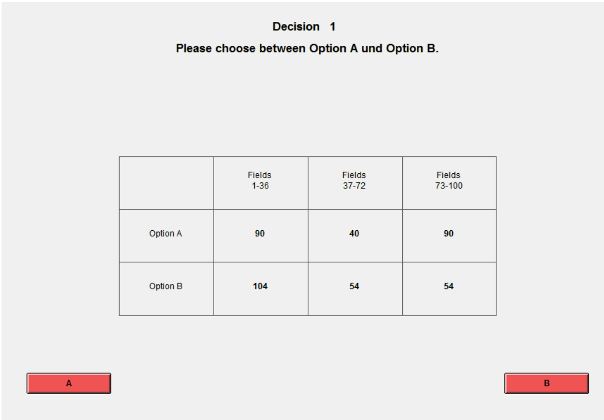
Example 1:



If the wheel of fortune stops on fields 1-90 (that corresponds to a 90% probability) with Option A you will receive exactly 120 Taler and with Option B exactly 96 Taler. If the wheel of fortune stops on fields 91-100 (that gives a 10% probability) you will receive with Option A exactly 0 Taler and with Option B exactly 216 Taler.

Figure 2.10: Instructions for Experiment 2, translated into English (first part).

Example 2:



If the wheel of fortune stops on fields 1-36 (that corresponds to a 36% probability) with Option A you will receive exactly 90 Taler and with Option B exactly 104 Taler. If the wheel of fortune stops on fields 37-72 (that gives a 36% probability) you receive with Option A exactly 40 Taler and with Option B exactly 54 Taler. If the wheel of fortune stops on fields 73-100 (that corresponds to a 28% probability) you receive with Option A exactly 90 Taler and with Option B exactly 54 Taler.

Payoffs:

At the end of the experiment the computer will choose one of your 12 choice tasks randomly. This choice task is payoff relevant. Your payoff will be determined through the simulation of the turn of a wheel of fortune. Assume, for instance, the choice task given in Example 1 is payoff relevant and the wheel of fortune stops on field 93. If you have chosen option A you will receive 0 Taler. If you have chosen Option B you will receive 216 Taler.

Your payoff will be paid in cash at the end of the experiment.

Please look carefully at each of the 12 choice tasks. Between tasks the payoff probabilities and the corresponding payoffs change.

Figure 2.11: Instructions for Experiment 2, translated into English (second part).

D.2: Additional Results

Experiment 1. We report further results of Experiment 1. Table 2.8 presents the numbers underlying Figure 3, and Figure 2.12 illustrates additional results regarding our within-subjects predictions. In particular, we have estimated the Regression Model (1) of Table 4 for each subject separately and plot the point estimates for the coefficient on the lottery's skewness together with the corresponding 95%-confidence intervals. We observe that—in line with Proposition 3—the point estimates are positive for the majority of subjects.

Table 2.8: *Descriptives for Experiment 1.*

	Choice of lottery for $E = 30$		**Choice of lottery for** $E = 50$	
	# of choices	% of choices	# of choices	% of choices
$S = -1.5$	2	3%	7	11%
$S = -0.6$	3	5%	9	15%
$S = 0$	9	15%	15	24%
$S = 1.5$	22	35%	43	69%
$S = 2.7$	34	55%	47	76%
$S = 6.9$	49	79%	52	84%

Notes: *The table presents the number and share of risk seeking choices for each combination of skewness level and expected value.*

Experiment 2. In the following, we provide further results on Experiment 2. Table 2.9 presents the numbers underlying Figure 4, and Figure 2.13 illustrates the results for the combined data separately for each Mao pair. In particular, we find that the results are robust across Mao pairs.

Finally, we extend our analysis of relative skewness by regressing a binary indicator of whether the right-skewed lottery of a given Mao pair is chosen on the left-skewed lottery's relative skewness (see Table 2.10). We find that the average probability of choosing the right-skewed lottery of a Mao pair significantly decreases in the relative skewness of the left-skewed lottery. Figure 2.14 presents individual-level versions of our regression model using the combined data, and we observe that—in line with Proposition 5—the majority of point estimates are negative.

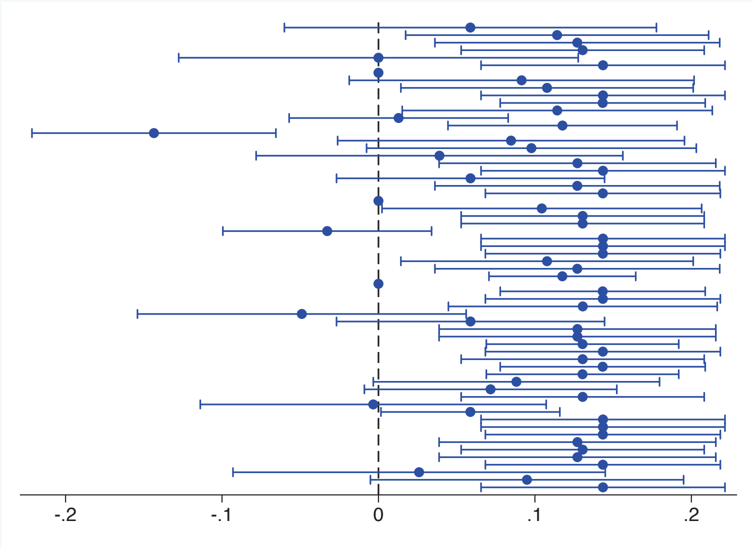


Figure 2.12: *The figure illustrates the point estimates and 95%-confidence intervals for the coefficient on the risky option's skewness in individual-level versions of Regression Model (1) presented in Table 4. Each of the 62 point estimates corresponds to a specific subject and is based on twelve observations at six different skewness levels. A positive coefficient implies that the average probability that this subject chooses the lottery over its expected value increases with the lottery's skewness. Due to the small number of observations per subject the confidence intervals should be interpreted with caution.*

Table 2.9: *Descriptives for Experiment 2.*

		Perfectly Negative Correlation		**Maximal Positive Correlation**	
		# of choices	% of choices	# of choices	% of choices
Initial Study	$S = 2.7$	224	95%	214	90%
	$S = 0.6$	145	61%	113	48%
Replication	$S = 2.7$	313	92%	318	94%
	$S = 0.6$	216	64%	175	52%
Combined	$S = 2.7$	537	93%	532	92%
	$S = 0.6$	361	63%	288	50%

Notes: *The table presents the number and share of choices of the right-skewed lottery of a Mao pair pooled over all symmetric and skewed Mao pairs, respectively, as introduced in Table 3. We present results separately for the initial study and the replication as well as the combined results for both studies.*

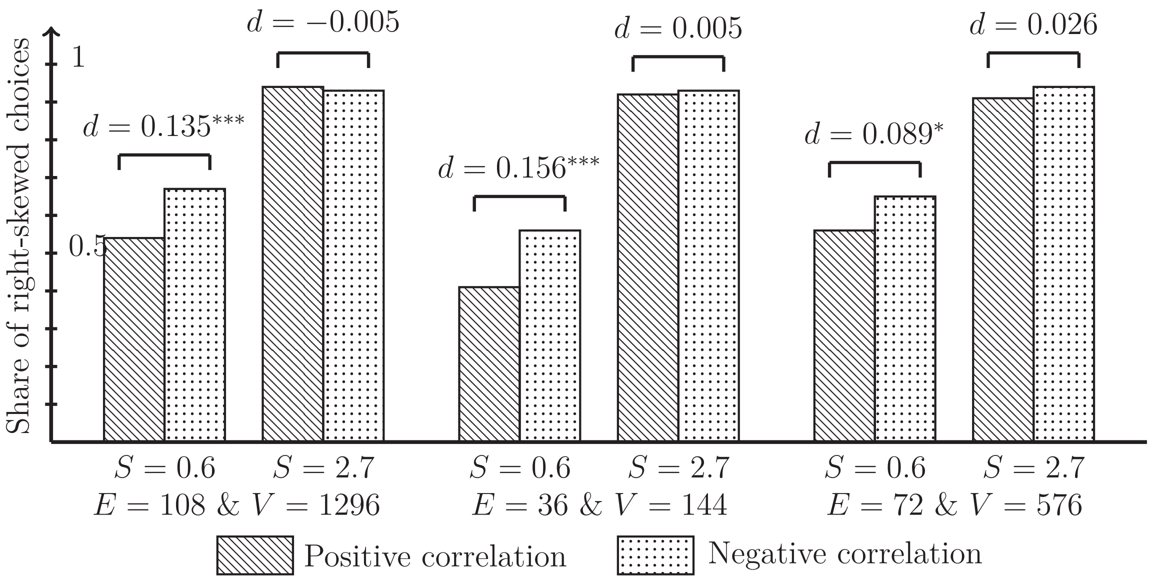


Figure 2.13: *The figure illustrates the share of choices of the right-skewed lottery under the maximal positive and the perfectly negative correlation, respectively. We present the combined results for both studies separately for each Mao pair. We further report the results of paired t-tests. Significance level: *: 10%, **: 5%, ***: 1%.*

Table 2.10: *Additional regressions for Experiment 2.*

Parameter	**Initial Study**	**Replication**	**Combined**
Constant	0.602***	0.635***	0.622***
	(0.025)	(0.021)	(0.016)
Relative Skewness	-0.140***	-0.126***	-0.132***
	(0.011)	(0.010)	(0.007)
# Subjects	79	113	192
# Choices	948	1,356	2,304

Notes: *The table presents the results of OLS regressions of a dummy indicating the choice between the lotteries of a Mao pair (which takes a value of one if the subject chooses the right-skewed lottery and a value of zero otherwise) on the relative skewness of the left-skewed lottery. All standard errors are clustered at the subject level and provided in parenthesis. Significance level: *: 10%, **: 5%, ***: 1%.*

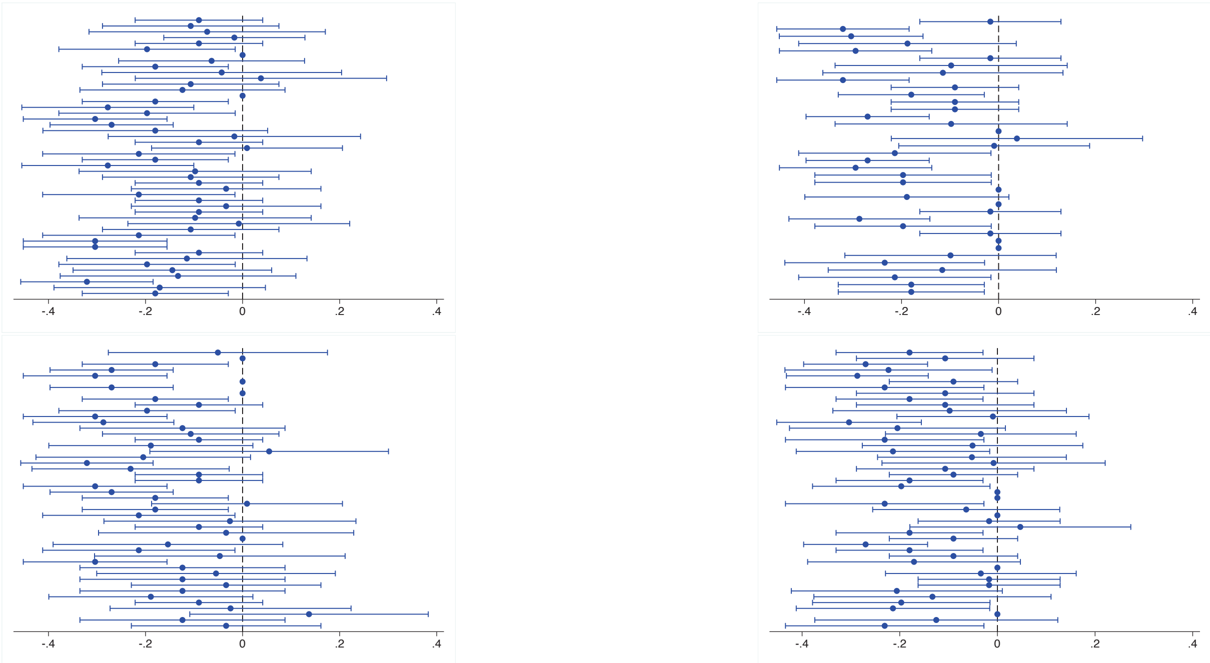


Figure 2.14: *The figure illustrates the point estimates and 95%-confidence intervals for the coefficient on the left-skewed lottery's relative skewness in individual-level versions of the regression model presented in Table 2.10 using the combined data. Each of the 192 point estimates corresponds to a specific subject and is based on twelve observations. A negative coefficient implies that the average probability that this subject chooses the right-skewed lottery decreases with the left-skewed lottery's relative skewness. But due to the small number of observations per subject the confidence intervals should be interpreted with caution.*

Declaration of Contribution

Hereby I, Mats Köster, declare that the chapter “Salience and Skewness Preferences” is co-authored by Markus Dertwinkel-Kalt. It has been published in the Journal of the European Economic Association (Dertwinkel-Kalt and Köster, 2020b). Both authors contributed equally to the chapter.

Signature of coauthor (Markus Dertwinkel-Kalt): M. Dertwinkel-Kalt

Chapter 3

Salient Cues and Complexity

Co-authored by Markus Dertwinkel-Kalt

3.1 Introduction

Most economic choices — whether it is investing in stocks, buying a durable good, or voting in a general election — have in common that these decisions are complex along at least two dimensions: there are many options that have to be evaluated and compared, and the consequences of choosing one option over the others are not readily observable to consumers. A growing literature argues that the complexity of a choice problem can fundamentally change the way in which people approach the problem and eventually reach a decision (see, e.g., Gigerenzer and Gaissmaier, 2011; Abeler and Jäger, 2015; Chernev, Böckenholt and Goodman, 2015; Banovetz and Oprea, 2020; Oprea, 2020). A separate body of work shows that economic decisions are often influenced by *salient cues* that stand out in the choice context and attract attention (e.g. Chetty, Looney and Kroft, 2009; Bordalo *et al.*, 2012, 2013b; Bordalo, Gennaioli and Shleifer, 2020a; Kőszegi and Szeidl, 2013; Gabaix, 2014, 2019; Li and Camerer, 2019). Intuitively, salience thereby reduces the "dimensionality" of the decision problem: people focus their attention on the most salient aspects of the available options, at the cost of less salient ones. We, thus, hypothesize that there is a systematic relationship between complexity and salience in shaping behavior: people are more likely to behave consistently across differently complex problems if there is a common salient cue guiding their attention and, consequently, behavior.

We theoretically develop and experimentally test this hypothesis in the context of choice under risk. Specifically, we consider *portfolio selection problems* of how to allocate a given budget across two assets. Here, we not only have a well-established notion of a

Portfolio 1: 25% in X_1 and 75% in X_2.
Portfolio 2: 50% in X_1 and 50% in X_2.

	90%	10%
Asset X_1	96	216
Asset X_2	120	0

Portfolio 1 Portfolio 2

Portfolio 1: 25% in Y_1 and 75% in Y_2.
Portfolio 2: 50% in Y_1 and 50% in Y_2.

	50%	50%
Asset Y_1	120	48
Asset Y_2	72	96

Portfolio 1 Portfolio 2

Figure 3.1: *This figure illustrates two portfolio selection problems with either skewed assets (left panel) or symmetric assets (right panel). The columns of the each table refer to the different states of the world, and the rows indicate the outcomes of the respective assets in these states. While in the left panel the large contrast in the second state of the world (216 versus 0) is particularly salient, in the right panel the contrast in outcomes is the same in both states, so neither state is particularly salient.*

salient cue, but we can also easily manipulate the complexity of the problem. Our notion of salience builds on research from psychology and economics documenting and modeling *contrast effects*: attention is directed to the states of the world in which the available outcomes differ the most (e.g. Rubinstein, 1988; Simonson and Tversky, 1992; Schkade and Kahneman, 1998; Bordalo *et al.*, 2012; Kőszegi and Szeidl, 2013). Along these lines, we think of a salient cue as an asymmetric distribution of contrasts in outcomes that guides attention toward a certain state of the world. Whether or not such a salient cue exists, depends on the skewness of the two assets (see Figure 1 for an illustration). By the means of a controlled lab experiment, we identify an interaction between two natural layers of complexity (see Figure 3.2) and the skewness of the assets (and thereby salience): while revealed attitudes toward skewed risks are consistent across differently complex problems, revealed attitudes toward symmetric risks are not. When the assets are symmetric, subjects behave as if they were risk averse in simple problems, but *diversify naively* (Benartzi and Thaler, 2001) by investing equal shares in both assets — irrespective of the monetary consequences — in complex ones. This indicates that absent a salient cue complexity can fundamentally change the way in which people approach a problem and eventually reach a decision. We propose a variant of Bordalo *et al.*'s (2012) salience theory that rationalizes these findings: When confronted with a complex problem, subjects look for a salient cue — i.e. an asymmetric contrast in outcomes — indicating that one of the two assets is the "better" investment. If such a cue is present, it guides a subject's behavior. Without a salient piece of information to base her decision on, however, the subject reverts to a "default heuristic" of naive diversification.[1]

[1]The idea underlying our theoretical framework is reminiscent of Shafir, Simonson and Tversky's (1993) informal notion of reason-based choice: People need a reason — similar to a salient cue — for

	90%	10%
Portfolio 1	114	54
Portfolio 2	108	108

Portfolio 1 Portfolio 2

Let { 0%, 1%, ..., 99%, 100% } be the feasible investments in Asset X_1.

	90%	10%
Asset X_1	96	216
Asset X_2	120	0

What share to invest in X_1? ☐ %

Figure 3.2: *This figure illustrates modifications of the portfolio selection problem with skewed assets from Figure 3.1. In the left panel, for each portfolio, the final outcomes are readily presented ("simple" choice). In the right panel, the number of portfolios is increased to 101, and final outcomes have to be computed.*

Building on *salience theory* (Bordalo *et al.*, 2012, 2013b), we derive in Section 3.2 an interaction between the salience of outcomes and what we call "comparative" complexity, that is, the number of options to be evaluated and compared. When applied to choice under risk, salience theory assumes that a subject evaluates a lottery (such as a portfolio in the examples above) relative to some *reference point*, and then inflates the probabilities of the states of the world in which the lottery's outcomes differ the most from this reference point. The reference point is given by the state-wise average over all alternative options and, thus, varies with the comparative complexity of the problem. We show that predicted attitudes toward skewed risks — which have extreme and salient outcomes — do not vary with comparative complexity, while attitudes toward symmetric risks — where such a salient cue is missing — are predicted to be less robust.

As a baseline, we study the "simple" choice between a non-negative binary lottery and the safe option that pays the lottery's expected value with certainty. As argued in Dertwinkel-Kalt and Köster (2020b), salience theory makes two qualitative predictions: *skewness preferences* (i.e. a preference for sufficiently right-skewed and an aversion to sufficiently left-skewed risks) and an *aversion toward symmetric risks*. Consider, for example, the right-skewed lottery that pays either 94 with 90% probability or 154 with 10% probability. The upside of 154 differs by a lot more from the expected value of 100 — in this case also the corresponding reference point — than does the downside of 94. By the contrast effect, people thus focus too much on the upside of this right-skewed lottery, and behave as if they were overweighting the probability that it occurs. More generally, the contrast effect induces skewness preferences. For symmetric lotteries, on the other hand, the contrast on the up- and downside is identical: compare, for instance, the binary lot-

investing more in one of the assets. If such a cue or reason is missing, however, they may abstain from making a choice, and simply invest equal shares in both assets.

tery that pays with equal probability either 82 or 118 to its expected value of 100. Thus, by the contrast effect, neither the up- nor downside of a symmetric lottery is particularly salient. Bordalo *et al.* (2012) deal with this by further imposing a *level effect*: a given contrast in outcomes is more salient at a lower outcome level, which results in an aversion toward symmetric risks.

When making the problem comparatively more complex through enlarging the choice set, salience theory still predicts skewness preferences, but no longer an aversion to symmetric risks. To understand these predictions, take the exemplary lotteries from above. For the right-skewed lottery, the upside of 154 differs by a lot more from the expected value of 100 — the reference point in the "simple" problem — than does the downside of 94. Hence, comparative complexity would have to change the reference point by a lot for the downside to become salient. For the symmetric lottery, on the other hand, the contrast on the upside (118 vs. 100) and downside (82 vs. 100) is identical, so that already a relatively small change to the reference point could render the upside salient. As an illustration, let us expand the choice set by another symmetric lottery that pays with equal probability either 50 or 150. The symmetric lottery that pays either 82 or 118 is then evaluated relative to the reference lottery that pays with equal probability 75 or 125. If the lotteries are perfectly negatively correlated, the two states of the world are (82, 125) and (118, 75). While the contrast in outcomes is still identical in both states of the world, the payoff level is now lower in the state where the symmetric lottery from above pays its upside of 118. Hence, by the level effect, the upside of this symmetric lottery is salient, which induces risk-seeking behavior. We conclude that salience theory predicts skewness preferences to be more robust to an increase in comparative complexity than the aversion to symmetric risks.

Motivated by the predictions of salience theory, we develop in Section 3.3 an experimental design that allows us to test for potential interactions between the salience of outcomes and different layers of complexity. We implement portfolio selection problems of how to allocate a given budget across two assets (see Figure 3.1). Since there is no commonly accepted definition of complexity, we focus on two natural layers of complexity in arguably uncontroversial cases (see Figure 3.2): "computational" complexity refers to the difficulty in observing the consequences of an option, while "comparative" complexity refers to the number of options to be evaluated and compared. Building on the contrast effect, we use the skewness of the assets as a proxy for salience, and compare problems with symmetric assets to problems with skewed assets. Salience theory predicts an interaction between skewness and comparative complexity, but it predicts no effect of computational complexity, which is outside of the model. To obtain a clean test of these predictions, we assume that the two assets as well as all resulting portfolios are binary lotteries with the same expected value, and that subjects could always choose the *diversified portfolio* that pays this expected value with certainty. This last feature provides us with a first behavioral benchmark: any expected utility maximizer with a concave utility function chooses the diversified portfolio. A second behavioral benchmark, which is empirically relevant in portfolio selection problems with symmetric assets (e.g. Eyster and Weizsäcker, 2016), is that of *naive diversification*: subjects mechanically invest equal shares in the two underlying assets. Building on the experimental design proposed in Eyster and Weizsäcker (2016), we then construct a class of portfolio selection problems that allows us to identify both skewness preferences and an aversion toward symmetric risks from the behavioral benchmarks of actual and naive diversification. We provide details on the implementation of the experiment in Section 3.4.

Section 3.5 presents our main experimental result: revealed attitudes to skewed risks are robust to both computational and comparative complexity, while revealed attitudes toward symmetric risks are not robust to either layer of complexity. When the assets are skewed, subjects consistently reveal skewness preferences across the differently complex problems. When the assets are symmetric, however, subjects behave as if they were risk-averse in simple problems, but diversify naively in complex ones, thereby often revealing risk-seeking behavior. Salience theory can account for the interaction between skewness and comparative complexity, but it cannot explain naive diversification in problems with symmetric assets that are only computationally complex. Our experimental results do confirm, however, the broader intuition that subjects behave consistently across differently complex portfolio selection problems if there is an asymmetric contrast in outcomes guiding their behavior, but not if such a salient cue is missing.

In Section 3.6, we document two stylized facts that shed light on the mechanism underlying our main experimental result: (1) asymmetric contrasts in outcomes shape memory of and, arguably, the attention paid to these outcomes; and (2) the degree of naive diversification with symmetric assets is sensitive to cognitive skills, while the strength of skewness preferences is not. The first observation suggests that skewness preferences are driven by the contrast effect (see also Dertwinkel-Kalt and Köster, 2020b). The second observation indicates that choices involving skewed risks are cognitively less demanding than choices involving symmetric risks.

Based on this additional evidence, we propose in Section 3.7 a variant of salience theory that rationalizes all results. Incorporating ideas from the literature on deliberate inattention (e.g. Gabaix, 2014, 2019), we introduce the distinction between *stimulus-driven* (or more active) and *default-driven* (or more passive) choices. In the context of portfolio selection problems, we think of this distinction as follows: Without any information on the joint distribution of the underlying assets, with equal probability each of them is the better investment. In this sense, naive diversification is a natural starting point when thinking about which portfolio to choose. Upon learning the joint distribution of the assets, the subject looks for a salient cue — i.e. an asymmetric contrast in outcomes — regarding which asset is the "better" investment. If the problem is neither computationally nor comparatively complex and, therefore, relatively easy to solve, she makes an active choice even in the absence of a salient cue. But, if the problem is more complex *and* a salient cue is missing, the subject may avoid doing so, and instead stick with her first "guess" of naive diversification. We formalize this idea by assuming that if, for a given complexity of the problem, the distribution of contrasts in outcomes is sufficiently close to symmetric (i.e. if a salient cue is missing), subjects apply the $1/N$-heuristic (Benartzi and Thaler, 2001) as a default. If, on the other hand, a subject makes an active choice, then her behavior can be described by salience theory (Bordalo *et al.*, 2012). While being ex-post, this framework — which can be microfounded via the contrast effect — makes testable predictions on the conditions under which choices will be stimulus- or default-driven. A direct test of these predictions reveals that subjects are indeed more likely to behave in line with salience theory the more asymmetric is the distribution of contrasts in outcomes and the lower is the complexity of the portfolio selection problem. As suggested by our framework, an increase in the asymmetry of contrasts in outcomes operates through mitigating the effect of complexity on portfolio selection.

In Section 3.8, we discuss the related literature. While there are growing literatures on how complexity and salience can affect economic behavior, no previous paper has systematically studied the interaction between the two. We conclude in Section 3.9 by sketching applications of our framework beyond the domain of choice under risk (e.g. to address the role of political polarization in voting). While choice under risk allows us to cleanly test our hypothesis on the interaction between complexity and the availability of a salient cue, the general idea — namely, that people are more likely to behave consistently across differently complex problems if there is a common salient cue guiding their behavior — is much broader than this specific application.

3.2 The Role of Complexity in Salience Theory of Choice under Risk

Building on *salience theory* (Bordalo *et al.*, 2012, 2013b), we derive an interaction between the salience of outcomes and the "comparative" complexity of a problem, that is, the number of options to be evaluated and compared. For "simple" problems with just two options, salience theory makes two qualitative predictions (Section 3.2.2): an aversion toward symmetric risks and skewness preferences. But only skewness preferences are predicted to be robust to comparative complexity, while the aversion to symmetric risks is not (Section 3.2.3). A complete analysis of this interaction, alongside the proof of

Proposition 1, can be found in Appendix A. Other natural layers of complexity are outside of the model and should, thus, not affect behavior (Section 3.2.4).

3.2.1 Salience Theory of Choice under Risk

Consider a choice set $\mathcal{C} = \{X_i\}_{i=1}^n$. The random variables (or *lotteries*) X_1 to X_n are non-negative with a joint cumulative distribution function $F : \mathbb{R}^n_{\geq 0} \to [0,1]$. A state of the world refers to a tuple of outcomes, $(x_1, \ldots, x_n) \in \mathbb{R}^n_{\geq 0}$. If a random variable is degenerate, we call it a *safe* option.

A *salient thinker* evaluates a lottery X_i relative to a *reference point* R_i given by the state-wise average over all alternative options: $R_i := \frac{1}{n-1}\sum_{j\neq i} X_j$. When evaluating X_i, a salient thinker assigns a subjective probability to the state $(x_1, \ldots, x_n) \in \mathbb{R}^n_{\geq 0}$ that depends on the state's objective probability, and on how salient the realization x_i is relative to the realized reference point $r_i = \frac{1}{n-1}\sum_{j\neq i} x_j$. The salience of $(x_i, r_i) \in \mathbb{R}^2_{\geq 0}$ is measured via a *salience function* as follows:

Definition 1 (Salience Function)**.** *A symmetric, bounded, and continuous function σ : $\mathbb{R}^2_{\geq 0} \to \mathbb{R}_{>0}$ is a* salience function *if and only if it satisfies the following two properties:*[2]

1. Ordering*. Let $x \geq y$. Then, for any $\epsilon, \epsilon' \geq 0$ with $\epsilon + \epsilon' > 0$, we have*

$$\sigma(x + \epsilon, y - \epsilon') > \sigma(x, y).$$

2. Diminishing sensitivity*. For any $\epsilon > 0$, we have*

$$\sigma(x + \epsilon, y + \epsilon) < \sigma(x, y).$$

We say that $(x_i, r_i) \in \mathbb{R}^2_{\geq 0}$ is the *more salient* the larger its salience value $\sigma(x_i, r_i)$ is. Ordering implies that a pair of outcomes is the more salient the more these outcomes differ, thereby capturing the well-known *contrast effect* (e.g., Tversky and Kahneman, 1992; Schkade and Kahneman, 1998). Diminishing sensitivity reflects *Weber's law* of perception and can be understood as a *level effect*: a given contrast in outcomes is more salient at a lower outcome level. The implications of contrast effect and level effect are not aligned in general. Hence, for some results, we need to impose additional structure on the salience function (e.g., Bordalo *et al.*, 2013b, 2016b).

Assumption 1. *For any $x, y \geq 0$ and $\lambda > 0$, $\sigma(x, y) = \sigma(\lambda x, \lambda y)$.*

[2] Bordalo *et al.* (2012) allow for random variables with negative outcomes and add a third property to ensure that diminishing sensitivity (with respect to zero) reflects to the negative domain: by the *reflection* property, for any $w, x, y, z \geq 0$, it holds that $\sigma(x, y) > \sigma(w, z)$ if and only if $\sigma(-x, -y) > \sigma(-w, -z)$.

A salient thinker evaluates outcomes via a linear value function, $u(x) = x$, and chooses from the set $\mathcal{C} = \{X_j\}_{j=1}^n$ as to maximize her salience-weighted utility defined as follows:

Definition 2 (Salience-Weighted Utility)**.** *The* salience-weighted utility *of a random variable X_i evaluated in the choice set $\mathcal{C} = \{X_j\}_{j=1}^n$ equals*

$$U^s(X_i|\mathcal{C}) = \frac{1}{\int_{\mathbb{R}^n_{\geq 0}} \sigma(x_i, r_i)\ dF(x_1, \dots, x_n)} \int_{\mathbb{R}^n_{\geq 0}} x_i\ \sigma(x_i, r_i)\ dF(x_1, \dots, x_n)$$

for $r_i = \frac{1}{n-1} \sum_{j \neq i} x_j$ and a salience function $\sigma(\cdot, \cdot)$ that is bounded away from zero.

Notice that the salience-weighted probabilities are normalized so that they sum to one, which implies that a salient thinker's valuation of a safe option $x \in \mathbb{R}_{\geq 0}$ is undistorted.

3.2.2 Two Qualitative Predictions on Simple Choices between Binary Lotteries

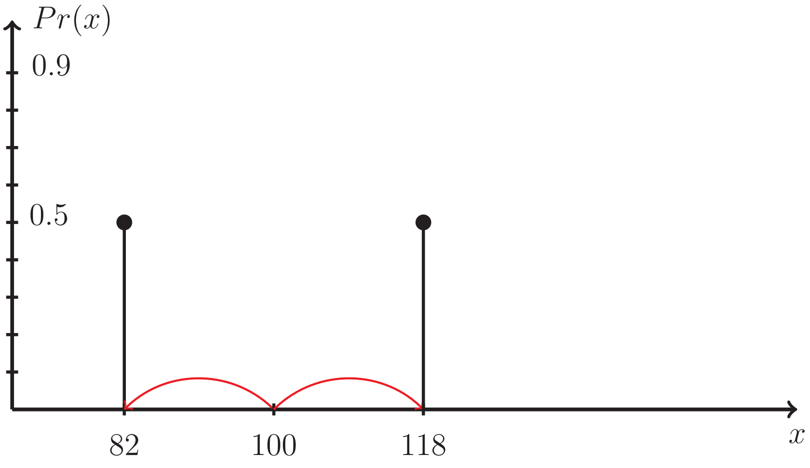
As a baseline, we study the "simple" choice between a binary lottery and the safe option that pays the lottery's expected value with certainty. We make use of the fact that any binary lottery can be uniquely characterized through its expected value, variance, and skewness (see Ebert, 2015): for any $E \in \mathbb{R}$, $V \in \mathbb{R}_{>0}$, and $S \in \mathbb{R}$, there exists exactly one binary lottery $L = L(E, V, S)$ with expected value $\mathbb{E}[L] = E$, variance $\mathrm{Var}(L) = V$, and skewness $S(L) = S$. The skewness of the binary lottery L is here defined as

$$S(L) := \mathbb{E}\left[\left(\frac{L - \mathbb{E}[L]}{\sqrt{\mathrm{Var}(L)}}\right)^3\right];$$

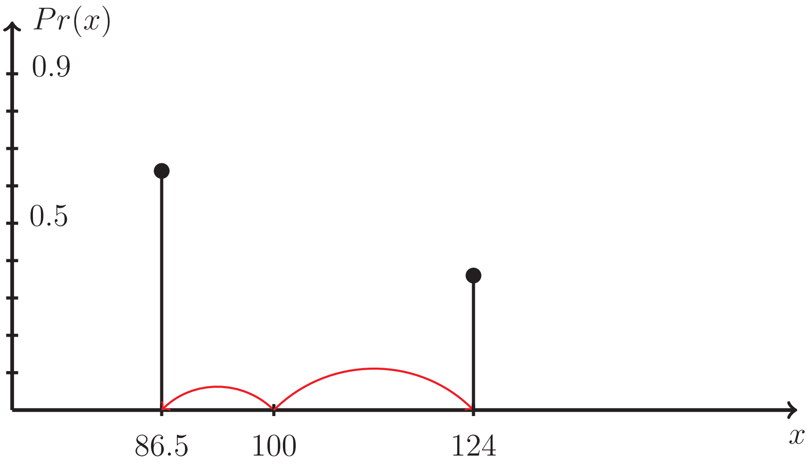
that is, its third standardized central moment. Other, in the case of binary risks equivalent, notions of skewness refer to the length or thickness of the tails of the probability distribution (see, e.g., Ebert, 2015, for a discussion).

For conciseness, denote the outcomes of lottery $L(E, V, S)$ as $x_1 = x_1(E, V, S)$ and $x_2 = x_2(E, V, S)$, and let $p = p(S) \in [0, 1]$ be the probability with which the downside $x_1 < x_2$ is realized. The binary lottery $L(E, V, S)$ can be parameterized as follows:

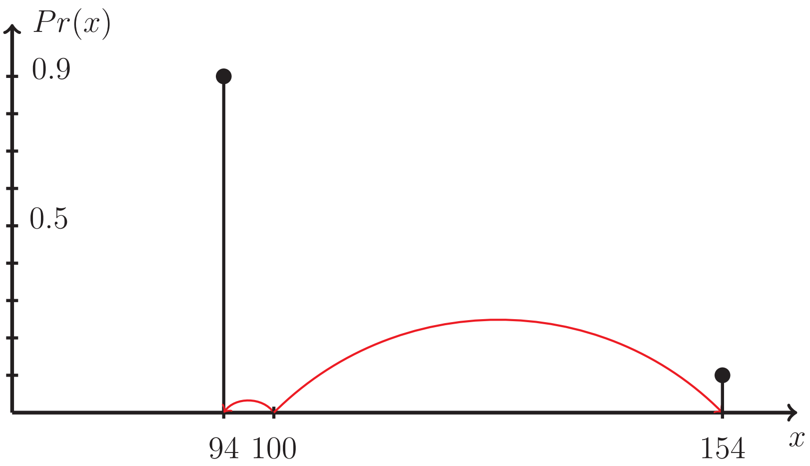
$$x_1 = E - \sqrt{\frac{V(1-p)}{p}},\ x_2 = E + \sqrt{\frac{Vp}{1-p}},\ \text{ and } p = \frac{1}{2} + \frac{S}{2\sqrt{4 + S^2}}. \tag{3.1}$$



$E = 100$, $V = 324$, and $S = 0$.



$E = 100$, $V = 324$, and $S = 0.6$.



$E = 100$, $V = 324$, and $S = 2.7$.

Figure 3.3: *The figure depicts the probability mass functions of the binary lotteries with an expected value $E = 100$, variance $V = 324$, and skewness $S \in \{0, 0.6, 2.7\}$. In addition, we depict in red for each of these three lotteries by how much the up- and downside, respectively, differ from the expected value. Holding expected value and variance fixed, we observe that the contrast on the upside monotonically increases in skewness, while the contrast on the downside monotonically decreases in skewness.*

A salient thinker chooses the binary lottery $L(E, V, S)$ over the safe option paying its expected value if and only if, relative to the reference point, the upside of this lottery is more salient than its downside. In this case, the reference point corresponds to the expected value. As illustrated in Figure 3.3, for any given expected value and variance, the upside of a binary lottery is salient if and only if its skewness exceeds a certain threshold. Precisely, the contrast in upside payoff and expected value monotonically increases in the skewness of the lottery, while the contrast on its downside monotonically decreases in skewness. Hence, by the contrast effect, there exists some threshold value $\hat{S} = \hat{S}(E, V)$ such that the binary lottery $L(E, V, S)$ is chosen over its expected value if and only if $S > \hat{S}$ (Dertwinkel-Kalt and Köster, 2020b, Proposition 3). Moreover, since we restrict attention to non-negative lotteries, the level effect implies that $\hat{S} > 0$; that is, all symmetric and, as a consequence, also all left-skewed lotteries are unattractive to a salient thinker. In sum, we obtain two qualitative predictions on simple choices: an *aversion toward symmetric risks* (via the level effect) and *skewness preferences* (via the contrast effect).

3.2.3 Skewness Preferences Are More Robust than the Aversion to Symmetric Risks

We are interested in how robust our qualitative predictions on "simple" choice problems are to increasing the number of available options. Suppose that the choice set is given by $\mathcal{C} = \{L(E_1, V_1, S_1), \ldots, L(E_n, V_n, S_n)\}$ for some $n \in \mathbb{N}_{\geq 3}$. To obtain clean and simple results, we impose additional structure on the available options. First, we assume that all options are binary lotteries with the same expected value, $E_i = E$ for some $E \in \mathbb{R}_{>0}$, and that one option pays this expected value with certainty. Second, we assume that all lotteries (except for the safe option) have the same absolute value of skewness; namely, $S_i \in \{-S, S\}$ for some $S \in \mathbb{R}$. Third, to keep the state space as small as possible, we assume that all lotteries (except for the safe option) are perfectly correlated.

Our first result is illustrated in Figure 3.4: a salient thinker might choose a symmetric binary lottery over its expected value. Intuitively, for a symmetric binary lottery the contrast on up- and downside is identical when compared to its expected value, so that already a relatively small change to the reference point — through adding a third option to the choice set — can be sufficient to render the upside of a symmetric lottery salient. For skewed risks, on the other hand, contrasts on up- and downside are highly asymmetric to start with, so that the reference point would have to change by a lot for salience theory to no longer predict skewness preferences.

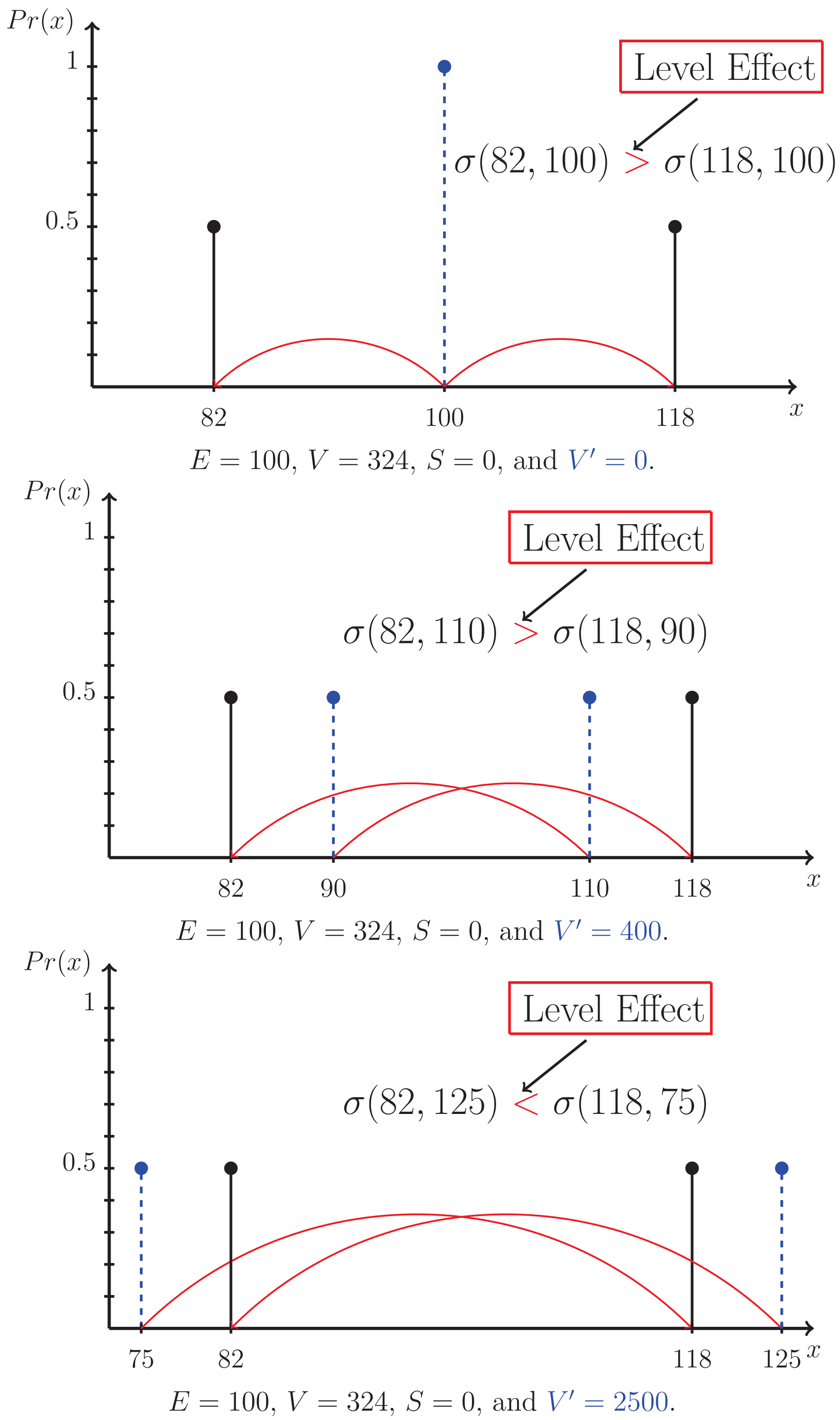


Figure 3.4: *This figure depicts the probability mass function of the symmetric binary lottery with an expected value $E = 100$, variance $V = 324$, and skewness $S = 0$. In blue we depict the reference point for this lottery, $L(E, V, S)$, assuming that the choice set is given by $\mathcal{C} = \{L(E, V, S), E, L(E, V', S)\}$ with $V' \in \{0, 400, 2500\}$. In red we further depict the contrast in outcomes in the two states of the world under the assumption that the two lotteries $L(E, V, S)$ and $L(E, V', S)$ are perfectly negatively correlated.*

Proposition 1 (Robust Skewness Preferences)**.** *Let $n \in \mathbb{N}_{\geq 3}$, and suppose that Assumption 1 holds. There exist threshold values S' and S'' with $S' < 0 < S''$, such that the following two statements hold:*

(a) If $\max\{S_1, \ldots, S_n\} > S''$, a salient thinker chooses one of the right-skewed lotteries.

(b) If $\max\{S_1, \ldots, S_n\} < S'$, a salient thinker chooses the safe option E.

Proposition 1 (for a proof see Appendix A) says that, even when making the problem comparatively more complex by enlarging the choice set, salience theory predicts a preference for sufficiently right-skewed risks and an aversion to sufficiently left-skewed risks. In combination with the example provided in Figure 3.4, we thus conclude that skewness preferences are predicted to be more robust to comparative complexity than the aversion toward symmetric risks.

The robustness of skewness preferences relies — to some extent at least — on the structure imposed on the available options. Without any distributional restrictions on the available lotteries, we could create a situation where a salient thinker avoids even highly right-skewed risks or chooses highly left-skewed ones.[3] But, due to the extreme outcomes of sufficiently skewed risks, the reference point would have to move a lot in the "right" direction for salience theory to no longer predict skewness preferences. In this sense, the statement that skewness preferences are more robust than the aversion toward symmetric risks is valid also in more general settings.

3.2.4 Discussion

What is complexity? Salience theory predicts an interaction between "comparative" complexity and how salient different outcomes are (as captured by skewness). To study more broadly the interaction between salience (or skewness) and how complex a problem is, we introduce in Section 3.3 an additional layer of "computational" complexity, which is outside of salience theory.

What is salience? When thinking about salience, we have in mind aspects of the available options that *stand out* in the choice context (e.g. the upside of a right-skewed risk). Other notions of salience (as in Chetty *et al.*, 2009) have a different flavour: people may neglect *non*-salient aspects of an option such as a tax hidden in the fine print, which would, however, not be particularly salient even when not hidden. What we have in mind instead is a kind of salience — like asymmetric contrasts in outcomes — that channels a

[3]Consider, for instance, the left-skewed binary lottery $L(E, V, S)$ with $E = 108$, $V = 324$, and $S = -2.7$, which yields 54 with 10% probability or 114 with 90% probability. Let $\mathcal{C} = \{L(E, V, S), E, E'\}$ with $E' = 0$. The reference point for this left-skewed lottery is thus the safe option that pays 54 with certainty, so that its upside of 114 is salient.

subject's attention to certain information — like a certain state of the world —, thereby reducing the "dimensionality" of the problem.

3.3 Experimental Design

Motivated by the predictions of salience theory, we develop in this section an experimental design to test for an interaction between the salience of outcomes and different layers of complexity. We will implement portfolio selection problems of how to allocate a given budget across two assets, which allow us to easily manipulate both complexity and salience.

3.3.1 Portfolio Selection Problems and Behavioral Benchmarks

Let X_1 and X_2 be non-negative, binary lotteries (henceforth: assets). A *portfolio* is defined as a convex combination, $X(\alpha) := \alpha X_1 + (1-\alpha)X_2$ for some $\alpha \in [0,1]$, of these two assets. We denote as $\mathcal{A} \subseteq [0,1]$ the set of admissible investments in asset X_1, and as $\mathcal{C}(\mathcal{A}) := \{X(\alpha) : \alpha \in \mathcal{A}\}$ the choice set. A *portfolio selection problem* is given by a tuple $(\mathcal{A}, \mathcal{C}) = (\mathcal{A}, \mathcal{C}(\mathcal{A}))$. We say that two portfolio selection problems $(\mathcal{A}, \mathcal{C})$ and $(\mathcal{A}', \mathcal{C}')$ are *essentially equivalent* if $\mathcal{C} = \mathcal{C}'$ and $\mathcal{A} \neq \mathcal{A}'$.

Layers of complexity. First, motivated by evidence on people struggling with the reduction of compound lotteries (e.g., Kahneman and Tversky, 1979; Wilcox, 1993; Camerer and Ho, 1994), we think of a problem's *computational complexity* as the difficulty of determining the consequences of an option.[4] This is illustrated in Figure 3.1 and Figure 3.2: The two problems presented in the left panels of the respective figures are equivalent in terms of the monetary consequences induced by the two available portfolios. But, while in the left panel of Figure 3.2 the outcomes of the two portfolios are readily observable, in Figure 3.1 these outcomes have to be computed.

Definition 3 (Computational Complexity)**.** *A portfolio selection problem* $(\mathcal{A}, \mathcal{C})$ *is said to be* computationally complex *if and only if* $\mathcal{A} \neq \{0,1\}$*.*

Second, building on the vast literature on choice overload (see Chernev *et al.*, 2015, for a meta-study), we identify a problem's *comparative complexity* with the number of options to be evaluated and compared (e.g., compare the right panels of Figures 3.1 and 3.2 for an illustration).

[4] As demonstrated by Halevy (2007) and more recently by Enke and Graeber (2020), subjects choose between compound lotteries in a similar way as they choose between ambiguous lotteries, which suggests that compounding of lotteries might do more than simply complicating the assessment of final outcomes. To abstract from attitudes to ambiguity, we use portfolios with *deterministic* shares invested in two assets instead of compound lotteries.

Definition 4 (Comparative Complexity)**.** *A portfolio selection problem* $(\mathcal{A}, \mathcal{C})$ *is said to be* comparatively complex *if and only if the choice set* $\mathcal{C}$ *includes more than two options.*

A problem that is neither computationally nor comparatively complex is said to be *simple*. As we only consider portfolios given by the convex combination of just two assets, a (trivial) observation is that comparative complexity implies computational complexity.[5] We, thus, compare simple to computationally and to computationally and comparatively complex problems.

Behavioral benchmarks. We evaluate behavior relative to the benchmarks of *diversification* and *naive diversification*. The idea of diversification goes back to Markowitz (1952): a portfolio is said to be *diversified* if, for given expected value, the variance in portfolio outcomes is minimized. Whenever a portfolio does not minimize the variance in outcomes (for a given expected value), we will say that it is *under-diversified*. Markowitz (1952) also discusses a second, in his words "wrong" type of diversification: in the literal sense of the word, diversification refers to an equal allocation of resources across different investments. Along these lines, we will say that a portfolio is *naively diversified* if it invests equal shares in the two underlying assets.

Why portfolio selection problems? We are interested in the interaction between complexity and salience in shaping the way in which people reach a decision. While in general choice complexity can manifest itself in various ways — some of which are non-trivial and economically meaningful and some of which are not —, portfolio selection problems allow us to study the effects of two uncontroversial layers of complexity that have been discussed in the literature and can be precisely defined. These two layers of complexity refer to the two different components, $\mathcal{A}$ and $\mathcal{C}$, of a portfolio selection problem. Moreover, following Bordalo *et al.* (2012), we can easily manipulate the salience of outcomes through varying the skewness of the underlying assets. Finally, we have two clear behavioral benchmarks relative to which we can evaluate behavior.

3.3.2 How to Identify Skewness Preferences from the Behavioral Benchmarks?

Our experimental design should allow us to identify skewness preferences from the benchmarks of actual and naive diversification. For that, we use a class of binary lotteries that was first introduced by Mao (1970) and later formalized by Ebert and Wiesen (2011, Definition 2).

[5] With more than two underlying assets, we could create problems that are comparatively, but not computationally complex (e.g. with three assets and three portfolios each of which invests everything in one of the assets).

Definition 5 (Mao Pair)**.** *Let $S \in \mathbb{R}_{>0}$. The binary lotteries $L(E, V, S)$ and $L(E, V, -S)$ with a correlation coefficient $\rho = \rho\Big(L(E, V, -S), L(E, V, S)\Big)$ denote a* Mao pair.

The two lotteries of a Mao pair have the exact same expected value and variance, but the opposite skewness, which makes these lotteries ideal for eliciting preferences over skewed risks. Indeed, when assuming that the two lotteries are perfectly negatively correlated (i.e. $\rho = -1$), also the set of portfolios that can be derived from Mao pairs has useful statistical properties.

Lemma 1. *Consider a portfolio selection problem with the assets $X_1 = L(E, V, S)$ and $X_2 = L(E, V, -S)$ that form a Mao pair, and assume that the two assets are perfectly negatively correlated (i.e. $\rho = -1$).*

(a) For any $\alpha \in [0, 1]$, we have $X(\alpha) = L(E, (2\alpha - 1)^2 V, \mathrm{sgn}(2\alpha - 1)S)$.

(b) Actual and naive diversification coincide: $\mathrm{Var}(X(\alpha)) = 0$ *if and only if $\alpha = \frac{1}{2}$.*

(c) For any $\alpha > 1/2$, the two portfolios $X(\alpha)$ and $X(1 - \alpha)$ constitute a Mao pair.

First, assuming a perfectly negative correlation, any portfolio consisting of Mao lotteries is again a binary lottery. Second, naive and actual diversification coincide, and yield the expected value of the assets with certainty. Third, the two portfolios $X(\alpha)$ and $X(1 - \alpha)$ constitute a Mao pair again. By comparing choices from the choice set $\{X(\alpha), X(\frac{1}{2})\} = \{L(E, (2\alpha - 1)^2 V, S), E\}$ to choices from the set $\{X(1 - \alpha), X(\frac{1}{2})\} = \{L(E, (2\alpha - 1)^2 V, -S), E\}$, we can thus identify skewness preferences from the benchmarks of actual and naive diversification, while properly controlling for the other moments (expected value and variance) of the portfolios.

3.3.3 How to Identify an Aversion to Symmetric Risks from Naive Diversification?

The goal of the experimental design is to identify an interaction between complexity and the salience of outcomes, along the lines suggested by salience theory of choice under risk. Hence, we are interested in whether the aversion to symmetric risks that salience theory predicts for simple, but not comparatively complex problems is robust to different layers of complexity.

To identify an aversion toward symmetric risks from the benchmark of naive diversification — which seems to be empirically relevant in this case —, we build on the experimental design proposed in Eyster and Weizsäcker (2016). To guide our analysis, we first uncover the mathematical structure underlying the specific portfolio selection problems that Eyster and Weizsäcker (2016) implemented. The following definition can be understood as a generalization of their paired portfolio selection problems (henceforth: *twin problems*), which captures the qualitative features of the numerical examples that they used in their experiment.[6]

Definition 6 (Twin Problems)**.** *Let* $\gamma \in [0,3]$. *Consider the perfectly negatively correlated assets* $X_1 = L(E,V,0)$ *and* $X_2 = L(E,\gamma^2 V,0)$, *and denote* $X(\alpha,\gamma) := \alpha X_1 + (1-\alpha)X_2$.

A twin problem *consists of the following two portfolio selection problems:*

(P.1) $\mathcal{C} = \{X(\alpha,1) : \alpha \in \mathcal{A}\}$ *and* $\mathcal{A} = [0,1]$.

(P.2) $\mathcal{C} = \{X(\alpha,\gamma) : \alpha \in \mathcal{A}\}$ *and* $\mathcal{A} = \left[\max\left\{0, \frac{\gamma-1}{\gamma+1}\right\}, 1\right]$ *for some* $\gamma \neq 1$.

Analogously, a binary twin problem *consists of the two portfolio selection problems:*

(P.1) $\mathcal{C} = \{X(\alpha,1) : \alpha \in \mathcal{A}\}$ *and* $\mathcal{A} = \left\{\frac{3-\gamma}{4}, \frac{1}{2}\right\}$.

(P.2) $\mathcal{C} = \{X(\alpha,\gamma) : \alpha \in \mathcal{A}\}$ *and* $\mathcal{A} = \left\{\frac{1}{2}, \frac{\gamma}{1+\gamma}\right\}$ *for some* $\gamma \neq 1$.

The main feature of a twin problem — in the spirit of Eyster and Weizsäcker (2016) — is that the set of available portfolios in (P.1) and (P.2) is exactly the same, but that the diversified portfolio coincides with the naively diversified portfolio only in (P.1). This is illustrated in Figure 3.5 for the case of a binary twin problem: Portfolio 1 in the left panel implements the binary lottery that yields with equal probability either 96 or 72, which is exactly the same distribution over outcomes that Portfolio 2 in the right panel implements. Portfolio 2 in the left panel as well as Portfolio 1 in the right panel yield the assets' expected value of 84 with certainty. Hence, by diversifying naively, the subject chooses the portfolio with less variance in outcomes in the left panel, but the portfolio with more variance in outcomes in the right panel.

Portfolio 1: 67% in X_1 and 33% in X_2.
Portfolio 2: 50% in X_1 and 50% in X_2.

	50%	50%
Asset X_1	120	48
Asset X_2	48	120

Portfolio 1 Portfolio 2

Portfolio 1: 25% in Y_1 and 75% in Y_2.
Portfolio 2: 50% in Y_1 and 50% in Y_2.

	50%	50%
Asset Y_1	120	48
Asset Y_2	72	96

Portfolio 1 Portfolio 2

Figure 3.5: *The binary twin problem with* $E = 84$, $V = 1296$, *and* $\gamma = \frac{1}{3}$.

The following lemma summarizes these observations.

[6] To be precise, Eyster and Weizsäcker (2016) implement joint distributions with four states of the world, and they have eight pairs of portfolio selection problems where only six of these pairs are covered by our definition.

Lemma 2. *Consider a (binary) twin problem. Then, the following two statements hold:*

(a) (P.1) and (P.2) are essentially equivalent portfolio selection problems.

(b) The diversified portfolio coincides with the naively diversified portfolio in (P.1). In (P.2), however, the diversified portfolio is given by $\alpha = \frac{\gamma}{1+\gamma}$ and, therefore, differs from the naively diversified portfolio.

The main difference compared to the setup in Eyster and Weizsäcker (2016) is the introduction of *binary* twin problems, which allow us to identify the effect of computational complexity on portfolio selection with symmetric assets. In addition, as we discuss in Section 3.3.5, the binary twin problems also allow us to test salience theory against naive decision rules such as the $1/N$-heuristic by Benartzi and Thaler (2001).

3.3.4 Summary of the Design: Relation Between Mao Pairs and Twin Problems

We conclude this section by deriving the connection between the portfolio selection problems based on Mao pairs and the twin problems in the spirit of Eyster and Weizsäcker (2016). First, to identify skewness preferences from our behavioral benchmarks, we need to control for the other moments (i.e., expected value and variance) of the available portfolios, which can be achieved by using Mao pairs. Second, to identify the interaction between skewness and complexity, we need to compare problems with skewed assets to problems with symmetric assets that are otherwise identical. A straightforward way to achieve exactly this is by taking the limit of a Mao pair for the skewness of its lotteries approaching zero, which gives exactly (P.1) of a twin problem.

Remark 2. *The assets in (P.1) of a (binary) twin problem constitute the limit of a Mao pair for $S \searrow 0$.*

The reason for additionally implementing (P.2) of a twin problem is twofold: First, in (P.1) of a twin problem, the diversified and the naively diversified portfolio coincide. This makes it impossible to identify an aversion toward symmetric risks from the benchmark of naive diversification. Second, in (P.1) of a twin problem, the two assets, X_1 and X_2, have the exact same distribution, which could have an effect on behavior. Since we want to study portfolio selection problems with assets that constitute symmetric binary lotteries (i.e. the skewness is fixed to zero), we are left with varying either the assets' expected value or their variance, both of which would be a change beyond the variation in skewness relative to the corresponding Mao pair. Following Eyster and Weizsäcker (2016), we vary the assets' variance, as this allows for a clean test of naive diversification. Figure 3.6 summarizes the idea of the experimental design.

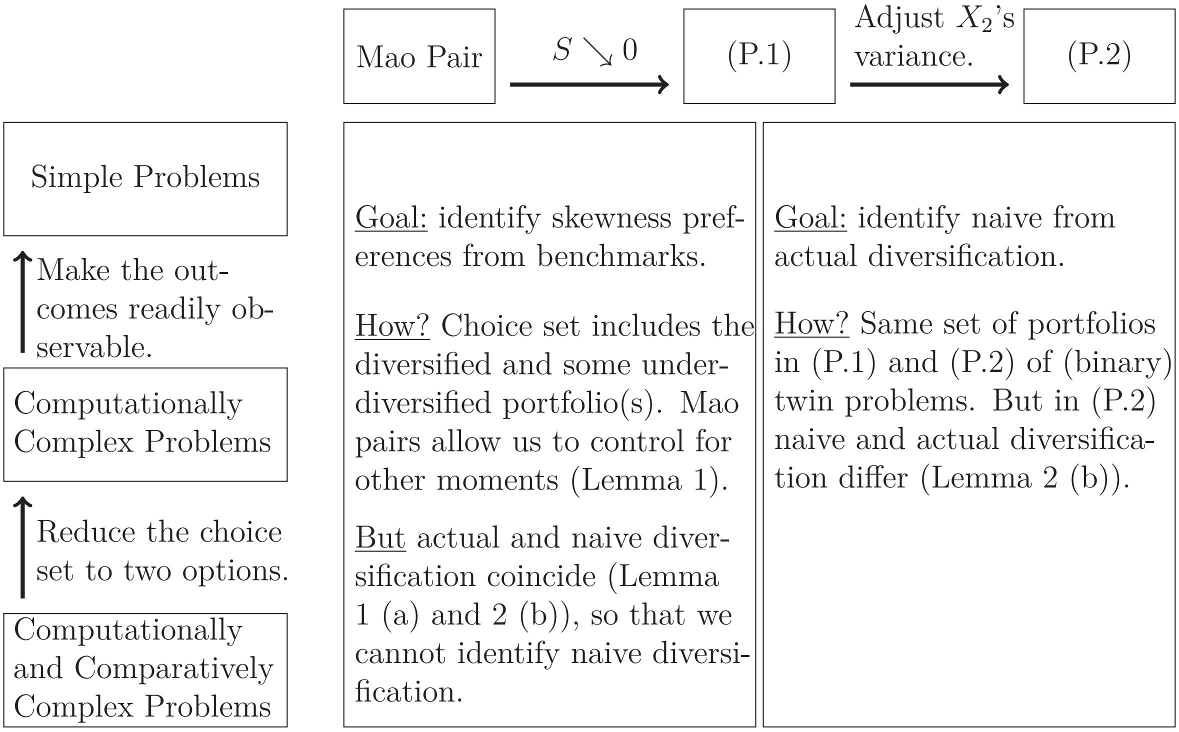


Figure 3.6: *Summary of the experimental design.*

3.3.5 Predictions of (Behavioral) Approaches to Choice under Risk

Expected Utility Theory. Assuming a concave utility function, expected utility theory (henceforth: EUT) predicts that subjects always choose the diversified portfolio, which yields the expected value of the under-diversified portfolio(s) with certainty. Allowing for a general utility function, EUT makes the clear-cut prediction that subjects choose the exact same portfolio in (P.1) and (P.2) of any (binary) twin problem, because the problems are essentially equivalent.[7]

1/N-Heuristic. According to the 1/N-heuristic (Benartzi and Thaler, 2001), subjects invest the exact same amount in both underlying assets; that is, they choose the naively diversified portfolio. For what we call simple problems, however, the 1/N-heuristic does not make a prediction.

Salience Theory of Choice under Risk. Building on the analysis in Section 3.2, we derive in Appendix C the precise predictions of salience theory for our experimental setting. For the problems with skewed assets that form a Mao pair, irrespective of the complexity, salience theory predicts that subjects choose an under-diversified portfolio if

[7] In Appendix B, we derive slightly more general results on the behavior of an expected utility maximizer.

and only if it is (i) right-skewed and (ii) the skewness of the underlying Mao lotteries exceeds a certain threshold. For simple and computationally complex twin problems, salience theory predicts that subjects always choose the diversified portfolio. For computationally and comparatively complex twin problems, on the other hand, salience theory predicts that behavior varies between (P.1) and (P.2): subjects should choose the diversified portfolio in (P.1), but not necessarily in (P.2). While some of the predictions derived in Section 3.2 rely on the additional structure of Assumption 1, the precise predictions for our experimental design hold for *any* salience function.

Other Behavioral Models. The most prominent behavioral model of choice under risk is still (cumulative) prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). In the context of our experiment, (cumulative) prospect theory (CPT) can rationalize similar behavior as salience theory (see Appendix D for a formal analysis). We briefly discuss in Appendix H how to distinguish salience theory from CPT, which is not what the experiment in this paper was designed for. At least for the problems with a binary choice set, also (generalized) regret theory (Loomes and Sugden, 1982, 1987) can rationalize the same behavior as predicted by salience theory. But, in contrast to salience theory and CPT, it cannot be easily extended to larger choice sets, and has, therefore, nothing to say on the role of comparative complexity.

3.4 Implementation of the Experiment

The design of the experiment — including the fully specified salience model — and its implementation were pre-registered in the AEA RCT registry as trial AEARCTR-0005311.[8]

We use a between-subjects design with multiple decisions per subject and two randomly assigned treatments: In the treatment *Skewed*, subjects have to choose between portfolios that consist of Mao lotteries (as illustrated in Table 3.1). In the treatment *Symmetric*, subjects have to choose between the portfolios available in (binary) twin problems (as illustrated in Table 3.2). To identify skewness preferences — as predicted by salience theory — from our behavioral benchmarks, we focus in the treatment *Skewed* on perfectly negatively correlated Mao lotteries. But, as a robustness check, we implement all Mao pairs illustrated in Table 3.1 also under the maximal positive correlation (see Appendix C.4 for the corresponding salience predictions). We use in both treatments only assets with non-negative outcomes to ensure incentive compatibility.[9]

Each treatment consists of three (main) blocks: (1) subjects make 12 choices between exactly two portfolios; (2) subjects make 12 choices from a larger set of portfolios, includ-

[8]The pre-registration can be found at `https://doi.org/10.1257/rct.5311-1.0`.

[9]As we show in Appendix G, the specific Mao pairs that we use for our experiment are not chosen arbitrarily, but can be deduced from simple, experimentally convenient principles.

Left-skewed Asset	Right-skewed Asset	E	V	S	$\mathcal{A}$
(120, 90%; 0, 10%)	(96, 90%; 216, 10%)	108	1296	2.7	$\{\alpha, 0.50\}$ or $[0, 1]$
(135, 64%; 60, 36%)	(81, 64%; 156, 36%)	108	1296	0.6	$\{\alpha, 0.50\}$ or $[0, 1]$
(40, 90%; 0, 10%)	(32, 90%; 72, 10%)	36	144	2.7	$\{\alpha, 0.50\}$ or $[0, 1]$
(45, 64%; 20, 36%)	(27, 64%; 52, 36%)	36	144	0.6	$\{\alpha, 0.50\}$ or $[0, 1]$
(80, 90%; 0, 10%)	(64, 90%; 144, 10%)	72	576	2.7	$\{\alpha, 0.50\}$ or $[0, 1]$
(90, 64%; 40, 36%)	(54, 64%; 104, 36%)	72	576	0.6	$\{\alpha, 0.50\}$ or $[0, 1]$

Table 3.1: *Mao pairs with an expected value E, variance V, and skewness S used in the treatment Skewed, where the share $\alpha \in \{0.10, 0.25, 0.75, 0.90\}$ that the under-diversified portfolio invests in the right-skewed asset is randomized at the subject-Mao-pair level. To be precise, subjects did not face a continuous, but a discretized choice set. We use $\mathcal{A} = [0, 1]$ as a shorthand for $\mathcal{A} = \{0, 0.01, ..., 0.99, 1\}$. Each Mao pair is implemented under the perfectly negative and the maximal positive correlation.*

ing those in the first block; and (3) subjects make the exact same choices as in the first block, with the only exception that the outcomes of the two portfolios are made readily observable to the subjects.[10] The problems in Block 1 are only computationally complex, the problems in Block 2 are computationally and comparatively complex, and the problems in Block 3 are neither computationally nor comparatively complex. When computing the final outcomes of the portfolios in Block 3, we round all numbers to the next integer. In Block 1 of the treatment *Skewed*, we randomize the share $\alpha \in \{0.10, 0.25, 0.75, 0.90\}$ that the under-diversified portfolio invests in the right-skewed asset at the subject-Mao-pair level. We further randomize in Block 1, at the subject level, the order in which the diversified and the under-diversified portfolio are displayed (see Figure 3.20 in Appendix G for a screenshot). When making the outcomes of the portfolios readily observable, we implement not only the same $\alpha \in \{0.10, 0.25, 0.75, 0.90\}$ as in Block 1, but also use the same order of diversified and under-diversified portfolio. Finally, we randomize the order of blocks and the order of tasks within each block at the subject level.

At the beginning of Block 1 and Block 2, subjects have to answer control questions to check for comprehension of the tasks. After the last decision of the last block (which could be either Block 1 or Block 2 or Block 3), subjects are confronted with a surprise memory task, in which they have to recall the joint distribution of the assets presented on the preceding screen (see Figure 3.23 in Appendix G for a screenshot). Subsequently,

[10]Since (P.1) and (P.2) of a twin problem are essentially equivalent portfolio selection problems, in the treatment *Symmetric* the third block includes only 6 decisions.

X_1	X_2 in (P.2)	E	V	γ	$\mathcal{A}$ in (P.1)	$\mathcal{A}$ in (P.2)
(96; 48)	(72; 72)	72	576	0	$\{0.75, 0.50\}$ or $[0, 1]$	$\{0.50, 0\}$ or $[0, 1]$
(79; 51)	(65; 65)	65	196	0	$\{0.75, 0.50\}$ or $[0, 1]$	$\{0.50, 0\}$ or $[0, 1]$
(120; 48)	(72; 96)	84	1296	0.33	$\{0.67, 0.50\}$ or $[0, 1]$	$\{0.50, 0.25\}$ or $[0, 1]$
(82; 58)	(66; 74)	70	144	0.33	$\{0.67, 0.50\}$ or $[0, 1]$	$\{0.50, 0.25\}$ or $[0, 1]$
(128; 72)	(16; 184)	100	784	3	$\{0.50, 0\}$ or $[0, 1]$	$\{0.75, 0.50\}$ or $[0.5, 1]$
(71; 49)	(27; 93)	60	121	3	$\{0.50, 0\}$ or $[0, 1]$	$\{0.75, 0.50\}$ or $[0.5, 1]$

Table 3.2: *Twin problems used in the treatment Symmetric, where (x; y) := (x, 50%; y, 50%). The expected value of the assets is denoted by E and their variance is denoted by V. The parameter γ scales the difference in variance between the two assets (see Definition 6). Again, subjects did not face a continuous choice set, but a discretized one with 1 p.p. steps from the lower to the upper bound. As before, $\mathcal{A} = [x, 1]$ serves as a shorthand for the discretized set $\mathcal{A} = \{x, x + 0.01, ..., 0.99, 1\}$.*

subjects answer five financial literacy questions (as in van Rooij, Lusardi and Alessie, 2011; Lusardi and Mitchell, 2011) and five questions of a modified cognitive reflection test (CRT; closely aligned to Primi, Morsanyi, Chiesi, Donati and Hamilton, 2016), and they solve five math problems similar to computing the outcomes of a portfolio. Details on the additional questions can be found in Appendix G. Finally, we ask for demographics (incl. age, gender, field of study).

During the experiment, the outcomes of the assets are presented in units of an experimental currency (ECU). At the end of the experiment, for each subject, one of the 36 decisions in *Skewed* or 30 decisions in *Symmetric*, respectively, is randomly drawn by the computer to be payoff-relevant. Also the outcome of the payoff-relevant choice is randomly determined by the computer. This implies, in particular, that subjects do not receive any feedback on their earnings before they have done the memory task and answered the financial literacy, modified CRT, and maths questions. In addition, subjects are rewarded for their performance in the memory task (1 ECU per correctly remembered outcome) as well as for correct answers to the control, financial literacy, modified CRT, and math questions (1 ECU per correctly answered question). Experimental earnings are converted at a ratio of 4 ECU : 1 Euro. Additionally, the subjects receive a show-up fee of 4 Euro for their willingness to participate in the experiment.

We conducted 10 sessions with a total number of $n = 296$ subjects (146 subjects in *Skewed* and 150 subjects in *Symmetric*). The sessions took place in January 2020 at the experimental laboratory of the University of Cologne. The experiment lasted for around one hour and average earnings were slightly higher than 25 Euro. Subjects were invited to the lab via ORSEE (Greiner, 2015) and the experiment was implemented with the software z-Tree (Fischbacher, 2007).

3.5 Main Experimental Result

Main result: *skewness preferences are robust to both, computational and comparative complexity (see Figure 3.7), while the aversion toward symmetric risks revealed in simple problems is not robust to either layer of complexity (see Figure 3.8).*

Portfolio selection problems with skewed assets. If the problem is simple — i.e. if the outcomes of the just two portfolios are readily observable —, subjects reveal an aversion toward left-skewed and a preference for sufficiently right-skewed risks. Both findings are consistent with salience theory, but the latter is inconsistent with our benchmarks of actual and naive diversification. As predicted by salience theory, we observe basically the same behavioral patterns for the essentially equivalent, but computationally complex portfolio selection problems, where the final outcomes of the two portfolios have to be computed. Finally, also for the computationally and comparatively complex portfolio selection problems with a larger choice set, we observe a large and significant increase in the share of right-skewed portfolios chosen, when increasing the skewness of the underlying Mao lotteries. This is again consistent with salience theory, but inconsistent with our behavioral benchmarks of actual and naive diversification. In sum, revealed skewness preferences are not only robust regarding computational complexity, but also regarding comparative complexity. These findings are summarized in Figure 3.7.

As illustrated in Figure 3.26 in Appendix H.1, revealed skewness preferences are similar for low-variance (i.e. $\alpha \in \{0.25, 0.75\}$) and high-variance portfolios (i.e. $\alpha \in \{0.1, 0.9\}$). By splitting the sample in this way, we can control for the variance in portfolio returns and identify skewness preferences. Moreover, subjects with above- and below-median cognitive skills, as measured by the number of correct answers to the financial literacy, modified CRT, and math questions reveal similar skewness preferences (see Figure 3.27 in Appendix H.1). Finally, we observe similar, but more nuanced, patterns for positively correlated Mao pairs (see Figure 3.24).

Portfolio selection problems with symmetric assets. In simple problems — with the outcomes of the two portfolios being readily observable —, a large majority of subjects chooses the diversified portfolio, thereby revealing risk-averse behavior. The share of subjects choosing the diversified portfolio drops, however, from around 75% in simple problems to slightly more than 40% in (P.2) of computationally complex twin problems

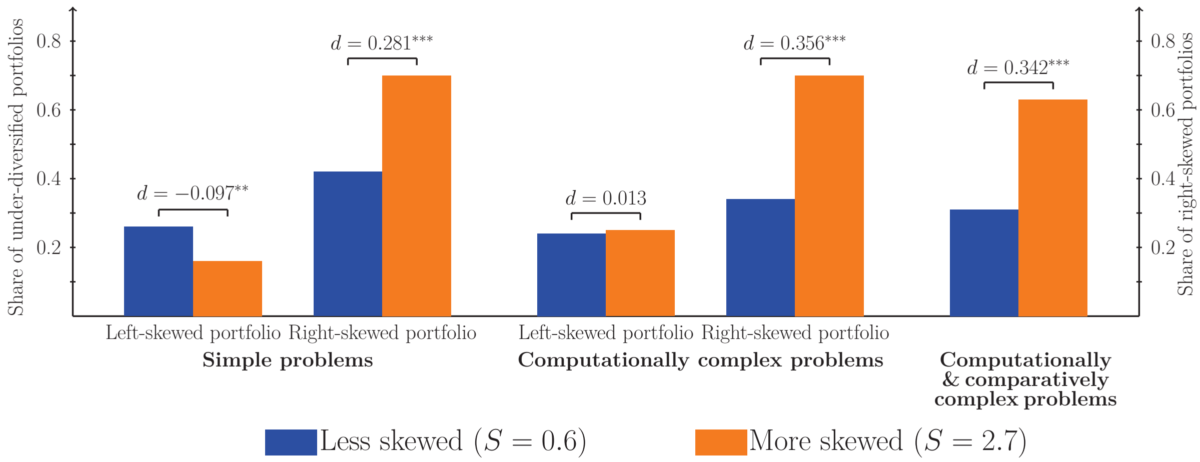


Figure 3.7: *This figure illustrates our main experimental result for the treatment Skewed. We depict the share of under-diversified portfolios chosen in simple and computationally complex problems as well as the share of right-skewed portfolios chosen in computationally and comparatively complex problems. The data for problems with a binary choice set is presented separately for right- and left-skewed under-diversified portfolios. We further split the sample by less and more skewed Mao pairs. We report the results of t-tests with standard errors being clustered at the subject level. Significance level: **: 5%, ***: 1%.*

and to less than 15% in (P.2) of computationally and comparatively complex twin problems. While the latter finding could, in principle, be rationalized via salience theory,[11] the former cannot, since computational complexity is outside of the model. The same is true for expected utility theory and any other consequentialist theory of choice under risk that models decisions based on final outcomes. In sum, the revealed aversion toward symmetric risks in simple problems is not robust to increasing the problem's comparative and/or computational complexity. These findings are summarized in Figure 3.8.

While in (P.2) of the (binary) twin problems subjects behave, on average, as if they were risk-seeking, they behave as if being risk averse in (P.1): in (P.1), on average, 75% of the subjects choose the diversified portfolio, which is consistent with the revealed aversion toward symmetric risks in simple problems. The crucial difference between (P.1) and (P.2) is that in (P.1) the diversified portfolio coincides with the naively diversified portfolio, which seems to drive the revealed risk aversion in (P.1). Taken together, our results suggest that an increase in the comparative and/or computational complexity of a portfolio selection problem with symmetric assets results in subjects switching from the diversified to the naively diversified portfolio.

The findings presented in Figure 3.8 are robust to splitting the sample by cognitive skills (Figure 3.28 in Appendix H.2). In addition, we observe the same patterns of naive diversification for all different parameterizations (Figure 3.29 in Appendix H.2). This

[11]As illustrated in Figure 3.14 in Appendix E, behavior in the computationally and comparatively complex twin problems are mostly inconsistent with salience theory (Proposition 5 in Appendix C).

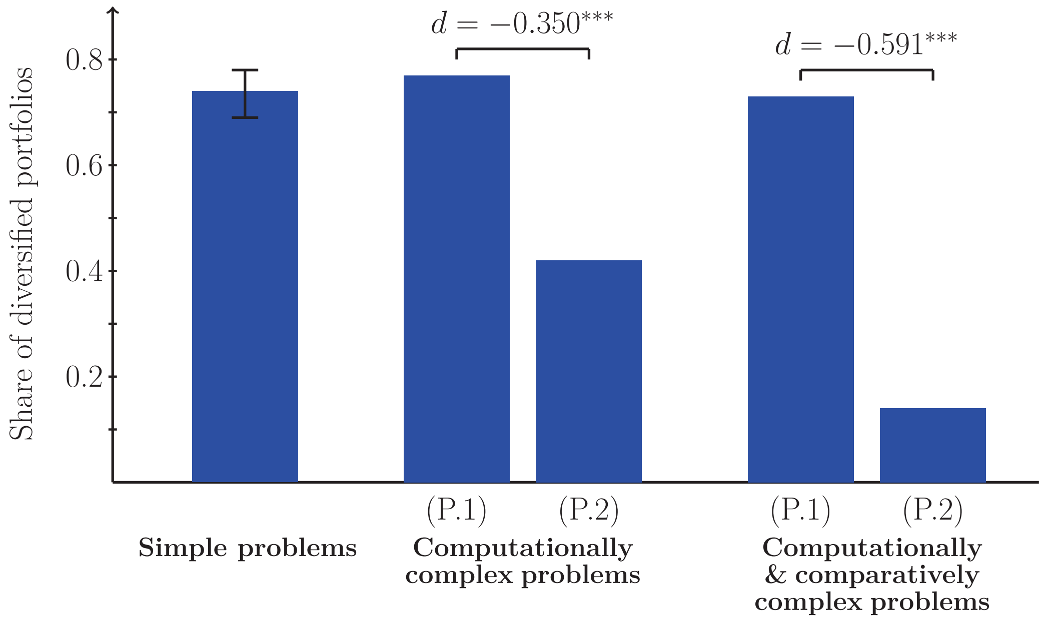


Figure 3.8: *This figure illustrates our main experimental result for the treatment Symmetric. We depict the share of diversified portfolios chosen in differently complex problems. The data for comparatively and/or computationally complex problems is presented separately for (P.1) and (P.2). For simple problems, we report the 95%-confidence interval with standard errors clustered at the subject level. For complex problems, we present the results of t-tests with clustered standard errors. Significance level: ***: 1%.*

last finding is particularly striking, because for the twin problems that are parameterized by $\gamma = 0$, choosing the diversified portfolio in (P.2) does not require any computations: for $\gamma = 0$, asset X_2 in (P.2) yields the portfolios' expected value with certainty, but still subjects choose to diversify naively.

Summary. We document an interaction between the skewness of the available portfolios and the complexity of the problem: while revealed preferences over skewed portfolios are robust to both computational and comparative complexity, subjects choosing between symmetric portfolios switch from diversification in simple problems to naive diversification in complex ones. Salience theory can explain large parts, but not all of the experimental results. In particular, salience theory performs well in predicting behavior if the underlying assets are skewed; that is, if there is a salient cue attracting attention and guiding choices. While salience theory can also rationalize naive diversification in the comparatively complex twin problems, our experimental results clearly suggest that salience theory is not the "right" model for thinking about naive diversification. As summarized in Table 3.3, expected utility theory and naive decision rules such as the $1/N$-heuristic perform even worse in rationalizing our experimental findings.

	Skewed Assets	Symmetric Assets
Simple Problems	EUT: ✗ ST: ✓ $1/N$: –	EUT: ✓ ST: ✓ $1/N$: –
Compu. Complex Problems	EUT: ✗ ST: ✓ $1/N$: ✗	EUT: ✗ ST: ✗ $1/N$: ✓
Compa. Complex Problems	EUT: ✗ ST: ✓ $1/N$: ✗	EUT: ✗ ST: ✗/✓ $1/N$: ✓

Table 3.3: *This table summarizes whether the predictions of expected utility theory (EUT), salience theory (ST), and the 1/N-heuristic fit different parts of our main experimental result. The sign ✓ indicates that the prediction of a theory fits the data and the sign ✗ indicates that it does not. The sign ✗/✓ indicates that a prediction partially fits the data, and – indicates that a theory/decision rule does not make a prediction.*

3.6 On the Mechanism Underlying the Main Experimental Result

Next, we try to shed some light on the mechanism underlying the observed behavior.

3.6.1 Contrast Effects Shape Memory of and Arguably Attention Paid to Outcomes

Using the memory data elicited toward the end of the experiment, we can establish a more direct link between the asymmetry of contrasts in outcomes and the attention paid to different states of the world. Obviously, paying attention to an outcome is necessary to recall it later, which is a key feature of models on selective attention (e.g., Gagnon-Bartsch, Rabin and Schwartzstein, 2020). But it also seems sensible to assume that better recall of an outcome indicates that more attention was paid in the first place (see Hartzmark, Hirshman and Imas, 2020; Valletti and Veiga, 2020, for supporting evidence).

To pin down in how far (asymmetric) contrasts in outcomes affect memory, we introduce the notions of *relative variance* $\mathbb{E}[(X_1 - X_2)^2]$, and *relative skewness* $\frac{|\mathbb{E}[(X_1-X_2)^3]|}{\text{Var}(X_1-X_2)}$. The relative variance measures the size of contrasts in outcomes across the different states of the world. The relative skewness, on the other hand, measures the asymmetry of contrasts in outcomes across states.[12] We regress the percentage share of correctly recalled outcomes on the assets' relative variance and relative skewness as well as several controls using OLS (see Table 3.4 for the results).

Due to a coding error regarding one of the outcomes, the memory results presented in this section should be interpreted with a little caution: we did not force subjects to do

[12] The concept of relative skewness has been formally introduced in Dertwinkel-Kalt and Köster (2020b) and shown to be an important driver of choice under risk. In that paper, we have argued that much of the evidence that documents skewness-seeking choices is, in fact, evidence for subjects seeking relative skewness.

the memory task (see Figure 3.23 in Appendix G), and all missing answers were coded as zero, which is unfortunately also the lower outcome of the left-skewed lottery for the more skewed Mao pairs. To address this problem, we count a correct answer of "zero" if and only if a subject has entered non-zero answers for all the other outcomes, which indicates that the subject made these entries actively.[13] While this adjustment might even bias the results against our hypothesis that subjects better recall the outcomes of more skewed assets, we cannot conclusively rule out that we overestimate the number of correctly recalled outcomes for the more skewed Mao pairs.

	% of correctly recalled outcomes			
Constant	21.220***	57.200***	51.930***	54.270***
	(2.259)	(9.929)	(10.050)	(10.040)
Relative Variance	0.001	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Relative Skewness	0.133***	0.156***	0.158***	0.149***
	(0.042)	(0.042)	(0.042)	(0.041)
# States	-	-15.760***	-15.560***	-16.120***
	-	(4.239)	(4.200)	(4.202)
Above-Median Skills	-	-	8.837**	8.654**
	-	-	(3.471)	(3.434)
Decision Time for Memory Task	-	-	-	-0.127
	-	-	-	(0.114)
Decision Time for Recalled Decision	-	-	-	0.507***
	-	-	-	(0.167)
Adjusted R^2	0.044	0.084	0.101	0.123
# Subjects	296	296	296	296

Table 3.4: *This table presents the results of OLS regressions with the percentage share of correctly recalled outcomes as the dependent variable and relative variance and skewness of the assets as the main independent variables. We add controls for the number of states, a subject's skills, and the time a subject spent on the memory task and on the decision task that is to be recalled. Significance level: **: 5%, ***: 1%.*

Across all specifications, we find that a more asymmetric contrast in outcomes — as measured by a higher relative skewness — predicts the share of correctly recalled outcomes with a positive sign. The relative variance (i.e. the simple size of contrasts), on the other hand, has no predictive power. Precisely, increasing the relative skewness from zero to the 75th-percentile of the empirical distribution (conditional on the relative

[13]We adjusted the number of correct answers for 8 out of 49 subjects for whom this issue arises. Overall, 4 out of these 8 subjects "entered" a zero for all outcomes, which indicates that they completely skipped the task.

skewness being positive) increases the share of correctly recalled outcomes by 8.51 to 10.11 p.p., on average. This is roughly the same average performance difference as for subjects with below- and above-median skills. We further observe that, even when controlling for how much time a subject spends on the portfolio selection problem that she has to recall, relative skewness has predictive power. Put differently, asymmetric contrasts in outcomes improve the recall of the respective outcomes in a way that is not captured by the time spent on the decision screen. This last finding speaks to the idea that asymmetric contrasts indeed come to mind more easily or, in other words, are more salient.

The fact that the assets' relative skewness, but not their relative variance, is predictive of a subject's recall is particularly interesting in the light of models on deliberate or "rational" inattention (e.g., Sims, 2003; Gabaix, 2014; Caplin and Dean, 2015). While models of deliberate inattention typically assume that the attention devoted to a certain dimension of a problem (here, a state of the world) depends on the variance or the size of contrasts in outcomes, we find that it is rather the asymmetry in contrasts that shapes memory and arguably also attention.

3.6.2 Cognitive Skills Drive Naive Diversification, Not Skewness Preferences

Many studies (e.g. Graham, Harvey and Huang, 2009) have found that sophistication improves financial performance (for surveys of the literature see Hastings, Madrian and Skimmyhorn, 2013; Lusardi and Mitchell, 2014). In our experiment, however, low financial literacy and low cognitive skills are associated with under-diversification only in the case of symmetric assets, but not in the case of skewed assets.

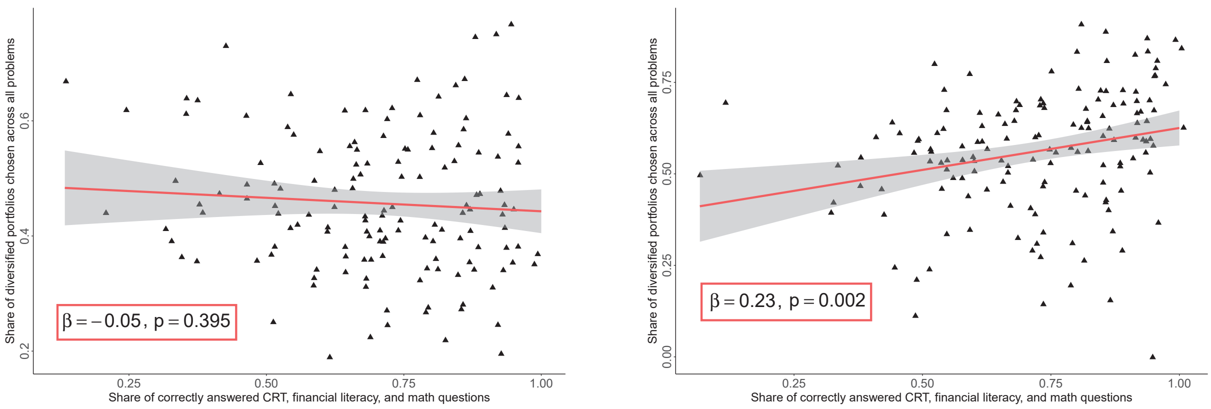


Figure 3.9: *This figure illustrates the correlation between the share of diversified portfolios chosen by a subject and his or her cognitive skills, as measured by the share of correct answers to the financial literacy, modified CRT, and math questions. In the left panel, we depict the results for the treatment Skewed, while in the right panel, we depict the results for the treatment Symmetric. The red lines depict the fitted values of an OLS regression with 95%-confidence intervals in grey. The black triangles depict the data.*

As depicted in the left panel of Figure 3.9, in the treatment *Skewed*, we find a zero correlation between a subject's cognitive skills — as measured by the share of correctly answered financial literacy, modified CRT, and math questions — and the share of diversified portfolios chosen by this subject. In the treatment *Symmetric*, we do observe a positive correlation between our measure of cognitive skills and the share of diversified portfolios chosen during the experiment (right panel of Figure 3.9).[14] But this correlation is mostly driven by the subjects' performance in the CRT and math questions (see Figure 3.35 in Appendix H.6).[15] In sum, we conclude that naive diversification, but not skewness-driven under-diversification, is sensitive to cognitive skills.

One interpretation of these results is that choices involving skewed risks are cognitively less demanding than choices involving symmetric risks. This interpretation is consistent with the idea that asymmetric contrasts in outcomes serve as a salient cue that helps subjects in making a decision. In Appendix H.4 we provide further evidence supporting this interpretation.

3.7 Salient Cues and Complexity: Theoretical Framework

A useful model should not only rationalize our main experimental result, but also account for the evidence presented in the preceding section: asymmetric contrasts in outcomes attract attention and drive skewness preferences, and choices between skewed risks are cognitively less demanding than choices between symmetric ones (see Appendix H for additional evidence).

Borrowing ideas from the literature on deliberate inattention (Gabaix, 2014, 2019), we assume that the process of selecting a portfolio can be divided into two steps. In a first step, the agent "decides" whether to react to the monetary stimuli by making an "active" decision. We think of this first step as the agent unconsciously screening the problem for a salient cue that would help her in making a decision. In the second step, the agent either makes an active choice — if she actually found some cue to base her decision on — or she relies on a default heuristic to reach a decision. This has a similar spirit as models of "rational" or deliberate inattention (e.g., Sims, 2003; Gabaix, 2014), according to which the agent thinks through the problem if and only if a deliberate choice results in a sufficiently large (expected) utility gain. The mechanism that we have in mind, however, differs from these models in that attention allocation is not based on a (pure) cost-benefit

[14] When neglecting the two outliers, who have a share of correctly answered questions well below 0.25, the regression coefficient increases to $\beta = 0.31$ (p-value < 0.001).

[15] In the treatment *Skewed*, all three measures of skills are equally unpredictive of the share of diversified portfolios chosen by a subject (see Figure 3.35 in Appendix H.6).

trade-off, but relies on the availability of a salient cue. In particular, we do not postulate that active choices have to be normatively more appealing than default-driven choices; that is, the perceived benefit of making an active choice must not be equal to the true benefit of doing so (see Handel and Schwartzstein, 2018, for a similar discussion).

Step 1: The role of salient cues in making a stimulus-driven choice. Our framework is motivated by the observations that asymmetric, not necessarily large, contrasts in outcomes (i) attract attention and (ii) seem to make the problem cognitively less demanding. Intuitively, we think of a mechanism along the following lines: an asymmetric distribution of contrasts in outcomes can be interpreted as a salient cue in as much as it guides attention to a certain state of the world, thereby reducing the "dimensionality" of the decision problem.

More formally, consider the portfolio selection problem $(\mathcal{A}, \mathcal{C})$. Define as $\kappa(\mathcal{A}, \mathcal{C})$ a measure of complexity, which is assumed to be zero for simple problems. As before, we denote the underlying assets as X_1 and X_2, with a joint cumulative distribution function F and the same expected value. We assume that the agent thinks through the extreme cases of either getting X_1 or X_2 — even if these extreme portfolios are not available —, and asks herself whether she likes one of the two assets more and, if so, by how much. In doing so, the decision-maker puts an excessive weight on asymmetrically large contrasts in outcomes, which is determined by a salience function $\hat{\sigma} : \mathbb{R}^2_{\geq 0} \to \mathbb{R}_{>0}$ that satisfies ordering, but not necessarily diminishing sensitivity. Since the level effect is more subtle than the contrast effect, it seems sensible to assume that the contrast effect plays an even more pronounced role in forming a "first impression" of a problem. We then assume that the agent makes a stimulus-driven choice if and only if

$$\int_{\mathbb{R}^2_{\geq 0}} (x_1 - x_2) \cdot \frac{\hat{\sigma}(x_1, x_2)}{\int_{\mathbb{R}^2_{\geq 0}} \hat{\sigma}(y_1, y_2)\, dF(y_1, y_2)}\, dF(x_1, x_2) \geq \kappa(\mathcal{A}, \mathcal{C});$$

that is, if and only if she perceives one of the two assets to be sufficiently more attractive in order to justify making an active decision. If we assume that the agent's first impression of the problem is purely shaped by the contrasts in outcomes and, in addition, impose the functional form $\hat{\sigma}(x, y) = (x - y)^2$, then the preceding condition simplifies to[16]

$$\frac{\left|\mathbb{E}\left[(X_1 - X_2)^3\right]\right|}{\mathrm{Var}(X_1 - X_2)} \geq \kappa(\mathcal{A}, \mathcal{C}), \tag{3.2}$$

where the left-hand side constitutes the relative skewness of the two assets.

[16]Let $\hat{\sigma}(x, y) = (x - y)^2$. Since $\mathbb{E}[X_1] = \mathbb{E}[X_2]$, the normalization factor, $\int_{\mathbb{R}^2_{\geq 0}} \hat{\sigma}(x_1, x_2)\, dF(x_1, x_2)$, simplifies to $\mathbb{E}[(X_1 - X_2)^2] = \mathbb{E}[(X_1 - X_2)^2] - \mathbb{E}[X_1 - X_2]^2 = \mathrm{Var}(X_1 - X_2)$, since $\mathbb{E}[X_1 - X_2]^2 = 0$ by assumption.

Step 2a: Default-driven choice. Following Gabaix (2014, 2019), we assume that an agent, who does not make a stimulus-driven choice, acts on a default perception of the problem. Without any information on the assets' joint distribution, with equal probability each asset is the better investment. It thus seems sensible to assume that agents, when not really thinking through the problem, revert to the $1/N$-heuristic as a default. A default heuristic of naive diversification is also similar, in spirit, to the "ignorance prior" that Enke and Graeber (2020) propose as a default in their model of cognitive uncertainty.[17] In a sense, our framework is reminiscent of Shafir *et al.*'s (1993) informal notion of reason-based choice: the agent looks for a "reason" — like a salient cue — to choose one of the available options, and if no convincing reason exists, she abstains from making a choice, which we capture through a default of naive diversification.

Step 2b: Stimulus-driven choice. If the agent makes a stimulus-driven choice, we assume that she behaves as if she was maximizing the salience-weighted utility introduced in Definition 2. This way our model can account for our main experimental result as well as the stylized facts presented in Section 3.6. This modeling choice further acknowledges that the salience of outcomes may also distort active choices, which is reminiscent of the idea that salient cues make people jump to conclusions (e.g. Tversky and Kahneman, 1974). In particular, by assuming that subjects make an active choice in simple problems, the model accounts for existing evidence on salience effects in choice under risk (e.g. Bordalo *et al.*, 2012; Dertwinkel-Kalt and Köster, 2020b).

Testing the framework with the available data. The framework presented in this section was not pre-registered, but using the data that we have, we can still test for its main implications: namely, subjects should be the more likely to behave as if maximizing a salience-weighted utility (i) the larger is the relative skewness of the underlying assets and (ii) the less complex is the portfolio selection problem. As we show in Appendix H.4, both predictions are consistent with the data. More specifically, we find that the effect of relative skewness operates through mitigating the impact of complexity on behavior: while for simple problems that are neither computationally nor comparatively complex the relative skewness has no significant effect on the average probability to behave as predicted by salience theory, for more complex problems subjects are, on average, more

[17] Notice that, while Gabaix (2014) and, in parts, also Enke and Graeber (2020) assume that the default applies to how inputs (e.g. probabilities or outcomes) are perceived, we assume that the default directly applies to the output (i.e. the choice or, more specifically, in the context of our experiment the share invested in each of the two assets). To put it differently, while Enke and Graeber (2020), for instance, assume that complexity affects how people perceive inputs like probabilities and, thereby, indirectly how they behave, we assume that complexity has a direct effect on the "rules" subjects use to come up with a decision. The idea that subjects sometimes base their choices in complex portfolio selection problems on simple rules is supported by experimental evidence (Halevy and Mayraz, 2020).

likely to behave in line with salience theory the more skewed are the assets in relative terms. This finding is exactly in line with the intuition underlying Eq. (3.2), and the corresponding distinction between stimulus-driven and default-driven choices.

3.8 Related Literature

Our paper builds on the growing literatures on the role of complexity and the implications of salience-driven attention for economic behavior. No previous paper, however, systematically studies our main question: what are the interactions between choice complexity and the availability of a salient cue? By studying behavior in portfolio selection problems, we further contribute to the (experimental) finance literature on portfolio under-diversification.

Complexity and economic choice. A growing literature studies the implications of complexity for economic choice. Abeler and Jäger (2015) and Breunig, Huck, Schmidt and Weizsacker (2019) investigate how people react to complex incentive schemes, and they find that complexity could lead to an adoption of simpler choice rules and, consequently, lower the responsiveness to incentives. Similarly, Banovetz and Oprea (2020) have shown that if the optimal choice procedure becomes more complex — e.g. more demanding in terms of the information subjects have to track and memorize — subjects turn to much simpler choice rules. Several studies find that the demand effects of shrouding surcharges are more pronounced for multiplicative rather than additive surcharges (see, e.g. Morwitz, Greenleaf and Johnson, 1998; Kalaycı and Serra-Garcia, 2016). This points to an interaction between computational complexity (which is arguably larger for multiplicative surcharges) and the salience of prices — a hypothesis that also Kalaycı and Serra-Garcia (2016) state and confirm. We formalize and systematically test this hypothesis in the context of choice under risk.

There is no universal definition of choice complexity. Oprea (2020) measures the complexity of a choice procedure by eliciting a subject's willingness-to-pay to avoid implementing it again in the future. Abeler, Huffman and Raymond (2020) argue that an incentive scheme is the more complex the less subjects react to the provided incentives. We, on the contrary, focus on the notions of computational and comparative complexity in arguably uncontroversial cases. Our notion of computational complexity is motivated by the observation that people often struggle with the reduction of compound lotteries (e.g., Kahneman and Tversky, 1979; Wilcox, 1993; Camerer and Ho, 1994). But, to isolate the effect of computational complexity and to abstract from attitudes to ambiguity,[18]

[18] As demonstrated by Halevy (2007) and more recently also by Enke and Graeber (2020), subjects choose between compound lotteries in a similar way as they choose between ambiguous lotteries, which suggests that compounding of lotteries might do more than simply complicating the assessment of final outcomes.

we use portfolios with *deterministic* shares invested in two lotteries instead of compound lotteries. Our notion of comparative complexity builds on the large literature on choice overload (for a meta-study see Chernev *et al.*, 2015), which documents that an increase in the number of available options can result in subjects switching to simpler choice rules (e.g. to choosing easily justifiable options). There are certainly other relevant notions of complexity, such as the number of distinct dimensions that an option has (see Herman and Bahrick, 1966, for an early treatment), which we do not focus on in this study.[19] Instead, using the precisely defined notions of computational and comparative complexity, we document an interaction between the complexity of a choice problem and the availability of a salient cue.

Salience effects in economic choice. The importance of salient cues that attract attention and guide behavior has been documented across various economic domains, such as public economics (e.g., Chetty *et al.*, 2009; Taubinsky and Rees-Jones, 2018), finance (e.g. Frydman and Wang, 2020), industrial organization (e.g. Bordalo *et al.*, 2016b; Dertwinkel-Kalt, Köster and Peiseler, 2019b), political economy (e.g. Bordalo, Tabellini and Yang, 2020b; Nunnari and Zápal, 2020) as well as in basic strategic interactions (e.g. Crawford and Iriberri, 2007; Li and Camerer, 2019). We focus on salience theory (Bordalo *et al.*, 2012), which has been shown to explain most risk puzzles in one coherent framework. Our study adds both to the theoretical (Bordalo *et al.*, 2012; Dertwinkel-Kalt and Köster, 2020b; Lanzani, 2020) and experimental (Dertwinkel-Kalt and Köster, 2020b; Frydman and Mormann, 2018) literature that studies salience effects in choice under risk. While the previous literature has mostly confirmed the predictions of salience theory, we document its limits by showing that it cannot account for portfolio selection with symmetric risks. This confirms the conclusions in Dertwinkel-Kalt and Köster (2020b) and Dertwinkel-Kalt *et al.* (2020a) that salience theory of choice under risk is mainly a model of (relative) skewness preferences.

Salience-driven versus deliberate (in)attention. The variant of Bordalo *et al.*'s (2012) salience theory that we propose in Section 3.7 incorporates features of models on deliberate inattention (e.g. Sims, 2003; Gabaix, 2014; Matějka and McKay, 2015). But we replace the trade-off between the costs and benefits of being attentive, which determines whether or not a deliberately inattentive agent makes an "active" choice, by a mechanism based on the availability of a salient cue. This mechanism is also closely related to the similarity-based model proposed by Rubinstein (1988), who assumes that a decision-maker neglects dimensions in which the options are similar; that is, dimensions in which the options

[19]When choosing between monetary risks, the number of dimensions is typically associated with the number of distinct states of the world. The number of states can be manipulated in an economically meaningful way by changing the correlation of the assets underlying a portfolio, or in an economically trivial way, by simply splitting up certain states of the world (as in Starmer and Sugden, 1993; Birnbaum, 2005; Bernheim and Sprenger, 2020). In Dertwinkel-Kalt and Köster (2015) we discuss such event-splitting effects in the context of a salience model.

differ only by a sufficiently small amount. While classical models of deliberate inattention cannot be directly applied to our setting,[20] when extending the intuition underlying these models — that is, the agent pays attention if and only if there is enough variation in outcomes — one might arrive at a condition similar to Eq. (3.2), where the left-hand side is replaced by a *symmetric* measure of the variation in outcomes (e.g. the relative variance of the assets). Our experimental results suggest, however, that it is the asymmetry of contrasts in outcomes rather than the size of contrasts that shapes attention.

Evidence on portfolio under-diversification. Our study also adds to the (experimental) literature on portfolio selection. Existing experiments have exclusively looked at portfolio selection problems with symmetric assets, and have documented quite robust evidence for the (conditional) $1/N$-heuristic (e.g., Eyster and Weizsäcker, 2016; Ackert, Church and Qi, 2015; Kallir and Sonsino, 2009; Baltussen and Post, 2011; Ungeheuer and Weber, forthcoming). We qualify these findings by showing that subjects in our experiment deviate from the $1/N$-heuristic if the assets are sufficiently skewed. We also contribute to the debate on how to model skewness preferences. As we argue in Appendix H.7, our experimental design allows us to test models on a variance-skewness trade-off (e.g. Mitton and Vorkink, 2007), which assume that people seek positive skewness, but also dislike variance per se. Our subjects, however, seem to strictly like variance for positively skewed portfolios, which yields additional support for a salience-based explanation of skewness preferences. Finally, we add to a literature that associates portfolio under-diversification with a lack of financial literacy (e.g. Lusardi and Mitchell, 2011, 2014). While in our experiment more literate subjects are less likely to diversify naively, we find no effect of literacy on skewness-driven under-diversification. This suggests that financial training may not reduce under-diversification in many instances, namely, when a salient cue is present.

3.9 Conclusion

We theoretically and experimentally study simple and complex choices under risk, and identify an interaction between the skewness of the available options and the complexity of the choice problem. While revealed attitudes toward skewed risks, which have extreme and salient outcomes, are robust to computational and comparative complexity, the revealed attitudes toward symmetric risks, where such a salient cue is missing, are not robust to either layer of complexity. We propose a variant of Bordalo *et al.*'s (2012) salience theory that introduces the distinction between stimulus-driven choices — based on salient cues

[20] These models need some uncertainty about the parameters of the problem that could be resolved by paying enough attention. This uncertainty is missing in the type of portfolio selection problems that we study in this paper.

guiding behavior —, and default-driven choices in case a salient cue is missing. The model rationalizes our experimental results.

While our framework builds on the idea that people behave more consistently in the presence of a salient cue, in contrast to a common view held in economics, it does not equate consistent with normatively appealing behavior.[21] When asymmetric contrasts in outcomes indicate that one or another option truly is the better choice, the contrast effect *can* be consistent with evidence on attention modulating "the information most relevant for the goal of the decision-maker" (Sepulveda, Usher, Davies, Benson, Ortoleva and De Martino, 2020). But we do not postulate that stimulus-driven choices are normatively better than default-driven ones. Instead, by modeling stimulus-driven choices via salience theory, we acknowledge that the salience of outcomes might not only affect whether or not someone makes an active choice, but may also "distort" the active choice itself. This is consistent with the idea that, upon attracting attention, salient cues can make people jump to conclusions (an idea that goes back to at least Tversky and Kahneman, 1974).

The general message — namely, that people deal with the complexity inherent to most economic decisions by relying on salient cues that guide their behavior — is much broader than the specific application to choice under risk that we consider in this paper. Our modeling framework can be directly applied to other domains including intertemporal choice and political engagement (for a formal analysis see Appendix I). First, in the case of intertemporal choice, it is the relative skewness of the distribution of contrasts in outcomes *across the different dates* that determines whether or not a subject makes an active choice and also shapes active choices (see Dertwinkel-Kalt, Gerhardt, Riener, Schwerter and Strang, 2019a, for evidence on contrast effects in intertemporal choice). Second, with respect to political engagement, it would be the relative skewness of the (perceived) distribution of differences *in the positions that the parties hold on various issues* that shapes behavior. When assuming a default of no political engagement, our model could rationalize evidence on increased political participation in response to parties becoming more polarized or at least being perceived as more polarized (e.g. Westfall, Van Boven, Chambers and Judd, 2015; Bordalo *et al.*, 2020b):[22] as parties become more polarized — i.e. as they take on more extreme positions regarding just a few issues — the distribution of contrasts in positions becomes more skewed and, thus, people should become more willing to deviate from their default of no engagement, such as not to vote in an election.

While in this paper we have argued that in many economic problems asymmetric

[21] See Halevy and Mayraz (2020, p.2) for a discussion of how in economics "Rationality is equated with consistency."

[22] Building on Bordalo, Coffman, Gennaioli and Shleifer's (2016a) model of stereotypes, Bordalo *et al.* (2020b) argue that perceived polarization is much larger than actual polarization as people amplify existing differences through the stereotypes they hold. In this sense, strong stereotypes — through increasing the relative skewness of the perceived outcome distribution — could contribute to the existence of salient cues that guide behavior.

contrasts in outcomes can serve as a salient cue that attracts attention and guides choices, there are likely to be other salient cues — which might, for instance, relate to size or color — that affect economic behavior and may interact with the complexity of a choice problem. Developing a general notion of a salient cue is beyond the scope of this paper and is left for future research.

The same is true for developing a fully fledged theory of the complexity of a choice problem and the (cognitive) costs associated with having to deal with increased complexity. We have focused on two uncontroversial notions of complexity, and have identified an interaction between complexity and the availability of a salient cue. Precisely, subjects in our experiment changed their behavior in response to an increase in complexity if a salient cue was missing, but not if a common salient cue was present. This implies, in particular, that a fully empirical approach — based on when subjects change their behavior — to answer the question of what makes a problem complex is unlikely to be successful. Our results instead suggest that we need a proper theoretical notion of complexity to start with; a problem that we leave for future research.

Appendix A: Proofs

A.1: Complexity in Salience Theory of Choice under Risk

The goal of this section is to prove Proposition 1. We provide some additional results on the way. First, we restate Ebert's (2015) moment characterization of binary risks for further reference.

Lemma 3 (Moment Characterization of Binary Risks). *Consider the binary lottery $L = L(E, V, S)$. Its outcomes $x_1 = x_1(E, V, S)$ and $x_2 = x_2(E, V, S)$, and the probability $p = p(S)$ of realizing the downside $x_1 < x_2$ can be written in terms of expected value E, variance V, and skewness S as follows:*

$$x_1 = E - \sqrt{\frac{V(1-p)}{p}},\ x_2 = E + \sqrt{\frac{Vp}{1-p}}, \quad \textit{and } p = \frac{1}{2} + \frac{S}{2\sqrt{4+S^2}}.$$

Second, we delineate an important implication of assuming that the salience function is homogeneous of degree zero (i.e. Assumption 1).

Lemma 4. *Let $w, x, y, z > 0$. If the salience function satisfies Assumption 1, then:*

$$\sigma(w, x) > \sigma(y, z) \quad \textit{if and only if} \quad \max\left\{\frac{w}{x}, \frac{x}{w}\right\} > \max\left\{\frac{y}{z}, \frac{z}{y}\right\}.$$

Proof. Let $w \geq x$ and $y \geq z$. By homogeneity of degree zero (Assumption 1), we have

$$\sigma(w,x) > \sigma(y,z) \quad \text{if and only if} \quad \sigma(w/x,1) > \sigma(y/z,1).$$

By ordering, we thus conclude that

$$\sigma(w,x) > \sigma(y,z) \quad \text{if and only if} \quad \frac{w}{x} > \frac{y}{z}.$$

The remaining cases follow by the exact same arguments. □

Lottery / Probability	p	$1-p$
$L(E,V,S)$	$E+\sqrt{V}\pi$	$E-\sqrt{V}\frac{1}{\pi}$
E	E	E
$L(E,V',-S)$	$E-\sqrt{V'}\pi$	$E+\sqrt{V'}\frac{1}{\pi}$
$\frac{1}{2}E+\frac{1}{2}L(E,V',-S)$	$E-\frac{\sqrt{V'}}{2}\pi$	$E+\frac{\sqrt{V'}}{2}\frac{1}{\pi}$
$\frac{1}{2}E+\frac{1}{2}L(E,V,S)$	$E+\frac{\sqrt{V}}{2}\pi$	$E-\frac{\sqrt{V}}{2}\frac{1}{\pi}$

Table 3.5: *This table illustrates the joint distribution of the available options when the choice set is given by $C = \{L(E,V,S), L(E,V',-S), E\}$. Since the lotteries $L(E,V,S)$ and $L(E,V',-S)$ have the opposite skewness, they have to be perfectly negatively correlated. We further depict in the last two lines for each of the two lotteries the corresponding reference point (i.e. the state-wise average over alternative options).*

Now consider the choice set $\mathcal{C} = \{L(E_1,V_1,S_1),\ldots,L(E_n,V_n,S_n)\}$. We impose some structure: First, $E_i = E$ for some $E \in \mathbb{R}_{>0}$, and one option pays this expected value with certainty. Second, for all lotteries (except for the safe option), let $S_i \in \{-S,S\}$ for some $S \in \mathbb{R}$. Third, to keep the state space as small as possible, all lotteries (except for the safe option) are perfectly correlated with each other. By Lemma 3, the upside of any (weakly) right-skewed lottery is realized with the same probability $p = p(S) \leq \frac{1}{2}$, which at the same time equals the probability of each of the (weakly) left-skewed lotteries realizing its downside. Hence, any two lotteries with the same (non-zero) skewness have to be perfectly positively correlated, while any two lotteries with the opposite skewness have to be perfectly negatively correlated. Symmetric lotteries, on the other hand, can be perfectly positively or negatively correlated. This is illustrated in Table 3.5 for $n=3$, whereby we denote as $\pi = \pi(p) := \sqrt{(1-p)/p}$ the square-root of the likelihood ratio.

We proceed in two steps. In a first step, we consider the case of just three options, and derive precise predictions on a salient thinker's behavior. In a second step, we generalize the results for the case of three options to an arbitrary number of options. In particular, we directly apply the results derived for the case of three options to eventually prove Proposition 1.

The case of $n = 3$. Let $\mathcal{C} = \{L(E, V, S), L(E, V', S'), E\}$ with $S' = \pm S$. We distinguish two cases depending on whether the two lotteries have the same or opposite skewness.

1. Case: $S' = -S \leq 0$. A salient thinker prefers the (weakly) right-skewed lottery $L(E, V, S)$ — which is evaluated relative to $\frac{1}{2}E + \frac{1}{2}L(E, V', -S)$ — over the safe option E if and only if

$$\sigma\left(E + \sqrt{V}\pi, E - \frac{\sqrt{V'}}{2}\pi\right) > \sigma\left(E - \sqrt{V}\frac{1}{\pi}, E + \frac{\sqrt{V'}}{2}\frac{1}{\pi}\right);$$

that is, if and only if, relative to the reference point, the upside of the lottery $L(E, V, S)$ is salient. By Lemma 4, under Assumption 1, this is indeed the case if and only if

$$\frac{E + \sqrt{V}\pi}{E - \frac{\sqrt{V'}}{2}\pi} > \frac{E + \frac{\sqrt{V'}}{2}\frac{1}{\pi}}{E - \sqrt{V}\frac{1}{\pi}} \quad \text{or, equivalently,} \quad S > \max\left\{\frac{\sqrt{V} - \sqrt{V'}/2}{E}, 0\right\} =: S_r.$$

Analogously, under Assumption 1, the (weakly) left-skewed lottery $L(E, V', -S)$ — which is evaluated relative to $\frac{1}{2}E + \frac{1}{2}L(E, V, S)$ — is more attractive than the safe option if and only if

$$S < \max\left\{\frac{\sqrt{V}/2 - \sqrt{V'}}{E}, 0\right\} =: S_l.$$

Clearly, $S_l \leq S_r$, so that for any parameter combination at most one of the two lotteries is more attractive to a salient thinker than the safe option. This gives the first lemma.

Lemma 5. *Let $n = 3$ and $S' = -S \leq 0$. In addition, suppose that Assumption 1 holds.*

(a) If $S \in [0, S_l)$, a salient thinker chooses the (weakly) left-skewed lottery $L(E, V', -S)$.

(b) If $S \in [S_l, S_r]$, a salient thinker chooses the safe option E.

(c) If $S \in (S_r, \infty)$, a salient thinker chooses the (weakly) right-skewed lottery $L(E, V, S)$.

2. Case: $S' = S$. The two lotteries $L(E, V, S)$ and $L(E, V', S)$ have the same skewness and are, therefore, assumed to be perfectly positively correlated. We conclude that salience theory predicts skewness preferences as well as an aversion toward symmetric risks. Formally:

Lemma 6. *Let $n = 3$ and $S' = S$. In addition, suppose Assumption 1 holds. The salient thinker chooses one of the two — in this case, right-skewed — lotteries if and only if*

$$S > \min\left\{\frac{\sqrt{V}/2 + \sqrt{V'}}{E}, \frac{\sqrt{V} + \sqrt{V'}/2}{E}\right\}.$$

Proof. First, let $S > 0$. A salient thinker prefers $L(E,V,S)$ — which is evaluated (state-by-state) relative to the reference point $\frac{1}{2}E + \frac{1}{2}L(E,V',S)$ — over the safe option E if and only if

$$\sigma\left(E + \sqrt{V}\pi, E + \frac{\sqrt{V'}}{2}\pi\right) > \sigma\left(E - \sqrt{V}\frac{1}{\pi}, E - \frac{\sqrt{V'}}{2}\frac{1}{\pi}\right).$$

By Lemma 4, under Assumption 1, this is indeed the case if and only if

$$\max\left\{\frac{E + \sqrt{V}\pi}{E + \frac{\sqrt{V'}}{2}\pi}, \frac{E + \frac{\sqrt{V'}}{2}\pi}{E + \sqrt{V}\pi}\right\} > \max\left\{\frac{E - \frac{\sqrt{V'}}{2}\frac{1}{\pi}}{E - \sqrt{V}\frac{1}{\pi}}, \frac{E - \sqrt{V}\frac{1}{\pi}}{E - \frac{\sqrt{V'}}{2}\frac{1}{\pi}}\right\} \quad \Longleftrightarrow \quad S > \frac{\sqrt{V} + \frac{\sqrt{V'}}{2}}{E}.$$

A salient thinker, thus, chooses one of the lotteries over the safe option E if and only if

$$S > \min\left\{\frac{\sqrt{V} + \frac{\sqrt{V'}}{2}}{E}, \frac{\sqrt{V'} + \frac{\sqrt{V}}{2}}{E}\right\}.$$

Second, let $S \leq 0$. A salient thinker prefers $L(E,V,S)$ — which is evaluated (state-by-state) relative to the reference point $\frac{1}{2}E + \frac{1}{2}L(E,V',S)$ — over the safe option E if and only if

$$\sigma\left(E + \sqrt{V}\frac{1}{\pi}, E + \frac{\sqrt{V'}}{2}\frac{1}{\pi}\right) > \sigma\left(E - \sqrt{V}\pi, E - \frac{\sqrt{V'}}{2}\pi\right).$$

Again, under Assumption 1, this is the case if and only if

$$\max\left\{\frac{E + \sqrt{V}\frac{1}{\pi}}{E + \frac{\sqrt{V'}}{2}\frac{1}{\pi}}, \frac{E + \frac{\sqrt{V'}}{2}\frac{1}{\pi}}{E + \sqrt{V}\frac{1}{\pi}}\right\} > \max\left\{\frac{E - \frac{\sqrt{V'}}{2}\pi}{E - \sqrt{V}\pi}, \frac{E - \sqrt{V}\pi}{E - \frac{\sqrt{V'}}{2}\pi}\right\} \quad \Longleftrightarrow \quad S > \frac{\sqrt{V} + \frac{\sqrt{V'}}{2}}{E}.$$

Since $S \leq 0$ by assumption, the above inequality cannot hold, so that the (weakly) left-skewed lottery $L(E,V,S)$ is unattractive to a salient thinker. By the exact same arguments, also the (weakly) left-skewed lottery $L(E,V',S)$ is unattractive to a salient thinker. □

The general case of $n \in \mathbb{N}_{\geq 3}$. We can easily extend the preceding arguments to a general number of available options $n \in \mathbb{N}_{\geq 3}$, which then immediately yields a proof of Proposition 1.

Proof of Proposition 1. Let $S > 0$ and $k \in \{0, \dots, n-1\}$. We order the options in such a way that options $i \in \{1, \dots, k\}$ are left-skewed with $S_i = -S$, options $i \in \{k+1, \dots, n-1\}$ are right-skewed with $S_i = S$, and option $i = n$ is the safe option paying the expected value E with certainty. For a given option j, we denote the total standard deviation of alternative left-skewed options as $V_{j,l} := \sum_{i \neq j, i \leq k} \sqrt{V}_i$, and the total standard deviation of alternative right-skewed options as $V_{j,r} := \sum_{i \neq j, k+1 \leq i \leq n-1} \sqrt{V}_i$. It is easily verified that lottery $L(E, V_j, S_j)$ is evaluated relative to the reference point $L\Big(E, V_j', \mathrm{sgn}(V_{j,r} - V_{j,l})S\Big)$ with $V_j' := (V_{j,r} - V_{j,l})^2/(n-1)^2$. We can, therefore, simply invoke Lemma 5 and Lemma 6 to prove the proposition. □

A.2: Experimental Design

Preliminaries on Mao pairs. Notice that the joint distribution of the two lotteries $X_1 = L(E, V, S)$ and $X_2 = L(E, V, -S)$ can be fully parameterized by the correlation coefficient

$$\rho = \rho\Big(L(E, V, S), L(E, V, -S)\Big) \in [-1, \overline{\rho}(S)] \quad \text{with} \quad \overline{\rho}(S) := \frac{\sqrt{4+S^2} - S}{\sqrt{4+S^2} + S}.$$

The correlation coefficient is bounded from above, and the upper bound $\overline{\rho}(S)$ is strictly decreasing in S. Let $p = p(S) \in (0, \frac{1}{2})$ be the probability of the right-skewed lottery's larger payoff, which is, by construction, identical to the probability of the left-skewed lottery's smaller payoff. For a correlation coefficient $\rho \in [-1, \overline{\rho}(S)]$, Table 3.6 depicts the joint distribution of a Mao pair.

	$(1-p)p(1+\rho)$	$p^2 - (1-p)p\rho$	$(1-p)^2 - (1-p)p\rho$	$(1-p)p(1+\rho)$
$X_1 - E$	$-\sqrt{\frac{Vp}{(1-p)}}$	$\sqrt{\frac{V(1-p)}{p}}$	$-\sqrt{\frac{Vp}{(1-p)}}$	$\sqrt{\frac{V(1-p)}{p}}$
$X_2 - E$	$-\sqrt{\frac{V(1-p)}{p}}$	$-\sqrt{\frac{V(1-p)}{p}}$	$\sqrt{\frac{Vp}{(1-p)}}$	$\sqrt{\frac{Vp}{(1-p)}}$
$X(\alpha) - E$	$-\frac{\sqrt{V}[(1-p)-\alpha(1-2p)]}{\sqrt{p(1-p)}}$	$-\frac{\sqrt{V}(1-p)[1-2\alpha]}{\sqrt{p(1-p)}}$	$\frac{\sqrt{V}p[1-2\alpha]}{\sqrt{p(1-p)}}$	$\frac{\sqrt{V}[p+\alpha(1-2p)]}{\sqrt{p(1-p)}}$

Table 3.6: *Joint distribution of the lotteries of a Mao pair.*

Proof of Lemma 1. Let $X_1 = L(E, V, S)$ and $X_2 = L(E, V, -S)$, and let $\rho(X_1, X_2) = -1$.

PART (a). Using Table 3.6, we conclude that investing some share $\alpha \in [0, 1]$ in asset X_1 and the remaining share $1 - \alpha$ in asset X_2 results in a portfolio $X(\alpha)$ that pays either

$$E - \mathrm{sgn}(1-2\alpha)\sqrt{(1-2\alpha)^2 V}\sqrt{\frac{(1-p)}{p}}$$

with probability $p \in (0, \frac{1}{2})$ or

$$E + \mathrm{sgn}(1-2\alpha)\sqrt{(1-2\alpha)^2 V}\sqrt{\frac{p}{1-p}}$$

with probability $1-p \in (\frac{1}{2}, 1)$. By Lemma 3, $X(\alpha) = L(E, (2\alpha-1)^2 V, \mathrm{sgn}(2\alpha-1)S)$. Importantly, the resulting portfolio $X(\alpha)$ is right-skewed if and only if $\alpha > \frac{1}{2}$.

PART (b) and PART (c). Follow immediately from Part (a). □

Proof of Lemma 2. We prove the statements for the twin problems with a continuous choice set. The statement for binary twin problems follows by the exact same arguments.

PART (a). Notice that in (P.1) of a twin problem, for any $x \in [0, \frac{1}{2}]$, the investments $\alpha_1 = \frac{1}{2} - x$ and $\alpha_1' = \frac{1}{2} + x$ implement the same portfolio, and that, for any $\alpha_1' \in [\frac{1}{2}, 1]$, we can implement the exact same portfolio in (P.2) of the twin problem by setting

$$\alpha_2 = \frac{\gamma + 2\alpha_1' - 1}{\gamma + 1}.$$

This also implies that we have to restrict the set of admissible investments in (P.2) to be

$$\mathcal{A} = \left[\max\left\{ \frac{\gamma - 1}{\gamma + 1}, 0 \right\}, 1 \right],$$

as otherwise in (P.2) we would offer a strictly larger set of portfolios than in (P.1).

PART (b). Obviously, in (P.1), the diversified portfolio coincides with the naively diversified portfolio. Hence, consider (P.2) from now on. The variance of the portfolio $X(\alpha)$ is given by

$$\mathrm{Var}(\alpha X_1 + (1-\alpha)X_2) = \alpha^2 \mathrm{Var}(X_1) + (1-\alpha)^2 \mathrm{Var}(X_2) + 2\alpha(1-\alpha)\mathrm{Cov}(X_1, X_2).$$

Since $\mathrm{Var}(X_1) = V$ as well as $\mathrm{Var}(X_2) = \gamma^2 V$ and $\mathrm{Cov}(X_1, X_2) = -\gamma V$, it follows that

$$\mathrm{Var}(\alpha X_1 + (1-\alpha)X_2) = V\Big(\alpha^2 + (1-\alpha)^2\gamma^2 - 2\alpha(1-\alpha)\gamma\Big).$$

Taking the derivative of the variance of $X(\alpha)$ with respect to α, we obtain

$$\begin{aligned}\frac{\partial}{\partial \alpha}\mathrm{Var}(\alpha X_1 + (1-\alpha)X_2) &= 2\alpha - 2(1-\alpha)\gamma^2 - 2(1-\alpha)\gamma + 2\alpha\gamma \\ &= -2\gamma(1+\gamma) + 2\alpha(1 + 2\gamma + \gamma^2) \\ &= -2\gamma(1+\gamma) + 2\alpha(1+\gamma)^2,\end{aligned}$$

which is equal to zero if and only if $\alpha = \frac{\gamma}{1+\gamma}$. The claim follows from the fact that the variance $\mathrm{Var}(\alpha X_1 + (1-\alpha)X_2)$ is a convex function of α, so that it is minimized at $\alpha = \frac{\gamma}{1+\gamma}$. □

Appendix B: Predictions of Expected Utility Theory

B.1: Portfolio Selection Problems with Skewed Assets

We first derive the precise prediction of expected utility theory in the context of our experiment under the assumption of a concave utility function.

Lemma 7. *Let $\mathcal{A} = \{\frac{1}{2}\} \cup A$ with $A \subseteq [0, 1]$, and suppose that the underlying assets form a perfectly negatively correlated Mao pair: $X_1 = L(E, V, S)$ and $X_2 = L(E, V, -S)$ with $\rho(X_1, X_2) = -1$. Any expected utility maximizer with a concave utility function chooses the diversified portfolio.*

Proof. By Lemma 1 (a), we have $X(\frac{1}{2}) = E$ as well as $X(\alpha) = L(E, (2\alpha - 1)^2 V, \operatorname{sgn}(2\alpha - 1)S)$ for any $\alpha \neq \frac{1}{2}$. The claim follows by Jensen's Inequality. □

Next, we derive a slightly more general result to justify why we focus on the case of a perfectly negative correlation. If the lotteries of a Mao pair are *not* perfectly negatively correlated, expected utility theory can explain under-diversification even when assuming concave utility.

Lemma 8. *Let $\mathcal{A} = \{\frac{1}{2}\} \cup A$ with $A \subseteq [0, 1]$, and suppose that the underlying assets form a Mao pair: $X_1 = L(E, V, S)$ and $X_2 = L(E, V, -S)$. Let $\rho(X_1, X_2) > -1$. Then, the following statements hold:*

(a) The diversified portfolio coincides with the naively diversified portfolio.

(b) Any expected utility maximizer with a concave, thrice-differentiable utility function $u : \mathbb{R}_{\geq 0} \to \mathbb{R}$ chooses the diversified portfolio if $u'''(\cdot) = 0$. Otherwise, there exists some Mao pair with moments $(E, V, S) \in \mathbb{R}^3_{>0}$ for which she chooses an under-diversified portfolio.[23]

Proof. PART (a). The variance of the portfolio $X(\alpha)$ is given by

$$\begin{aligned}
\operatorname{Var}(\alpha X_1 + (1 - \alpha)X_2) &= \alpha^2 V + (1 - \alpha)^2 V + 2\alpha(1 - \alpha)\operatorname{Cov}(X_1, X_2) \\
&= V - 2\alpha(1 - \alpha)\Big(V - \operatorname{Cov}(X_1, X_2)\Big) \\
&= V\Big[1 - 2\alpha(1 - \alpha)\Big(1 - \rho(X_1, X_2)\Big)\Big],
\end{aligned}$$

where the second equality follows from the definition of the correlation coefficient and the fact that X_1 and X_2 have the same variance. Since the correlation of assets that form a

[23] Any expected utility maximizer with a concave utility function that satisifies $u'''(\cdot) > 0$, over-invests in the right-skewed asset; in particular, she chooses $\alpha \in (\frac{1}{2}, 1)$. Any expected utility maximizer with a concave utility function that satisifies $u'''(\cdot) < 0$, chooses a portfolio with $\alpha \in (0, \frac{1}{2})$ instead.

Mao pair cannot exceed $\overline{\rho} < 1$, it follows that the variance of the portfolio is minimized at $\alpha = \frac{1}{2}$.

PART (b). Fix a correlation coefficient $\rho \in (-1, \overline{\rho}]$, and consider an expected utility maximizer with a concave and thrice-differentiable utility function $u : \mathbb{R}_{\geq 0} \to \mathbb{R}$. Using Table 3.6, we conclude that the expected utility of a portfolio $X(\alpha)$ is given by

$$\begin{aligned}\mathbb{E}[u(X(\alpha))] = \big(p^2 - (1-p)p\rho\big)u\Big(E - \frac{\sqrt{V}(1-p)[1-2\alpha]}{\sqrt{p(1-p)}}\Big) + \big((1-p)^2 - (1-p)p\rho\big)u\Big(E + \frac{\sqrt{V}p[1-2\alpha]}{\sqrt{p(1-p)}}\Big) \\ + (1-p)p(1+\rho)\Big[u\Big(E - \frac{\sqrt{V}[(1-p) - \alpha(1-2p)]}{\sqrt{p(1-p)}}\Big) + u\Big(E + \frac{\sqrt{V}[p + \alpha(1-2p)]}{\sqrt{p(1-p)}}\Big)\Big]\end{aligned}$$

First, since the utility function is concave, it is easily verified that

$$\begin{aligned}\frac{\partial^2}{\partial\alpha^2}\mathbb{E}[u(X(\alpha))] =& 4\big(p^2 - (1-p)p\rho\big)\Big(\frac{1-p}{p}\Big)Vu''\Big(E - \frac{\sqrt{V}(1-p)[1-2\alpha]}{\sqrt{p(1-p)}}\Big) \\ &+ 4\big((1-p)^2 - (1-p)p\rho\big)\Big(\frac{p}{1-p}\Big)Vu''\Big(E + \frac{\sqrt{V}p[1-2\alpha]}{\sqrt{p(1-p)}}\Big) \\ &+ (1-2p)(1+\rho)\Big[u''\Big(E - \frac{\sqrt{V}[(1-p) - \alpha(1-2p)]}{\sqrt{p(1-p)}}\Big) + u''\Big(E + \frac{\sqrt{V}[p + \alpha(1-2p)]}{\sqrt{p(1-p)}}\Big)\Big] < 0.\end{aligned}$$

Hence, the portfolio that maximizes the expected utility is unique.

Next, looking at the first derivative of the expected utility with respect to α, which is

$$\begin{aligned}\frac{\partial}{\partial\alpha}\mathbb{E}[u(X(\alpha))] =& 2\big(p^2 - (1-p)p\rho\big)\sqrt{\frac{1-p}{p}}\sqrt{V}u'\Big(E - \frac{\sqrt{V}(1-p)[1-2\alpha]}{\sqrt{p(1-p)}}\Big) \\ &- 2\big((1-p)^2 - (1-p)p\rho\big)\sqrt{\frac{p}{1-p}}\sqrt{V}u'\Big(E + \frac{\sqrt{V}p[1-2\alpha]}{\sqrt{p(1-p)}}\Big) \\ &+ (1-2p)(1+\rho)\sqrt{p(1-p)V}\Big[u'\Big(E - \frac{\sqrt{V}[(1-p) - \alpha(1-2p)]}{\sqrt{p(1-p)}}\Big) + u'\Big(E + \frac{\sqrt{V}[p + \alpha(1-2p)]}{\sqrt{p(1-p)}}\Big)\Big],\end{aligned}$$

we observe that, by the concavity of the utility function,

$$\lim_{\alpha\to 0}\frac{\partial}{\partial\alpha}\mathbb{E}[u(X(\alpha))] = (1-\rho)\Big[u'\Big(E - \sqrt{\frac{V(1-p)}{p}}\Big) - u'\Big(E + \sqrt{\frac{Vp}{1-p}}\Big)\Big] > 0$$

holds. Analogously, we have $\lim_{\alpha\to 1}\frac{\partial}{\partial\alpha}\mathbb{E}[u(X(\alpha))] < 0$, which, in turn, implies that the optimal portfolio is characterized via the first-order condition $\frac{\partial}{\partial\alpha}\mathbb{E}[u(X(\alpha))] = 0$.

By evaluating $\frac{\partial}{\partial\alpha}\mathbb{E}[u(X(\alpha))] = 0$ at $\alpha = \frac{1}{2}$ and re-arranging, we conclude that the diversified portfolio maximizes the expected utility (for a concave utility function) if and only if

$$\frac{1}{2}u'\Big(E - \frac{1}{2}\frac{V}{\sqrt{p(1-p)}}\Big) + \frac{1}{2}u'\Big(E + \frac{1}{2}\frac{V}{\sqrt{p(1-p)}}\Big) = u'(E).$$

If $u'''(\cdot) = 0$, the first-order condition is satisfied at $\alpha = \frac{1}{2}$. If there exists some $x \in \mathbb{R}_{>0}$

with $u'''(x) \neq 0$, however, then, by Jensen's Inequality, for any Mao pair with an expected value $E = x$ and a sufficiently small variance V, the first-order condition does not hold at $\alpha = \frac{1}{2}$. □

B.2: Portfolio Selection Problems with Symmetric Assets

Lemma 9. *Consider a (binary) twin problem. Any expected utility maximizer with a concave utility function chooses the diversified portfolio. Moreover, allowing for a general utility function, any expected utility maximizer chooses the same portfolio in (P.1) and (P.2).*

Proof. The first claim follows immediately from Jensen's Inequality. The second claim follows from the fact the problems (P.1) and (P.2) are essentially equivalent. □

Appendix C: Predictions of Salience Theory of Choice under Risk

C.1: An Application of Salience Theory to Portfolio Selection Problems

Consider a portfolio selection problem $(\mathcal{A}, \mathcal{C})$, and denote as F the joint cumulative distribution function of the assets X_1 and X_2. The *reference point* for portfolio $X(\alpha)$ is then given by

$$R(\alpha) := \int_{\mathcal{A}\setminus\{\alpha\}} X(\beta)\ dG_\alpha(\beta), \tag{3.3}$$

where G_α is the cumulative distribution function of the uniform distribution over the augmented choice set $\mathcal{C}_\alpha := \{X(\beta) : \beta \in \mathcal{A},\ \beta \neq \alpha\}$.[24]

For a portfolio selection problem $(\mathcal{A}, \mathcal{C})$, the salience-weighted utility of portfolio $X(\alpha)$ is

$$U^s(X(\alpha)|\mathcal{C}, \mathcal{A}) = \int_{\mathbb{R}^2_{\geq 0}} (\alpha x_1 + (1-\alpha)x_2) \cdot \frac{\sigma(\alpha x_1 + (1-\alpha)x_2, r(\alpha))}{\int_{\mathbb{R}^2_{\geq 0}} \sigma(\alpha y_1 + (1-\alpha)y_2, r(\alpha))\ dF(y_1, y_2)}\ dF(x_1, x_2),$$

where $\sigma : \mathbb{R}^2_{\geq 0} \to \mathbb{R}_{>0}$ is a salience function that is bounded away from zero. This implicitly assumes that a salient thinker can deal with the computational complexity arising from portfolio selection problems; that is, the salient thinker evaluates any portfolio based on its correct distribution of outcomes. Thus, for any two essentially equivalent portfolio selection problems, $(\mathcal{A}, \mathcal{C})$ and $(\mathcal{A}', \mathcal{C}')$, a salient thinker chooses the exact same

[24]Precisely, the (augmented) choice set is a multiset; that is, duplicates $X(\alpha) = X(\beta)$ with $\alpha \neq \beta$ appear twice.

distribution over outcomes. Comparative complexity, however, can affect the behavior of a salient thinker through changing the reference point (see Section 3.2 for an illustration). Whether these are a sensible assumptions is eventually an empirical question, which we address in Sections 4.5 – 4.7 of the main text.

In Appendix E, we sketch an alternative variant of the salience model in which only computational complexity matters. One could also argue that comparative and/or computational complexity results in people using a "simpler" reference point instead of the state-wise average over all alternative options (e.g. they evaluate each portfolio relative to the (naively) diversified portfolio). We derive the predictions of such a variant of salience theory in Appendix F.

C.2: Portfolio Selection Problems with Skewed Assets

Suppose that the two assets are perfectly negatively correlated and form a Mao pair; that is, $X_1 = L(E, V, S)$ and $X_2 = L(E, V, -S)$. By Lemma 1, the diversified and naively diversified portfolio coincide.

Portfolio selection problems with a binary choice set. We consider a salient thinker's choice between the (naively) diversified portfolio $X(\frac{1}{2})$ and an under-diversified portfolio $X(\alpha)$ with $\alpha \neq \frac{1}{2}$. By Lemma 1 (a), due to the perfectly negative correlation, the (naively) diversified portfolio eliminates all variance in outcomes, and the problem boils down to the question of whether a salient thinker chooses the non-negative, binary lottery $L(E, (2\alpha - 1)^2 V, \mathrm{sgn}(2\alpha - 1)S)$ over the safe option that pays its expected value E with certainty.

Proposition 2. *Suppose that the two assets are perfectly negatively correlated and form a Mao pair. Consider the portfolio selection problem* $(\mathcal{A}, \mathcal{C})$ *with* $\mathcal{A} = \{\alpha, \frac{1}{2}\}$ *for* $\alpha \neq \frac{1}{2}$.

(a) A salient thinker chooses $X(\alpha)$ *over* $X(\frac{1}{2})$ *only if* $\alpha > \frac{1}{2}$ *holds; that is, only if the under-diversified portfolio is a right-skewed binary lottery.*

(b) For any $\alpha > \frac{1}{2}$*, any expected value* E*, and any variance* V*, there exists some threshold value* $\hat{S} = \hat{S}(\alpha, E, V) \in \mathbb{R}$*, such that a salient thinker chooses* $X(\alpha)$ *over* $X(\frac{1}{2})$ *if and only if* $S > \hat{S}$.

Proof. Follows from Lemma 1 (a) in combination with Proposition 3 in Dertwinkel-Kalt and Köster (2020b). □

First, by ordering and diminishing sensitivity, a salient thinker never chooses a left-skewed portfolio with $\alpha < \frac{1}{2}$: the downside of the binary lottery $L(E, (2\alpha - 1)^2 V, -S)$ differs by more from its expected value E than the lottery's upside, and also the payoff level is lower in the state of the world where the lottery yields its downside payoff. As a consequence, the downside of any left-skewed binary lottery with non-negative outcomes is salient, and any left-skewed portfolio is, therefore, unattractive to a salient thinker. Second, when fixing some $\alpha > \frac{1}{2}$ and thereby the variance of $L(E, (2\alpha-1)^2 V, S)$, it follows from Proposition 3 in Dertwinkel-Kalt and Köster (2020b) that a salient thinker chooses the under-diversified portfolio if and only if the skewness of the underlying Mao pair exceeds a certain threshold: Increasing the skewness of a binary lottery, while keeping its expected value and variance constant, results in a smaller contrast on the lottery's downside and a larger contrast on its upside (see also Figure 3.3). This implies that the relative salience of a right-skewed portfolio's upside monotonically increases in the skewness of the underlying Mao lotteries, which then gives Proposition 2 (b).

Portfolio selection problems with a continuous choice set. Next, we increase the comparative complexity of the problem by replacing the binary choice set with a continuous set of portfolios.

Proposition 3. *Suppose that the two assets are perfectly negatively correlated and form a Mao pair. Consider the portfolio selection problem* $(\mathcal{A}, \mathcal{C})$ *with* $\mathcal{A} = [0, 1]$*, and let* $\check{S}(E, V) := \inf_{\alpha > \frac{1}{2}} \hat{S}(\alpha, E, V)$*.*

(a) A salient thinker chooses a portfolio $X(\alpha)$ *with* $\alpha \geq \frac{1}{2}$*; that is, she invests (weakly) more in the right-skewed lottery.*

(b) A salient thinker chooses an under-diversified portfolio if and only if $S > \check{S}(E, V)$ *holds.*

Proof. Follows immediately from Proposition 2. □

We observe that a salient thinker's behavior is basically the same as in the case of a binary choice set analyzed above: she chooses an under-diversified portfolio if and only if it is sufficiently right-skewed. In addition, Proposition 3 shows that a salient thinker does not simply trade off the skewness of portfolio returns against the variance in portfolio returns, as it is typically assumed in the empirical finance literature (e.g., Mitton and Vorkink, 2007). Precisely, by Lemma 1 (a), under the perfectly negative correlation, the variance in portfolio returns monotonically increases in α on the interval $(\frac{1}{2}, 1]$, while the skewness of returns is independent of the exact share α. By choosing a portfolio with $\alpha > \frac{1}{2}$, a salient thinker, thus, does not minimize variance for a given level of skewness in the distribution of returns. As we show in Appendix H.7, subjects indeed choose a substantial amount of variance for sufficiently right-skewed portfolios, which supports the salience-based explanation of skewness preferences.

The exact threshold value on the assets' skewness, $\check{S} = \check{S}(E, V)$, for a salient thinker to choose an under-diversified portfolio depends not only on the other two moments of the underlying Mao lotteries, but also on the functional form of the salience function. If the salience function is, for instance, homogeneous of degree zero (as it is assumed in Bordalo *et al.*, 2013b), a salient thinker *always* chooses an under-diversified portfolio. This follows from the fact that the variance in portfolio returns continuously approaches zero as α approaches $\frac{1}{2}$, while its skewness sharply drops at $\frac{1}{2}$. As long as the level effect is not too strong relative to the contrast effect, which is guaranteed by homogeneity of degree zero, this sharp increase in the asymmetry of contrasts in outcomes is sufficient to induce under-diversification by a salient thinker.

C.3: Portfolio Selection Problems with Symmetric Assets

Suppose the assets form a (binary) twin problem in the spirit of Eyster and Weizsäcker (2016).

Twin problems with a binary choice set. A binary twin problem consists of two essentially equivalent portfolio selection problems, (P.1) and (P.2), both of which allow for the agent to choose the diversified or the naively diversified portfolio. While in (P.1) the diversified and the naively diversified portfolio coincide, they differ in (P.2).

Proposition 4. *A salient thinker picks the diversified portfolio in (P.1) and (P.2) of a binary twin problem.*

Proof. Since the choice set is binary and identical in both problems, the decision is exactly the same from a salient thinker's perspective and she, therefore, chooses the same portfolio in (P.1) and in (P.2). Hence, it is sufficient to look at a salient thinker's choice in (P.1).

In (P.1), for $\gamma \in [0, 3]$, the choice is between the safe option $X(\frac{1}{2}) = E$ and the binary lottery $X(\frac{3-\gamma}{4}) = L(E, \frac{1}{2}(1-\gamma)^2 V, 0)$. The salient thinker thus prefers $X(\frac{1}{2})$ over $X(\frac{3-\gamma}{4})$ if and only if

$$\frac{1}{2}\sqrt{V}\big(1-\gamma\big)\left[\sigma\left(E, E - \frac{1}{2}\sqrt{V}(1-\gamma)\right) - \sigma\left(E, E + \frac{1}{2}\sqrt{V}(1-\gamma)\right)\right] > 0,$$

which is indeed the case due to diminishing sensitivity. □

Twin problems with a continuous choice set. In contrast to the *binary* twin problems, a salient thinker does not necessarily choose the same portfolio in (P.1) and (P.2) of a comparatively complex twin problem with a continuous choice set. While a salient thinker always chooses the diversified portfolio in (P.1) of a twin problem with a continuous choice set, this is not true in general for (P.2). Consider (P.2), and let us start with the less interesting case where $\gamma \geq 1$, which also covers (P.1). Here, the reference point coincides with the diversified portfolio and, thus, yields the assets' expected value with certainty. But then — by the same arguments as in the case of a binary twin problem — choosing this reference point is more attractive to a salient thinker than choosing any other portfolio. Intuitively, since the reference point does not change compared to the binary twin problems, neither does the behavior predicted by salience theory.

Portfolio / Probability	$\frac{1}{2}$	$\frac{1}{2}$
$X(\alpha)$	$E + \sqrt{V}[\alpha - (1-\alpha)\gamma]$	$E - \sqrt{V}[\alpha - (1-\alpha)\gamma]$
$X(\frac{1}{2})$	$E + \frac{1}{2}\sqrt{V}[1-\gamma]$	$E - \frac{1}{2}\sqrt{V}[1-\gamma]$

Table 3.7: *This table depicts the joint distribution of $X(\alpha)$ and $X(\frac{1}{2})$ in (P.2) for $\gamma < 1$.*

Next, suppose that $\gamma < 1$, in which case the reference point coincides with the naively diversified portfolio and, thus, changes relative to the binary twin problems. Since the salience-weighted utility of a safe option is undistorted, by choosing the diversified portfolio, a salient thinker could always secure a salience-weighted utility of at least the portfolios' expected value E. But a salient thinker chooses an under-diversified portfolio if it yields a salience-weighted utility strictly larger than E. This happens to be the case if and only if the upside of such an under-diversified portfolio is more salient than its downside (relative to the reference point). Table 3.7 depicts the joint distribution of a general portfolio $X(\alpha)$ and the reference point $X(\frac{1}{2})$. For any $\alpha > \frac{\gamma}{1+\gamma}$, the upside of $X(\alpha)$ occurs in the first state of the world, while for any $\alpha < \frac{\gamma}{1+\gamma}$ it occurs in the second state. Since the underlying assets are symmetric, the contrast in the outcomes of the two portfolios is the same across the two states. Hence, by diminishing sensitivity, the first state is more salient than the second one if and only if the payoff level is lower in the first state. This is indeed the case if and only if $\alpha < \frac{3\gamma-1}{2(1+\gamma)}$, in which case the downside of the under-diversified portfolio is realized in the first state. As a consequence, the upside of the portfolio $X(\alpha)$ is salient if and only if $\alpha \in \left(\frac{3\gamma-1}{2(1+\gamma)}, \frac{\gamma}{1+\gamma}\right)$. It follows that for $\gamma = 0$ there is no portfolio that yields a strictly larger salience-weighted utility than the diversified portfolio. But, depending on the functional form of the salience function, the naively diversified portfolio also yields a salience-weighted utility of E, so that $\alpha \in \{0, \frac{1}{2}\}$ in this case.

The following proposition summarizes the predictions of salience theory.

Proposition 5. *Let $\gamma \in [0, 3]$ with $\gamma \neq 1$. In (P.1) of the corresponding twin problem, a salient thinker chooses the diversified portfolio with $\alpha = \frac{1}{2}$, while in (P.2) she chooses a* *portfolio with*

$$\alpha \in \begin{cases} \{0, \frac{1}{2}\} & \text{if } \gamma = 0, \\ (0, \frac{\gamma}{1+\gamma}) & \text{if } 0 < \gamma < \frac{1}{3}, \\ (\frac{3\gamma-1}{2(1+\gamma)}, \frac{\gamma}{1+\gamma}) & \text{if } \frac{1}{3} \leq \gamma < 1, \\ \{\frac{\gamma}{1+\gamma}\} & \text{otherwise,} \end{cases}$$

which is under-diversified for any $\gamma \in (0,1)$ *and diversified for any* $\gamma \in (1,3]$.

Proof. Consider (P.1): Here, we have $U^s(X(\alpha)|\mathcal{C}) \geq E$ if and only if

$$(2\alpha-1)\Big[\sigma(E+\sqrt{V}(2\alpha-1),E) - \sigma(E-\sqrt{V}(2\alpha-1),E)\Big] \geq 0.$$

By diminishing sensitivity, we have

$$\sigma(E+\sqrt{V}(2\alpha-1),E) > \sigma(E-\sqrt{V}(2\alpha-1),E) \quad \Longleftrightarrow \quad 2\alpha-1<0,$$

which in turn implies that $U^s(X(\alpha)|\mathcal{C}) \geq E$ if and only if $\alpha = \frac{1}{2}$.

Consider (P.2): If $\gamma < 1$, then $U^s(X(\alpha)|\mathcal{C}) > E$ if and only if

$$\left[\alpha - \frac{\gamma}{1+\gamma}\right]\left[\sigma\left(E+\sqrt{V}(\alpha-(1-\alpha)\gamma), E+\frac{1-\gamma}{2}\sqrt{V}\right) - \sigma\left(E-\sqrt{V}(\alpha-(1-\alpha)\gamma), E-\frac{1-\gamma}{2}\sqrt{V}\right)\right] > 0,$$

or, equivalently, $\alpha \in \left(\frac{3\gamma-1}{2(1+\gamma)}, \frac{\gamma}{1+\gamma}\right)$. Thus, if $\gamma = 0$, a salient thinker chooses either the diversified portfolio with $\alpha = 0$ or, depending on the properties of her salience function, the naively diversified portfolio with $\alpha = \frac{1}{2}$. The former always yields a salience-weighted utility of E, while the latter does so if and only if $\sigma(E+\frac{1-\gamma}{2}\sqrt{V}, E+\frac{1-\gamma}{2}\sqrt{V}) = \sigma(E-\frac{1-\gamma}{2}\sqrt{V}, E-\frac{1-\gamma}{2}\sqrt{V})$ (e.g. if the salience function is homogeneous of degree zero). For any $\gamma \in (0,1)$, in contrast, a salient thinker chooses an under-diversified portfolio with $\alpha \in \left(\frac{3\gamma-1}{2(1+\gamma)}, \frac{\gamma}{1+\gamma}\right)$.

If $\gamma \geq 1$, then $U^s(X(\alpha)|\mathcal{C}) > E$ holds if and only if

$$\left[\alpha - \frac{\gamma}{1+\gamma}\right] \times \underbrace{\left[\sigma\left(E+\sqrt{V}(\alpha-(1-\alpha)\gamma), E\right) - \sigma\left(E-\sqrt{V}(\alpha-(1-\alpha)\gamma), E\right)\right]}_{>0 \text{ if and only if } \alpha<\frac{\gamma}{1+\gamma} \text{ due to diminishing sensitivity}} > 0,$$

which cannot hold due to diminishing sensitivity. This implies that $U^s(X(\alpha)|\mathcal{C})$ is maximized by choosing the diversified portfolio with $\alpha = \frac{\gamma}{1+\gamma}$, which was to be proven. $\square$

C.4: Additional Predictions on Positively Correlated Mao Pairs

Throughout this section, we assume that the two assets form a Mao pair and that they are maximally positively correlated (as we have seen in Appendix A.2, there is an upper bound on the correlation coefficient). We further assume that the salience function satisfies Assumption 1.

Portfolio selection problems with a binary choice set. We observe that for sufficiently skewed Mao pairs a salient thinker chooses an under-diversified portfolio if and only if it is right-skewed, while for sufficiently symmetric Mao pairs the opposite is true.

Proposition 6. *Suppose the assets are maximally positively correlated, and let $\mathcal{A} = \{\alpha, \frac{1}{2}\}$ for $\alpha \neq \frac{1}{2}$. Under Assumption 1, there exist threshold values $\underline{\alpha} = \underline{\alpha}(E, V, S)$ and $\overline{\alpha} = \overline{\alpha}(E, V, S)$ with $\underline{\alpha} \leq \overline{\alpha}$ as well as threshold values $\underline{S} = \underline{S}(E, V)$ and $\overline{S} = \overline{S}(E, V)$ with $0 < \underline{S} < \overline{S}$, such that the following holds:*

(a) For any $\alpha \in [\overline{\alpha}, \frac{1}{2}) \cup (\frac{1}{2}, \underline{\alpha}]$, a salient thinker chooses $X(\alpha)$ over $X(\frac{1}{2})$.

(b) If $\alpha \leq \min\{\frac{1}{2}, \underline{\alpha}\}$ or $\alpha \geq \max\{\frac{1}{2}, \overline{\alpha}\}$, a salient thinker chooses $X(\frac{1}{2})$ over $X(\alpha)$.

(c) A salient thinker chooses any portfolio $X(\alpha)$ with $\alpha > \frac{1}{2}$ over the diversified portfolio $X(\frac{1}{2})$, but the diversified portfolio $X(\frac{1}{2})$ over any portfolio $X(\alpha)$ with $\alpha < \frac{1}{2}$, if and only if $S \geq \overline{S}$ holds.

(d) A salient thinker chooses any portfolio $X(\alpha)$ with $\alpha < \frac{1}{2}$ over the diversified portfolio $X(\frac{1}{2})$, but the diversified portfolio $X(\frac{1}{2})$ over any portfolio $X(\alpha)$ with $\alpha > \frac{1}{2}$, if and only if $S \leq \underline{S}$ holds.

Proof. Using Table 3.6, it is easily verified that a salient thinker chooses the under-diversified portfolio $X(\alpha)$ over the diversified portfolio $X(\frac{1}{2})$ if and only if

$$(2\alpha-1)\Bigg[\sigma\bigg(E-\frac{\sqrt{V}[\alpha(2p-1)+(1-p)]}{\sqrt{p(1-p)}}, E-\frac{\sqrt{V}}{2\sqrt{p(1-p)}}\bigg) \\ +\sigma\bigg(E+\frac{\sqrt{V}[p+\alpha(1-2p)]}{\sqrt{p(1-p)}}, E+\frac{\sqrt{V}}{2\sqrt{p(1-p)}}\bigg) - 2\sigma\bigg(E-\frac{\sqrt{V}p[2\alpha-1]}{\sqrt{p(1-p)}}, E\bigg)\Bigg] > 0 \tag{3.4}$$

holds. By diminishing sensitivity, we have

$$\sigma\bigg(E-\frac{\sqrt{V}[\alpha(2p-1)+(1-p)]}{\sqrt{p(1-p)}}, E-\frac{\sqrt{V}}{2\sqrt{p(1-p)}}\bigg) > \sigma\bigg(E+\frac{\sqrt{V}[p+\alpha(1-2p)]}{\sqrt{p(1-p)}}, E+\frac{\sqrt{V}}{2\sqrt{p(1-p)}}\bigg),$$

so that a sufficient condition for the bracket in Eq. (3.4) being positive is given by

$$\sigma\bigg(E+\frac{\sqrt{V}[p+\alpha(1-2p)]}{\sqrt{p(1-p)}}, E+\frac{\sqrt{V}}{2\sqrt{p(1-p)}}\bigg) \geq \sigma\bigg(E-\frac{\sqrt{V}p[2\alpha-1]}{\sqrt{p(1-p)}}, E\bigg), \tag{3.5}$$

while a sufficient condition for the bracket in Eq. (3.4) being negative is given by

$$\sigma\left(E - \frac{\sqrt{V}p[2\alpha - 1]}{\sqrt{p(1-p)}}, E\right) \geq \sigma\left(E - \frac{\sqrt{V}[\alpha(2p-1) + (1-p)]}{\sqrt{p(1-p)}}, E - \frac{\sqrt{V}}{2\sqrt{p(1-p)}}\right). \quad (3.6)$$

PART (a). By Assumption 1, the sufficient condition (3.5) holds for $\alpha > \frac{1}{2}$ if and only if

$$\frac{E + \frac{\sqrt{V}[p+\alpha(1-2p)]}{\sqrt{p(1-p)}}}{E + \frac{\sqrt{V}}{2\sqrt{p(1-p)}}} \geq \frac{E}{E - \frac{\sqrt{V}p[2\alpha-1]}{\sqrt{p(1-p)}}},$$

or, equivalently,

$$\alpha \leq \frac{E}{\sqrt{V}}\sqrt{\frac{1-p}{p}}\left(\frac{1-4p}{2-4p}\right) - \frac{p}{1-2p} =: \underline{\alpha}.$$

We conclude that, for $\alpha > \frac{1}{2}$, a salient thinker chooses $X(\alpha)$ over $X(\frac{1}{2})$ for any $\alpha \in (\frac{1}{2}, \underline{\alpha}]$.

Analogously, the sufficient condition (3.6) holds for $\alpha < \frac{1}{2}$ if and only if

$$\frac{E - \frac{\sqrt{V}p[2\alpha-1]}{\sqrt{p(1-p)}}}{E} \geq \frac{E - \frac{\sqrt{V}}{2\sqrt{p(1-p)}}}{E - \frac{\sqrt{V}[\alpha(2p-1)+(1-p)]}{\sqrt{p(1-p)}}},$$

or, equivalently,

$$\alpha \geq \frac{E}{\sqrt{V}}\sqrt{\frac{1-p}{p}}\left(\frac{1-4p}{2-4p}\right) + \frac{1-p}{1-2p} =: \overline{\alpha}.$$

We conclude that, for $\alpha < \frac{1}{2}$, a salient thinker chooses $X(\alpha)$ over $X(\frac{1}{2})$ for any $\alpha \in [\overline{\alpha}, \frac{1}{2})$.

PART (b). By the same arguments as in Part (a), for any $\alpha > \frac{1}{2}$, (3.6) holds if and only if $\alpha \geq \overline{\alpha}$, while, for any $\alpha < \frac{1}{2}$, (3.5) holds if and only if $\alpha \leq \underline{\alpha}$. In sum, a salient thinker chooses $X(\frac{1}{2})$ over $X(\alpha)$ for any $\alpha \leq \min\{\frac{1}{2}, \underline{\alpha}\}$ and any $\alpha \geq \min\{\frac{1}{2}, \overline{\alpha}\}$.

PART (c). By Part (a), a salient thinker chooses $X(\alpha)$ over $X(\frac{1}{2})$ for any $\alpha \in (\frac{1}{2}, \underline{\alpha}]$, while she chooses $X(\frac{1}{2})$ over $X(\alpha)$ for any $\alpha \leq \min\{\frac{1}{2}, \underline{\alpha}\}$. Hence, we have to show that there exists some $\overline{S} > 0$ so that, for any $S \geq \overline{S}$, we have $\underline{\alpha} \geq 1$. We obtain that $\underline{\alpha} \geq 1$ holds if and only if

$$\frac{1-4p}{\sqrt{p(1-p)}} \geq 2\frac{\sqrt{V}}{E}.$$

By Lemma 3.1, $1 - 4p = \frac{2S - \sqrt{4+S^2}}{\sqrt{4+S^2}}$ and $\sqrt{p(1-p)} = \frac{1}{\sqrt{4+S^2}}$, so that the inequality is equivalent to

$$2S - \sqrt{4 + S^2} \geq 2\frac{\sqrt{V}}{E}.$$

The claim holds since the left-hand side above increases in S and approaches infinity as $S \to \infty$.

PART (d). Analogous to Part (c). $\square$

As illustrated in Figure 3.10, we can numerically solve for the threshold values on the assets' skewness in Part (c) and Part (d) of Proposition 6. We can thus calibrate the Mao pairs for our experiment such that we can test for the implications of Assumption 1.

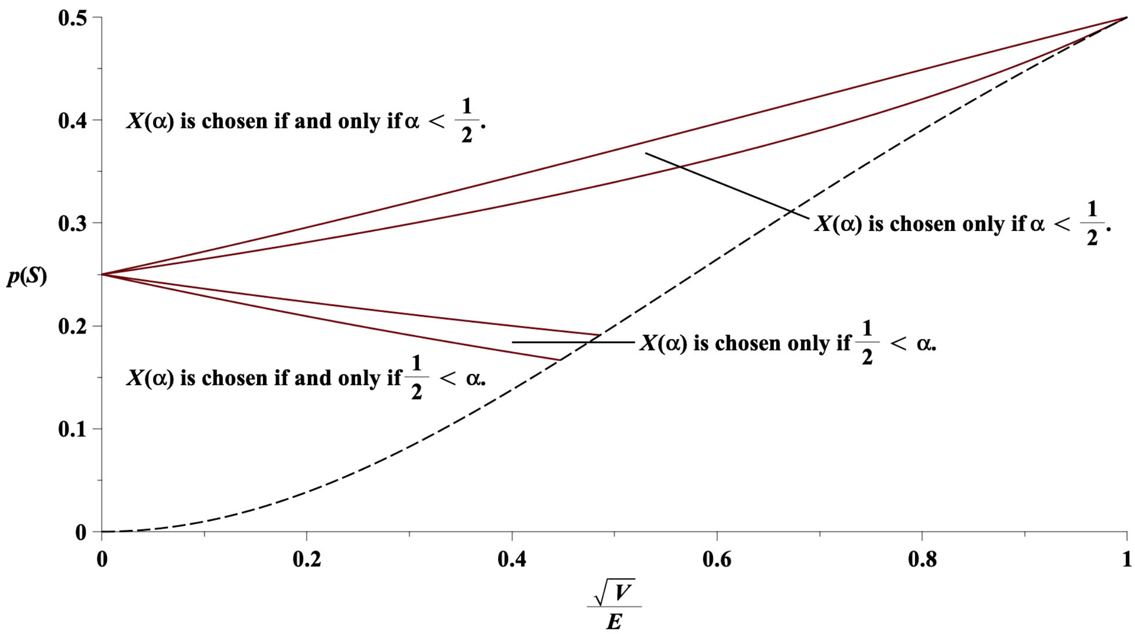


Figure 3.10: *This figure illustrates the predictions of salience theory, under the additional assumption of homogeneity of degree zero, for the binary portfolio selection problems with maximally positively correlated Mao pairs. On the x-axis, we depict the probability with which the less likely outcome of either Mao lottery is realized, which is a monotonic function of the lotteries' skewness. On the y-axis, we depict the lotteries' standard deviation normalized by the expected value. The lowest red line indicates the threshold* $\underline{S}$*, while the highest red line indicates the threshold* $\overline{S}$*, as functions of the normalized standard deviation.*

Portfolio selection problems with a continuous choice set. Under Assumption 1, a salient thinker never selects a left-skewed portfolio for sufficiently skewed Mao pairs, while she never selects a right-skewed portfolio for sufficiently symmetric Mao pairs.

Proposition 7. *Suppose that the assets are maximally positively correlated, and let* $\mathcal{A} = [0, 1]$. *Under Assumption 1, there exist threshold values* $\underset{\sim}{S} = \underset{\sim}{S}(E, V)$ *and* $\tilde{S} = \tilde{S}(E, V)$ *with* $0 < \underset{\sim}{S} < \tilde{S}$ *such that a salient thinker chooses a portfolio with*

$$\alpha \in \begin{cases} [0, \frac{1}{2}] & \text{if } S < \underset{\sim}{S}, \\ \{\frac{1}{2}\} & \text{if } \underset{\sim}{S} \leq S \leq \tilde{S}, \\ [\frac{1}{2}, 1] & \text{if } S > \tilde{S}. \end{cases}$$

Proof. The proof goes along the same lines as the proof of Proposition 6. □

Figure 3.11 illustrates the threshold values from Proposition 7. As in the case of a binary choice set, we can solve for these thresholds numerically, which helps us in calibrating the Mao pairs for the experiment. Importantly, the threshold values derived for a binary and a continuous choice set are sufficiently similar, so that we can test for the implications of Assumption 1 in both cases with the same set of Mao pairs.

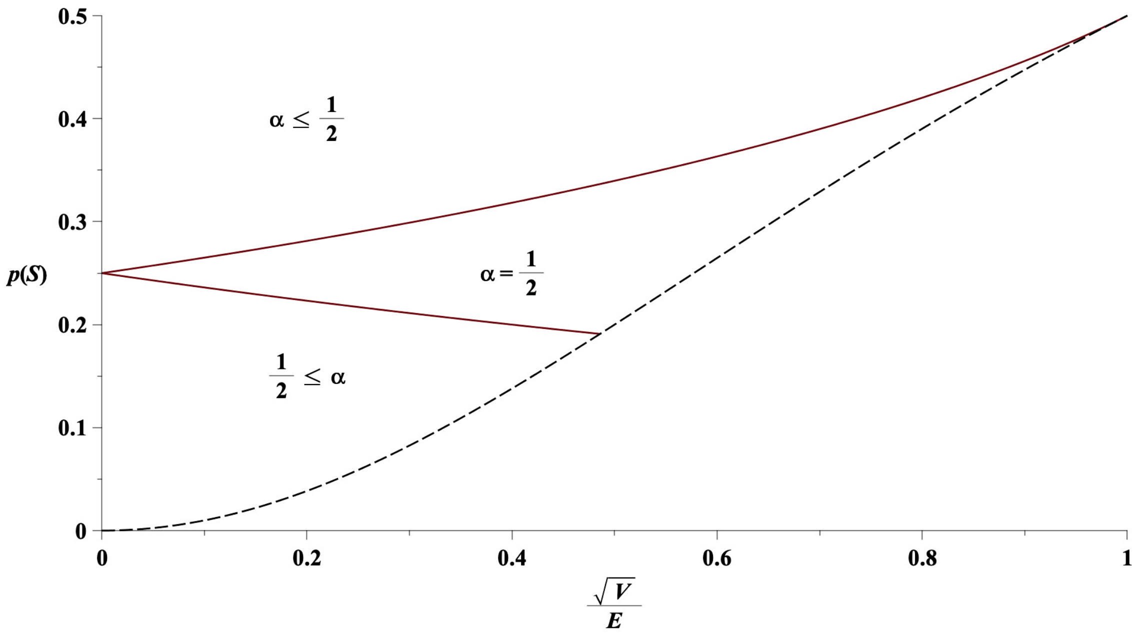


Figure 3.11: *This figure illustrates the predictions of salience theory, under the additional assumption of homogeneity of degree zero, for the continuous portfolio selection problems with maximally positively correlated Mao lotteries. On the x-axis, we depict the probability with which the less likely outcome of either Mao lottery is realized, which is a monotonic function of the lotteries' skewness. On the y-axis, we depict the lotteries' standard deviation normalized by the expected value. The red lines indicate the threshold values from Proposition 7, as functions of the normalized standard deviation.*

Appendix D: Predictions of Cumulative Prospect Theory

In this section, we analyze when does a CPT-agent (Tversky and Kahneman, 1992) choose the non-negative, binary lottery $L(E, V, S)$ over the safe option paying its expected value E with certainty. We will show that CPT is consistent with the revealed skewness preferences in our experiment, but does not unambiguously predict the patterns we observe.

To obtain clean and simple results, we make a few parametric assumptions. Since allowing for a more flexible specification would make the "predictions" of CPT even more ambiguous, making these assumptions is conservative given the goal of this exercise. First, we assume that the reference point is given by $r = 0$, so that all outcomes in the experiment represent gains, and that the CPT-agent evaluates these gains via a (weakly) concave value function $v(\cdot)$. Second, we assume that the CPT-agent weights cumulative probabilities via a strictly increasing and differentiable function $w(\cdot)$ that satisfies the three properties outlined below.

Assumption 2. *The weighting function* $w : [0,1] \to [0,1]$ *satisfies the following three properties:*

(i) $w(0) = 0$ *and* $w(1) = 1$.

(ii) There exists some $p' \in (0,1)$ *so that* $w(p) > p$ *for any* $p \in (0, p')$ *and* $w(p) < p$ *for any* $p \in (p', 1)$.

(iii) $\lim_{p\to 0} w(p)/\sqrt{p} = 0$.

Parts (i) and (ii) of Assumption 2 imply that events occurring with certainty are perceived as certain and that the weighting function is inverse-S-shaped. Part (iii) of Assumption 2 ensures that a CPT-agent's willingness-to-pay for the lottery $L(E, V, S)$ approaches zero as its skewness approaches infinity, thereby imposing a bound on the strength of skewness preferences. This last part is satisfied, for instance, by the weighting function proposed in Tversky and Kahneman (1992) for a large set of parameters (including the ones estimated in the paper). Other weighting functions proposed in the literature induce stronger probability weighting, violating Part (iii), which can result in implausible predictions (see, e.g., Azevedo and Gottlieb, 2012).

Denote as p the probability with which the upside of the lottery $L(E, V, S)$ is realized, and define $\pi(p) := \sqrt{\frac{1-p}{p}}$. The CPT-agent chooses the lottery over its expected value if and only if

$$w(p)\Big[v(E + \sqrt{V}\pi(p)) - v(E)\Big] - \Big(1 - w(p)\Big)\Big[v(E) - v(E - \sqrt{V}/\pi(p))\Big] > 0. \tag{3.7}$$

Since the value function is concave, it follows immediately from Assumption 2 (ii) that for any $p > p'$ Inequality (3.7) is violated. Hence, the CPT-agent avoids sufficiently left-skewed risks. To assess a CPT-agent's attitude toward sufficiently right-skewed risks, denote the left-hand side of (3.7) as $\Gamma(p)$, and take the derivative with respect to the upside probability p. Since $\Gamma(p)$ approaches zero as p approaches zero, the CPT-agent chooses sufficiently right-skewed risks if $\lim_{p\to 0} \Gamma'(p) > 0$, but avoids these risks if $\lim_{p\to 0} \Gamma'(p) < 0$. We have $\Gamma'(p) < 0$ if and only if

$$\overbrace{\frac{v(E+\sqrt{V}\pi(p)) - v(E-\sqrt{V}/\pi(p))}{\sqrt{V}\big[\pi(p)+1/\pi(p)\big]}}^{\text{average value increase when moving from downside to upside}} \times w'(p)$$
$$< \frac{w(p)}{2p} \times \underbrace{v'(E+\sqrt{V}\pi(p))}_{\text{marginal value of upside increase}} + \frac{1-w(p)}{2(1-p)} \times \underbrace{v'(E-\sqrt{V}/\pi(p))}_{\text{marginal value of downside increase}}. \tag{3.8}$$

If the CPT-agent weights probabilities linearly and her value function is concave, then (3.8) is clearly satisfied on an open interval to right of $p = 0$. Hence, a CPT-agent who weights probabilities linearly dislikes sufficiently right-skewed risks (as well as all other risks). If, in contrast, her value function is linear and her weighting function satisfies Assumption 2, then (3.8) is strictly violated for sufficiently small upside probabilities. Precisely, a CPT-agent agent with a linear value function and an inverse-S-shaped weighting function chooses any binary lottery with an upside probability less than p' over its expected value. In general, the effect of an increase in the skewness of a binary lottery on a CPT-agent's behavior depends the functional forms of her value and weighting function. This is summarized in the following proposition.

Proposition 8. *Suppose Assumption 2 holds, and let the choice set be* $\mathcal{C} = \{E, L(E, V, S)\}$.

(a) There exists some $S' \in \mathbb{R}$ *such that, for any* $S < S'$*, the CPT-agent chooses the safe option* E.

(b) There exists some $S'' \in \mathbb{R}$ *such that, for any* $S > S''$*, the following holds: If* (3.8) *holds in the limit of* p *approaching zero, then the CPT-agent chooses the safe option* E*. If* (3.8) *is strictly violated in this limit, however, the CPT-agent chooses the lottery* $L(E, V, S)$ *instead.*

While salience theory predicts that people choose a binary risk over its expected value if and only if it is sufficiently skewed (see Propositions 2 and 3), a CPT-agent might — depending on how concave her value function is and on how her weighting function is shaped — choose slightly right-skewed risks, but avoid extremely right-skewed ones. Since portfolios consisting of perfectly negatively correlated Mao pairs can be represented by binary lotteries, it follows that CPT — in contrast to salience theory — does not make clear-cut predictions on how behavior varies with skewness in the context of our experiment. Importantly, as shown in Appendix D of Dertwinkel-Kalt *et al.* (2020a), the salience predictions do not rely on the fact that we assume a linear value function, but indeed hold for any (weakly) concave value function. In this sense it is fair to compare salience theory with the version of CPT outlined in this section.

Appendix E: A Variant of Salience Theory Based on Visual Salience

E.1: The Visual Salience Model

In this section, we study an alternative extension of salience theory to portfolio selection problems, which assumes that salience distorts not the outcomes of a portfolio, but those of the underlying assets. We refer to this model as the visual salience model.

Definition 7 (Visual Salience-Weighted Utility)**.** *Fix a portfolio selection problem* $(\mathcal{A}, \mathcal{C})$*. The* visual salience-weighted utility *of the portfolio* $X(\alpha)$ *is given by*

$$U^s(X(\alpha)|\mathcal{C}, \mathcal{A}) = \int_{\mathbb{R}^2_{\geq 0}} (\alpha x_1 + (1-\alpha)x_2) \cdot \frac{\sigma(x_1, x_2)}{\int_{\mathbb{R}^2_{\geq 0}} \sigma(y_1, y_2)\ dF(y_1, y_2)}\ dF(x_1, x_2),$$

where $\sigma : \mathbb{R}^2_{\geq 0} \to \mathbb{R}_{>0}$ *is a salience function that is bounded away from zero.*

According to the visual salience model, only the computational, but not the comparative, complexity of a portfolio selection problems affects behavior in a non-trivial way. This is the exact opposite compared to the model proposed in Appendix C. By comparing these two different versions of salience theory, we add to the literature studying the fundamentals of salience theory (e.g., Dertwinkel-Kalt and Wenner, 2020). In the following subsections, we derive and test the predictions of the visual salience model for the portfolio selection problems in our experiment.

E.2: Portfolio Selection Problems with Skewed Assets

Throughout this subsection, we consider assets that form a Mao pair, and, to begin with, we assume that these two assets are perfectly negatively correlated.

Proposition 9. *Suppose the assets form a perfectly negatively correlated Mao pair, and apply the visual salience model.*

(a) Let $\mathcal{A} = \{\alpha, \frac{1}{2}\}$ *with* $\alpha \neq \frac{1}{2}$*. A salient thinker chooses* $X(\alpha)$ *over* $X(\frac{1}{2})$ *if and only if* $\alpha > \frac{1}{2}$ *holds.*

(b) *Let $\mathcal{A} = [0,1]$. A salient thinker chooses the portfolio $X(\alpha)$ with $\alpha = 1$.*

(c) *Consider the portfolio selection problem with $\mathcal{A}' = \{0,1\}$ that is essentially equivalent to the problem in Part (a). Here, a salient thinker chooses $X(\alpha)$ over $X(\frac{1}{2})$ only if $\alpha > \frac{1}{2}$, and, for any $\alpha > \frac{1}{2}$, she chooses $X(\alpha)$ over $X(\frac{1}{2})$ if and only if $S > \hat{S}$, where $\hat{S}$ is introduced in Proposition 2.*

Proof. PART (a). The agent chooses $X(\alpha)$ if and only if

$$(2\alpha-1)\left[\sigma\left(E-\sqrt{V}\sqrt{\frac{1-p}{p}}, E+\sqrt{V}\sqrt{\frac{1-p}{p}}\right)-\sigma\left(E+\sqrt{V}\sqrt{\frac{p}{1-p}}, E-\sqrt{V}\sqrt{\frac{p}{1-p}}\right)\right] > 0.$$

By the ordering property, the term in brackets is strictly positive, which yields the claim.

PART (b). As it is evident from the proof of Part (a), the difference in visual salience-weighted utility of an under-diversified portfolio and the diversified portfolio is monotonically increasing in α. Hence, a salient thinker invests as much as possible in the right-skewed asset.

PART (c). Follows immediately from Proposition 2. □

A first observation is that, if we apply the visual salience model to problems that are comparatively and/or computationally complex, the upside of the right-skewed asset and the downside of the left-skewed asset are salient, irrespective of how skewed the assets are. This implies that, when choosing between portfolios, any right-skewed portfolio is more attractive than the diversified one, while any left-skewed portfolio is less attractive than the diversified one. As a consequence, the skewness of the lotteries should affect behavior only if the outcomes of the lotteries are readily observable, which is clearly inconsistent with the data presented in Figure 3.7. Second, under the visual salience model, we would expect that, for rather symmetric Mao pairs, the right-skewed portfolios become more attractive due to computational complexity. This is clearly inconsistent with the evidence presented in Figure 3.37 in Appendix H.8.

Next, consider the case of the maximal positive correlation, which gives the following result.

Proposition 10. *Suppose that the assets form a maximally positively correlated Mao pair, and apply the visual salience model. There exists a threshold value $S' > 0$ such that the following holds:*

(a) *Let $\mathcal{A} = \{\alpha, \frac{1}{2}\}$ with $\alpha \neq \frac{1}{2}$. For any $S > S'$, a salient thinker chooses $X(\alpha)$ over $X(\frac{1}{2})$ if and only if $\alpha > \frac{1}{2}$ holds. For any $S < S'$, in contrast, a salient thinker chooses $X(\alpha)$ over $X(\frac{1}{2})$ if and only if $\alpha < \frac{1}{2}$. If $S = S'$, then a salient thinker is indifferent between both portfolios.*

(b) *Let $\mathcal{A} = [0,1]$. For any $S > S'$, a salient thinker chooses the portfolio $X(1)$, while for any $S < S'$ she chooses the portfolio $X(0)$. If $S = S'$, a salient thinker is indifferent between all portfolios.*

Suppose, in addition, that the salience function is homogeneous of degree zero.

(c) *Consider the portfolio selection problem with $\mathcal{A}' = \{0,1\}$ that is essentially equivalent to the problem in Part (a). There exist thresholds $\underline{S}, \overline{S} > 0$ with $\overline{S} > \underline{S}$, so that, for any $S \geq \overline{S}$, a salient thinker chooses $X(\alpha)$ over $X(\frac{1}{2})$ if and only if $\alpha > \frac{1}{2}$, while, for any $S \leq \underline{S}$, the opposite is true.*

Proof. PART (a). The salient thinker chooses $X(\alpha)$ over $X(\frac{1}{2})$ if and only if

$$(2\alpha-1)\Bigg[\sigma\left(E-\sqrt{V}\sqrt{\frac{1-p}{p}}, E-\sqrt{V}\sqrt{\frac{p}{1-p}}\right)+\sigma\left(E+\sqrt{V}\sqrt{\frac{p}{1-p}}, E+\sqrt{V}\sqrt{\frac{1-p}{p}}\right) \\ -\sigma\left(E+\sqrt{V}\sqrt{\frac{p}{1-p}}, E-\sqrt{V}\sqrt{\frac{p}{1-p}}\right)\Bigg] > 0.$$

By Proposition 5 (c) in Dertwinkel-Kalt and Köster (2020b), there exists some threshold $S' > 0$, so that the term in brackets is positive if and only if $S > S'$, which yields the claim.

PART (b). By the proof of Part (a), the difference in visual salience-weighted utility of an under-diversified and the diversified portfolio is monotonically increasing in α if and only if $S > S'$. Hence, for any $S > S'$, a salient thinker invests as much as possible in the right-skewed lottery, while, for any $S < S'$, she invests as much as possible in the left-skewed lottery.

PART (c). Follows immediately from Proposition 6. □

The visual salience model predicts that, in simple or computationally complex problems with rather symmetric Mao pairs, right-skewed portfolios are significantly more attractive under the perfectly negative than under the maximal positive correlation. This prediction is inconsistent with the data presented in Figure 3.12: if there is a correlation effect at all, then right-skewed portfolios are more attractive under the maximal positive than under the perfectly negative correlation. While the model introduced in Appendix C is consistent with this empirical fact,[25] it predicts that the correlation effect is the same for simple and computationally complex problems. While we cannot reject the null-hypothesis of identical effects (p-value $= 0.119$, OLS with clustered standard errors), Figure 3.12 suggests that this test may lack statistical power.

[25] As we show in Appendix C, however, this finding imposes restrictions on the functional form of the salience function. If we want to account for the correlation effects, the salience function cannot be homogeneous of degree zero (Assumption 1).

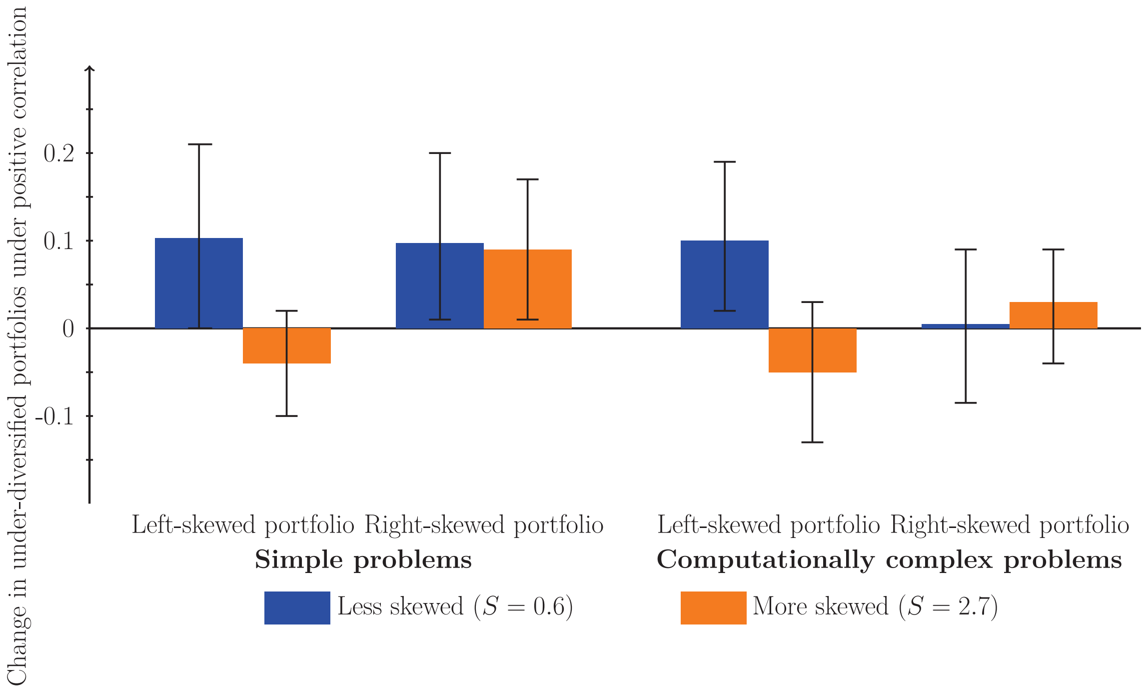


Figure 3.12: *This figure depicts the share of under-diversified portfolios chosen under the maximal positive correlation minus the share of under-diversified portfolios chosen under the perfectly negative correlation for simple and computationally complex problems in the treatment Skewed. We report 95%-confidence intervals with standard errors being clustered at the subject level.*

The most striking difference compared to the model in Appendix C is that the visual salience model predicts corner solutions in the computationally and comparatively complex problems: Under the perfectly negative correlation, subjects are predicted to invest everything in the right-skewed asset, irrespective of how skewed the Mao lotteries are. Under the maximal positive correlation, we obtain the exact same prediction for sufficiently skewed Mao pairs, while for sufficiently symmetric ones, the visual salience model predicts that subjects invest everything in the left-skewed asset. Both predictions are inconsistent with the data presented in Figure 3.13.

E.3: Portfolio Selection Problems with Symmetric Risks

Applied to (binary) twin problems the visual salience model makes the following predictions.

Proposition 11. *Consider a (binary) twin problem, and apply the visual salience model.*

(a) A salient thinker is indifferent between both portfolios in (P.1) of any binary twin problem, while she strictly prefers the diversified portfolio in (P.2) of any binary twin problem.

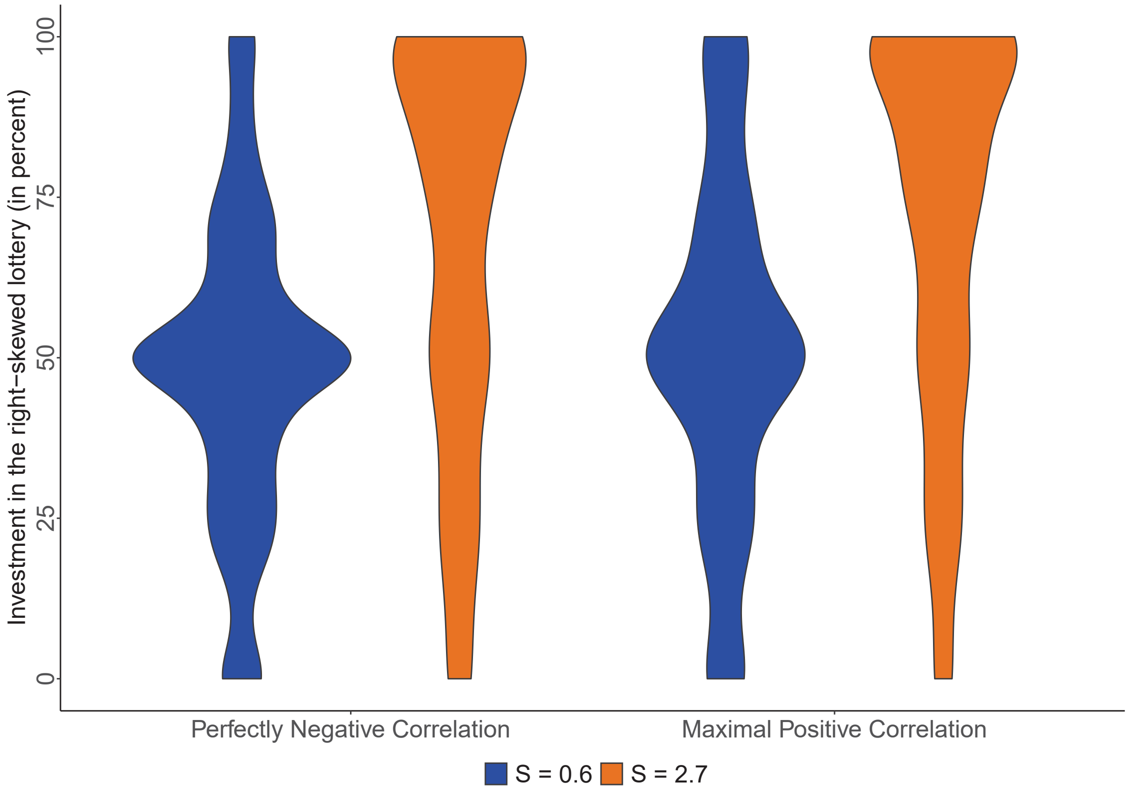


Figure 3.13: *This figure depicts the kernel density of the percentage share invested in the right-skewed lottery of a Mao pair in the computationally and comparatively complex portfolio selection problems.*

(b) *A salient thinker is indifferent between all portfolios in (P.1) of any twin problem. In (P.2) of a twin problem she chooses either $\alpha = 1$ if $\gamma > 1$ or $\alpha = \min \mathcal{A}$ if $\gamma < 1$.*

(c) *Consider the portfolio selection problem with $\mathcal{A}' = \{0, 1\}$ that is essentially equivalent to the problems in Part (a). Here, a salient thinker always chooses the diversified portfolio.*

Proof. For $\gamma \in [0, 3]$, a salient thinker chooses a portfolio with $\alpha \neq \frac{\gamma}{1+\gamma}$ if and only if

$$\sqrt{V}\frac{1}{1+\gamma}\left[\alpha - \frac{\gamma}{1+\gamma}\right]\left[\sigma(E+\sqrt{V}, E-\gamma\sqrt{V}) - \sigma(E-\sqrt{V}, E+\gamma\sqrt{V})\right] > 0.$$

By diminishing sensitivity, the term in brackets is positive if and only if $\gamma > 1$ and, by symmetry, it is zero for $\gamma = 0$. Hence, for $\gamma > 1$ only portfolios with $\alpha > \frac{\gamma}{1+\gamma}$ are more attractive than the diversified one, while for $\gamma < 1$ only portfolios with $\alpha < \frac{\gamma}{1+\gamma}$ are more attractive than the diversified one. Part (a) follows from the fact that $\frac{\gamma}{1+\gamma} > \frac{1}{2}$ if and only if $\gamma > 1$. Part (b) holds as the difference in salience-weighted utility monotonically increases in α if and only if $\gamma > 1$. Finally, Part (c) follows from Proposition 4, as for simple problems both versions of salience theory are identical. □

Most strikingly, the visual salience model predicts that in the binary twin problems we should observe more diversification in (P.2) than in (P.1), which stands in stark contrast to the actually observed behavior (see Figure 3.8 in the main text). Moreover, the visual salience model again predicts corner solutions in the computationally and comparatively complex problems, which is inconsistent with the data presented in Figure 3.14

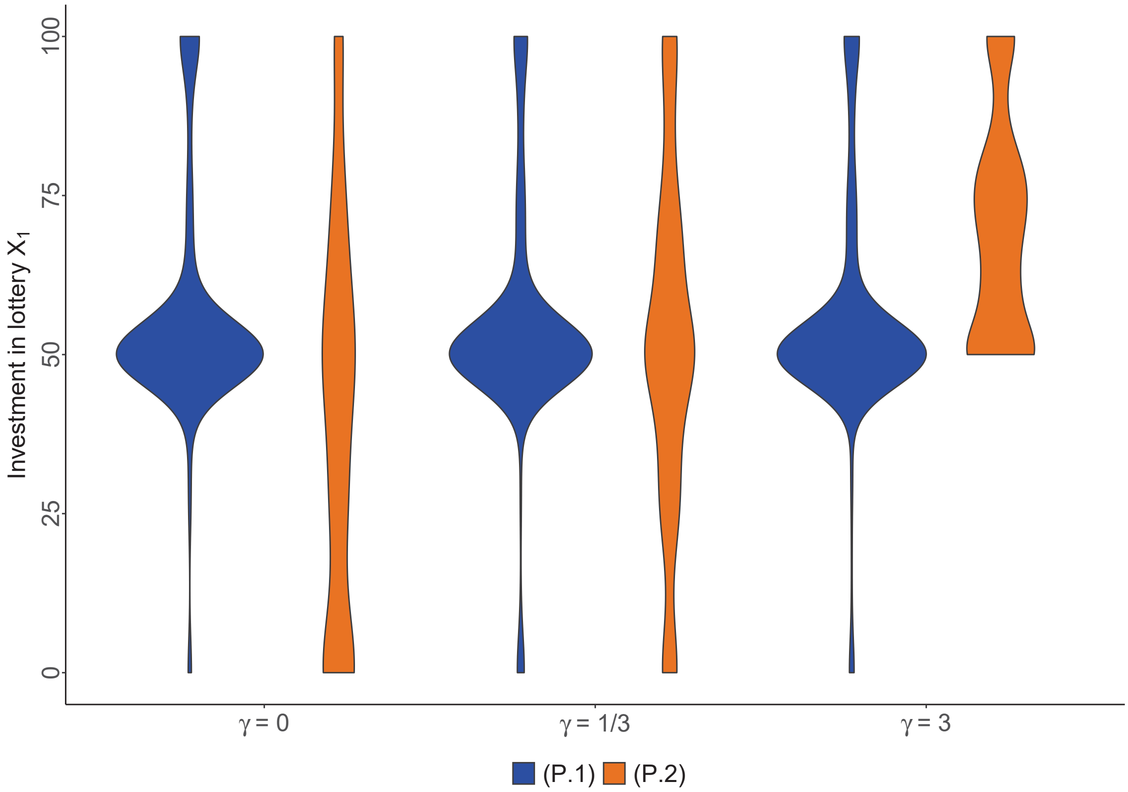


Figure 3.14: *This figure depicts the kernel density of the percentage share invested in* X_1 *for the computationally and comparatively complex twin problems.*

Appendix F: (Naive) Diversification as the Reference Point

Consider the salience model from Section 3.2 and Appendix C, respectively, with the one exception that in comparatively and/or computationally complex problems the reference point is given by the (naively) diversified portfolio. As we argue in the following, this adjustment is not sufficient to reconcile salience theory with the data.

Portfolio selection problems with skewed assets. Since the naively diversified portfolio coincides with the diversified portfolio and since it constitutes the "center" of the choice set, the reference point does not change and all propositions presented in Appendix C remain valid.

Portfolio selection problems with symmetric assets. By the same argument as for the problems with skewed assets, the results on (P.1) of a (binary) twin problem remain to hold.

Consider (P.2) of a binary twin problem. If the reference point is given by the diversified portfolio, then the salient thinker still chooses the diversified portfolio. If, in contrast, the reference point is given by the naively diversified portfolio, a salient thinker might be indifferent between the diversified portfolio and the naively diversified portfolio. Precisely, if the salience function satisfies $\sigma(E+\frac{1-\gamma}{2}\sqrt{V}, E+\frac{1-\gamma}{2}\sqrt{V}) = \sigma(E-\frac{1-\gamma}{2}\sqrt{V}, E-\frac{1-\gamma}{2}\sqrt{V})$ (e.g. if it is homogeneous of degree zero), then the salience-weighted utility derived from the naively diversified portfolio is also E. Hence, when assuming that the reference point is given by the naively diversified portfolio, salience theory can rationalize naive diversification in binary twin problems. But even this variant of the salience model cannot explain why a large majority of subjects breaks this tie in favor of the naively diversified portfolio.

Consider (P.2) of a twin problem with a continuous choice set. If the reference point is given by the diversified portfolio, a salient thinker would always choose the diversified portfolio, which is clearly inconsistent with the data. Now let the reference point be given by the naively diversified portfolio. If $\gamma < 1$, the reference point does not change and the results derived in Appendix C still hold. For $\gamma > 1$, the reference point changes, and a salient thinker — who evaluates a given portfolio relative to the naively diversified portfolio — would choose

$$\alpha \in \left(\frac{\gamma}{1+\gamma}, \frac{3\gamma-1}{1+\gamma}\right).$$

In particular, if $\gamma = 3$, then the model predicts that subjects choose $\alpha \in (\frac{3}{4}, 1)$, which is clearly inconsistent with the data presented in Figure 3.14.

Appendix G: Experimental Details

G.1: Derivation of the Experimental Mao Pairs

Each of the six Mao pairs, which we implemented in the experiment, is given by two lotteries $L(E_i, V_i, S_j)$ and $L(E_i, V_i, -S_j)$ with $i \in \{1, 2, 3\}$ and $j \in \{h, l\}$ that are uniquely constructed as follows. Take two skewness values $S_h > S_l$ such that the difference $S_h - S_l$ is maximized subject to the constraint that the corresponding probabilities (expressed in percentages) represent strictly positive integers. For these two skewness values, S_h and S_l, we take those three pairs of expected values and variances, (E_i, V_i) for $i \in \{1, 2, 3\}$, so that the resulting outcomes of the lotteries $L(E_i, V_i, S_h)$ and $L(E_i, V_i, S_l)$ are non-negative integers with the expected values E_1, E_2, and E_3 being as small as possible. The conversion rate is then selected in a way that gives subjects (for our incentivization) slightly above-average expected earnings.

Now, we obtain the Mao pairs that we used as follows. For each pair of expected value and variance, (E_i, V_i), we need two different skewness levels S_h and S_l and therefore two different probabilities p_h and p_l, such that

$$100p_j = \frac{1}{2} + \frac{S_j}{2\sqrt{4+S_j^2}} \in \mathbb{N} \quad \text{for } j \in \{h, l\}.$$

By Lemma 3, for a given p_j, the outcomes of the corresponding lottery, x_1 and x_2, are given by

$$x_1 = E_i - \sqrt{\frac{V_i(1-p_j)}{p_j}} \qquad \text{and} \qquad x_2 = E_i + \sqrt{\frac{V_i p_j}{1-p_j}}$$

These two outcomes are therefore integers if and only if $\frac{V_i(1-p_j)}{p_j}$ and $\frac{V_i p_j}{1-p_j}$ are quadratic integers for $i \in \{1, 2, 3\}$ and $j \in \{h, l\}$. Then, also V_i has to be a quadratic integer, and by prime factorization it can be seen that $\frac{p_j}{1-p_j}$ have to be quadratic rational numbers for both $j \in \{h, l\}$. In addition, $100p_j$ needs to be a natural number. It is straightforward to see that there are only three values for p_j admissible, namely, $p_j \in \{0.36, 0.2, 0.1\}$.

By the criterion to maximize the difference in skewness levels, we take the values $p_l = 0.36$ and $p_h = 0.1$. Now, V_i needs to a quadratic integer that is divisible by 6 and 8 as x_1 and x_2 need to be integers for $p_l = 0.36$ as well as $p_h = 0.1$. This gives that V_i has to be a multiple of 144. As E_i should be smallest possible, we take for V_i the three smallest quadratic integers that are multiples of 144, that is, $V_1 = 144$, $V_2 = 4 \cdot 144$, and $V_3 = 9 \cdot 144$. The corresponding expected values directly follow from the constraint that the payoffs x_1 and x_2 need to be non-negative and smallest possible for each $p_l = 0.36$ and $p_h = 0.1$, which is equivalent to the constraint that for each of the variances, the lower payoff of the left-skewed lottery equals zero for the lower downside probability, that is, $x_2 = 0$ for $p_h = 0.1$. This yields $E_1 = 36$, $E_2 = 72$, and $E_3 = 108$.

Finally, we decided to use a conversion rate that results — given that the experiment takes about one hour — in a wage per hour that is slightly above the average hourly wage paid in the laboratory of the University of Cologne, which lies between 15 and 20 Euro. Then, 4ECU=1 Euro is the only candidate, which yields an average wage of 18 Euro per subject for participation.

G.2: Instructions

Information about the experiment

Please do not use your mobile nor pan and paper from now on and throughout the entire experiment, and do not talk with other participants. Please read the following instructions carefully. Should you have any questions at any point in time please raise your hand. An experimenter will then answer your questions at your seat.

In this experiment you can earn an experimental currency (Taler) which will be converted into Euro at the end of the experiment. The conversion rate is

1 Euro = 4 Taler.

This experiment is about investment decisions. These decisions regard your personal preferences, that is, there are no right or wrong choices.

You will have to choose between different portfolios, which consist of two options, A and B. Both Option A and B represent a lottery, that is, the outcome of each option depends on a turn of the wheel of fortune with 100 fields (simulated by your computer), where each field is hit with the same probability.

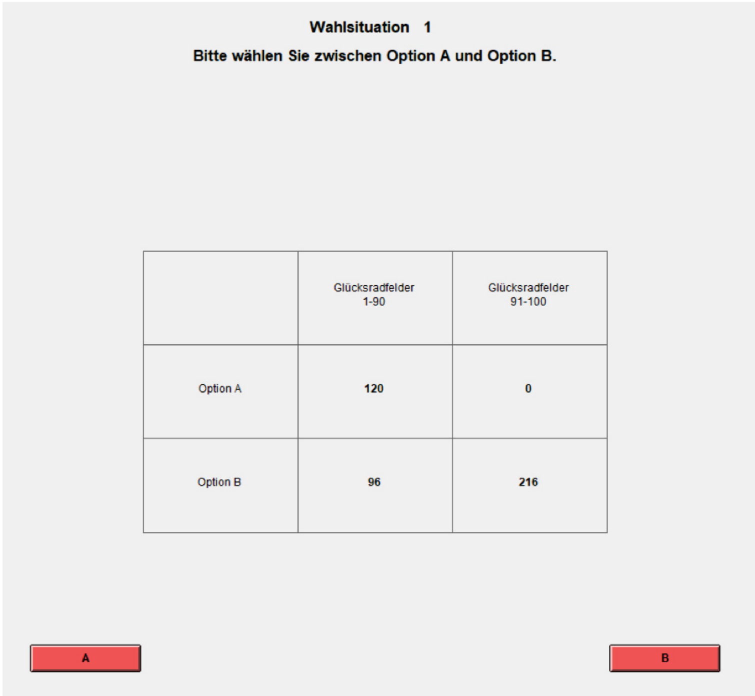
Each portfolio is described by how many of 100 points are invested into Option A and how many are invested into Option B. If all 100 points are invested in Option A, you exactly receive the lottery, which is described by Option A. If all 100 points are invested in Option B, you exactly receive the lottery, which is described by Option B. If half of the points are invested in Option A and half of the points are invested in Option B, you receive for each field of the wheel of fortune half of the respective sum Option A gives and half of the respective sum that Option B gives. If, in general, X points are invested in Option A and 100-X in Option B, you receive for each field of the wheel of fortune exactly X% of the sum Option A and (100-X)% of the sum Option B pays.

Throughout the experiment, you will make three types of investment choice: type-1-decisions, type-2-decisions, and type-3-decisions. In a type-1-decision, you choose between two options that invest everything into Option A or into Option B, respectively, that means you simply choose between Option A and Option B. In a type-2-decision, you again choose between two Portfolios, 1 and 2, but which consist to different parts of Options A and B. In a type-3-decision, you can freely choose how many points X you would like to invest into Option A and how many points 100-X you would like to invest into Option B. Here, you can invest any integer number between 0 and 100 into each of the options. Throughout the experiment, you will overall make between 30 and 40 investment decisions, which are divided into three blocks.

In the following we show you three exemplary choices. Please study them carefully.

Figure 3.15: Instructions, translated into English (first page).

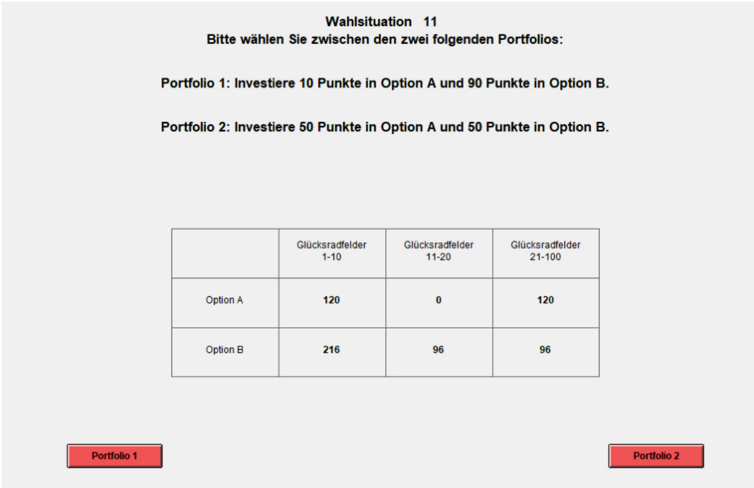
Example of a type-1-decision:



If the wheel of fortune stops on fields 1-90 (that corresponds to a 90% probability), you obtain when selecting Option A exactly 120 Taler and when selecting B exactly 96 Taler. If the wheel of fortune stops on fields 91-100 (that corresponds to a 10% probability, you obtain when selecting Option A exactly 0 Taler and when selecting B exactly 216 Taler.

Figure 3.16: Instructions, translated into English (second page).

Example of a type-2-decision:



If the wheel of fortune stops on fields 1-10, Option A pays exactly 120 Taler and Option B pays exactly 216 Taler. If the wheel of fortune stops on one of the fields 11-20, Option A pays exactly 0 Taler and Option B pays exactly 96 Taler. If the wheel of fortune stops on fields 21-100, Option A pays exactly 120 Taler and Option B pays exactly 96 Taler.

You can choose between two portfolios, and your payoff depends on your portfolio choice and on a rotation of the wheel of fortune. If the wheel of fortune stops on fields 1-10, you obtain with Portfolio 1 exactly 120*10% + 216*90% = 206,40 Taler and with Portfolio 2 exactly 120*50% + 216*50% = 168 Taler.

If the wheel of fortune stops on fields 11-20, you obtain with Portfolio 1 exactly 0*10% + 96*90% = 86,40 Taler and with Portfolio 2 exactly 0*50% + 96*50% = 48 Taler.

If the wheel of fortune stops on fields 11-20, you obtain with Portfolio 1 exactly 120*10% + 96*90% = 98,40 Taler and with Portfolio 2 exactly 120*50% + 96*50% = 108 Taler.

Figure 3.17: Instructions, translated into English (third page).

Example of a type-3-decision:

Wahlsituation 8

Bitte entscheiden Sie wie viele Ihrer 100 Punkte Sie in Option A investieren möchten. Alle übrigen Punkte werden in Option B investiert.

	Glücksradfelder 1-36	Glücksradfelder 37-72	Glücksradfelder 73-100
Option A	**90**	**40**	**90**
Option B	**104**	**54**	**54**

Wie viele Punkte möchten Sie in Option A investieren?

If the wheel of fortune stops on fields 1-36, Option A pays exactly 90 Taler and Option B pays exactly 104 Taler. If the wheel of fortune stops on fields 37-72, Option A pays exactly 40 Taler and Option B pays exactly 54 Taler. If the wheel of fortune stops on fields 73-100, Option A pays exactly 90 Taler and Option B pays exactly 54 Taler.

Your payoff depends on your investment decision and on a rotation of the wheel of fortune. If you enter "100" in the field above, you get with Option A the following: for fields 1-36 and 72-100 you get 90 Taler and for fields 37-72 you get 40 Taler.

If you enter "0" in the field above, you get with Option B the following: for fields 1-36 you get 104 Taler and for fields 37-100 you get 54 Taler.

If you enter "50" in the field above, you obtain half of Option A plus half of Option B: for fields 1-36 you obtain 50%*90 + 50%*104 = 97 Taler, for fields 37-72 ypu obtain 50%*40 + 50%*54 = 47 Taler, and for fields 73-100 you obtain 50%*90 + 50%*54 = 72 Taler.

If, in general, you enter number "X" above, you get for fields 1-36 payoff X%*90 + (100-X)%*104 Taler, for fields 37-72 payoff X%*40 + (100-X)%*54 Taler, and for fields 73-100 payoff X%*90 + (100-X)%*54 Taler.

Figure 3.18: Instructions, translated into English (fourth page).

Payoff for the investment decisions: At the end of the experiment your computer will randomly select one of your investment decisions. This decision could be a type-1-decision, a type-2-decision, or a type-3-decision. This randomly selected decision is payoff-relevant. You payoff will be then determined by a rotation of the wheel of fortune.

Suppose for instance that the decision from the second example is payoff-relevant, and that the wheel of fortune stops on field 3. If you have chosen portfolio 1, you earn 206.40 Taler, that is 51.60 Euro. If you have selected portfolio 2, you receive 42 Euro.

Or suppose the decision from the third example is payoff-relevant, and suppose the wheel of fortune stops on field 74. Depending on how many points X you have invested into Option A, you receive X%*90 + (100-X)%*54 Taler. If you have, for instance, invested X = 25 into Option A, you get 15.75 Euro.

Practice questions and additional questions: At the end of each block you have to answer between zero and three practice questions: for each correctly solved practice question you earn an additional Taler.

At the end of the experiment you have to answer some additional questions with which you can earn some additional money. How much exactly you can earn here you will learn directly on your computer.

In addition, for your participation in this experiment you earn a show-up fee of 4 Euro.

End of the Experiment: At the end of the experiment you will see a short questionnaire asking for, for instance, your age. As soon as you have completed this, please wait until all participants have finished the experiment. From this point on you are allowed to use your mobile phones, but not before. In case you use it before we might exclude you from the experiment.

As soon as all participants have completed the experiment, we call you and you receive your overall payoff, which consist of the 4 Euro for participation, your payoff of the randomly selected investment decisions, and the payoff of the practice questions and the additional questions.

Notes:

1) On your computer in the upper right corner you see for each question the remaining time. Please answer all questions within the remaining time.

2) Instead of a comma, please use a point for decimal numbers. Thus, please write 2.5 instead of 2,5.

Figure 3.19: Instructions for Experiment 2, translated into English (fifth part).

G.3: Screenshots

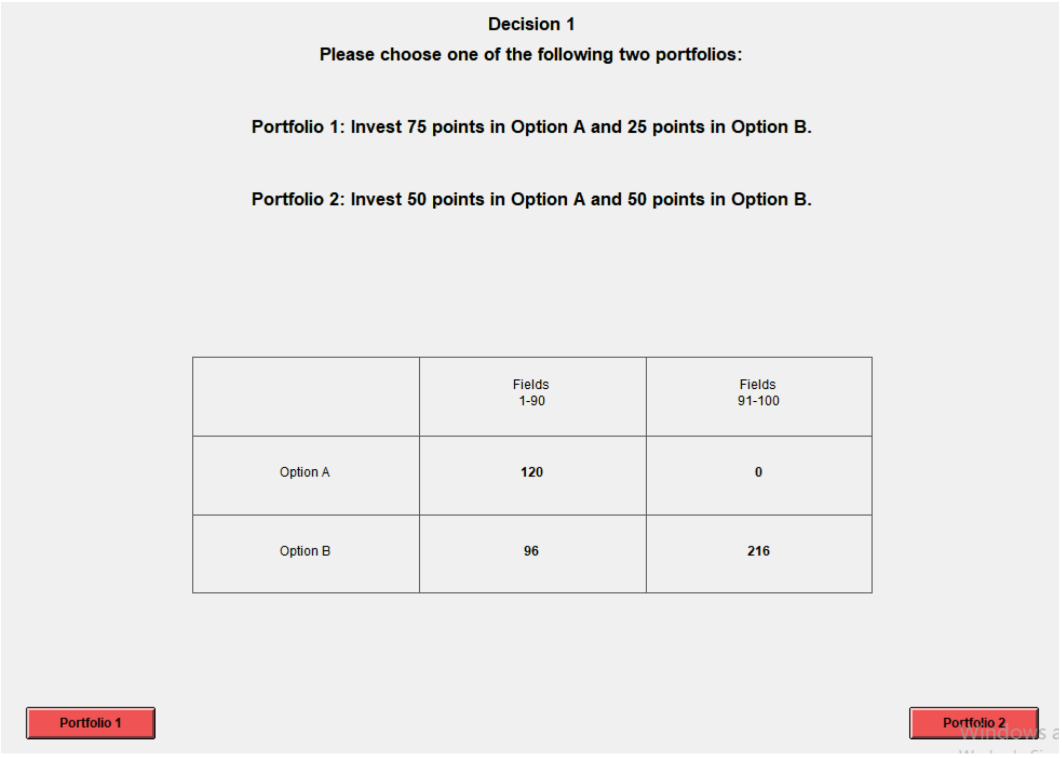


Figure 3.20: *Decision screen in Block 1 of the treatment Skewed.*

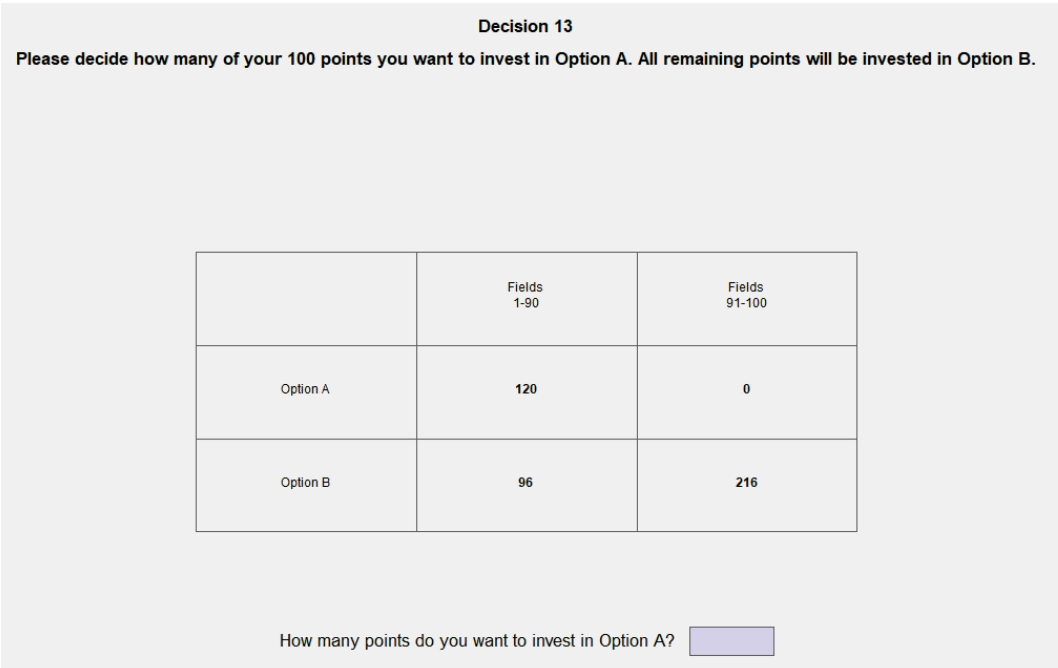


Figure 3.21: *Decision screen in Block 2 of the treatment Skewed.*

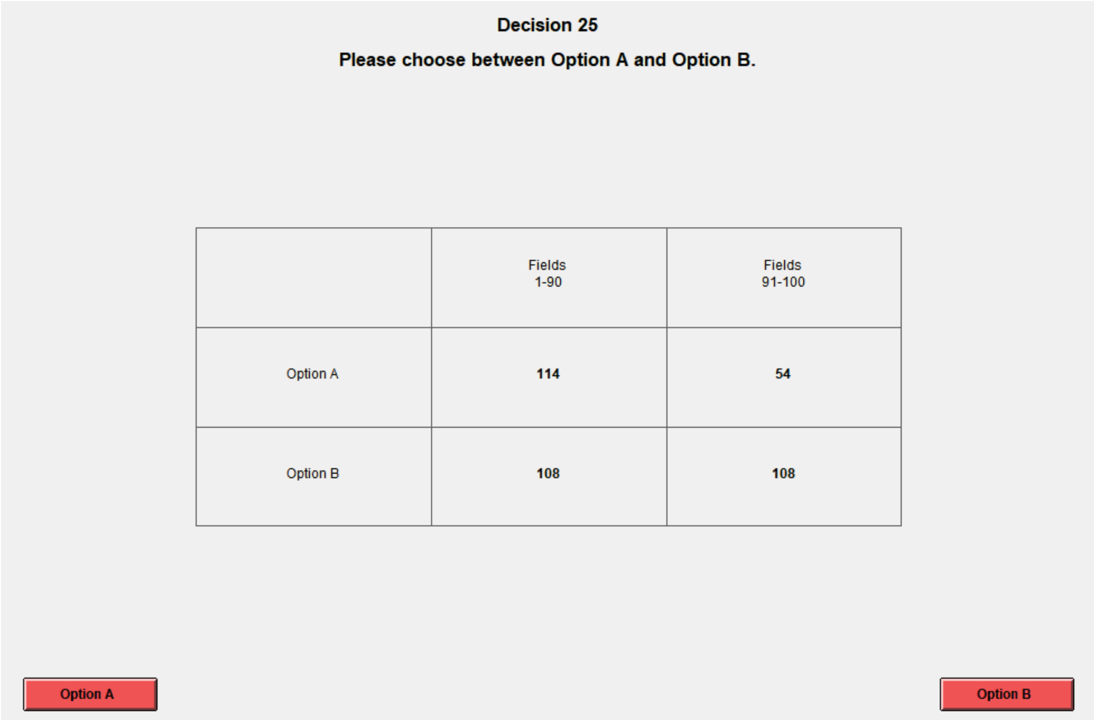


Figure 3.22: *Decision screen in Block 3 of the treatment Skewed.*

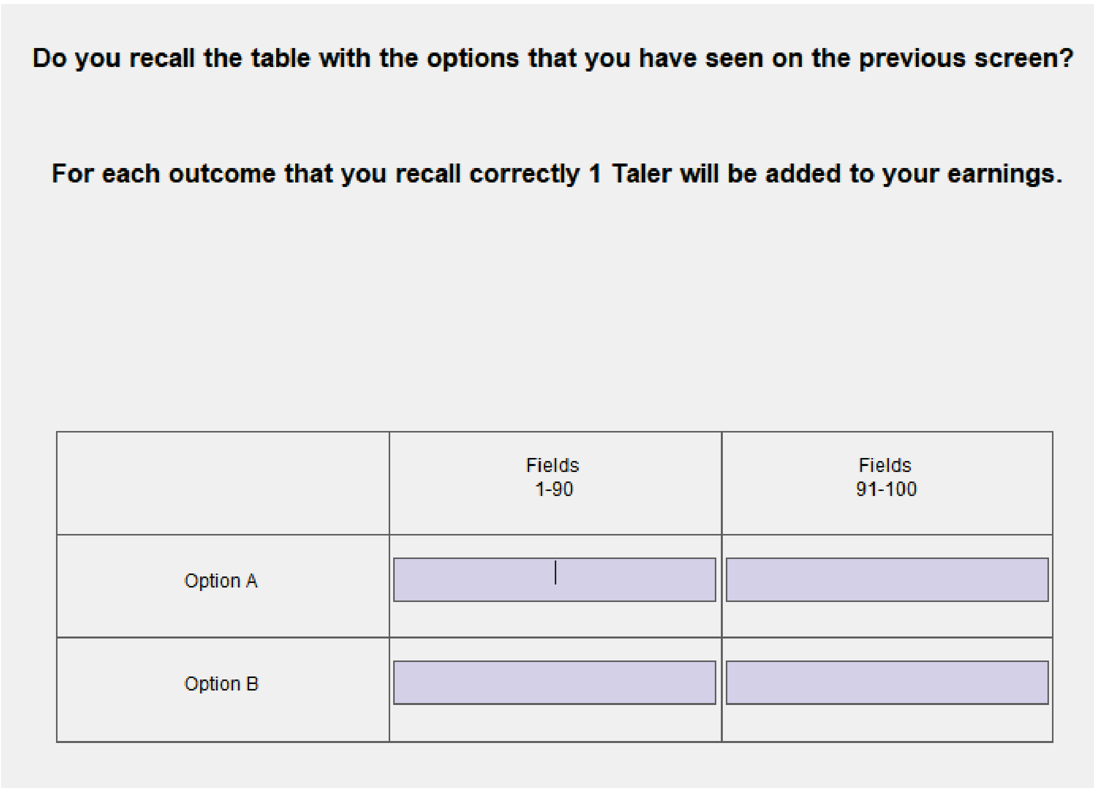


Figure 3.23: *Decision screen of the surprise memory task.*

G.4: Additional Questions Used in the Experiment

Questions of a modified cognitive reflection test. The following questions represent slight modifications of the cognitive reflection tests as used by Frederick (2005) and Primi *et al.* (2016).

1. If 10 machines take 10 minutes to make 10 nails, how many minutes do 100 machines need to make 100 nails?

2. A part of a pond is covered with water lilies. Every day the area covered with water lilies doubles. If it takes 24 days until the whole pond is covered with water lilies, how many days does it take until half of the pond is covered with water lilies?

3. If three elves can wrap three presents in one hour, how many elves does it take to wrap six presents in two hours?

4. Jerry has both the 15th best and the 15th worst grade in his class. How many students are in the class?

5. In a sports team tall members are three times as likely to win medals as short members. This year the team won 60 medals in total. How many medals were won by short team members?

Financial literacy questions. The questions are taken from Lusardi and Mitchell (2011).

1. Suppose you had 100 Euro in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?

[Options: “More than 102 Euro”, “Exactly 102 Euro”, “Less than 102 Euro”.]

2. Suppose you had 100 Euro in a savings account and the interest rate was 20% per year and you never withdraw money or interest payments. After 5 years, how much would you have on this account in total?

[Options: “More than 200 Euro”, “Exactly 200 Euro”, “Less than 200 Euro”.]

3. Imagine that the interest rate on your savings account was 1% per year and the inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account?

[Options: "More than today", "As much as today", "Less than today".]

4. Assume a friend inherits 10.000 Euro today and his brother inherits 10.000 Euro three years from now. There is a positive interest rate. Who is richer because of the inheritance?

[Options: "My friend", "Her brother", "Both are equally rich".]

5. Suppose that your income and all prices double in the next year. How much will you be able to buy with your income?

[Options: "More than today", "As much as today", "Less than today".]

Appendix H: Additional Experimental Results

H.1: Additional Results on Portfolio Selection Problems with Skewed Assets

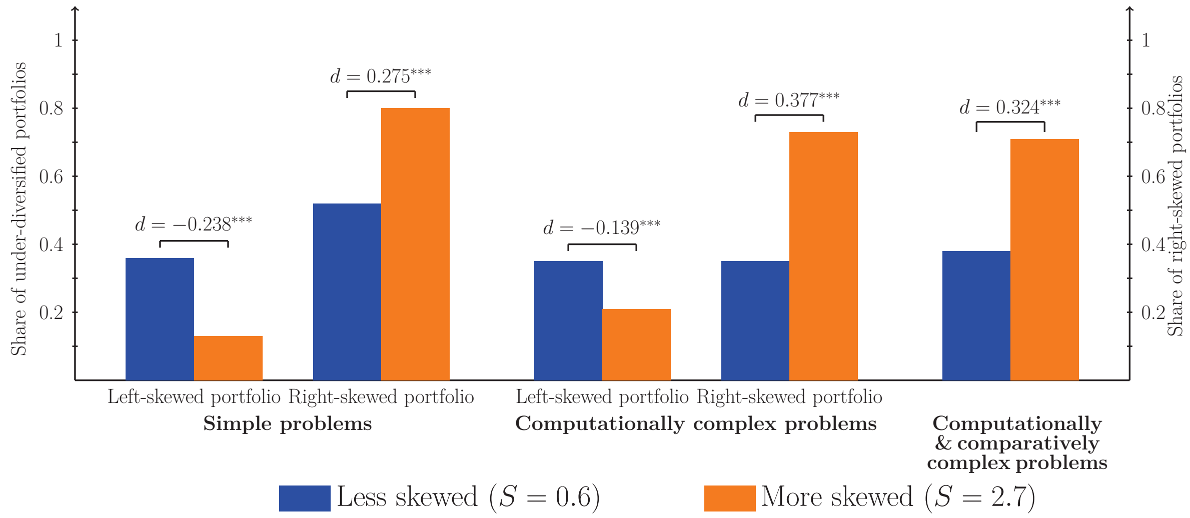


Figure 3.24: *This figure illustrates the experimental results for the treatment Skewed under the maximal positive correlation. We depict the share of under-diversified portfolios chosen in simple and computationally complex problems as well as the share of right-skewed portfolios chosen in computationally and comparatively complex problems. The data for problems with a binary choice set is presented separately for right- and left-skewed under-diversified portfolios. We further split the sample by less and more skewed Mao pairs. We also report the results of t-tests with standard errors being clustered at the subject level. Significance level: ***: 1%.*

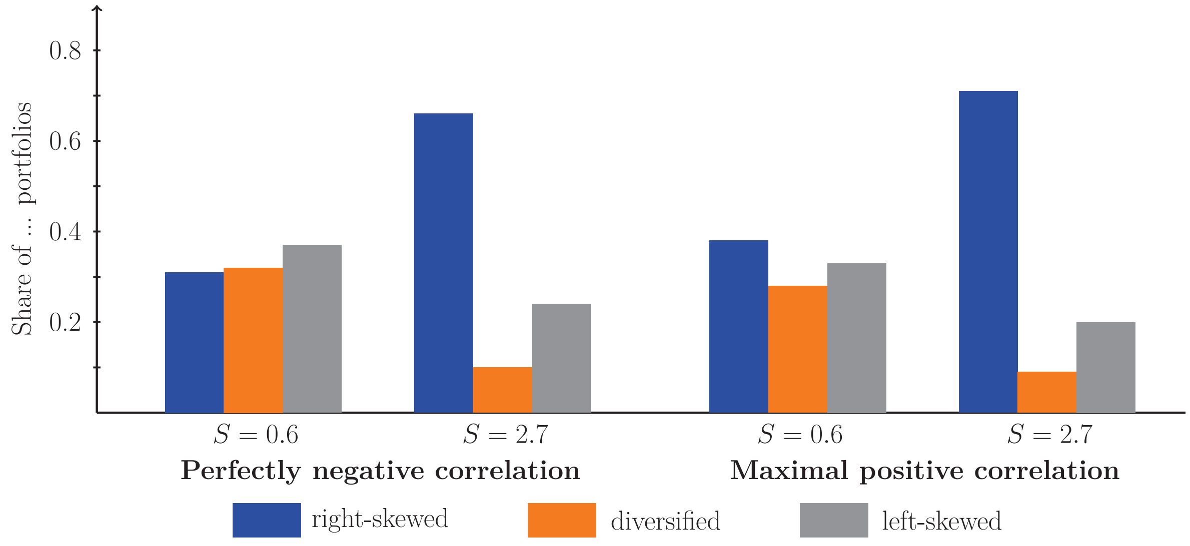


Figure 3.25: *This figure illustrates the share of the different portfolio types chosen in the computationally and comparatively problems in the treatment Skewed.*

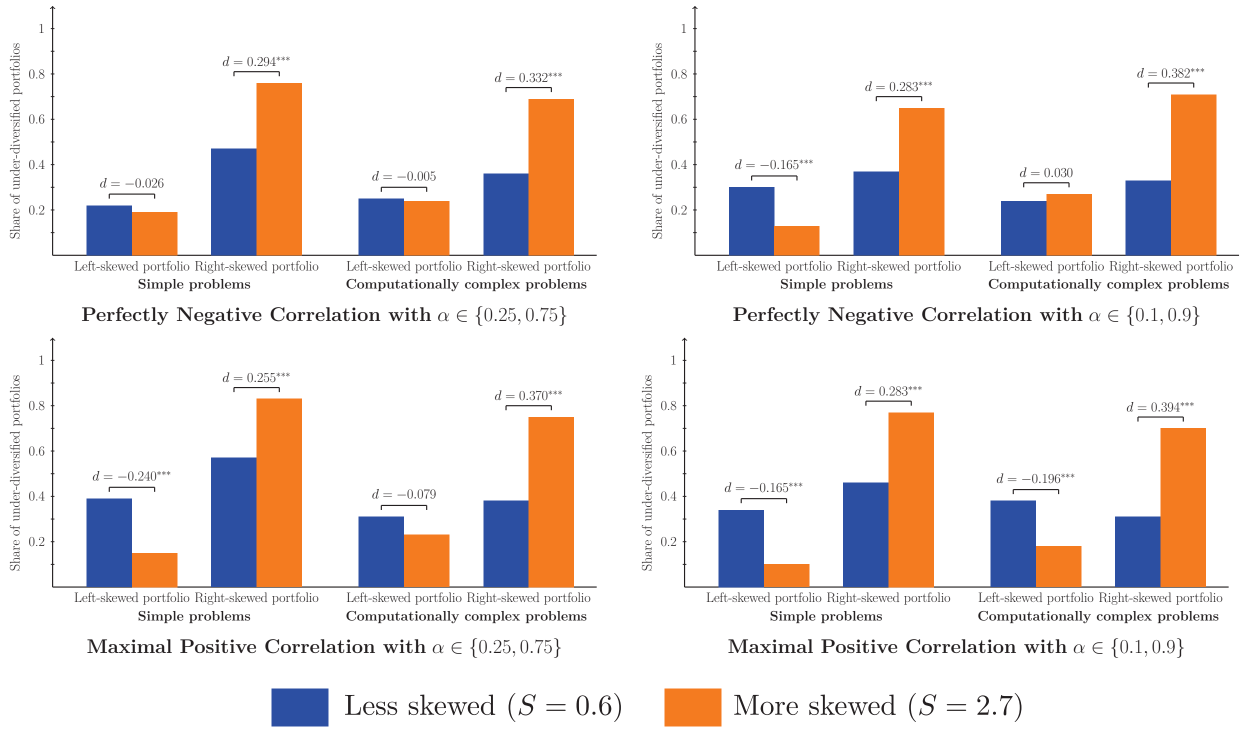


Figure 3.26: *This figure illustrates the share of under-diversified portfolios chosen the simple and computationally complex problems of the treatment Skewed for moderate portfolios with* $\alpha \in \{0.25, 0.75\}$ *and extreme portfolios with* $\alpha \in \{0.1, 0.9\}$. *The data is presented separately for right-skewed and left-skewed under-diversified portfolios consisting of less or more skewed Mao lotteries. We report the results of t-tests with standard errors clustered at the subject level. Significance level: ***: 1%.*

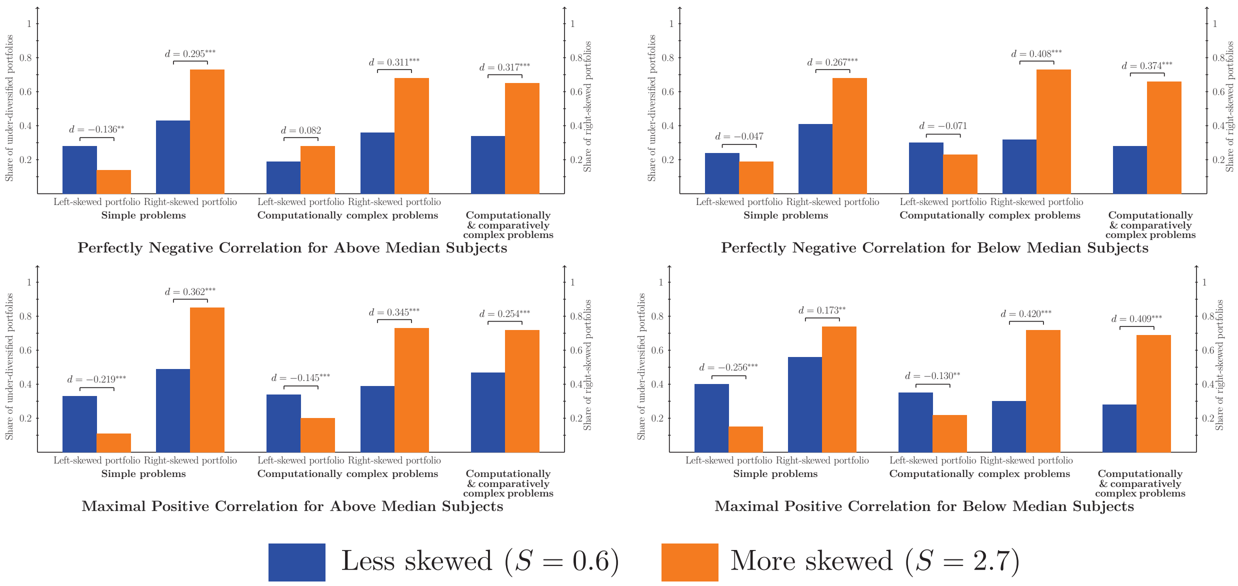


Figure 3.27: *This figure illustrates the share of under-diversified portfolios chosen in the simple and computationally complex problems as well as the share of right-skewed portfolios chosen in the computationally and comparatively complex problems of the treatment Skewed for subjects with above and below median cognitive skills. The data is presented separately for right-skewed and left-skewed under-diversified portfolios consisting of less or more skewed Mao lotteries. We report the results of t-tests with standard errors clustered at the subject level. Significance level: **: 5%, ***: 1%.*

H.2: Additional Results on Problems with Symmetric Assets

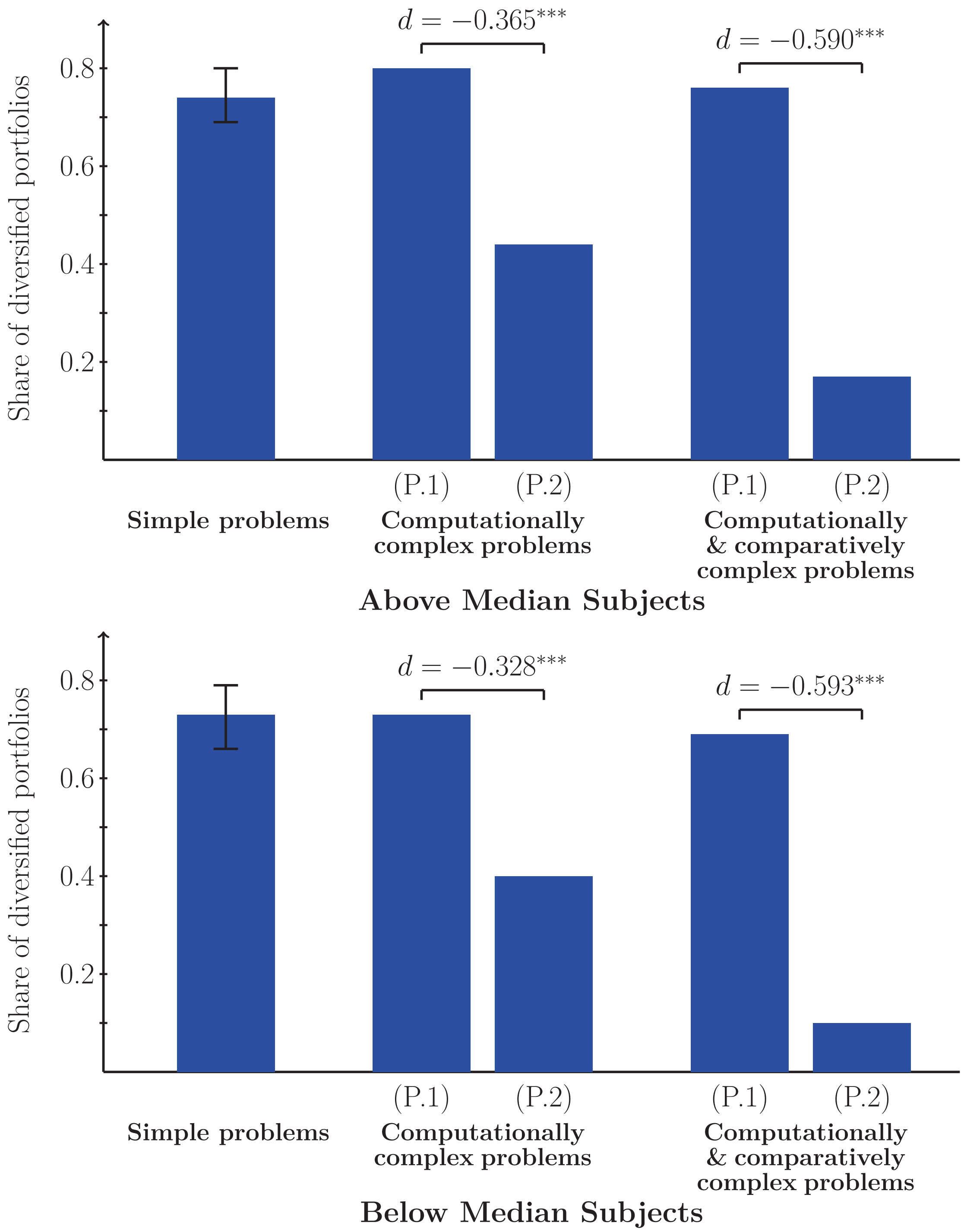


Figure 3.28: *This figure illustrates the share of diversified portfolios chosen in the different problems of the treatment Symmetric for above and below median subjects in terms of cognitive skills. The data for the comparatively and/or computationally complex problems is presented separately for (P.1) and (P.2) of a twin problem. For simple problems, we report the 95%-confidence interval with standard errors being clustered at the subject level. For complex problems, we present the results of t-tests, again with clustered standard errors. Significance level: ***: 1%.*

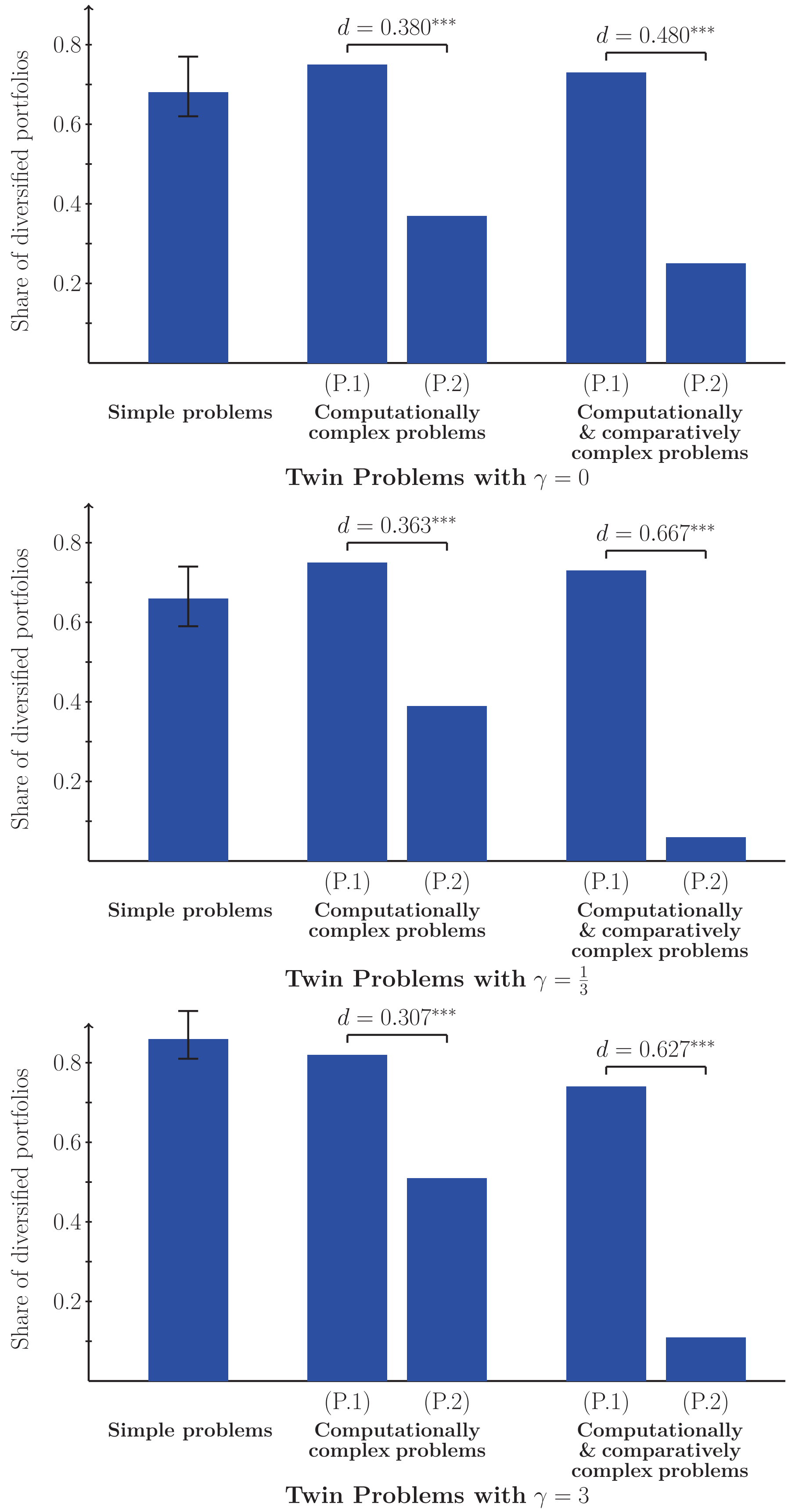


Figure 3.29: *This figure illustrates the share of diversified portfolios chosen in the treatment Symmetric for different types of twin problems. The data for the comparatively and/or computationally complex problems is presented separately for (P.1) and (P.2) of a twin problem. For simple problems, we report the 95%-confidence interval with standard errors being clustered at the subject level. For complex problems, we present the results of t-tests, again with clustered standard errors. Significance level: ***: 1%.*

H.3: Response Times and Complexity

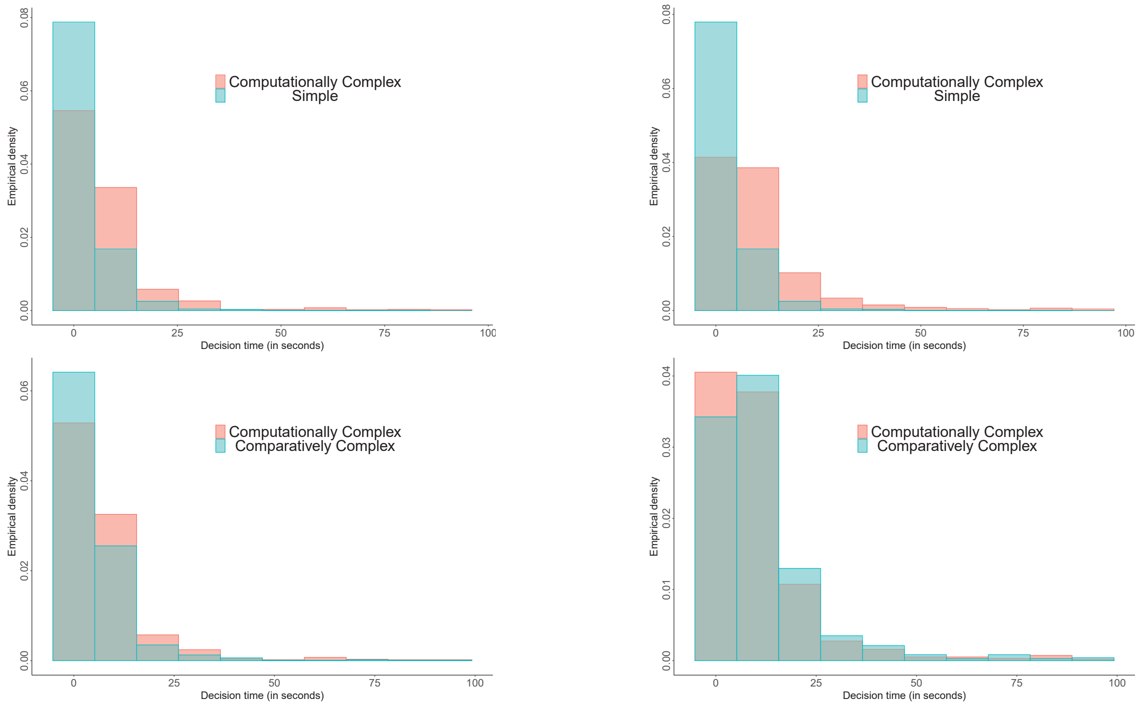


Figure 3.30: *The figure illustrates the distribution of decision times in Symmetric. The graphs in the left column refer to (P.1) and the graphs in the right column refer to (P.2).*

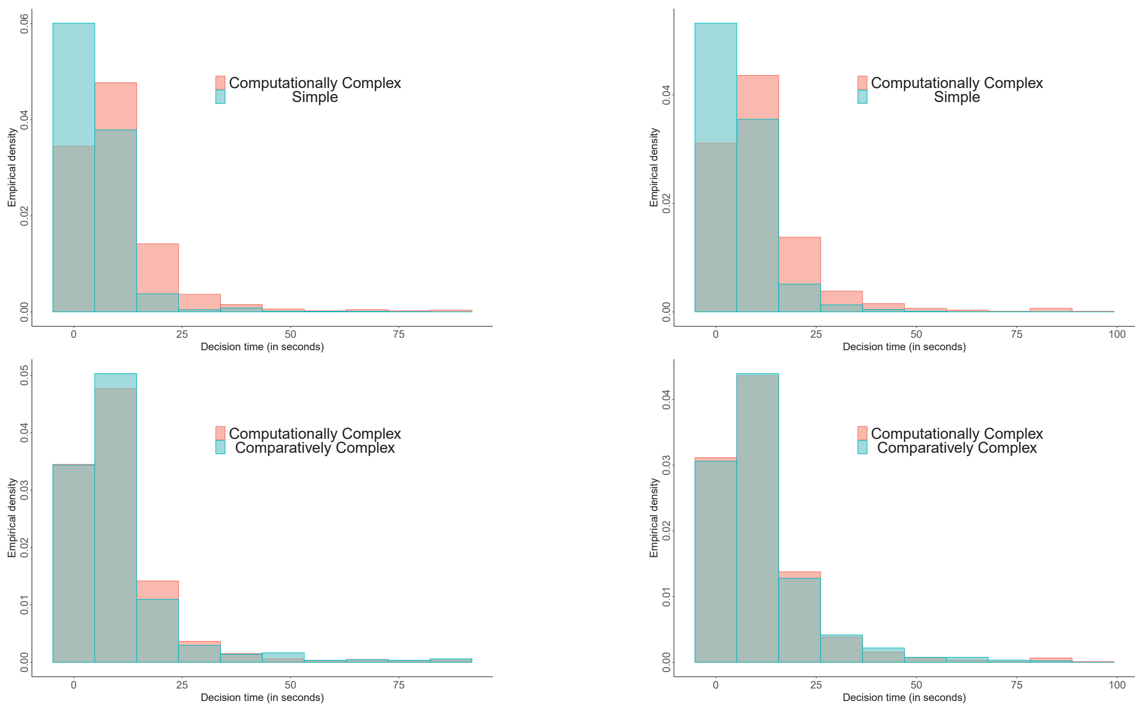


Figure 3.31: *The figure illustrates the distribution of decision times in Skewed. The graphs in the left column refer to the more skewed Mao pairs with* $S = 2.7$ *and the graphs in the right column refer to the less skewed Mao pairs with* $S = 0.6$.

H.4: Testing the Modeling Framework with the Available Data

The framework presented in Section 3.7 was not pre-registered, but using the data that we have, we can still test its main implications on portfolio selection: *namely, subjects are more likely to behave as if maximizing a salience-weighted utility (i) the larger the relative skewness of the underlying assets and (ii) the less complex the portfolio selection problem.* First, since $\frac{|\mathbb{E}[(X_1-X_2)^3]|}{\mathrm{Var}(X_1-X_2)} = 0$ for all (binary) twin problems, the model can explain why subjects in the treatment *Symmetric* diversify naively, even if the choice set comprises only two portfolios. This is not surprising, though, as the model was set up in a way to fit this observation. Precisely, due to the lack of asymmetric contrasts in outcomes, the agent will not make an active choice in complex portfolio selection problems with symmetric assets, but instead reverts to the default of naive diversification. Second, for the portfolio selection problems in the treatment *Skewed*, we have $\frac{|\mathbb{E}[(X_1-X_2)^3]|}{\mathrm{Var}(X_1-X_2)} > 0$, which creates independent variation in the left-hand side of Eq. (3.2), and allows our model to account for the fact that subjects behave consistently across the differently complex problems. Moreover, having both variation in the left-hand side and in the right-hand side of Eq. (3.2), we can formally test for the two predictions (i) and (ii) outlined above.

To test the predictions stated above, we need two requirements to be fulfilled: salience theory needs to make a clear-cut prediction, so that choices can be classified as being consistent with the model in a non-trivial way, and it needs to predict behavior different from naive diversification for at least some choices. As a consequence, we restrict attention to the negatively correlated Mao pairs, where salience theory makes a clear-cut prediction, and to (P.2) of the (binary) twin problems — as in (P.1) salience theory always predicts naive diversification — as well as to the corresponding simple problems. For all of these choices (i.e., 18 choices per subject), we can construct a binary indicator that takes a value of one if a subject's decision is consistent with salience theory and a value of zero otherwise.[26] Using OLS with standard errors being clustered at the subject level, we then regress this binary indicator of behaving in line with salience theory on a measure of relative skewness (to test for the first prediction) and on dummy variables indicating whether the problem is computationally or computationally and comparatively complex (to test for the second prediction). Since the empirical distribution of relative skewness is very skewed (half of the values are zero), we use dummy variables that indicate whether the relative skewness is positive and, conditional on being positive, below or above median. We further present an alternative specification with the exact relative skewness as an independent variable, which yields basically the same results (see Table 3.9).

As predicted by our framework, we observe that subjects are more likely to behave

[26] We classify a choice of an under-diversified portfolio in the treatment *Skewed* as consistent with salience theory if and only if $S = 2.7$ and the portfolio is right-skewed. If $S = 0.6$, we classify only choices of the diversified portfolio as consistent with salience, which assumes that diminishing sensitivity is not too weak.

	$\mathbb{1}_{\{\text{Choice consistent with salience theory}\}}$		
Constant	0.653***	0.737***	0.728***
	(0.019)	(0.021)	(0.024)
Computational Complexity	-0.192***	-0.313***	-0.313***
	(0.020)	(0.025)	(0.025)
Both Layers of Complexity	-0.428***	-0.558***	-0.558***
	(0.020)	(0.023)	(0.023)
Relative Skewness			
- Below median (conditional on positive)	0.107***	-0.035	-0.034
	(0.022)	(0.029)	(0.029)
- Above median (conditional on positive)	0.269***	0.013	0.014
	(0.022)	(0.039)	(0.039)
Relative Skewness # Computational Complexity			
- Below median (conditional on positive)	-	0.314***	0.314***
	-	(0.043)	(0.043)
- Above median (conditional on positive)	-	0.284***	0.284***
	-	(0.044)	(0.044)
Relative Skewness # Both Layers of Complexity			
- Below median (conditional on positive)	-	0.188***	0.188***
	-	(0.042)	(0.042)
- Above median (conditional on positive)	-	0.373***	0.373***
	-	(0.044)	(0.044)
Above-Median Skills	-	-	0.015
	-	-	(0.019)
Decision Time	-	-	-0.000
	-	-	(0.001)
Adjusted R^2	0.155	0.175	0.175
# Choices	5,328	5,328	5,328
# Subjects	296	296	296

Table 3.8: *This table presents the results of OLS regressions using the data on negatively correlated Mao pairs and (P.2) of the twin problems as well as the corresponding simple problems. The dependent variable is a binary indicator that takes a value of one if a choice is consistent with salience theory and a value of zero otherwise. The main independent variables of interest are indicators of the assets' relative skewness (where the problems with a relative skewness of zero serve as the base category) and binary indicators for problems that are only computationally complex or computationally and comparatively complex (where the simple problems serve as the base category). We add controls for a subject's cognitive skills and the decision time (in seconds). Standard errors are clustered at the subject level. Significance level: ***: 1%.*

in line with salience theory the larger is the relative skewness of the assets and the less complex is the portfolio selection problem (first column of Table 3.8). We further interact our measures of relative skewness and complexity (second and third column of Table 3.8), and we find that the effect of relative skewness operates through mitigating the impact of complexity on behavior: while for problems that are neither computationally nor comparatively complex the relative skewness has no significant effect on the average probability to behave as predicted by salience theory, for more complex problems subjects are, on average, more likely to behave in line with salience theory the more skewed are the assets in relative terms. This is exactly in line with the intuition underlying Eq. (3.2) and the corresponding distinction between stimulus-driven and default-driven choices.

	$\mathbb{1}_{\{\text{Choice consistent with salience theory}\}}$		
Constant	0.699***	0.717***	0.707***
	(0.014)	(0.017)	(0.020)
Relative Skewness	0.002***	0.001	0.001
	(0.000)	(0.001)	(0.001)
Computational Complexity	-0.213***	-0.210***	-0.210***
	(0.019)	(0.022)	(0.022)
Both Layers of Complexity	-0.449***	-0.494***	-0.494***
	(0.018)	(0.020)	(0.020)
Relative Skewness # Computational Complexity	-	0.001	0.001
	-	(0.001)	(0.001)
Relative Skewness # Both Layers of Complexity	-	0.002***	0.002***
	-	(0.001)	(0.001)
Above-Median Skills	-	-	0.012
	-	-	(0.019)
Decision Time	-	-	0.000
	-	-	(0.001)
Adjusted R^2	0.149	0.154	0.154
# Choices	5,328	5,328	5,328
# Subjects	296	296	296

Table 3.9: *This table presents the results of OLS regressions using the data on negatively correlated Mao pairs and (P.2) of the twin problems as well as the corresponding simple problems. The dependent variable is a binary indicator that takes a value of one if a choice is consistent with salience theory and a value of zero otherwise. The main independent variables of interest are indicators of the assets' relative skewness and binary indicators for problems that are only computationally complex or computationally and comparatively complex (where the simple problems serve as the base category). We add controls for a subject's cognitive skills and the decision time (in seconds). Standard errors are clustered at the subject level. Significance level: ***: 1%.*

Our model on salient cues and complexity does not only rationalize naive diversification in binary twin problems, but it can also account for large parts of the additional findings on diminished skewness preferences due to computational complexity. For illustrative purposes, we first focus again on the case of perfectly negatively correlated Mao lotteries, but the arguments for the maximally positively correlated Mao pairs go along the same lines. Conditional on making an active choice, a salient thinker chooses a right-skewed portfolio, which invests a share $\alpha \in (\frac{1}{2}, 1]$ in the right-skewed asset, over the diversified portfolio if and only if the Mao pair is sufficiently skewed; that is, if and only if $S > \hat{S}$ holds (see Proposition 2). Moreover, a salient thinker would never choose a left-skewed portfolio over the diversified one.

Imposing the reduced-form in (3.2), the agent makes an active choice if and only if

$$2\sqrt{V}\,\underbrace{\frac{(1-2p)}{\sqrt{p(1-p)}}}_{=S} \geq \kappa(\mathcal{A}, \mathcal{C}),$$

where the left-hand side is strictly increasing in the skewness of the underlying Mao lotteries. Fixing the complexity of the portfolio selection problem, there exists a threshold value $S' \in \mathbb{R}_{>0}$ such that the agent makes an active choice if and only if the Mao pair satisfies $S \geq S'$.

Since $\kappa(\mathcal{A}, \mathcal{C}) = 0$ for problems that are neither computationally nor comparatively complex, we conclude that a salient thinker chooses a right-skewed portfolio in the binary choice between simple lotteries if and only if $S > \hat{S}$, while she does so in the computationally complex problems if and only if $S > \max\{\hat{S}, S'\}$ holds. Hence, computational complexity reduces the demand for right-skewed portfolios, which is consistent with the evidence provided in Figure 3.37. The model cannot explain, however, why for rather skewed Mao pairs left-skewed portfolios become more attractive due to computational complexity.

The model further suggests that behavior is more robust to computational complexity the more skewed the lotteries of a Mao pair are, not only under the perfectly negative correlation, but also for the maximal positive correlation. Figure 3.32 shows that, consistent with our model, the positive correlation between the share of *skewness-seeking* choices[27] in simple and computationally complex problems is indeed more pronounced for more skewed Mao pairs.

Our model makes no clear-cut prediction on the effect of comparative complexity, however. When facing a computationally and comparatively complex portfolio selection problem with $\mathcal{A} = [0, 1]$, the agent chooses a right-skewed portfolio if and only if $S > \check{S}$

[27] A choice is classified as skewness-seeking if and only if the portfolio with a higher third moment is chosen.

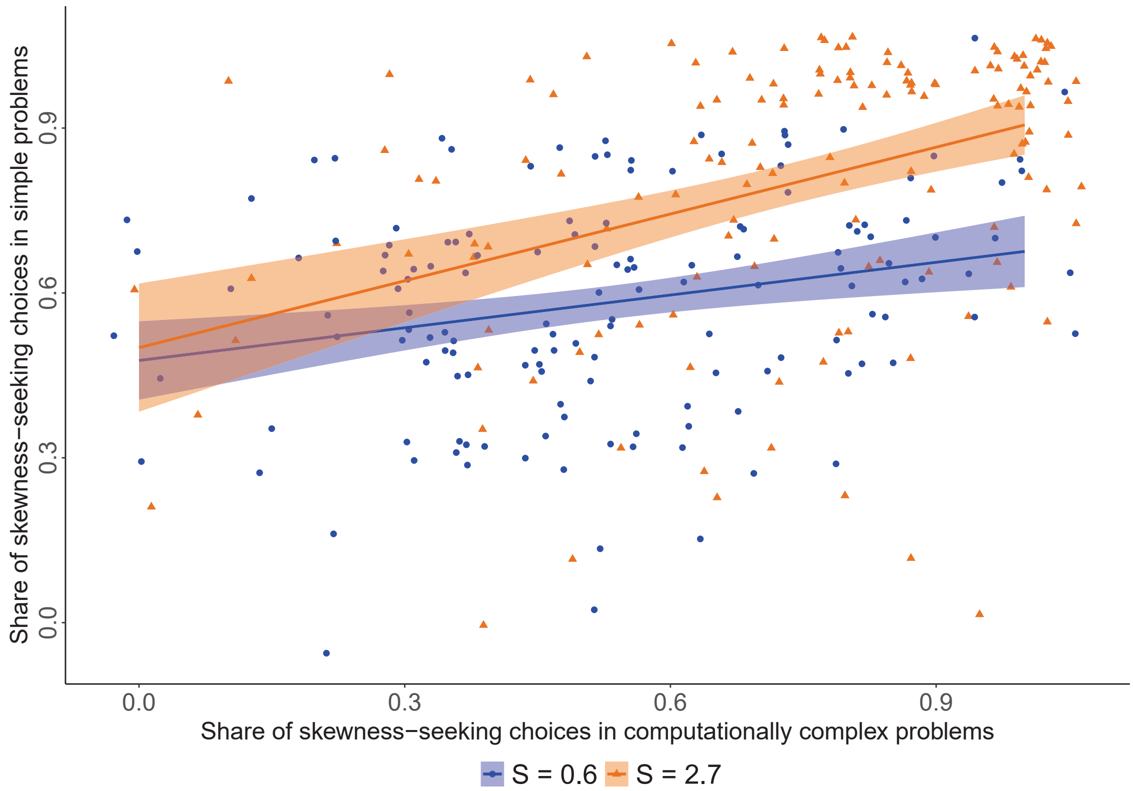


Figure 3.32: *This figure illustrates the correlation between the share of skewness-seeking choices in simple and computationally complex problems for the treatment Skewed.*

holds, where the threshold satisfies $\check{S} \leq \hat{S}$ (see Proposition 3). In addition, the agent makes an active choice in such a problem if and only if $S \geq S''$ holds for some $S'' > S'$, as the larger choice set increases the complexity of the problem. The model thus predicts that the agent chooses some right-skewed portfolio if and only if $S > \max\{\check{S}, S''\}$. Since $\check{S} \leq \hat{S}$ and $S'' > S'$, we do not obtain a clear-cut prediction on how revealed skewness preferences change due to comparative complexity.

H.5: Additional Memory Data

As illustrated in Figure 3.33, subjects are more likely to remember the outcomes of the rather skewed Mao lotteries correctly than those of the less skewed or even symmetric lotteries. If we restrict attention to computationally and/or comparatively complex problems, either under the perfectly negatively correlation in *Skewed* or for (P.1) of the twin problems in *Symmetric*, which makes the results most comparable as discussed in Section 3.3.4, the distribution of correctly remembered outcomes for the more skewed Mao pairs first-order stochastically dominates the distributions for the less skewed or symmetric assets (left panel of Figure 3.33). When taking all the memory data into account (right panel of Figure 3.33), we obtain a similar picture.[28]

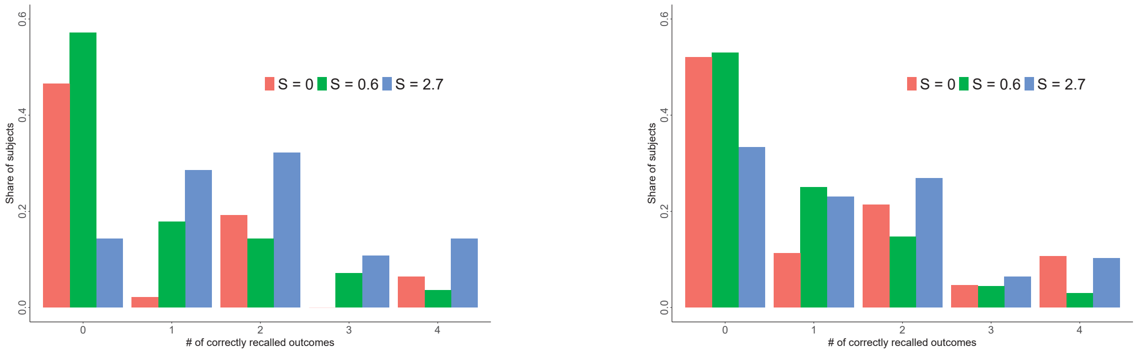


Figure 3.33: *This figure depicts the number of correctly recalled outcomes by the skewness of the assets. In the left panel, we consider only data that refers to comparatively and/or computationally complex problems, and we further restrict attention to subjects who had to recall the joint distribution of a perfectly negatively correlated Mao pair (in Skewed) or the joint distribution of the assets in (P.1) of a twin problem (in Symmetric). In the right panel, we consider all subjects, but those with either 5 or 6 correctly recalled outcomes (in Skewed under the maximal positive correlation) are set to 4 correctly recalled outcomes.*

	$\frac{64}{100}$	$\frac{36}{100}$
$L(E,V,-0.6)$	29%	14%
$L(E,V,0.6)$	25%	14%
$\Delta(V,0.6)$	$\frac{3}{8}\sqrt{V}$	$-\frac{8}{3}\sqrt{V}$

	$\frac{90}{100}$	$\frac{10}{100}$
$L(E,V,-2.7)$	32%	79%
$L(E,V,2.7)$	32%	39%
$\Delta(V,2.7)$	$\frac{2}{3}\sqrt{V}$	$-6\sqrt{V}$

	$\frac{36}{100}$	$\frac{36}{100}$	$\frac{28}{100}$
$L(E,V,-0.6)$	25%	0%	20%
$L(E,V,0.6)$	10%	5%	10%
$\Delta(V,0.6)$	$-\frac{7}{12}\sqrt{V}$	$-\frac{7}{12}\sqrt{V}$	$\frac{3}{2}\sqrt{V}$

	$\frac{10}{100}$	$\frac{10}{100}$	$\frac{80}{100}$
$L(E,V,-2.7)$	29%	62%	38%
$L(E,V,2.7)$	14%	14%	19%
$\Delta(V,2.7)$	$-\frac{8}{3}\sqrt{V}$	$-\frac{8}{3}\sqrt{V}$	$\frac{2}{3}\sqrt{V}$

Figure 3.34: *This figure illustrates, for memory tasks that correspond comparatively and/or computationally complex problems in Skewed, the share of subjects who correctly recalled each of the outcomes. The top row refers the perfectly negatively and the bottom row to the maximal positive correlation.*

Since in (P.1) of a (binary) twin problem as well as in simple problems there are fewer distinct outcomes to be remembered (namely, two or three instead of four), the cleanest

[28]Subjects recall the outcomes in (P.1) of a twin problem better than the outcomes of slightly skewed Mao pairs. This might be driven by the fact that in (P.1) of a twin problem there are only two distinct outcomes to remember, while Mao pairs have four distinct outcomes, which makes the memory task harder.

comparison that we can make is the one between the differently skewed Mao pairs, when the outcomes of the portfolios are *not* readily observable. As illustrated in Figure 3.34, for the case of the perfectly negative correlation, subjects are more likely to correctly recall the outcomes of the more skewed Mao pairs, in particular, those in the state with the larger (absolute) contrast in outcomes, $\Delta(V, S) = L(E, V, -S) - L(E, V, S)$. With the caveat of the coding problem in mind, we observe that, for $S = 2.7$, a large majority of subjects correctly recalls the zero outcome of the left-skewed asset. This might be unsurprising, as zero is not only salient in terms of our model, but it is also the only single-digit number on the screen. Reassuringly, a substantial share of subjects correctly recalls also the non-zero outcome in this state, namely, the right-skewed asset's large upside. We find similar patterns for the case of the maximal positive correlation.

H.6: Additional Data on Diversification and Skills

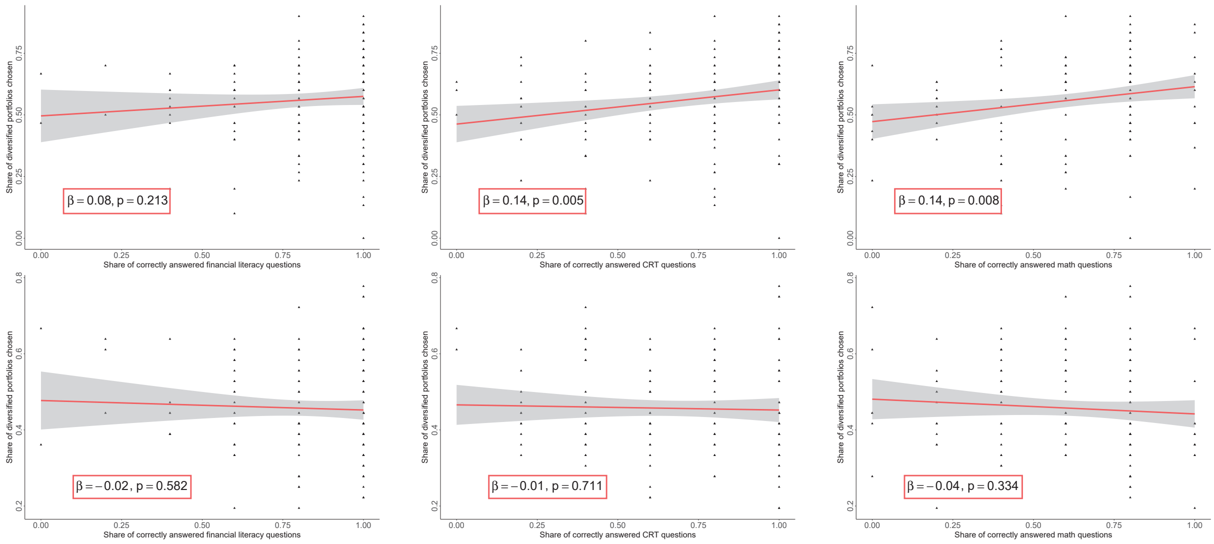


Figure 3.35: *This figure illustrates the correlation between the share of diversified portfolios chosen by a subject and his or her cognitive skills, as measured by the share of correct answers to the financial literacy (left), modified CRT (middle), or maths questions (right). In the top row, we depict the results for the treatment Symmetric, while in the bottom row, we depict the results for the treatment Skewed.*

H.7: Skewness-Dependent Attitude toward Variance

With our results on comparatively complex portfolio selection problems, we can contribute to the broader debate on how to model skewness preferences. The typical assumption in the empirical finance literature is that people trade off variance for skewness — e.g. Mitton and Vorkink's (2007) extension of the mean-variance framework to also incorporate a preference for skewness. Behavioral models of choice under risk, in contrast, predict that the attitude toward variance varies with the skewness of a risk (Section 3.2.2 or Barberis and Huang, 2008; Bordalo *et al.*, 2013a). Using perfectly negatively correlated Mao pairs, we can precisely test both approaches.

By Lemma 1 (a), for a perfectly negatively correlated Mao pair, the variance of the resulting portfolios depends on the exact share α invested in the right-skewed asset, while their skewness only depends on whether α is smaller or larger than $\frac{1}{2}$. If subjects indeed minimize the variance in portfolio returns for a given level of skewness, we should observe that, conditional on choosing a right-skewed portfolio, they invest just slightly more than 50% in the right-skewed asset. Behavioral models like salience theory, on the other hand, predict that subjects invest significantly more than 50% in the right-skewed asset if the Mao pair is sufficiently skewed. This follows from the fact that, according to these models, positive skewness renders (some) variance attractive. Mao pairs thus allow for a clean test of these two competing approaches.

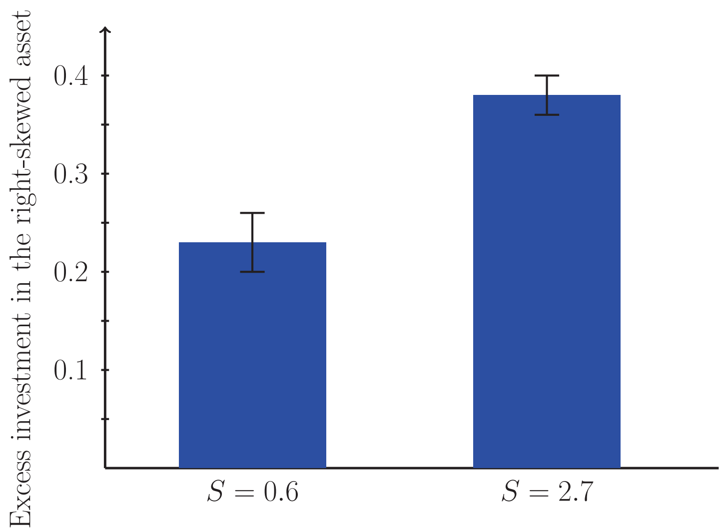


Figure 3.36: *This figure depicts, for the computationally and comparatively complex problems in Skewed, the average excess investment, $\alpha - \frac{1}{2}$, in the right-skewed asset, conditional on choosing a right-skewed portfolio. We report 95%-confidence intervals with standard errors being clustered at the subject level.*

As illustrated in Figure 3.36, subjects do not minimize variance for a given level of skewness, but instead seek (some) variance when choosing a positively skewed portfolio. Strikingly, for the more skewed Mao pairs, subjects, who choose a right-skewed portfolio, almost maximize the variance in portfolio returns. Even for less skewed Mao pairs, subjects choose portfolios with a substantial level of variance. The results are consistent with salience theory and other behavioral models (such as CPT), but inconsistent with a variance-skewness trade-off.

H.8: Are Skewness Preferences Driven by Outcomes or Probabilities?

The leading behavioral theories of choice under risk suggest different mechanisms for why people seek positively skewed and avoid negatively skewed risks (see Ebert and Karehnke, 2019, for a survey of the classical and non-classical literature on skewness preferences). According to salience theory, skewness preferences are driven by large and asymmetric contrasts in outcomes attracting our attention. Other models such as (cumulative) prospect theory (henceforth: CPT; Kahneman and Tversky, 1979; Tversky and Kahneman, 1992), in contrast, explain skewness preferences through the mechanical overweighting of extreme probabilities. This raises the question whether it is outcomes or probabilities that drive skewness preferences.

Skewness preferences under CPT. As we show in Appendix D, CPT can explain why subjects reveal skewness preferences in our experiment. But, in contrast to salience theory, it does not make clear-cut predictions on how behavior varies with skewness (see Proposition 8 in Appendix D). More importantly, CPT is inconsistent with experimental evidence on skewness preferences in other settings. First, as shown in Ebert and Strack (2015, 2018), skewness preferences implied by common CPT-specifications are so strong that they result in unrealistic predictions for optimal stopping problems. Dertwinkel-Kalt *et al.* (2020a) show that salience theory, in contrast, makes sensible predictions on the strength of skewness preferences in static problems, and, therefore, predicts also more sensible stopping behavior. Second, CPT cannot account for the fact that subjects seem to seek *relative* rather than *absolute* skewness (Dertwinkel-Kalt and Köster, 2020b; Frydman and Mormann, 2018): In simple problems, observed behavior varies in a systematic way with the distribution of contrasts in outcomes, which in turn depends on the joint distribution of the available options. While salience theory can account for such "correlation effects", CPT predicts that the value of a lottery is determined by its marginal distribution.

Microfoundation of probability weighting: cognitive uncertainty. Enke and Graeber (2020) argue that non-linear probability weighting is the result of people not understanding probabilities, in particular, if the probability distribution is non-uniform. Precisely, they assume that people deal with their *cognitive uncertainty* about probabilities by compressing any probability distribution to a uniform one (i.e., they over-weight small and under-weight large probabilities), the more so, the more complicated the choice problem is. This mechanism suggests that skewness preferences should be more pronounced in more complex problems, in particular, that skewness preferences become stronger due to computational complexity.[29]

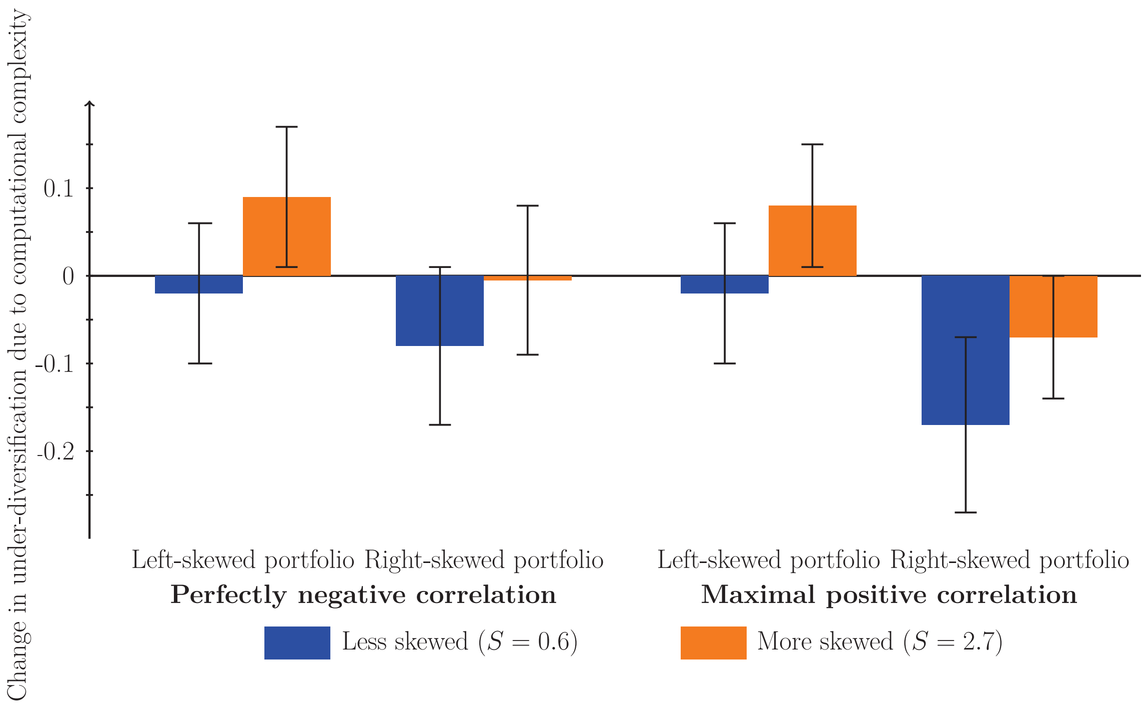


Figure 3.37: *This figure depicts, for the treatment Skewed, the share of under-diversified portfolios chosen in computationally complex problems minus the share of under-diversified portfolios chosen in simple problems. We present the results separately for left- and right-skewed under-diversified portfolios under the perfectly negative and maximal positive correlation. We split the sample by more and less skewed Mao pairs. We further report 95%-confidence intervals with standard errors being clustered at the subject level.*

As illustrated in Figure 3.37, in our experimental data, skewness preferences become, if at all, weaker due to computational complexity. While the differences in revealed skewness preferences are admittedly small, this result is inconsistent with the basic idea underlying the model of cognitive uncertainty. While the fact that revealed skewness preferences vary with computational complexity cannot be explained by the salience model introduced in Section 3.2 either, it can be reconciled with the broader idea of the contrast effect: since outcomes and, thus, also contrasts in outcomes are directly observable only in simple problems, one might argue that skewness preferences driven by the contrast effect are weakened by computational complexity.

[29] As the methodology developed by Enke and Graeber (2020) cannot be easily applied to binary choice problems, we could not directly measure the cognitive uncertainty associated with the different portfolio selection problems. But the assumption that cognitive uncertainty increases with computational complexity is supported by the fact that subjects are much faster in making a decision in simple problems (see Figure 3.31 in Appendix H.3), which is, according to Wilcox (1993), an indication of a change in complexity.

Appendix I: Applications Beyond Choice under Risk

The idea of combining stimulus-driven and default-driven choices in a model building on the availability of salient cues is not limited to the domain of choice under risk, but might help us to understand behavior in different domains as well. Depending on the specific domain, however, the appropriate default heuristic as well as the model of stimulus-driven choice might vary. We briefly sketch two applications of the model to intertemporal choice and voting.

I.1: Intertemporal Choice

Consider the choice between deterministic consumption streams $(x_t)_{t\in\{1,\ldots,T\}}$ and $(y_t)_{t\in\{1,\ldots,T\}}$ that yield outcomes at $T \in \mathbb{N}$ distinct points in time. We abstract from comparative complexity here, and assume that the (computational) complexity increases with the number of points in time T at which the two options yield outcomes. Denote the complexity of the problem by $\kappa(T)$. We then say that the agent makes an active choice if and only if

$$\left| \sum_{t=1}^{T} (x_t - y_t) \frac{\hat{\sigma}(x_t, y_t)}{\sum_{s=1}^{T} \hat{\sigma}(x_s, y_s)} \right| \geq \kappa(T).$$

Imposing the same functional form as in the main text, the above expression simplifies to

$$\frac{\left| \sum_{t=1}^{T} (x_t - y_t)^3 \right|}{\sum_{t=1}^{T} (x_t - y_t)^2} \geq \kappa(T),$$

where the left-hand side captures the relative skewness of the distribution of intertemporal contrast in outcomes. The agent is therefore the more likely to make an active choice the more asymmetric the contrasts in outcomes are across the different points in time.

Based on the evidence on self-control problems and present-biased behavior (e.g., Mischel and Ebbesen, 1970; Cohen, Ericson, Laibson and White, 2020; Augenblick, Niederle and Sprenger, 2015; Augenblick and Rabin, 2019; Falk, Kosse and Pinger, 2020), one natural default heuristic might be that the agent goes for immediate gratification and picks the option with a higher outcome at $t = 1$. Alternatively, an agent, who does not make an active choice, might randomly choose one of the options. Based on recent lab evidence by Dertwinkel-Kalt *et al.* (2019a), an appropriate model of active choice might be Kőszegi and Szeidl's (2013) focusing model, which builds just like salience theory on the contrast effect.

Suppose that for at least some subjects, the default heuristic and model of stimulus-driven choice make different predictions. Increasing the (computational) complexity should alter the subject's choice if the relative skewness of the distribution of intertemporal contrasts in outcomes is zero, but not if it is large. Insofar, our model predicts choices to be weakly more consistent between simple and complex problems if the options' relative skewness is higher.

I.2: Voting

Building on Nunnari and Zápal (2020), we consider a voter who has to decide which of two candidates to choose. Each candidate $n \in \{1, 2\}$ proposes a policy $p_n = (p_1^n, p_2^n, \ldots, p_I^n)$ on how to allocate a fixed budget $W > 0$ across $I \in \mathbb{N}$ issues; that is, $p_i^n \in [0, W]$ and $\sum_{i=1}^{I} p_i^n = W$. The voter's consumption value from a policy $p_n = (p_1^n, p_2^n, \ldots, p_I^n)$ is $V(p_n) = \sum_{i=1}^{I} \theta_i p_i^n$ for some $\theta_i > 0$. Denote as $\tilde{I} = \tilde{I}(p_1, p_2)$ the number of issues on which at least one candidate wants to spend a positive amount of money. We assume that the complexity $\kappa(\tilde{I})$ of the problem is an increasing function of the number of relevant issues $\tilde{I}$; that is, ceteris paribus, the voter is the less likely to make an active choice the more issues there are on which at least one candidate wants to spend some money. Also an increase in the number of candidates could make the problem more complex, but for the sake of illustration we again abstract from comparative complexity.

Given the complexity of the problem, the voter now asks herself whether she likes one of the policies more and whether the difference in perceived consumption values justifies to actually vote. Formally, the voter makes an active choice if and only if

$$\left| \sum_{i=1}^{I} \theta_i (p_i^1 - p_i^2) \frac{\hat{\sigma}(\theta_i p_i^1, \theta_i p_i^2)}{\sum_{j=1}^{I} \hat{\sigma}(\theta_j p_j^1, \theta_j p_j^2)} \right| \geq \kappa(\tilde{I}).$$

Imposing again the functional form assumed in the main text, the inequality simplifies to

$$\frac{\left| \sum_{i=1}^{I} \theta_i^3 (p_i^1 - p_i^2)^3 \right|}{\sum_{i=1}^{I} \theta_i^2 (p_i^1 - p_i^2)^2} \geq \kappa(\tilde{I}),$$

where the left-hand side captures the relative skewness of the distribution of positions that the candidates hold across issues. The voter is therefore the more likely to make an active choice the more asymmetric are the contrasts in positions that the two candidates hold.

Assuming a default of no political engagement, this model could rationalize evidence on increased political participation in response to parties becoming more polarized or at least being perceived as more polarized (e.g. Westfall *et al.*, 2015; Bordalo *et al.*, 2020b). As argued in Nunnari and Zápal (2020), by assuming that voters focus too much on issues in which the candidates hold very different positions, the focusing model by Kőszegi and Szeidl (2013) could be a sensible candidate for describing the effect of salient cues on active voting.

Declaration of Contribution

Hereby I, Mats Köster, declare that the chapter "Salient Cues and Complexity" is co-authored by Markus Dertwinkel-Kalt. Both authors contributed equally to the chapter.

Signature of coauthor (Markus Dertwinkel-Kalt): M. Dertwinkel-Kalt

Chapter 4

Attention-Driven Demand for Bonus Contracts

Co-authored by Markus Dertwinkel-Kalt and Florian Peiseler

4.1 Introduction

Supply contracts (e.g., for electricity, telephony, or banking services) typically include many payments, one of which often represents a bonus payment to consumers. More specifically, such *bonus contracts* involve a series of small, regular payments to be made by subscribers, and a single, large bonus (e.g., a monetary payment, or a premium such as a smartphone) that is paid to consumers at some point during the contractual period. As the eligibility for a bonus often needs to be verified and as transfers are to be put and tracked, each of these payments generates transaction costs, including quasi-fixed costs of hiring employees to do these tasks. Bonus payments, in addition, involve checks that need to be sent out and redeemed, imposing costs both on firms and consumers. Non-monetary bonuses may involve other inefficiencies, for instance, if the consumer values the bonus below its production costs. Thus, abandoning bonuses and reducing the number of transfers to be made by consumers may increase efficiency.[1] In this sense, the predominant use of bonus contracts appears puzzling through the lens of the classical model.

We offer a novel explanation for the frequent occurrence of bonus contracts that builds on a recent model of attentional focusing by Kőszegi and Szeidl (2013). Accordingly, consumers select an option which performs particularly well in those choice dimensions

[1] In particular situations, spreading payments over time can serve other purposes such as relaxing the budget or credit constraints, so that a contract with several regular payments is not inefficient per se.

where the available alternatives differ a lot, while dimensions along which the available options are rather similar tend to be neglected in the decision-making process. In our setup, the choice dimensions correspond to the different payments specified in a contract. For illustrative reasons, suppose a consumer decides whether to sign some bonus contract for a certain good. Here, the large bonus payment attracts a great deal of attention as the difference between obtaining the bonus if the contract is signed and not getting the bonus otherwise is large. In contrast, regular fees (at least if sufficiently small) play only a minor role. The difference between paying one rate if the contract is signed and paying zero otherwise is relatively small, so that none of the regular payments attract much attention. Thus, the inclusion of a large bonus at the cost of slightly higher monthly payments can persuade a consumer to sign a contract which she might otherwise abandon.

In this paper, we derive a firm's optimal contract choice if consumers are focused thinkers. Irrespective of the market structure, this contract exhibits two general features. On the one hand, payments to be made by consumers are equally dispersed over the contractual period in order to minimize the consumers' focus on costs. This contract feature is in line with anecdotal evidence on the attractiveness of installments.[2] On the other hand, the contract involves at most one bonus payment, and if this bonus is non-zero, it will always be maximal.[3] These features create a decision situation that is highly imbalanced with respect to the dispersion of the costs and benefits of the contract. In general, the more imbalanced a decision situation is, the stronger the distortion of a focused thinker's valuation for a good is, so that a consumer's willingness to pay for a subscription can be maximized by concentrating its benefits and equally dispersing its costs. This focusing mechanism has found strong support in a recent lab experiment by Dertwinkel-Kalt *et al.* (2019a). In conclusion, firms should offer contracts that include (at most) a single, maximal bonus payment as well as dispersed and rather small regular payments.

In a first step, we analyze a monopolistic market and show that a monopolist offers a bonus contract, if at all, only for low-value goods. If consumers already have a high valuation for the product, the payments to be made by consumers are relatively high, even absent a bonus payment. In this case, setting a bonus at the cost of increased regular payments cannot shift the consumer's attention solely toward the bonus, but draws attention also to the increased regular payments. Thus, offering a bonus contract does not pay off for high-value products.

---

[2] See, for instance, `https://smallbiztrends.com/2018/11/monthly-installments.html`, or `https://www.retailtouchpoints.com/topics/pos-payments-emv/54-of-shoppers-prefer-installment-payment-plans-over-free-shipping`, or `https://www.nickkolenda.com/psychological-pricing-strategies/#pricing-t7`, all accessed on February 11, 2019.

[3] As we discuss in more detail in Section 2, it seems reasonable to assume that bonus payments are bounded.

In a second step, we consider a perfectly competitive market and show that, independent of the consumers' valuation for the product, competition forces firms to offer bonus contracts (at least in a symmetric equilibrium). If none of the firms pay a bonus, competition drives down regular payments to cost. Relative to these low regular payments the maximal bonus would attract much attention and each firm could obtain a competitive advantage by offering it. Thus, in any (symmetric) competitive equilibrium, consumers sign a bonus contract.

Our results mirror a practice that is common, among others, in markets for electricity, telephony, and bank accounts. As an illustration, consider the electricity retail market. Competition authorities in the European Union regard this market as split into two separate markets, one of which consists of *loyal consumers* who stay with their default provider, and the other one consists of *switching consumers* (see, for instance, Haucap, Kollmann, Nöcker, Westerwelle and Zimmer, 2013, pp. 282). This view is supported by recent empirical studies suggesting that a substantial share of consumers do not even consider *switching the provider* as a viable alternative, so that their default provider de facto serves as a monopolist for this group (e.g., Handel, 2013; Hortaçsu, Madanizadeh and Puller, 2017). Since electricity is essential for running most devices, consumers can be assumed to have a high valuation. Our model predicts—as it is observed in practice—that electricity providers will offer their loyal consumers not a bonus contract, but a contract that involves only (relatively high) monthly fees. In contrast, firms fiercely compete for switching consumers, who are searching for the best deal in the market. As electricity is a homogeneous good, we predict that firms compete for consumers' limited attention by offering bonus contracts. That is indeed ongoing practice: on German price-comparison websites, for instance, virtually every power provider offers a large bonus payment instead of a reduction in regular fees in order to attract new customers.

Under standard assumptions, the common design of bonus contracts—that is, small regular payments uniformly dispersed over the contractual period and a single bonus paid at *some* point in time—is hard to reconcile with the classical model or established behavioral approaches such as (quasi-)hyperbolic discounting. According to the classical model, consumers should be indifferent between a bonus payment and a reduction of regular payments as long as the contract's net present value stays the same. As a consequence, inefficient bonuses should not occur in equilibrium. If consumers are (quasi-)hyperbolic discounters and therefore present-biased, it is suboptimal for a firm to pay a bonus at *some* point during the contractual period, since a present-biased agent prefers to obtain the bonus payment as soon as possible. Also, (quasi-)hyperbolic discounters would prefer a back-loaded instead of a uniform payment stream. In practice, however, the bonus is often paid at *some* point during the contractual period, and regular payments are small and constant (a thorough discussion is provided in Section 4.5.1).[4]

Our study adds to a growing body of theoretical and empirical research that has investigated and supported the importance of attentional focusing for economic choice. Accordingly, a decision maker automatically focuses on eye-catching choice features. These salient aspects of an option obtain an over-proportionate weight in the decision-making process, while less prominent attributes tend to be neglected. A key implication of attentional focusing is *a bias toward concentration* (Kőszegi and Szeidl, 2013) whereby a decision maker pays disproportionately more attention to concentrated rather than dispersed outcomes; which has been supported by recent lab evidence (Dertwinkel-Kalt *et al.*, 2019a). Attentional focusing further provides a unified account for puzzling behavior in a wide range of domains, such as consumer choice (e.g., the attraction effect and the efficacy of misleading sales, see Bordalo *et al.*, 2013b), choice under risk (e.g., the four-fold pattern of risk attitudes, the Allais paradox, and preference reversals, see Bordalo *et al.*, 2012), and financial decision making (e.g., the equity premium puzzle and skewness preferences, see Bordalo *et al.*, 2013a; Dertwinkel-Kalt and Köster, 2020b). Applied to industrial organization, attentional focusing can explain, for instance, why drastic (minor) innovations yield decommoditized (commoditized) markets (Bordalo *et al.*, 2016b). We apply attentional focusing in order to understand how firms design contracts to attract focused thinkers.

4.2 Model

Suppose there are L firms offering a homogeneous product at zero production costs, and a unit mass of homogeneous consumers who value the good at $v \geq 0$ and purchase at most one unit.

Contract Space. Each firm $k \in \{1, \ldots, L\}$ can offer an $M + N$-part tariff consisting of

(i) $M \geq 1$ bonus payments $b_1^k, \ldots, b_M^k \geq 0$ to be paid to consumers, and

(ii) $N \geq 2$ regular payments $p_1^k, \ldots, p_N^k \geq 0$ to be made by consumers.

While we interpret the regular payments by consumers, p_i^k, as payments to be made at different points in time, we stay agnostic in regard to the timing of different bonus payments. As we will show in the next section, the assumption of a fixed number of bonus payments is without loss of generality. In contrast, without imposing further restrictions, a fixed number of payments to be made by consumers entails a loss. But, on the one hand, it seems plausible to assume that consumers aggregate payments they have to make for

[4]See, for instance, `https://www.marktwaechter-energie.de/aerger-mit-energieversorgern/boni/`, accessed on October 1, 2018.

a specific good within a short time period, so that firms may not be able to increase the perceived number of payments beyond a certain threshold.[5] And, on the other hand, if each additional payment to be made by consumers is accompanied by an increasing transaction cost (i.e., transaction costs incurred by consumers are a convex function of the number of regular payments), an "optimal" number of regular payments exists and N could be understood as being optimally chosen by the firms. We discuss implications of this interpretation in the next section when analyzing the robustness of our results.

We also limit the maximum bonus that firms can pay. In other words, we impose a floor on the total price a firm could charge (see Heidhues and Kőszegi, 2018, for a broader discussion).

Assumption 1. *The sum of bonus payments is bounded from above by some* $\bar{b} > 0$.

Since even large firms face financial constraints, in practice firms cannot afford very large bonus payments. More importantly, a very large bonus may create incentives for the consumers to betray the firm and to not fulfill the contract. Finally, offering too large bonus payments might make consumers suspicious in that they believe something fishy to be going on. In this sense, setting a bonus beyond some level $\bar{b}$ may never pay off for a firm.

Timing of the Game. In a first stage, each firm $k \in \{1, \dots, L\}$ chooses a contract

$$\boldsymbol{c}^k := (v, b_1^k, \dots, b_M^k, p_1^k, \dots, p_N^k).$$

In a second stage, consumers decide whether and from which firm to buy the product. Formally, each consumer chooses a contract from the set

$$\mathcal{C} := \{\boldsymbol{c}^k \mid 0 \leq k \leq L\},$$

where $\boldsymbol{c}^0 := (0, \dots, 0) \in \mathbb{R}^{M+N+1}$ refers to the outside option of not buying the product.

For simplicity, firms and consumers adopt the same discount factor which may be determined by the market interest rate. Throughout our analysis we assume that all payments refer to present values (i.e., real instead of nominal sums). While this assumption is not crucial for our qualitative insights, it allows us to abstract from discounting.

[5] In an experimental study, Dertwinkel-Kalt *et al.* (2019a) find that subjects regard payments as separate that are dispersed over several weeks, but aggregate payments that are split within a day. In the context of supply contracts, for instance, it feels natural to assume that consumers aggregate all payments they have to cover with one salary. Then, there is no reason for firms to disperse payments between two paydays as this raises transaction costs, but does not affect a consumer's valuation for the contract. Supportive of this, in Europe where salaries are typically paid monthly, most supply contracts (i.e., mobile or electricity contracts) also involve monthly payments. In the US, where salaries are often paid weekly, supply contracts also often involve weekly payments.

A Firm's Problem. Each firm k designs a contract $\boldsymbol{c}^k$ in order to maximize her profits,

$$\pi_k(\boldsymbol{c}^k, \boldsymbol{c}^{-k}) := D_k \cdot \left(\sum_{i=1}^{N} p_i^k - \sum_{j=1}^{M} \left[\, b_j^k \, + \, \mathbb{1}_{\mathbb{R}_{>0}}(b_j^k) \cdot \kappa \, \right] \right),$$

where $D_k = D_k\Big(\boldsymbol{c}^k, \boldsymbol{c}^{-k}\Big)$ corresponds to the share of consumers choosing the contract offered by firm k from the set $\mathcal{C}$, where $\mathbb{1}_{\mathbb{R}_{>0}}$ is the indicator function on the interval of positive, real numbers, and where $\kappa > 0$ are per-customer transaction costs for each additional bonus payment. We discuss below why we regard it as a plausible assumption that bonuses cause an inefficiency.

A Consumer's Problem. We assume that consumers are *focused thinkers* (Kőszegi and Szeidl, 2013, henceforth: KS). Focused thinkers put an excessive weight on the salient choice dimension(s) of a contract, while they partly neglect less prominent attributes. Following KS, we assume that payments at different points in time as well as a good's quality (or its value to consumers) correspond to different choice dimensions.[6] Moreover, we assume that consumers also perceive the different bonus payments as distinct attributes.[7] Altogether, we assume that the N regular payments to be made by consumers, the M bonus payments offered by the firms, and the consumption value of the product all represent distinct choice dimensions.

Given these assumptions, a focused thinker chooses a contract from the choice set $\mathcal{C}$ in order to maximize her *focus-weighted utility* given by

$$U\Big(\boldsymbol{c}^k \Big| \, \mathcal{C}\Big) := \begin{cases} g(\Delta^v)v - \sum_{i=1}^N g(\Delta_i^p)p_i^k + \sum_{j=1}^M g(\Delta_j^b)b_j^k & \text{if } k > 0, \\ 0 & \text{if } k = 0, \end{cases}$$

whereby the weights on the different choice dimensions are determined by a *focusing function* $g : \mathbb{R}_+ \to \mathbb{R}_+$. According to KS, the weight on price component i, $g(\Delta_i^p)$, depends on the range of attainable utility along this choice dimension denoted as

$$\Delta_i^p := \max_{0 \le k \le L} p_i^k - \min_{0 \le k \le L} p_i^k = \max_{0 < k \le L} p_i^k,$$

where the equality follows from the fact that the outside option does not involve any non-zero regular payments. Analogously, the weight on bonus payment j depends on the

[6]While KS do not analyze a model of industrial organization, they point out that it is plausible to assume that in such models quality represents a choice dimension that is distinct from the price dimension(s); also the related model by Bordalo *et al.* (2013b) adopts the assumption that quality constitutes a separate choice dimension.

[7]This assumption is particularly plausible if some of the bonus payments refer to non-monetary premiums such as a smartphone or an other gadget while others refer to monetary payments. In addition, it is straightforward to show that our results would not change if consumers did not perceive the different bonus payments as distinct attributes, but aggregated them into a single bonus attribute.

range of attainable utility along this bonus attribute, which we denote as Δ_j^b, and the weight on the product's consumption value depends on the utility range in this choice dimension, Δ^v, which is spanned by v in the case that the consumer buys and 0 in the case that she does not buy.

Following KS, we assume that the weight assigned to a certain attribute increases in the utility range along this choice dimension that is attainable given $\mathcal{C}$. This captures the intuition that large contrasts are particularly salient (see, e.g., Schkade and Kahneman, 1998), so that choice dimensions along which the available options differ a lot attract a great deal of attention.

Assumption 2 (Contrast Effect)**.** *The focusing function g is strictly increasing with $g' > 0$.*

In addition, we assume that the contrast effect is sufficiently strong.

Assumption 3. *The function $h : \mathbb{R}_+ \to \mathbb{R}_+$ with $h(x) := g(x)x$ is convex.*

Notice that Assumption 3 is not very restrictive as it admits for convex, linear, and mildly concave focusing functions. In fact, it is violated only for strongly concave focusing functions.[8]

Contractual Inefficiencies. We assume that for each non-zero bonus payment, a firm bears per-customer transaction costs $\kappa > 0$.[9] In the case of a monetary bonus payment, these costs could come from issuing, sending, and tracking checks. More generally, the inefficiency could arise due to administrative processes related to bonus schemes: in order to receive the bonus payment consumers often have to make an inquiry, which implies costs both for the consumer and the firm that needs employees verifying and handling these inquiries. In the case of a non-monetary bonus, the inefficiency could come from an imperfect match between the bonus and the consumer's preferences in the sense that the costs to produce the bonus exceed the consumer's valuation for it. In this case the inefficiency could be exacerbated due to consumers opting for a new bonus product, such as a smartphone, inefficiently often: as consumers typically do not sell their old smartphones, the inefficiency increases by the remaining value of the old and now unused smartphone.[10]

---

[8] More formally, Assumption 3 holds if and only if $-\frac{g''(x)x}{g'(x)} < 2$ for any $x \in \mathbb{R}_+$, which is satisfied for any convex or linear focusing function and for concave focusing functions with a first derivative that is not too elastic. In particular, any increasing power function $g(x) = x^\alpha$, $\alpha > 0$, satisfies Assumption 3.

[9] For certain types of bonuses, it might be more plausible to assume a cost function that continuously increases in the size of the bonus, but most examples of bonuses that we discuss below have in common that the inefficiency is independent of the size of the bonus. By constraining the maximal bonus, we consider the limit case of an infinitely convex cost function. While this assumption substantially simplifies our analysis (e.g., we do not have to deal with implicitly defined interior solutions), economic intuition suggests that our results do not hinge on this cost specification, but should hold under the assumption of a continuous and convex cost function.

[10] See, for instance, `https://www.umweltbundesamt.de/sites/default/files/medien/`

In order to allow firms to increase a consumer's focus-weighted utility using a bonus payment and to break even at the same time, we assume that the costs of paying a bonus are not too large relative to the maximum bonus itself, that is,

$$\frac{\kappa}{\bar{b}} < \frac{g(\bar{b})}{g(\bar{b}/N + c/N)} - 1. \tag{4.1}$$

While this assumption allows firms to benefit from using bonus contracts, it is not very restrictive and becomes weaker for larger values of N or $\bar{b}$, respectively. Suppose, for instance, that the number of regular payments is $N = 24$, that the maximum bonus is $\bar{b} = 120$, and that the focusing function is the identity function. In this case, the costs of making a bonus payment could be much larger than the maximum bonus itself without violating (4.1).

4.3 Equilibrium Analysis

In this section, we first analyze under which conditions a monopolist offers a bonus contract. Second, we derive equilibrium contracts in a perfectly competitive market. Third, we discuss the robustness of our findings. All missing proofs can be found in Appendix A.

4.3.1 Monopolistic Market

Suppose that a single firm monopolizes the market (i.e., $L = 1$). For brevity, we drop the index k in this subsection. Then, the monopolist's maximization problem is given by

$$\max_{\boldsymbol{c}} \pi(\boldsymbol{c}) \quad \text{subject to} \quad \sum_{i=1}^{N} g(p_i)p_i \leq g(v)v + \sum_{j=1}^{M} g(b_j)b_j, \text{ and } \sum_{j=1}^{M} b_j \leq \bar{b},$$

and the optimal contract offer is characterized in the following lemma.

Lemma 1. *A contract $\boldsymbol{c} = (v, b_1, \ldots, b_M, p_1, \ldots, p_N)$ maximizes the monopolist's profit only if*

(i) the payments made by consumers are spread equally across periods, that is, $p_1 = \ldots = p_N$,

378/publikationen/texte_10_2015_einfluss_der_nutzungsdauer_von_produkten_auf_ihre_umwelt_obsoleszenz_17.3.2015.pdf or https://www.beuc.eu/documents/files/FC/durablegoods/articles/0913_Stiftung_Warentest_Germany.pdf or https://www.faz.net/aktuell/finanzen/meine-finanzen/geld-ausgeben/nachrichten/neue-smartphones-oft-schlechter-als-aeltere-geraete-13726913.html, all accessed on February 11, 2019.

(ii) if bonus payment(s) are made, the bonus is maximal, that is, $\sum_{j=1}^{M} b_j = \bar{b}$, *and*

(iii) the contract involves at most a single bonus payment, that is, if a bonus payment is made, then $b_j = \bar{b}$ *for some* $j \in \{1, \dots, M\}$ *and* $b_i = 0$ *for any* $i \neq j$.

Since the monopolist can fully extract the consumers' willingness to pay, he offers a contract that maximizes focus-weighted utility conditional on extracting it. According to the contrast effect a focused thinker's attention is directed to particularly large payments, so that the monopolist can minimize the consumers' perceived costs by dispersing the regular payments uniformly over the entire contractual period. More formally, suppose that one of the regular payments was larger than the others and, without loss of generality, let $p_1 > p_i$ for all $i \in \{2, \dots, N\}$. Then, since $g(p_i)p_i$ is convex (Assumption 3), decreasing p_1 by ϵ and increasing each of the other payments by $\epsilon/(N-1)$ lowers the perceived costs of the contract, while keeping revenue constant. As a result, a necessary condition for maximizing the consumers' willingness to pay conditional on extracting a fixed revenue (and therefore to maximize the monopolist's profit) is that all payments to be made by consumers are of equal size. In contrast, if the monopolist chooses to pay a bonus, it should attract as much attention as possible, which is achieved by setting the maximal bonus, $\bar{b}$, and concentrating it into a single payment.

Yet the monopolist will not always choose a bonus contract. A bonus will be offered if and only if the following *two* conditions are satisfied: (i) the consumers' valuation for the good is sufficiently low, and (ii) the inefficiency that arises from a bonus payment is sufficiently small.

Proposition 1. *There exists a threshold value* $\hat{\kappa} > 0$ *and, for any* $\kappa < \hat{\kappa}$, *a threshold value* $\hat{v}(\kappa) > 0$ *such that the monopolist offers a bonus contract if and only if* $\kappa < \hat{\kappa}$ *as well as* $v < \hat{v}(\kappa)$. *In addition, the threshold value* $\hat{v}$ *monotonically decreases in* κ *on* $(0, \hat{\kappa})$.

Even if the costs for paying a bonus are low, the monopolist offers a bonus only if the consumers' valuation for the product is sufficiently low as well. This follows from the fact that only if the consumers' valuation and therefore the regular payments are sufficiently low, the monopolist can increase its relatively small margin by setting a bonus that grabs attention. If the valuation is high, consumers are already willing to accept relatively high regular payments, even absent a bonus. Then, the focus on the bonus—although it is maximal—cannot outweigh the consumers' focus on the even higher regular payments that are necessary to make a bonus contract profitable. Thus, the monopolist cannot benefit from offering a bonus payment.

In order to put the preceding result into perspective, we consider an example.

Example 1. *Suppose that the focusing function is linear with* $g(x) = x$. *Then, we obtain a threshold value* $\hat{\kappa} = (\sqrt{N}-1)\bar{b}$ *and, for any* $\kappa < \hat{\kappa}$, *a threshold value* $\hat{v}(\kappa) = \frac{(N-1)\bar{b}^2-(2\bar{b}+\kappa)\kappa}{2\sqrt{N}(\bar{b}+\kappa)}$.

Since $\hat{v}(\kappa)$ strictly decreases with the inefficiency arising from bonus payments, Example 1 further suggests that the monopolist will offer a bonus contract only if the regular payments he would charge when not paying a bonus, lie strictly below $\frac{\bar{b}}{2}\frac{(N-1)}{N}$. Suppose, for instance, that the number of regular payments is $N = 24$, and that the maximum bonus is $\bar{b} = 120$. If the consumers' valuation is high enough, so that the monopolist would already charge regular payments $p > 57.5$ when not paying a bonus, then offering a bonus contract would not increase his profits. And with a concave focusing function—such as $g(x) = \sqrt{x}$, which is consistent with experimental evidence by Dertwinkel-Kalt *et al.* (2019a)—regular payments when not paying a bonus would have to be even lower to make a bonus contract profitable.

4.3.2 Competitive Market

Suppose that there are at least two firms trying to attract customers. As the product is homogeneous, firms fiercely compete for consumer attention and, as we will see below, bonus contracts play an even larger role than in a monopolistic market, despite the inefficiencies they produce. The (symmetric) equilibria of the game are characterized in the following proposition.

Proposition 2. *If $L = 2$, an equilibrium exists and any equilibrium has the following properties:*

(i) the market is covered and firms earn zero profits,

(ii) payments to be made by consumers are spread equally across periods (i.e., $p_1^k = \ldots = p_N^k$), and both firms charge the exact same regular payments (i.e., $p_i^1 = p_i^2$), and

(iii) both firms offer the maximum bonus (i.e., $\sum_{j=1}^{M} b_j = \bar{b}$) using a single bonus payment.

If $L \geq 3$, a symmetric equilibrium exists and any symmetric equilibrium satisfies (i)-(iii).[11]

As in the monopoly case, the payments to be made by consumers are equally spread over the contractual period. Given that the remaining firms offer the equilibrium contract, no firm can benefit from unilaterally decreasing some payments and increasing some others. In doing so, the firm would induce consumers to focus more on increased payments, that is, exactly on those choice dimensions along which it offers a worse deal compared to

[11]Note that for $L \geq 3$ also asymmetric equilibria exist, where at least two firms offer $(v, 0, \ldots, 0)$ and serve the market, while at least one firm offers a contract with regular payments exceeding the maximum bonus payment. Importantly, these asymmetric equilibria are neither robust to assuming that clearly dominated options do not affect a consumer's attention allocation nor to assuming that firms want to maximize demand for a given profit level. In this sense, we would argue that the symmetric equilibria delineated above are the only plausible equilibria.

the competitors. At the same time, the focus-weight attached to the decreased payment does not change, as it is determined by the other firms' high regular payments. To sum up, a price hike attracts more attention than the corresponding price cuts, so that the firm cannot benefit from such a contract adjustment.[12]

In contrast to the monopoly case, competing firms always offer a bonus contract (at least in the symmetric equilibrium), irrespective of the consumers' valuation for the product or service. For the sake of contradiction, suppose that firms do not offer a bonus in a symmetric equilibrium. Since firms must earn zero profits, the regular payments consumers make would have to be zero. But then, any firm could benefit from offering a single bonus $\bar{b}$ and increasing each regular payment to $\frac{\kappa+\bar{b}}{N}+\epsilon$ for some sufficiently small $\epsilon > 0$, since Assumption 2 together with Eq. (4.1) ensure that—given zero regular payments offered by the other firms—consumers focus more on the bonus payment than on the increase in regular payments. We have already discussed in the monopoly case that a bonus attracts most attention if it is concentrated into a single payment. In addition, raising the bonus to the maximal level increases the focus on the own contract's advantage—the large bonus—by more than it increases the focus on the also higher regular payments. Thus, in equilibrium, firms offer a single, but maximal bonus.

Importantly, even though paying a bonus creates an inefficiency, bonus contracts are more prevalent in competitive rather than monopolistic markets. This follows from the fact that firms standing in competition only care about beating the best offer of their competitors and not necessarily about maximizing the consumers' willingness to pay. More precisely, by increasing the regular payments in order to cover a bonus payment, a firm in a competitive market makes not only her own offer less attractive, but it also makes her competitors' offers look worse. Hence, since only the incremental change over and above the competitors matters, paying a bonus is indeed a good idea. In contrast, even if regular payments are relatively low, such an increase in regular payments may not pay off for a monopolist, since it can lower the consumers' willingness to pay due to the fact that inframarginal payments are also weighted more.

Corollary 1. *If the product's value to consumers is sufficiently high, the contractual inefficiencies are strictly lower in a monopolistic than in a competitive market.*[13]

[12]Notably, the idea of avoiding high and therefore attention-grabbing prices is also relevant in the model by de Clippel, Eliaz and Rozen (2014) where firms compete for consumers' inattention (to the own price) by making price components non-salient. Here, each firm avoids charging a sum that exceeds the other firms' payments as this would attract a great deal of attention, thereby deterring consumers from signing the respective contract.

[13]Notice that, if Eq. (1) is violated, then consumers will sign a contract without bonus payment(s) in any competitive equilibrium. In this sense Corollary 1 relies on the assumption that (1) is satisfied. But, as we discuss in Section 2, Eq. (1) seems to be an extremely weak restriction on the inefficiency arising from a bonus payment.

4.3.3 Robustness

In this subsection, we argue in how far our findings take over to more general setups.

Outside Option. So far, we have assumed that the outside option—a vector of zeros—affects attribute ranges and therefore decision weights. This seems particularly plausible in the monopoly case, where the consumer only decides whether or not to sign a certain contract. This way our model captures the intuitive feature of the focusing model that large payments attract more attention than smaller ones. For that reason, we regard it as a plausible assumption that the outside option determines attribute ranges.

Importantly, our main insight—that bonus contracts are more widely used in competitive than in monopolistic markets—does not depend on whether the outside option affects decision weights or not. Suppose the outside option does not impact on the focus-weighted utility derived from offered contracts. Independent of the consumers' valuation for the product, a monopolist will never set a bonus contract as it could not manipulate the relative weight on regular payments by setting a bonus: all attribute ranges are zero. Under competition, any equilibrium contract (that generates a positive demand) involves at least one bonus payment. For the sake of contradiction, suppose that firms do not offer any bonus in equilibrium. Then, as firms earn zero profits in any equilibrium, at least two firms offer the contract $(v, 0, \ldots, 0)$ and serve the market. But this gives rise to the same incentives to deviate to a bonus contract as if the outside option affects focus weights. Hence, by the same arguments as in the proof of Proposition 2, we cannot have an equilibrium without a bonus payment. In addition, it is easily verified that an equilibrium with the same properties as described in Proposition 2 exists.[14]

Maximal Bonus Payment. We have also assumed that the maximal bonus a firm can pay is bounded and that this upper bound does not depend on other primitives of the model. In the following, we extend Example 1 to show that our qualitative results do not change if we impose the plausible assumption that the maximal bonus increases mildly in the product's value v.

Example 2. *Suppose that $g(x) = x$ and that the maximal bonus is given by $\bar{b}(v) := \gamma v + \delta$*

[14]Suppose all firms offer the contract $\left(v, \bar{b}, 0, \ldots, 0, \frac{\kappa+\bar{b}}{N}, \ldots, \frac{\kappa+\bar{b}}{N}\right)$ in equilibrium. Decreasing the sum of regular payments by some amount $\epsilon > 0$ and at the same time decreasing the bonus—still using the same bonus attribute—by an amount $\epsilon' > \epsilon$ is not a profitable deviation, as the decrease in the bonus attracts more attention than the decrease in regular payments (irrespective of how ϵ is spread across regular payments). In addition, shifting a share $\alpha \in (0, 1]$ of the bonus to another bonus attribute does not increase focus-weighted utility relative to the competitors (and therefore does not allow a firm to attract consumers), since the range in both bonus dimensions would be exactly the same, namely: $\alpha\bar{b}$. Combining the two arguments implies that there does not exist a profitable deviation involving a positive bonus. Finally, setting all bonus payments equal to zero (in order to save κ and reduce regular payments) is not a profitable deviation either, since consumers always choose $\left(v, \bar{b}, 0, \ldots, 0, \frac{\kappa+\bar{b}}{N}, \ldots, \frac{\kappa+\bar{b}}{N}\right)$ over $(v, 0, \ldots, 0)$ by Eq. (4.1). Hence, even if the outside option does not affect focus weights, there exists a pure-strategy equilibrium with the same properties as described in Proposition 2.

with $\gamma, \delta > 0$*. This implies a threshold value* $\hat{\kappa} = (\sqrt{N} - 1)\delta$*. In addition, there exists some* $\hat{\gamma} > 0$ *such that, for any* $\gamma < \hat{\gamma}$*, the monopolist offers a bonus contract if and only if* $\kappa < \hat{\kappa}$ *as well as*

$$v < \hat{v}(\gamma, \kappa) = \frac{\gamma\Big[(N-1)\delta - \kappa\Big] + \sqrt{N}\left[\sqrt{(\kappa+\delta)^2 + \gamma\kappa\Big(\gamma\kappa - 2\sqrt{N}\delta\Big)} - (\kappa+\delta)\right]}{\gamma\Big[2\sqrt{N} - \gamma(N-1)\Big]},$$

whereby $\hat{v}(\gamma, \kappa) > 0$ *holds for any combination of* γ *and* κ *with* $\gamma < \hat{\gamma}$ *and* $\kappa < \hat{\kappa}$*.*

As examplified above, it depends on the product's value v, and on the curvature of the focusing function $g(\cdot)$ whether a monopolist offers a bonus contract if the maximal bonus increases in the value v. In contrast, our qualitative results on the structure of contracts offered by firms that stand in competition generalizes to the case where the maximal bonus is an increasing function of v. More specifically, also if the maximal bonus increases in v, in any (symmetric) competitive equilibrium firms offer a bonus contract (as in Proposition 2). Altogether, our finding that a bonus payment is always made under competition, but not necessarily in a monopolistic market, is robust to allowing for a maximal bonus that increases in v.

Attribute Space. In addition, we have assumed that each contract is characterized by M bonus payments, N regular payments, and the consumption value. We can relax this assumption in the following two ways, without changing our qualitative results.

Reduced Attribute Space. Consumers might not consider each regular payment as a separate attribute, but instead think about the monthly payments as one dimension. In this case, the fact that the monthly payments all have the same size would not be an endogenous equilibrium property anymore, but would represent an exogenous assumption. Besides this, our results carry over to this alternative specification of the attribute space. Indeed, all proofs remain the same, except for skipping the part in which we verify that all regular payments are of the same size.

Endogenous Attribute Space. So far, we have considered the case with a fixed number of bonus and regular payments, respectively. Obviously, as firms want to pay at most a single bonus, such a restriction on the number of bonus payments is without loss of generality. Just assuming a fixed number of regular payments without imposing further restrictions (e.g., transaction costs for regular payments) entails a loss of generality, however, as then firms would always want to increase the number of regular payments. Instead, we could assume that firms can freely choose the number of bonus and regular payments, but that consumers incur transaction costs that are increasing and convex in the number of non-zero regular payments.

As we prove in Appendix B, given these assumptions, the qualitative insights from

Propositions 1 and 2 remain valid. If in addition consumers' transaction costs are sufficiently convex, then also our result on the comparison of the monopolistic and the competitive outcome remains valid; that is, if transaction costs are sufficiently convex and the consumers' valuation for the product is sufficiently high, the contractual inefficiencies are strictly lower in a monopolistic than in a competitive market. In order to illustrate this result, assume a cost function that is relatively flat first and steep afterwards. In this case, it is easy to see that a monopolist would choose the same number of regular payments as competitive firms would do, so that the only welfare-relevant difference between the monopolistic and the competitive outcome refers to the question of whether the monopolist pays a bonus or not.

Heterogeneous Consumers. So far, we have assumed that consumers are homogeneous, both with respect to their valuation for the product, v, and their focusing function, $g(\cdot)$. In the following, we subsequently relax each of these assumptions.

First, suppose that consumers have the same valuation for the product, but are heterogeneous with respect to the curvature of their focusing function, and that firms can only offer a single contract. Then, a monopolist offers a contract that sets the consumer type that has, among those that should be attracted, the flattest focusing function indifferent between buying and not buying. Depending on v and the focusing function of the indifferent type, either no bonus or the maximal bonus will be set. Consumers with a stronger focusing bias (i.e., a steeper focusing function) will also be attracted by that contract as they appreciate the bonus even more. If there are at least two firms competing for consumers, the maximal bonus will be set (at least in the symmetric equilibrium) and the equal-sized regular payments will be chosen in such a way that firms earn zero profits. As an illustration, suppose that consumers were not susceptible to focusing and were therefore indifferent between a contract with a maximal bonus and no bonus, respectively, as long as the net payment (and hence firms' profits) was held constant. But this implies that, if a small fraction of consumers have focused-weighted utility, competing firms want to exploit this by offering a bonus contract.

Second, suppose that consumers differ only in their valuation for the product, and that each firm can offer a single contract. Also in this case, the main insights of Propositions 1 and 2 still hold. A monopolist makes the consumer type that has, among those that should be attracted, the lowest valuation indifferent between buying and not buying, and maximizes profits along the lines of Proposition 1. Competitive firms set the maximal bonus in any case, charging equal-sized regular payments that allow them to break even.

Third, suppose that consumers are heterogeneous (in some or both dimensions) and that firms can perfectly discriminate between them, that is, each firm can offer a different contract to each consumer type. In addition, assume that each consumer type can only see the contracts(s) tailored to it. Then, all of our preceding results apply separately to each consumer group.

4.4 Applications

In order to apply our results to three exemplary markets, we first extend our model by including two types of consumers. Subsequently, we discuss anecdotal evidence in line with our predictions.

4.4.1 Extending the Model: Loyal and Switching Consumers

Suppose there are $L \geq 2$ firms, where each firm $k \in \{1, \ldots, L\}$ has a share of *loyal consumers* $\alpha_k > 0$ with $\sum_{k=1}^{L} \alpha_k < 1$. A consumer who is loyal to firm k only considers her tailor-made contract offered by firm k, and buys as long as this contract gives her a non-negative focus-weighted utility. The remaining consumers, a share $1 - \sum_{k=1}^{L} \alpha_k$, we call *switching consumers*. They observe all contracts except for those tailored to the loyal consumers, and they choose among these contracts so as to maximize their focus-weighted utility. Finally, suppose that the consumers' valuation for the good is so high that a monopolist would not offer a bonus contract.

We then predict that firms offer different contracts for loyal and for switching consumers. Since each firm k acts as a monopolist for its loyal consumers and since the product's value is assumed to be high, it offers a contract without a bonus payment as defined in Lemma 1. In contrast, firms fiercely compete for switching consumers and offer them the bonus contract defined in Proposition 2. The following corollary summarizes this result.

Corollary 2. *Each firm serves its loyal consumers using a contract without a bonus payment, thereby making positive profits. In addition, at least two firms serve the switching consumers, offering them the maximal bonus via a single bonus payment, thereby making zero profits. Both contracts involve regular payments that are equally spread across all periods.*

4.4.2 Anecdotal Evidence

Application I: Electricity Supply Contracts. In practice, power consumption is not binary, but continuous. Consumers not only decide whether or not to consume, but also how much to consume. However, in many countries, such as Germany, electricity suppliers charge fixed monthly pre-payments that are based on a consumer's estimated power consumption. Arguably, even though the actual monthly fees are not fixed, this contract design (involving pre-payments) might make consumers ex-ante reason as if the regular payments were fixed, in particular because demand for electricity has been found to be highly inelastic with respect to the actual marginal price (e.g., Ito, 2014). Furthermore, in many countries (e.g., in Germany, the UK, or the US), the electricity market consists of local default providers and several smaller entrants. Empirical studies suggest that a substantial share of consumers do not consider switching from their default provider to a cheaper alternative as a feasible option (Hortaçsu *et al.*, 2017). As a consequence, our formal setup that distinguishes between loyal and switching consumers matches the market for electricity supply quite well.

As predicted by our model, loyal consumers are typically charged high regular payments and do not receive a bonus. In contrast, but in line with our model, electricity suppliers fiercely compete for the remaining consumers who are willing to switch (i.e., who compare offers across providers) by offering large bonus payments that are paid after the subscription (often around 60 days later).[15] As mentioned above, it is a common feature of electricity supply contracts that the (often monthly) payments to be made by consumers are constant over the contractual period, even though actual usage is measured and therefore also billed only once a year.[16] Such a contract design involving regular (pre-)payments that are equally dispersed over the contractual period is also optimal according to our model as—unlike contracts conditioning payments in each period on the actual per-period usage—it minimizes the consumers' focus on costs.

Application II: Telephony and Internet Contracts. Our model also fits the market for mobile phone contracts. Although in most OECD countries there are several providers of telecommunication services, a substantial share of consumers have never switched their provider.[17] While contracts designed for new customers typically include valuable features (such as smartphones, tablets, special discounts, or bonus payments), customers who extend an already existing contract usually obtain worse offers that do not involve such a bonus.[18] This observation is suggestive for our prediction that firms offer contracts without bonus payments to their loyal consumers, and compete for switching consumers with bonus contracts. In line with our model, firms often advertise flat-rate contracts, that is, contracts involving payments to be made by consumers that are equally spread over the contractual period and that do not depend on actual usage frequency. Analogous tariff structures are common in the market for Internet contracts.[19]

[15] See, for instance, `http://www.handelsblatt.com/politik/konjunktur/oekonomie/nachrichten/anbieterwechsel-die-teure-traegheit-der-verbraucher/3560414-all.html`, accessed on October 1, 2018.

[16] See, for instance, `https://www.gov.uk/guidance/gas-meter-readings-and-bill-calculation`, or `https://www.verbraucherzentrale.de/energieversorger-rechnungen`, both accessed on October 1, 2018.

[17] See, for instance, `https://www.oecd.org/sti/consumer/40679279.pdf`, pp. 32, accessed on October 1, 2018.

[18] See, for instance, `http://money.cnn.com/2015/03/18/smallbusiness/tmobile-uncarrier/index.html`, accessed on October 1, 2018.

[19] Our assumption of unit demand is particularly plausible for flat-rate contracts that are very common for telephony and Internet services. To be precise, however, not only flat-rate contracts exist, but also

Application III: Bank Accounts. The retail banking industry serves as another example that our setup applies to. In most EU countries, a considerable share of consumers do not even consider the option to switch their bank as a viable alternative, although there are several competitors in the market.[20] While account management fees are usually dispersed over the contractual period, banks try to attract new customers by offering a large switching bonus that is typically paid after the contract is signed and certain conditions (e.g., minimal monthly deposits) are satisfied.[21] As predicted by our model, banks offer bonus payments only to those consumers who are willing to switch and open a new account, but not to their existing customers.

Summary of Stylized Facts. In order to compare the explanatory power of our model and alternative approaches (that we discuss in the next section), we find it useful to sum up the stylized features of bonus contracts observed in the exemplary markets, as delineated above:

(a) If a firm pays a bonus to consumers, it offers a single, but relatively high bonus payment.

(b) Consumers receive the bonus not immediately when signing the contract, but with a delay.

(c) Regular payments made by consumers are equally dispersed over the contractual period.

(d) Only those consumers who search for the best deal are offered bonus contracts, others not.

4.5 Alternative Approaches and Related Literature

In this section, we review the related literature and discuss alternative explanations for bonus contracts, with an emphasis on predictions that allow us to distinguish between our model and these different approaches. For that, we first derive the predictions of models on time preferences (Section 4.5.1) and switching costs (Section 4.5.2) in our context, and second relate these alternative approaches to the anecdotal evidence that we have presented in the previous section. We argue that features (a), (b), and (d) remain largely unexplained by the existing literature. Finally, we relate our paper also to the literature on partitioned pricing and shrouding.

contracts that depend on actual usage, in which case the periodical payments differ to some degree.

[20] See, for instance, `http://www.ec.europa.eu/competition/sectors/financial_services/inquiries/sec_2007_106.pdf`, p. 66, accessed on October 1, 2018.

[21] See, for instance, `https://www.welt.de/finanzen/geldanlage/article126159643/Hohe-Praemien-fuer-Girokonten-bringen-Nachteile.html`, accessed on October 1, 2018.

4.5.1 Exponential and (Quasi-)Hyperbolic Discounting

Exponential Discounting. According to the classical model, as proposed by Samuelson (1937), an agent maximizes her expected intertemporal utility, which (i) is additively separable across payoffs received at different points in time, and (ii) satisfies exponential discounting (i.e., payoffs t periods ahead are discounted by δ^t for some discount factor $\delta < 1$). A classical agent should be indifferent between any allocation of payments across time that has the same net present value. Therefore, firms will avoid inefficient bonus payments. If in addition we impose the common assumption that the marginal utility from money decreases (i.e., preferences can be represented by a concave utility function over monetary wealth), the use of large bonus payments becomes even less attractive. Thus, the classical model can only explain (c).

(Quasi-)Hyperbolic Discounting. In order to match evidence on present-biased behavior, more recent approaches to intertemporal decision making have assumed that discounting is hyperbolic (for seminal contributions, see, Chung and Herrnstein, 1967, and Loewenstein and Prelec, 1992) or quasi-hyperbolic (Laibson, 1997). In principle, a model of quasi-hyperbolic discounting could explain the use of bonus contracts. In the following, we discuss its predictions on monopolistic and competitive outcomes in more detail. In order to give quasi-hyperbolic discounting the best chance to explain the anecdotal evidence, we assume that firms can indeed offer bonus payments that lie in the *present*, while they can shift all regular payments into the future. We discuss these assumptions subsequently. As usual, denote the short-term discount factor by $\beta < 1$ and, for the sake of comparability, let the long-term discount factor equal 1.

Monopolistic Market. If the monopolist offers a bonus payment, it will always choose to pay the bonus immediately, as otherwise the consumer would not overweight the bonus relative to the regular payments. More precisely, if consumers derive linear utility from money, a monopolist would move regular payments into the future and offer immediate bonus payments such that

$$\sum_{i=1}^{N} p_i = \frac{1}{\beta}\left(v + \sum_{j=1}^{M} b_j\right) \tag{4.2}$$

is satisfied. Since each bonus implies an additional cost of $\kappa > 0$, we conclude from Eq. (4.2) that the monopolist chooses at most one bonus payment. If the monopolist sets a bonus, then its maximization problem is given by

$$\max_{b\in(0,\bar{b}]} \frac{1}{\beta}(v+b) - b.$$

Since $\beta < 1$, the monopolist offers either the maximal bonus or no bonus at all, whereby a bonus contract is indeed optimal if and only if $\frac{\kappa}{\bar{b}} < \frac{1-\beta}{\beta}$. Given linear utility of money, a model of quasi-hyperbolic discounting makes no prediction on the structure of regular payments.

If we instead assume that consumers have a decreasing marginal utility from consumption, which implies that the marginal disutility from making a payment is increasing, the monopolist equally disperses the regular payments across the N periods. But given a decreasing marginal utility from consumption, a model of quasi-hyperbolic discounting does no longer make a clear-cut prediction on bonus payments. If consumers integrate the different bonus payments, the monopolist still prefers to offer at most one bonus. If consumers perceive bonus payments as separate (which seems particularly plausible for a mix of monetary and non-monetary bonuses), however, setting multiple bonuses might be more profitable than offering a single bonus payment.

Interestingly, given that consumers integrate bonus payments, the monopolist is more likely to offer a bonus the higher the valuation v is. As an illustration, denote utility from the bonus by $u(\cdot)$ with $u(0) = 0$, $u' > 0$, and $u'' < 0$, and utility from a regular payment by $-u(\cdot)$. Then, the monopolist sets the uniform regular payment, p, and the bonus payment, b, such that

$$Nu(p) = \frac{1}{\beta}\left(v + u(b)\right).$$

We denote as $p^*(v,b)$ the unique solution to this equation. Consequently, conditional on offering a positive bonus, the monopolist's problem is given by

$$\max_{b\in(0,\bar{b}]} Np^*(v,b) - b,$$

so that the marginal profit from increasing the bonus payment equals

$$\frac{1}{\beta}\frac{u'(b)}{u'\Big(p^*(v,b)\Big)} - 1.$$

Since $u(\cdot)$ is concave and $p^*(v,b)$ increases with v, the marginal profit from a bonus payment also increases with v, so that the monopolist is more likely to pay a bonus the higher v is.

Competitive Market. A standard Bertrand-type argument implies that firms earn zero profits in equilibrium. In addition, any equilibrium contract has to maximize consumers' perceived utility, as otherwise a firm could design a contract that consumers strictly prefer, thereby earning positive profits. If consumers derive linear utility from money, this immediately implies that firms offer at most one bonus payment. Then, conditional on firm k offering a bonus contract (and serving a positive share of consumers), the zero-profit condition implies $\sum_{i=1}^{N} p_i^k = \kappa + b^k$. If firm k does not offer a bonus but serves at least some consumers, the zero-profit condition yields $\sum_{i=1}^{N} p_i^k = 0$. Using these zero-profit conditions, it is easy to see that consumers prefer a bonus contract with bonus payment b if and only if $\frac{\kappa}{b} < \frac{1-\beta}{\beta}$. This further implies that firms offer either the maximal bonus or no bonus at all. Altogether, we conclude that competitive firms offer a bonus contract if and only if a monopolist would also do so. If we instead assume a decreasing marginal utility from consumption, a monopolist not only sets a bonus whenever competitive firms do so, but might even set a bonus when competitive firms do not.[22]

(Quasi-)Hyperbolic Discounting vs. Focusing. While a model of focusing predicts that a monopolist offers a bonus contract only for low-value goods, a model of quasi-hyperbolic discounting either makes the opposite prediction (under decreasing marginal consumption utility) or predicts that the choice to offer a bonus does not depend on the product's value (under linear consumption utility). In addition, according to quasi-hyperbolic discounting, the frequency of bonus contracts does not depend on the degree of competition (under linear consumption utility) or bonus contracts are even more frequent in monopolistic rather than in competitive markets (under decreasing marginal consumption utility). Thus, stylized fact (d) remains unexplained by quasi-hyperbolic discounting. In contrast and in line with anecdotal evidence, focusing predicts that bonus contracts are more prevalent in competitive markets.

Most importantly, quasi-hyperbolic discounting can only explain bonus payments that are perceived by the consumer as payments in the *present*. Recent studies have estimated $\beta \approx 1$ for payments that are obtained not immediately, but later on the same day (see, for instance, Andreoni and Sprenger, 2012). And even for immediate payments the extent of present bias is small or negligible (Andersen, Harrison, Lau and Rutström, 2014; Augenblick *et al.*, 2015; Balakrishnan, Haushofer and Jakiela, 2020).[23] Also a model of hyperbolic discounting suggests that any bonus payment is made immediately, while in practice bonus payments are often delivered with a substantial delay: in the case of power supply contracts, for instance, most bonuses are paid only after 60 days. Due to this inevitable delay in payments, the recent experimental literature on time-discounting

[22]For the sake of argument, suppose that both a monopolist and competitive firms set either a single, maximal bonus or no bonus at all, which is indeed the case as long as consumption utility is not too concave. Let $u^{-1}(\cdot)$ denote the inverse of $u(\cdot)$. Then, the monopolist will offer a bonus if and only if $N\big[u^{-1}\left(\frac{v+\bar{b}}{\beta N}\right) - u^{-1}\left(\frac{v}{\beta N}\right)\big] - \bar{b} > \kappa$ holds, while competitive firms offer a bonus contract in equilibrium if and only if $Nu^{-1}\left(\frac{\bar{b}}{\beta N}\right) - \bar{b} > \kappa$ is satisfied. Since $u(0) = 0$ and since $u(\cdot)$ is monotonically increasing and concave, $u^{-1}(\cdot)$ is super-additive. This gives

$$u^{-1}\left(\frac{v+\bar{b}}{\beta N}\right) > u^{-1}\left(\frac{v}{\beta N}\right) + u^{-1}\left(\frac{\bar{b}}{\beta N}\right),$$

which in turn implies that a monopolist offers a bonus contract whenever competitive firms do so.

[23]Notably, a comprehensive review of the literature on time-discounting not only questions whether subjects indeed discount future payoffs (quasi-)hyperbolically, but makes the point that the discount function framework in general might not be adequate for understanding intertemporal decisions (Cohen *et al.*, 2020).

suggests that (quasi-)hyperbolic discounting could not explain the observed design of bonus contracts.[24] Consequently, also stylized fact (b) is inconsistent with models of (quasi-)hyperbolic discounting.

4.5.2 Switching Costs and Automatic-Renewal Contracts

Switching Costs. Models on switching costs (e.g., Klemperer, 1995) can explain why consumers might be reluctant to switch providers. Accordingly, consumers need to be compensated for their switching costs, which may in principle be achieved by paying a bonus to consumers (see, Farrell and Klemperer, 2007, for a discussion). For the sake of comparability, we again abstract from discounting. In addition, we assume a decreasing marginal utility from consumption.

A Simple Model. Consider the following two-period game. In each period firms simultaneously offer a contract and consumers choose from the set of contracts. In the second period, consumers incur some cost $s > 0$ when switching to another firm. We solve for a subgame-perfect Nash equilibrium, whereby we assume that firms cannot commit to second-period prices. In the markets discussed in the previous section (i.e., electricity, telephony, and bank accounts), prices are typically fixed for the contractual period of one or two years. Thus, to meaningfully apply a model on switching costs without commitment to our setup, we need to assume that each period lasts for the duration of a contract.

Suppose that a positive share of consumers bought at firm i in the first period. Because of switching costs, firm i serves as a de-facto monopolist for these consumers, and since any bonus payment implies an inefficiency, it has no incentive to offer a bonus contract. In addition, due to decreasing marginal consumption utility, it equally disperses regular payments across the N payment dates in the second period. More specifically, firm i chooses its regular payments such that $Nu(p_i) = s$, and therefore earns $Nu^{-1}\left(\frac{s}{N}\right)$ per consumer.

In the first period, firms fiercely compete for consumers and—depending on the magnitude of switching costs—might indeed offer bonus payment(s). In fact, firms offer at least one bonus payment if and only if $\kappa < \frac{L-1}{L}Nu^{-1}\left(\frac{s}{N}\right)$, where the optimal number of bonus payments depends on the curvature of consumption utility relative to the bonus-related inefficiency. In any bonus contract, regular payments are equally spread across payment dates and are chosen in a way that, as long as $\overline{b}$ is large enough, second-period profits are handed back to consumers. If no bonus is paid, firms set regular payments equal to zero, thereby earning positive profits, because of the floor on the regular payments.

[24] Also the theoretical behavioral IO literature has argued that payments specified in common contracts necessarily affect future, not present consumption (see, Heidhues and Kőszegi, 2010, for a discussion of credit contracts).

Switching Costs vs. Focusing. In order to explain inefficient bonus payments, models on switching costs need to assume that firms cannot ex-ante commit to second-period prices (a point that has been made already in DellaVigna and Malmendier, 2004). Given this assumption, firms fiercely compete for a large customer base in the first period, which may result in consumers signing bonus contracts in equilibrium. In contrast, our approach explains bonus contracts also if firms can commit to second-period prices. In addition, when marginal production costs were positive, firms would rather charge lower regular payments in the first period than give a bonus payment to their consumers, due to the inherent inefficiency of bonus payments. Only if lowering regular payments does not suffice for attracting consumers, switching costs could explain why a firm might set a bonus payment. Then, the equilibrium bonus payment must exceed the sum of all regular payments, however, which we regard as implausible, as we are not aware of any markets where this is the case.[25] Thus, features (a) and (d) remain unexplained by models on switching costs.

Automatic-Renewal Contracts. Relatedly, Johnen (2019) studies a market in which firms offer automatic-renewal contracts to consumers who are inert in the sense that they forgo benefits from switching to another firm. If a consumer underestimates the probability of failing to cancel a contract (e.g., due to limited attention or a naive present-bias), firms can exploit this consumer by offering an attractive teaser rate that increases after the automatic renewal of the contract. Although his approach provides a plausible explanation for offering attractive teaser rates, it does not make specific predictions on whether firms should use a bonus payment to attract consumers or whether they should simply lower the regular payments, as in his model only the predicted net present value matters. Indeed, if the bonus payment is inefficient, firms will never offer it (i.e., feature (a) is not explained). In order to explain bonus contracts, therefore, the model by Johnen (2019) needs to be augmented, for instance, along the lines that we suggest in this paper (whereby consumers are assumed to be focused thinkers).[26] As a consequence, we regard the explanation for bonus contracts offered by our model and by Johnen (2019) as complementary. Interestingly, also in Johnen (2019), competitive firms focus more on

[25] Only under the specific assumption that consumers face high switching costs, but have very limited access to credit, inefficient bonus payments may be made if firms can commit to future prices. This scenario, however, does not fit to the examples that we have in mind, as switching power suppliers, banks, or providers of mobile phone services seems to be typically very cheap or even for free.

[26] In order to explain the use of bonus contracts one could augment Johnen (2019) also in other ways, for instance, by assuming that a consumer is more likely to remember to switch contracts if the per-period payments increase over time relative to the benchmark with constant payments. We thank a referee for suggesting this possible extension of Johnen (2019). If the change in regular payments occurs after the auto-renewal or the cancellation period, however, we would expect this effect to be rather small. More importantly, we are unaware of any direct empirical support for this mechanism. One could, of course, extend Johnen (2019) also in other ways, and we think of the focusing model as one natural candidate. In particular, the focusing mechanism that we explore is supported by lab evidence (Dertwinkel-Kalt *et al.*, 2019a). We regard our focusing-based explanation as the more parsimonious extension of Johnen (2019).

exploiting consumer mistakes than a monopolist does, so that similar to our findings also in his model the monopolistic outcome can be more efficient than the competitive one.[27]

4.5.3 Salience Theory of Consumer Choice

Salience theory, as proposed by Bordalo *et al.* (2013b), shares Kőszegi and Szeidl's central assumption that dimensions along which alternatives differ much attract much attention. Salience theory, however, has more degrees of freedom than the model by Kőszegi and Szeidl, as the salience of an attribute is determined by a bivariate *salience function* that compares an option's attribute value against the average value that this attribute takes in the choice set. Besides the contrast effect, a salience function exhibits a second fundamental property, that is the *level effect*, whereby a fixed contrast attracts less attention for larger overall attribute values. Also, unlike the focusing model, the salience model predicts that it depends on the specific option at hand which of its attributes grabs a consumer's attention; that is, the attribute that is most salient and therefore attracts most attention can vary across options.

Due to its greater flexibility, salience theory can, depending on the exact context, predict both concentration bias or its inverse, dispersion bias. Put differently, the results derived in this paper are consistent with salience theory, but also various other equilibrium structures could be consistent with the salience model. As a consequence, we build our analysis on the focusing model by Kőszegi and Szeidl (2013), which makes the clear-cut prediction of concentration bias.

4.5.4 Partitioned Pricing, Shrouding, and Socially Wasteful Products

We have shown that attentional focusing can explain why firms frequently partition a product's total price into several price components in order to increase a consumer's willingness to pay (for empirical evidence, see Morwitz *et al.*, 1998). Indeed, a number of older studies observe that consumers systematically underestimate a product's overall price if it is partitioned into several price components that the consumer is simultaneously (*partitioned pricing*) or sequentially (*drip pricing*) informed about (e.g., Carlson and Weathers, 2008; Ahmetoglu, Furnham and Fagan, 2014). More recent experimental evidence from the lab (Dertwinkel-Kalt *et al.*, 2019a) and from the field (Dertwinkel-Kalt, Köster and Sutter, 2020b) suggests, however, that simply splitting up the total price without dispersing the price components over time does not affect a consumer's willingness to pay. Hence,

[27]There is a small, but recently growing literature on the distorting effects of competition (e.g., Carlin, 2009; Gabaix, Laibson, Li, Li, Resnick and de Vries, 2016; Friedrichsen, 2018). See Johnen (2019) for a discussion of the mechanisms in these papers.

in line with our interpretation, it seems to be the correct timing of regular payments (e.g., monthly) that gives partioned pricing the potential to increase a consumer's willingness to pay.

While also the literature on shrouded price components (e.g., Gabaix and Laibson, 2006; Heidhues, Kőszegi and Murooka, 2016) can explain the success of partitioned pricing, it necessarily assumes that a share of consumers are not aware of some price components when making the purchase decision or the importance thereof when mispredicting their future behavior. In contrast, a model of focusing can also explain these effects if all information is readily available. Because several smaller prices attract less attention than a single, but large one, they are underweighted. As a consequence, the focusing model can account for the fact that a uniform dispersion of the total price over time increases a consumer's willingness to pay even if the consumer is fully informed about all price components.

Finally, our study connects to the literature demonstrating that even socially wasteful products can survive competition and may be sold in a competitive equilibrium. Heidhues *et al.* (2016) argue that, if a part of the product's price is shrouded, some consumers may not anticipate the product's total price at the moment of making the purchase decision, so that these consumers may purchase a good at a price that strictly exceeds their valuation. Thus, even a socially wasteful product—that is, a product for which the production costs lie above the consumers' valuation—may generate positive demand. Since firms typically have no incentive to unshroud the additional price, selling a socially wasteful product can be the equilibrium outcome in a perfectly competitive market. Also in our model socially wasteful products might be sold in a competitive equilibrium even if consumers are aware of the entire price.[28] By focusing on the contract's outstanding feature (i.e., the large bonus payment), consumers may overestimate the value of a deal and sign contracts for socially wasteful products.

4.6 Conclusion

Bonus contracts create two distinct inefficiencies. On the one hand, bonus payments create administrative costs, both for the issuing firm and for the consumer, and non-monetary bonuses such as included premiums may give an imperfect match between the bonus and the consumer's preferences. On the other hand, bonus contracts yield an imbalanced decision situation—benefits are concentrated in the form of a single, large bonus payment while costs are dispersed over many small payments—in which focused thinkers tend to

[28]Since our model assumes zero production costs, a product that consumers value at $v < 0$ is socially wasteful. While we assume $v \geq 0$, in fact our analysis also holds if v is negative but sufficiently close to zero.

make suboptimal decisions. In fact, a consumer chooses an inefficient bonus contract in the competitive equilibrium even if some firm offers an efficient contract with no bonus payment. If we follow Kőszegi and Szeidl (2013) in assuming that focus weights only affect the consumer's *decision utility* but not her *experienced utility* that defines her surplus, then this decision can be considered a mistake by the consumer: firms earn zero profits in any competitive equilibrium, so that consumers bear the additional costs that adjoin bonus payments (in the form of higher regular payments). Hence, under the assumption that focusing does not affect experienced utility, the use of bonus payments strictly lowers consumer welfare, both in monopolistic and competitive markets.

We have shown that these inefficiencies are not eliminated by competition, but can only be overcome by regulation. Indeed, firms *have to* exploit attentional focusing under competitive pressure, so that bonus contracts are even more frequent in competitive than in monopolistic markets. By enhancing the use of bonus payments, competition therefore exacerbates the inefficiencies arising from contracting with focused agents.[29]

From a policy perspective, our study suggests that a legal ban on bonus payments could have favorable consequences. On the one hand, a legal ban on bonus payments eliminates the inherent inefficiency of paying bonuses. On the other hand, it creates choice environments that are balanced, that is, where in equilibrium all payments receive the same amount of attention. Notably, making bonus payments is not necessary to encourage consumers to switch providers, as firms could instead lower the regular payments to attract consumers (see, Farrell and Klemperer, 2007, for a discussion of different modeling approaches). Hence, even if consumers incur costs for switching to another provider, a ban on bonus payments does not impair competitive forces. Altogether, we argue that prohibiting the use of bonus contracts not only reduces the direct inefficiencies arising from bonus payments, but could also induce better decisions by consumers.

Appendix A: Proofs

For brevity, we denote $\tilde{v} := g(v)v$ the focus-weighted consumption value of the product. In addition, we suppress the consumption value dimension of a contract throughout the Appendix.

Proof of Lemma 1. The proof proceeds in two steps. First, we rewrite the monopolist's maximization problem and characterize the optimal payments to be made by consumers. Second, we argue that the monopolist offers either the maximal bonus (i.e., $\sum_{j=1}^{M} b_j = \overline{b}$) or no bonus. Third, we show that the monopolist pays at most one bonus.

[29] As in other behavioral IO models (see, for instance, DellaVigna and Malmendier, 2004; Gabaix and Laibson, 2006), competition drives down firms' profits to zero, *even though* consumers' decision biases are fully exploited.

1. STEP: In order to solve the monopolist's maximization problem, we set up the Lagrangian

$$\mathcal{L}(\boldsymbol{c}, \boldsymbol{\mu}, \eta, \boldsymbol{\gamma}, \lambda) := \sum_{i=1}^{N} p_i - \sum_{j=1}^{M} b_j - \lambda \left(\sum_{i=1}^{N} g(p_i)p_i - \tilde{v} - \sum_{j=1}^{M} g(b_j)b_j \right) - \eta \left(\sum_{j=1}^{M} b_j - \bar{b} \right) - \sum_{j=1}^{M} \gamma_j(-b_j) - \sum_{i=1}^{N} \mu_i(-p_i),$$

where $\lambda, \eta, \gamma_j, \mu_i \geq 0$, which yields the following Karush-Kuhn-Tucker Conditions:

$$\frac{\partial \mathcal{L}}{\partial p_i} = 1 - \lambda \left[g(p_i) + g'(p_i)p_i \right] + \mu_i \leq 0, \tag{KKT^p_i-1}$$

holding with equality if $p_i > 0$, and

$$\frac{\partial \mathcal{L}}{\partial b_j} = -1 + \lambda \left[g(b_j) + g'(b_j)b_j \right] - \eta + \gamma_j \begin{cases} \leq 0 & \text{if } b_j = 0, \\ = 0 & \text{if } 0 < b_j < \bar{b}, \\ \geq 0 & \text{if } b_j = \bar{b}, \end{cases} \tag{KKT^b_j-1}$$

as well as the condition on the participation constraint

$$\lambda \cdot \left(\sum_{i=0}^{N} g(p_i)p_i - \tilde{v} - \sum_{j=1}^{M} g(b_j)b_j \right) = 0, \tag{KKT-PC}$$

and conditions on price constraints, that is,

$$\mu_i \cdot (-p_i) = 0 \tag{KKT^p_i-2}$$

for any $i \in \{1, \ldots, N\}$, and conditions on bonus constraints, that is,

$$\gamma_j \cdot (-b_j) = 0 \tag{KKT^b_j-2}$$

for any $j \in \{1, \ldots, M\}$ as well as

$$\eta \left(\sum_{j=1}^{M} b_j - \bar{b} \right) = 0. \tag{KKT-BC}$$

First, we characterize the optimal payments to be made by consumers (i.e., part (i) of our lemma). We observe that at least one p_i has to be larger than zero, as otherwise $\lambda = 0$ by (KKT-PC) and therefore $\frac{\partial \mathcal{L}}{\partial p_i} > 0$ by (KKT^p_i-1); a contradiction. Hence, from now on suppose $p_i > 0$ for some $i \in \{1, \ldots, N\}$. Then, since $1 + \mu_i > 0$, Condition (KKT^p_i-1) gives $\lambda > 0$. Together with (KKT-PC), this yields

$$\sum_{i=1}^{N} g(p_i)p_i = \tilde{v} + \sum_{j=1}^{M} g(b_j)b_j. \tag{4.3}$$

Next, we show that this implies $p_i > 0$ for any $i \in \{1, \dots, N\}$. For the sake of contradiction, suppose $p_j = 0$ for some $j \in \{1, \dots, N\}$. Then, Condition (KKT$_j^p$-1) yields $1 + \mu_j \leq \lambda g(0)$. As there is at least one $p_i > 0$, Conditions (KKT$_i^p$-1) and (KKT$_i^p$-2) yield $1 = \lambda\Big[g(p_i) + g'(p_i)p_i\Big]$. Together, these considerations give

$$1 \leq 1 + \mu_j \leq \lambda \underbrace{\Big[g(0) + g'(0)0\Big]}_{=g(0)} \overset{\text{A.3}}{<} \lambda \left[g(p_i) + g'(p_i)p_i\right] = 1, \tag{4.4}$$

a contradiction. Thus, we have $p_i > 0$ for any $i \in \{1, \dots, N\}$. In addition, Conditions (KKT$_i^p$-1) and (KKT$_j^p$-1) require $p_i = p_j$ for $i, j \in \{1, \dots, N\}$, which completes the proof of part (i).

2. STEP: Given the results derived above, we show that *the monopolist either offers the maximal bonus (i.e.,* $\sum_{j=1}^M b_j = \bar{b}$*) or no bonus at all.* For the sake of contradiction, suppose that the monopolist offers a contract with $0 < \sum_{j=1}^M b_j < \bar{b}$. Thus, we have $\eta = 0$ by (KKT-BC), and, for any $b_j > 0$, also $\gamma_j = 0$ by (KKT$_j^b$-2). Using the same arguments as in the first step, we conclude from Conditions (KKT$_i^p$-1) and (KKT$_j^b$-1) that either $b_j = p_i = p'$ or $b_j = 0$. Let $m \in \{1, \dots, M\}$ bonus payments be non-zero and notice that $m < N$, as otherwise profits would be zero. Then, (KKT-PC) implies that $g(p')p' = \tilde{v}/(N - m)$, so that the monopolist earns

$$\pi' = \frac{\tilde{v}}{g(p')} - m \cdot \kappa. \tag{4.5}$$

Suppose that the monopolist instead does not offer any bonus payments; that is, $b_j = 0$ for any $j \in \{1, \dots, M\}$. By the first step, we have $p_i = p''$ for any $i \in \{1, \dots, N\}$, and Condition (KKT-PC) yields $g(p'')p'' = \tilde{v}/N$, so that the monopolist earns

$$\pi'' = \frac{\tilde{v}}{g(p'')}. \tag{4.6}$$

Since $g(x)x$ is a strictly increasing function, by Assumption 2, we conclude $p'' < p'$ from

$$g(p'')p'' = \frac{\tilde{v}}{N} < \frac{\tilde{v}}{N - m} = g(p')p'.$$

Then, for any $\kappa \geq 0$, we obtain $\pi'' > \pi'$ by Assumption 2; a contradiction. As a consequence, the monopolist will never offer a contract with $0 < \sum_{j=1}^M b_j < \bar{b}$.

3. STEP: Given the results from the preceding steps, we next show that *the monopolist offers at most one bonus payment.* Suppose that $\sum_{j=1}^M b_j = \bar{b}$ and that $m \geq 1$ bonus payments are non-zero. By the first step, we have $p_i = p'''(m)$, $i \in \{1, \dots, N\}$, and

(KKT-PC) yields

$$g(p'''(m))p'''(m) = \frac{1}{N} \cdot \left[\tilde{v} + \sum_{j=1}^{m} g(b_j)b_j\right]. \tag{4.7}$$

For any $m > 1$, Assumption 2 immediately implies that

$$g(\bar{b})\bar{b} = \sum_{j=1}^{m} g(\bar{b})b_j \overset{A.2}{>} \sum_{j=1}^{m} g(b_j)b_j$$

holds. Thus, by Assumption 2 and Eq. (4.7), we have $p'''(1) > p'''(m)$ for any $m > 1$. As the bonus payment is fixed and as less bonus payments imply lower costs, the monopolist will choose at most one bonus payment, which was to be proven. □

Proof of Proposition 1. By Lemma 1, the monopolist offers either a bonus contract with a single bonus payment, $\boldsymbol{c}^{bon} := (\bar{b}, 0, \ldots, 0, p^{bon}, \ldots, p^{bon})$, or a contract without any bonus payments, $\boldsymbol{c}^{no} := (0, \ldots, 0, p^{no}, \ldots, p^{no})$. We have also seen in the proof of Lemma 1 that $p^{bon} = p^{bon}(v, \bar{b})$ is implicitly defined by $g(p^{bon})p^{bon} = \frac{1}{N}\left[\tilde{v} + g(\bar{b})\bar{b}\right]$, and that $p^{no} = p^{no}(v)$ is implicitly given by $g(p^{no})p^{no} = \frac{\tilde{v}}{N}$. We proceed in two steps. First, we neglect the cost of paying a bonus, that is, we set $\kappa = 0$. Second, we allow for positive costs of paying a bonus, that is, $\kappa > 0$.

1. STEP: Let $\kappa = 0$. Then, the monopolist offers a bonus contract $\boldsymbol{c}^{bon}$ if and only if

$$\frac{\tilde{v} + \left(g(\bar{b}) - g(p^{bon})\right)\bar{b}}{g(p^{bon})} > \frac{\tilde{v}}{g(p^{no})}$$

or, equivalently,

$$\bar{b}\left(g(\bar{b}) - g(p^{bon})\right) > \tilde{v}\left(\frac{g(p^{bon}) - g(p^{no})}{g(p^{no})}\right). \tag{4.8}$$

We proceed as follows: first, we verify that $\pi(\boldsymbol{c}^{bon}) - \pi(\boldsymbol{c}^{no})$ monotonically decreases in v, which implies that (4.8) is more likely to hold for small values of v. Second, we argue that (4.8) is violated as v approaches infinity while it is fulfilled as v approaches zero.

Recall that $p^{bon} > p^{no}$. Then, by the Implicit Function Theorem, we obtain

$$\begin{aligned}\frac{\partial}{\partial v}\left(\pi(\boldsymbol{c}^{bon}) - \pi(\boldsymbol{c}^{no})\right) &= N \cdot \left(\frac{\partial}{\partial v}p^{bon}(v, \bar{b}) - \frac{\partial}{\partial v}p^{no}(v)\right) \\ &= \frac{\partial \tilde{v}}{\partial v} \cdot \left(\frac{1}{g(p^{bon}) + g'(p^{bon})p^{bon}} - \frac{1}{g(p^{no}) + g'(p^{no})p^{no}}\right),\end{aligned}$$

which is strictly negative by Assumption 3 and $p^{bon} > p^{no}$. Thus, $\pi(\boldsymbol{c}^{bon}) - \pi(\boldsymbol{c}^{no})$ monotonically decreases in v, which was to be proven.

Next, suppose that v approaches infinity, and notice that the left-hand side of (4.8) is negative for sufficiently large values of v, while the right-hand side of (4.8) is non-negative for any $v \geq 0$. Hence, (4.8) is violated in the limit of v approaching infinity. Finally, we consider the limit for v approaching zero. By Assumption 2, this implies that also $\tilde{v} := g(v)v$ approaches zero. First, it is easy to see that in this limit the left-hand side of Inequality (4.8) is strictly larger than zero, as

$$\lim_{v\to 0} g\Big(p^{bon}(v,\bar{b})\Big)p^{bon}(v,\bar{b}) = \frac{g(\bar{b})\bar{b}}{N} < g(\bar{b})\bar{b},$$

and thus $\lim_{v\to 0} p^{bon}(v,\bar{b}) < \bar{b}$. Second, as $\lim_{v\to 0} \frac{\tilde{v}}{g(p^{no})} = N \lim_{v\to 0} p^{no}(v) = 0$ by definition of $p^{no}(v)$ and Assumption 2, the right-hand side of (4.8) is zero in the limit of v approaching zero.

Combining the above results and using the fact that $\pi(\boldsymbol{c}^{bon}) - \pi(\boldsymbol{c}^{no})$ is continuous in v, we conclude by the Intermediate Value Theorem that there exists some threshold value $v' > 0$ such that the monopolist offers a bonus contract if and only if $v < v'$.

2. STEP: Let $\kappa > 0$. Then, the monopolist offers a bonus contract $\boldsymbol{c}^{bon}$ if and only if

$$\frac{\tilde{v} + \Big(g(\bar{b}) - g(p^{bon})\Big)\,\bar{b}}{g(p^{bon})} - \kappa > \frac{\tilde{v}}{g(p^{no})}$$

or, equivalently,

$$\kappa < \underbrace{\frac{\bar{b}\left(g(\bar{b}) - g(p^{bon})\right)}{g(p^{bon})} - \tilde{v}\left(\frac{g(p^{bon}) - g(p^{no})}{g(p^{no})g(p^{bon})}\right)}_{=\pi(\boldsymbol{c}^{bon})\big|_{\kappa=0} - \pi(\boldsymbol{c}^{no})}. \tag{4.9}$$

We have already seen in the first step that the right-hand side of Inequality (4.9) monotonically decreases in v. Hence, the monopolist offers a bonus contract for some $v > 0$ only if

$$\begin{aligned}
\kappa &< \lim_{v\to 0}\left[\frac{\bar{b}\left[g(\bar{b}) - g\Big(p^{bon}(v,\bar{b})\Big)\right]}{g\Big(p^{bon}(v,\bar{b})\Big)} - \tilde{v}\left(\frac{g\Big(p^{bon}(v,\bar{b})\Big) - g\Big(p^{no}(v)\Big)}{g\Big(p^{no}(v)\Big)g\Big(p^{bon}(v,\bar{b})\Big)}\right)\right] \\
&= \lim_{v\to 0}\frac{\bar{b}\left[g(\bar{b}) - g\Big(p^{bon}(v,\bar{b})\Big)\right]}{g\Big(p^{bon}(v,\bar{b})\Big)} \\
&=: \hat{\kappa}.
\end{aligned}$$

By the same arguments as in the first step, for any $\kappa < \hat{\kappa}$, there exists some $\hat{v}(\kappa)$ such that the monopolist offers a bonus contract if and only if $v < \hat{v}(\kappa)$, which was to be proven.

Specifically, the function $\hat{v} : [0, \hat{\kappa}) \to \mathbb{R}_+$ is implicitly given by

$$\underbrace{\frac{\tilde{v}(\hat{v}) + \left[g(\bar{b}) - g\left(p^{bon}(\hat{v}, \bar{b})\right)\right]\bar{b}}{g\left(p^{bon}(\hat{v}, \bar{b})\right)} - \frac{\tilde{v}(\hat{v})}{g\left(p^{no}(\hat{v})\right)}}_{=:F(\hat{v},\kappa)} - \kappa = 0.$$

By construction, we have $\hat{v}(\hat{\kappa}) = 0$. In addition, the Implicit Function Theorem yields

$$\frac{\partial \hat{v}}{\partial \kappa} = -\frac{\frac{\partial}{\partial \kappa} F(\hat{v}, \kappa)}{\frac{\partial}{\partial \hat{v}} F(\hat{v}, \kappa)} = -\frac{(-1)}{\frac{\partial}{\partial \hat{v}}\left[\pi(\boldsymbol{c}^{bon})\Big|_{\kappa=0} - \pi(\boldsymbol{c}^{no})\right]} < 0,$$

since we have seen above that the denominator is strictly negative. This completes the proof. □

Proof of Proposition 2. For illustrative purposes, we only consider the case of $L = 2$, but we solve for the essentially unique equilibrium without restricting ourselves to symmetric equilibria. The generalization of the arguments to the case of $L > 2$ is straightforward when restricting the analysis to symmetric equilibria.

The proof proceeds in seven steps. First, we show that the standard Bertrand logic applies, so that in equilibrium firms earn zero profits, and consumers are indifferent between both offers. Second, we show that equilibrium payments made by consumers are equally spread across periods. Third, we argue that both firms charge the same regular payments, which also implies that both offer the same overall bonus. Fourth, we show that firms offer at most one bonus payment, which in turn implies that both firms offer essentially the same contract. Fifth, we argue that firms either offer the maximal bonus or no bonus. Sixth, we show that firms offer a bonus in equilibrium. Seventh, we prove that a unique equilibrium exists.

1. STEP: We show that *firms earn zero profits in any equilibrium.* For the sake of contradiction, suppose firm $k \in \{1, 2\}$ earns strictly positive profits in equilibrium, which implies

$$\tilde{v} + \sum_{i=1}^{M} g(\Delta_j^b) b_j^k \geq \sum_{i=1}^{N} g(\Delta_i^p) p_i^k, \text{ and } \sum_{i=1}^{M} g(\Delta_j^b)(b_j^k - b_j^{-k}) \geq \sum_{i=1}^{N} g(\Delta_i^p)(p_i^k - p_i^{-k}), \text{ and } \sum_{i=1}^{N} p_i^k > \sum_{j=1}^{M} b_j^k.$$

Hence, we have $p_i^k > 0$ for at least one $i \in \{1, \ldots, N\}$. Without loss of generality, let $p_1^k > 0$ and $\sum_{i=1}^N p_i^k - \sum_{j=1}^M \left[b_j^k + \mathbb{1}[b_j^k > 0] \cdot \kappa\right] \geq \sum_{i=1}^N p_i^{-k} - \sum_{j=1}^M \left[b_j^{-k} + \mathbb{1}_{\mathbb{R}_{>0}}(b_j^k) \cdot \kappa\right]$. This immediately implies that firm $-k$ earns at most

$$(1 - D_k) \cdot \left(\sum_{i=1}^{N} p_i^k - \sum_{j=1}^{M} \left[b_j^k + \mathbb{1}_{\mathbb{R}_{>0}}(b_j^k) \cdot \kappa\right]\right) \tag{4.10}$$

for some $D_k \leq 1$. By deviating to another contract $\boldsymbol{c}^{-k} = (b_1^k, \ldots, b_M^k, p_1^k - \epsilon, \ldots, p_N^k)$ for some $\epsilon > 0$, firm $-k$ can earn $\sum_{i=1}^N p_i^k - \sum_{j=1}^M \left[b_j^k + \mathbb{1}_{\mathbb{R}_{>0}}(b_j^k) \cdot \kappa\right] - \epsilon$, which exceeds (4.10) for ϵ sufficiently small. Hence, firm $-k$ has an incentive to deviate; a contradiction. As a consequence, firms earn zero profits in equilibrium. Finally, it is straightforward to see that, in equilibrium, consumers are indifferent between both firms' offers. Otherwise, the firm that serves the market could slightly adjust its contract and earn strictly positive profits.

2. STEP: We show that *all payments to be made by consumers are of the same size, that is, $p_i^k = p^k$ for any $i \in \{1, \ldots, N\}$*. For the sake of contradiction, suppose that there exist $i, j \in \{1, \ldots, N\}$ such that firm k offers a contract with $p_i^k \neq p_j^k$ in equilibrium. In this case, maximal payment $p_{\max} := \max\{p_1^k, \ldots, p_N^k\}$ strictly exceeds minimal payment $p_{\min} := \min\{p_1^k, \ldots, p_N^k\}$. Without loss of generality, let $p_{\max} = p_1^k$ and $p_{\min} = p_2^k$. As firms earn zero profits in equilibrium, firm $-k$ could profitably deviate to a contract $\tilde{\boldsymbol{c}}^{-k} = (b_1^k, \ldots, b_M^k, p_1^k - \epsilon, p_2^k + \epsilon + \epsilon', p_3^k, \ldots, p_N^k)$ for some $\epsilon, \epsilon' > 0$ such that $p_1^k > p_2^k + \epsilon + \epsilon'$. Obviously, all consumers choose contract $\tilde{\boldsymbol{c}}^{-k}$ if

$$g(p_1^k)[p_1^k - \epsilon] + g(p_2^k + \epsilon + \epsilon')[p_2^k + \epsilon + \epsilon'] + \sum_{l=3}^N g(\Delta_l^p)p_l^k < \sum_{l=1}^N g(\Delta_l^p)p_l^k,$$

or, equivalently,

$$g(p_1^k)p_1^k - g(p_1^k)[p_1^k - \epsilon] > g(p_2^k + \epsilon + \epsilon')[p_2^k + \epsilon + \epsilon'] - g(p_2^k + \epsilon + \epsilon')p_2^k.$$

Rearranging this inequality yields

$$g(p_1^k)\epsilon > g(p_2^k + \epsilon + \epsilon')[\epsilon + \epsilon'],$$

which is satisfied for ϵ' sufficiently small. Hence, firm $-k$ indeed has a profitable deviation; a contradiction. As a consequence, in equilibrium, we must have $p_i^k = p_j^k$ for any two payments $i, j \in \{1, \ldots, N\}$, and any firm $k \in \{1, 2\}$.

3. STEP: We show that *both firms offer the same regular payments, that is, $p_i^k = p$ for any $k \in \{1, 2\}$ and any $i \in \{1, \ldots, N\}$*. For the sake of contradiction, let $p_i^k = p^k > p^{-k} = p_i^{-k}$. Since consumers are indifferent between both contracts, we conclude that $\sum_{j=1}^M b_j^k > \sum_{j=1}^M b_j^{-k}$. Hence, at least one bonus payment of firm k exceeds the corresponding bonus payment of firm $-k$. Without loss of generality, let $b_1^k > b_1^{-k}$.

In a first step, we argue that firm k offers a single bonus payment. For the sake of contradiction, suppose further that firm k offers at least two bonus payments in equilibrium. Then, firm k could profitably deviate to a contract $\tilde{\boldsymbol{c}}^k = (\sum_{j=1}^M b_j^k, 0, \ldots, 0, p^k + \epsilon, p^k, \ldots, p^k)$ for some $\epsilon > 0$ since all consumers choose the contract $\tilde{\boldsymbol{c}}^k$ if

$$
\begin{aligned}
&-g(p^k)(N-1)p^k - g(p^k+\epsilon)[p^k+\epsilon] + g\left(\sum_{j=1}^{M} b_j^k\right)\left(\sum_{j=1}^{M} b_j^k\right) \\
&\quad = -g(p^k)Np^k + \sum_{j=1}^{M} g(\max\{b_j^k, b_j^{-k}\})b_j^k \\
&\qquad - p^k[g(p^k+\epsilon)-g(p^k)] - g(p^k+\epsilon)\epsilon + \sum_{j=1}^{M}\left[g\left(\sum_{j=1}^{M} b_j^k\right) - g(\max\{b_j^k, b_j^{-k}\})\right]b_j^k \\
&\quad > -g(p^k)Np^{-k} + \sum_{j=1}^{M} g(\max\{b_j^k, b_j^{-k}\})b_j^{-k} \\
&\qquad - p^{-k}[g(p^k+\epsilon)-g(p^k)] + \left[g\left(\sum_{j=1}^{M} b_j^k\right) - g(\max\{b_1^k, b_1^{-k}\})\right]b_1^{-k} + \sum_{j=2}^{M}\left[g(b_j^{-k}) - g(\max\{b_j^k, b_j^{-k}\})\right]b_j^{-k} \\
&\quad = -g(p^k)(N-1)p^{-k} - g(p^k+\epsilon)p^{-k} + g\left(\sum_{j=1}^{M} b_j^k\right)b_1^{-k} + \sum_{j=2}^{M} g(b_j^{-k})b_j^{-k}.
\end{aligned}
$$

As consumers must be indifferent between both contracts in equilibrium, we have

$$
-g(p^k)Np^k + \sum_{j=1}^{M} g(\max\{b_j^k, b_j^{-k}\})b_j^k = -g(p^k)Np^{-k} + \sum_{j=1}^{M} g(\max\{b_j^k, b_j^{-k}\})b_j^{-k},
$$

so that the above inequality holds if and only if

$$
\begin{aligned}
&\overbrace{g(p^k+\epsilon)\epsilon}^{\to 0 \text{ as } \epsilon\to 0} + [p^k - p^{-k}]\overbrace{[g(p^k+\epsilon)-g(p^k)]}^{\to 0 \text{ as } \epsilon\to 0} < \overbrace{\left[g\left(\sum_{j=1}^{M} b_j^k\right) - g(\max\{b_1^k, b_1^{-k}\})\right]}^{>0 \text{ by A.2}}\overbrace{[b_1^k - b_1^{-k}]}^{>0} \\
&\qquad + \underbrace{\sum_{j=2}^{M}\left[g\left(\sum_{j=1}^{M} b_j^k\right) - g(\max\{b_j^k, b_j^{-k}\})\right]b_j^k}_{>0 \text{ by our assumption towards a contradiction}} - \sum_{j=2}^{M}\underbrace{\left[g(b_j^{-k}) - g(\max\{b_j^k, b_j^{-k}\})\right]}_{\leq 0 \text{ by A.2}} b_j^{-k};
\end{aligned}
$$

that is, if and only if ϵ is sufficiently small; a contradiction. Thus, given our initial assumption that k charges higher regular payments than $-k$, firm k must offer a single bonus payment. Thus, from now on, let $b_j^k = 0$ for any $j \neq 1$, and notice that $b_1^k > p^k$, as otherwise firm k would earn positive profits

In a second step, we argue that—given that firm k offers a single bonus payment and higher regular payments—firm $-k$ could profitably deviate to a contract

$$
\tilde{\boldsymbol{c}}^{-k} = (b_1^{-k}+\epsilon, b_2^{-k}, \ldots, b_M^{-k}, p^{-k}+\epsilon+\epsilon', p^{-k}, \ldots, p^{-k})
$$

for some $\epsilon, \epsilon' > 0$ such that $b_1^k > b_1^{-k} + \epsilon$ and $p_1^{-k} + \epsilon + \epsilon' < p_1^k$ since all consumers choose

$\tilde{\boldsymbol{c}}^{-k}$ if

$$-g(p^k)[Np^{-k}+\epsilon+\epsilon']+g(b_1^k)[b_1^{-k}+\epsilon]+\sum_{j=2}^{M}g(b_j^{-k})b_j^{-k}>-g(p^k)Np^k+g(b_1^k)b_1^k$$

or, equivalently,

$$-g(p^k)[\epsilon+\epsilon']+g(b_1^k)\epsilon>\underbrace{g(p^k)N[p^{-k}-p^k]+g(b_1^k)[b_1^k-b_1^{-k}]-\sum_{j=2}^{M}g(b_j^{-k})b_j^{-k}}_{=0\text{ as consumers must be indifferent between contracts}}.$$

This inequality is satisfied for ϵ' sufficiently small since $g(b_1^k)>g(p^k)$; a contradiction. As a consequence, in equilibrium, both firms offer the same regular payments. This further implies that $\sum_{j=1}^{M}b_j^k=\sum_{j=1}^{M}b_j^{-k}$, as otherwise at least one firm would earn positive profits; that is, either both firms offer a bonus contract or none does so.

4. STEP: We show that *firms offer at most one bonus payment in equilibrium.* For the sake of contradiction, suppose that firm k offers at least two bonus payments. By STEP 3, we have $\sum_{j=1}^{M}b_j^k=\sum_{j=1}^{M}b_j^{-k}$, and therefore $\sum_{j=1}^{M}b_j^{-k}>b_j^k$ for any $j\in\{1,\dots,N\}$. Denote the payment to be made by consumers in each period by p, which is the same across periods by STEP 2 and the same across firms by STEP 3. Then, firm $-k$ could profitably deviate to $\tilde{\boldsymbol{c}}^{-k}=(\sum_{j=1}^{M}b_j^{-k},0,\dots,0,p+\epsilon,p,\dots,p)$ for some $\epsilon>0$ since all consumers choose $\tilde{\boldsymbol{c}}^{-k}$ if

$$\begin{aligned}-g(p)(N-1)p-g(p+\epsilon)[p+\epsilon]+g\left(\sum_{j=1}^{M}b_j^{-k}\right)\left(\sum_{j=1}^{M}b_j^{-k}\right)\\ >-g(p)(N-1)p-g(p+\epsilon)p+g\left(\sum_{j=1}^{M}b_j^{-k}\right)b_1^k+\sum_{j=2}^{M}g(b_j^k)b_j^k,\end{aligned}$$

or, equivalently,

$$g(p+\epsilon)\epsilon<\overbrace{g\left(\sum_{j=1}^{M}b_j^{-k}\right)\left(\sum_{j=1}^{M}b_j^{-k}\right)-\underbrace{\left[g\left(\sum_{j=1}^{M}b_j^{-k}\right)b_1^k+\sum_{j=2}^{M}g(b_j^k)b_j^k\right]}_{<\sum_{j=1}^{M}g\left(\sum_{j=1}^{M}b_j^{-k}\right)b_j^k\text{ by A.2}}}^{>0},$$

which holds if and only if ϵ is sufficiently small; a contradiction. Thus, firms offer at most one bonus payment in equilibrium.

5. STEP: We show that *firms either offer the maximal bonus or no bonus at all.* We already know that each firm offers at most one bonus and that bonus firms offer the same overall bonus payment. Without loss of generality, we can assume that both firms use

the same bonus attribute; that is, we can solve the game as if there is only one bonus attribute, say, $b_1^k = b_1$. Again, denote the payment to be made by consumers in each period by p.

For the sake of contradiction, suppose that $0 < b_1 < \bar{b}$ in equilibrium. Then, firm k could profitably deviate to a contract $\tilde{\boldsymbol{c}}^k = (b_1 + \epsilon, 0, \ldots, 0, p + \frac{\epsilon+\epsilon'}{N}, \ldots, p + \frac{\epsilon+\epsilon'}{N})$ for some $\epsilon, \epsilon' > 0$ such that $\bar{b} \geq b_1 + \epsilon > p + \frac{\epsilon+\epsilon'}{N}$ since all consumers choose the contract $\tilde{\boldsymbol{c}}^k$ if

$$-g\left(p + \frac{\epsilon+\epsilon'}{N}\right) N \left[p + \frac{\epsilon+\epsilon'}{N}\right] + g(b_1+\epsilon)[b_1+\epsilon] > -g\left(p + \frac{\epsilon+\epsilon'}{N}\right) Np + g(b_1+\epsilon)b_1,$$

or, equivalently,

$$g(b_1+\epsilon)\epsilon > g\left(p + \frac{\epsilon+\epsilon'}{N}\right)[\epsilon+\epsilon'].$$

This inequality is satisfied for ϵ' sufficiently small since $g(b_1+\epsilon) > g\left(p + \frac{\epsilon+\epsilon'}{N}\right)$. As a consequence, firms either pay the maximal bonus or no bonus at all.

6. STEP: Notice that there are only two equilibrium candidates left (again we assume that both firms use the same bonus attribute, which is in fact without loss): either both firms offer $\boldsymbol{c}^{no} = (0, \ldots, 0)$ or both firms offer $\boldsymbol{c}^{bon} = \left(\bar{b}, 0, \ldots, 0, \frac{\kappa+\bar{b}}{N}, \ldots, \frac{\kappa+\bar{b}}{N}\right)$. We show that *both firms offering* $\boldsymbol{c}^{no}$ *cannot be an equilibrium.*

For the sake of contradiction, suppose that both firms offer the contract $\boldsymbol{c}^{no}$ in equilibrium. Now, firm k could profitably deviate to a contract $\tilde{\boldsymbol{c}}^k = \left(\bar{b}, 0, \ldots, 0, \frac{\kappa+\bar{b}}{N} + \epsilon, \frac{\kappa+\bar{b}}{N}, \ldots, \frac{\kappa+\bar{b}}{N}\right)$ for some $\epsilon > 0$ since all consumers choose $\tilde{\boldsymbol{c}}^k$ if

$$\underbrace{-g\left(\frac{\kappa+\bar{b}}{N}\right)(N-1)\left(\frac{\kappa+\bar{b}}{N}\right) - g\left(\frac{\kappa+\bar{b}}{N}+\epsilon\right)\left(\frac{\kappa+\bar{b}}{N}+\epsilon\right)}_{\to -g\left(\frac{\kappa+\bar{b}}{N}\right)[\kappa+\bar{b}] \text{ as } \epsilon\to 0} + g(\bar{b})\bar{b} > 0,$$

which holds for sufficiently small ϵ by Eq. (4.1). Hence, both firms offering $\boldsymbol{c}^{no}$ is not an equilibrium, which was to be proven.

7. STEP: It remains to be proven that *both firms offering contract* $\boldsymbol{c}^{bon}$ *is indeed an equilibrium.* We show that firm k has no incentive to deviate. In order to attract consumers, firm k has to reduce some payment p_i^k for $i \in \{1, \ldots, N\}$ by an amount $\epsilon > 0$, as increasing the bonus payment is not feasible. In order to benefit from this deviation, it has to increase some other payments p_j^k, $j \neq i$, to be made by consumers, or decrease the bonus payment b_1^k by an overall amount $\epsilon' > \epsilon$. As $g(\cdot)$ is increasing by Assumption 2, the most effective way of increasing payments is to equally spread ϵ' over all payments to be made by consumers, namely p_j^k for $j \neq i$. Then, the price cut ϵ is weighted by $g\left(\frac{\kappa+\bar{b}}{N}\right)$, while each price increase $\frac{\epsilon'}{N-1}$ is weighted by $g\left(\frac{\kappa+\bar{b}}{N} + \frac{\epsilon'}{N-1}\right)$. Thus, this deviation attracts

consumers if and only if

$$g\left(\frac{\kappa+\bar{b}}{N}\right)\epsilon > g\left(\frac{\kappa+\bar{b}}{N}+\frac{\epsilon'}{N-1}\right)\epsilon',$$

which can only be satisfied for $\epsilon > \epsilon'$; a contradiction. Hence, firm k has no incentive to deviate, so that both firms offering the contract $\boldsymbol{c}^{bon}$ is an equilibrium. Since this was the last remaining equilibrium candidate, the equilibrium is unique.

Finally, consider the case of $L > 2$ firms. It is straightforward to show that in the essentially unique symmetric equilibrium all firms offer the contract $\boldsymbol{c}^{bon}$. This completes the proof. □

Appendix B: Endogenous Attribute Space

B.1: Model

We extend our baseline model from Section 2 in the following two ways: suppose first that the number of bonus payments and the number of regular payments are unbounded and second that consumers buying at firm k incur transaction costs, $\tau(N^k)$, depending on the number of non-zero regular payments specified in firm k's contract, which we denote as N^k.

We assume that consumers' transaction costs are strictly increasing and convex in the number of non-zero regular payments. For technical reasons and without loss of generality, we treat τ as a twice continuously differentiable function from $\mathbb{R}_+$ to $\mathbb{R}_+$ and we further assume that $\tau', \tau'' > 0$. In addition, we impose assumptions on the cost function: (i) $\tau(2) + 2g\left(\frac{\kappa+\bar{b}}{2}\right)\left(\frac{\kappa+\bar{b}}{2}\right) < g(\bar{b})\bar{b}$, and (ii) $\tau'(2) < g'\left(\frac{\kappa+\bar{b}}{2}\right)\left(\frac{\kappa+\bar{b}}{2}\right)^2$, and (iii) $\lim_{N\to\infty}\tau'(N) = \infty$. Notice that Assumptions (i) and (ii) are the natural extensions of Equation (4.1), which we imposed on the inefficiency arising from paying a bonus in order to allow firms to increase a consumer's focus-weighted utility using a bonus payment and to break even at the same time. Assumption (iii) is a typical Inada-Condition to ensure that a profit-maximizing number of regular payments exists.

The remainder of Appendix B is organized as follows. In Section B.2 we derive a monopolist's optimal contract offer. In Section B.3 we characterize (symmetric) competitive equilibria and compare the competitive outcome to the monopolistic one. Importantly, as long as transaction costs are sufficiently convex, our main result (i.e., Corollary 1) is robust to this extension.

B.2: Monopolistic Market

The monopolist's optimal contract offer is characterized in the following lemma.

Lemma 2. *A contract maximizes the monopolist's profit only if*

(i) the regular payments made by consumers are equally spread across N^{mon} periods, that is, $p_i(N^{mon}) = p(N^{mon})$ for any $i \in \{1, \ldots, N^{mon}\}$, whereby $N^{mon} \in \{\lfloor N^* \rfloor, \lceil N^* \rceil\}$ and N^* is the unique solution to

$$g'(p(N))p(N)^2 = \tau'(N),$$

(ii) and, if bonus payment(s) are made, the maximal bonus is paid using a single payment.

Proof. In order to prove the statement, we can make use of the insights derived in Lemma 1, where we have characterized the optimal contract offer for a fixed number of regular and bonus payments, respectively. Indeed, the second part immediately follows from Lemma 1 since, even if the number of potential bonus payments is fixed, the monopolist will not want to pay more than one bonus. Thus, it remains to be shown that also (i) holds.

Remember that we have seen in the proof of Lemma 1 that regular payments have to be of equal size and that there is either a single bonus payment that is maximal or none bonus at all. In addition, we know that for a given number of regular payments, $N \in \mathbb{N}$, it has to hold that

$$Ng(p(N))p(N) = \tilde{V} - \tau(N), \tag{4.11}$$

where $\tilde{V} = \tilde{v} + g(\bar{b})\bar{b}$ if a bonus is paid and $\tilde{V} = \tilde{v}$ otherwise. Notice that consumers are willing to buy at a price of zero only if $N \leq \tau^{-1}(\tilde{V}) =: \overline{N}^{mon}$, where τ^{-1} is the inverse of the transaction cost function, which indeed exists as τ is strictly increasing. This in turn implies that the optimal regular payments are characterized by (4.11) as long as N lies weakly below $\overline{N}^{mon}$.

Now ignore the integer constraint for a moment and suppose that $N \in (0, \overline{N}^{mon})$ holds. Then, when applying the Implicit Function Theorem to (4.11), we obtain

$$p'(N) = -\frac{1}{N}\frac{g(p(N))p(N) + \tau'(N)}{g(p(N)) + g'(p(N))p(N)} < 0. \tag{4.12}$$

Since the size of the bonus payment is independent of N and as both $N = 0$ and $N = \overline{N}^{mon}$ imply zero profit, the monopolist chooses N as to maximize $Np(N)$ subject to $N \in (0, \overline{N}^{mon})$. In addition, as the function $Np(N)$ is continuous in N on the interval $(0, \overline{N}^{mon})$ and also strictly positive by (4.11), it has at least one local maximum in this interval, so that—ignoring integer constraints—the optimal number of regular payments solves

$$
\begin{aligned}
0 &= p(N) + Np'(N) \\
&= p(N) - \frac{g(p(N))p(N) + \tau'(N)}{g(p(N)) + g'(p(N))p(N)} \\
&= \frac{1}{g(p(N)) + g'(p(N))p(N)} \cdot \Big[g'(p(N))p(N)^2 - \tau'(N) \Big].
\end{aligned}
\tag{4.13}
$$

Here, the second equality follows from (4.12) and the last equality is a simple re-arrangement. Hence, we conclude that the optimal number of payments has to solve

$$
g'(p(N))p(N)^2 = \tau'(N). \tag{4.14}
$$

Since Assumption 3 and Eq. (4.12) imply that the left-hand side of (4.14) strictly decreases in N and since $\tau'' > 0$ implies that the right-hand side of (4.14) strictly increases in N, there exists a unique solution to (4.14), which further implies that $Np(N)$ has a unique local maximum, N^*, on the interval $(0, \overline{N}^{mon})$. Finally, as $Np(N)$ strictly increases (decreases) for any $N < N^*$ ($N > N^*$), the statement follows immediately when taking the integer constraint into account. $\square$

Before we can prove the analogue to Proposition 1, the next lemma derives further properties of the monopolist's optimal contract that will be useful in the proof later on.

Lemma 3. *The monopolist's contract offer delineated in Lemma 2 satisfies:*

(i) $\frac{\partial N^*}{\partial v} > 0$ *and* $\lim_{v\to\infty} \frac{\partial N^*}{\partial v} = 0$.

(ii) There exists some $v' \in \mathbb{R}_+$ *such that for any* $v > v'$ *we have* $\frac{\partial N^{mon}}{\partial v} = 0$.

(iii) There exists some $v'' \in \mathbb{R}_+$ *such that for any* $v > v''$ *we have* $p(N^{mon}) > \overline{b}$.

(iv) There is some $\overline{\tau} \in \mathbb{R}_+$ *so that for any cost function with* $\tau''(\cdot) > \overline{\tau}$ *the monopolist chooses the same number of regular payments irrespective of whether she pays a bonus or not.*

Proof. First, we derive some preliminary results. Subsequently, we directly prove the statements.

PRELIMINARIES: First, when applying the Implicit Function Theorem to (4.11), we obtain

$$
\frac{\partial}{\partial \tilde{V}} p(N, \tilde{V}) = \frac{1}{N} \frac{1}{g'(p(N,\tilde{V}))p(N,\tilde{V}) + g(p(N,\tilde{V}))}. \tag{4.15}
$$

Second, when applying the Implicit Function Theorem to (4.14), we obtain

$$\begin{aligned}
\frac{dN^*}{d\tilde{V}} &= -\frac{g''(p(N^*,\tilde{V}))p(N^*,\tilde{V})^2\frac{\partial p}{\partial \tilde{V}} + 2g'(p(N^*,\tilde{V}))p(N^*,\tilde{V})\frac{\partial p}{\partial \tilde{V}}}{g''(p(N^*,\tilde{V}))p(N^*,\tilde{V})^2\frac{\partial p}{\partial N} + 2g'(p(N^*,\tilde{V}))p(N^*,\tilde{V})\frac{\partial p}{\partial N} - \tau''(N^*)} \\
&= -\left(\frac{\partial p/\partial \tilde{V}}{\partial p/\partial N}\right)\cdot\left(\frac{1}{1-\frac{\tau''(N^*)}{g''(p(N^*,\tilde{V}))p(N^*,\tilde{V})^2\frac{\partial p}{\partial N}+2g'(p(N^*,\tilde{V}))p(N^*,\tilde{V})\frac{\partial p}{\partial N}}}\right) \\
&= \left(\frac{1}{g(p(N^*,\tilde{V})p(N^*,\tilde{V})+\tau'(N^*)}\right)\cdot\left(\frac{1}{1-\frac{1}{\frac{\partial p}{\partial N}}\frac{\tau''(N^*)}{g''(p(N^*,\tilde{V}))p(N^*,\tilde{V})^2+2g'(p(N^*,\tilde{V}))p(N^*,\tilde{V})}}\right) \\
&= \frac{1}{g(p(N^*,\tilde{V})p(N^*,\tilde{V})+\tau'(N^*)+N^*\tau''(N^*)\cdot\left(\frac{g(p(N^*,\tilde{V}))+g'(p(N^*,\tilde{V}))p(N^*,\tilde{V})}{g''(p(N^*,\tilde{V}))p(N^*,\tilde{V})^2+2g'(p(N^*,\tilde{V}))p(N^*,\tilde{V})}\right)} \\
&> 0,
\end{aligned}$$

where the second equality is a simple re-arrangement, the third equality follows from inserting (4.12) and (4.15), and the last equality follows from inserting (4.12) once more.

PART (i): Since $\frac{dN^*}{d\tilde{V}} > 0$ and since $\frac{dN^*}{d\tilde{V}} < \frac{1}{\tau'(N^*)}$, we obtain (i) simply from the fact that $\tilde{V}$ increases with v and goes to infinity as v approaches infinity and that $\lim_{N\to\infty}\tau'(N) = \infty$.

PART (ii): Follows immediately from (i).

PART (iii): Follows immediately from (4.11), when taking the limit of v to infinity and keeping in mind that N^{mon} is constant for sufficiently large values of v by (ii).

PART (iv): Follows immediately from the fact that $N^* \geq 1$ and that $\lim_{\overline{\tau}\to\infty}\frac{dN^*}{d\tilde{V}} = 0$. □

Using the above lemmata, we can fully characterize the monopolist's contract offer. In particular, the following proposition shows that our previous result on the monopolistic outcome still holds if transaction costs are sufficiently convex.

Proposition 3. *The following statements hold true:*

(i) There exist a threshold value $\check{\kappa} > 0$ and, for any $\kappa < \check{\kappa}$, a threshold value $\check{v}_1(\kappa) > 0$ such that the monopolist offers a bonus contract if $\kappa < \check{\kappa}$ and $v < \check{v}_1(\kappa)$.

(ii) For any $\kappa > 0$, there exists a threshold value $\check{v}_2(\kappa) \geq 0$ such that the monopolist does not offer a bonus contract if $v > \check{v}_2(\kappa)$.

(iii) If transaction costs are sufficiently convex (i.e., if $\tau''(N)$ is sufficiently large for any N), then $\check{v}_1(\kappa) = \check{v}_2(\kappa) = \check{v}(\kappa)$ for any $\kappa < \check{\kappa}$ and $\check{v}$ monotonically decreases in κ on $[0, \check{\kappa})$.

Proof. PART (i): Obviously, if $v = 0$, the monopolist can earn positive profits only when offering a bonus contract. As $\tau(2) + g(\kappa/2)(\kappa/2) < g(\bar{b})\bar{b}$ by assumption, the monopolist can indeed earn strictly positive profits using a bonus contract even if $v = 0$. The statement then follows from the fact that the monopolist's profit is continuous in v conditional on offering a certain type of contract (i.e., a bonus contract or a contract without a bonus payment).

PART (ii): Follows immediately from Lemma 3 Part (iii) using basically the same arguments as in the proof of Proposition 1.

PART (iii): By Lemma 3 Part (iv), the monopolist chooses the same number of regular payments irrespective of whether she pays a bonus or not. Given this fact, the proof is analogous to that of Proposition 1. $\square$

B.3: Competitive Market

Next, we analyze the competitive outcome in our extended model with transaction costs.

Proposition 4. *If $L = 2$, an equilibrium exists and any equilibrium has the following properties:*

(i) the market is covered and firms earn zero profits,

(ii) the regular payments made by consumers are equally spread across N_k^{com} periods, that is, $p_i^k(N_k^{com}) = p^k(N_k^{com})$ for $i \in \{1, \dots, N_k^{com}\}$ and $k \in \{1, 2\}$, whereby $\underline{N} \leq N_k^{com} \leq \lceil N^{**} \rceil$ and N^{**} is the unique solution to

$$g'\left(\frac{\kappa + \bar{b}}{N}\right)\left(\frac{\kappa + \bar{b}}{N}\right)^2 = \tau'(N)$$

while $\underline{N}$ is the maximum of two and the smallest natural number N that satisfies

$$\frac{\kappa + \bar{b}}{N + 1} \leq \frac{\tau(N + 1) - \tau(N)}{g\left(\frac{\kappa+\bar{b}}{N}\right) - g\left(\frac{\kappa+\bar{b}}{N+1}\right)}, \tag{4.16}$$

(iii) both firms offer the maximum bonus using a single bonus payment, and

(iv) both firms provide the exact same focus-weighted utility to consumers.

If $L \geq 3$, a symmetric equilibrium exists and any such equilibrium satisfies properties (i) – (iv). For any $L \geq 2$, there exists a symmetric equilibrium with $N_k^{com} \in \{\lfloor N^{**} \rfloor, \lceil N^{**} \rceil\}$.

Proof. We prove the statement for $L = 2$, while the proof for $L \geq 3$ is a straightforward extension. Again, we can make use from the insights derived in the main text, namely, Proposition 2. For instance, we already know that firms earn zero profits in equilibrium and that consumers are indifferent between both offers, that is, Part (iv) immediately follows from Proposition 2. In addition, it follows directly from Proposition 2 that firms offer at most one bonus payment. Hence, without loss of generality, let $M = 1$ in the following.

The remainder of the proof proceeds in four steps. In a first step, we show that in any equilibrium $N_k^{com} \geq 2$, which in turn implies that firms offer bonus contracts. In a second step, we prove that in any equilibrium $N_k^{com} \leq \lceil N^{**} \rceil$. In a third step, we show that a symmetric equilibrium with $N^{com} \in \{\lfloor N^{**} \rfloor, \lceil N^{**} \rceil\}$ exists. In a fourth step, we show that an equilibrium with $N^k \in \{2, \ldots, \lfloor N^{**} \rfloor\}$ exists if and only if (4.16) holds at $N = N^k$ and that (4.16) is more likely to be fulfilled for larger values of N.

1. STEP: By Proposition 2, we know that for $M = 1$ and a fixed number of regular payments $N \geq 2$, there exists a unique equilibrium in which both firms offer the contract

$$\boldsymbol{c}^{bon}(M = 1, N) = \left(\bar{b}, \frac{\kappa + \bar{b}}{N}, \ldots, \frac{\kappa + \bar{b}}{N}\right).$$

Moreover, if firms choose at most one non-zero regular payment, they cannot profitably offer a bonus. But then the only other equilibrium candidate is setting all regular payments to zero.

For the sake of contradiction, suppose that firms do not offer a bonus payment in equilibrium, but set all regular payments to zero. Then, firm k could profitably deviate to a contract

$$\tilde{\boldsymbol{c}}^k(M = 1, N^k = 2) = \left(\bar{b}, \frac{\kappa + \bar{b} + \epsilon}{2}, \frac{\kappa + \bar{b} + \epsilon}{2}\right)$$

for some $\epsilon > 0$ since all consumers choose $\tilde{\boldsymbol{c}}^k(M = 1, N^k = 2)$ if

$$g(\bar{b})\bar{b} - 2g\left(\frac{\kappa + \bar{b} + \epsilon}{2}\right)\left(\frac{\kappa + \bar{b} + \epsilon}{2}\right) - \tau(2) > 0,$$

which holds for sufficiently small values ϵ by the assumption that $\tau(2) + 2g\left(\frac{\kappa+\bar{b}}{2}\right)\left(\frac{\kappa+\bar{b}}{2}\right) < g(\bar{b})\bar{b}$; a contradiction. Hence, we have $N_k^{com} \geq 2$ in any equilibrium.

2. STEP: For the sake of contradiction, suppose that $N_k^{com} > \lceil N^{**} \rceil$ holds in equilibrium. Then, firm k could profitably deviate to a contract

$$\tilde{\boldsymbol{c}}^k(M = 1, N^k) = \left(\bar{b}, \frac{\kappa + \bar{b} + \epsilon}{N^k}, \ldots, \frac{\kappa + \bar{b} + \epsilon}{N^k}\right)$$

for $N^k \in \{\lfloor N^{**} \rfloor, \lceil N^{**} \rceil\}$ and some $\epsilon > 0$ since all consumers choose $\tilde{\boldsymbol{c}}^k(M=1, N^k)$ if

$$\begin{aligned} -N^k g\left(\frac{\kappa+\bar{b}+\epsilon}{N^k}\right)\left(\frac{\kappa+\bar{b}+\epsilon}{N^k}\right) - \tau(N^k) \\ > -N^k g\left(\frac{\kappa+\bar{b}+\epsilon}{N^k}\right)\left(\frac{\kappa+\bar{b}}{N^{com}_{-k}}\right) - [N^{com}_{-k} - N^k] g\left(\frac{\kappa+\bar{b}}{N^{com}_{-k}}\right)\left(\frac{\kappa+\bar{b}}{N^{com}_{-k}}\right) - \tau(N^{com}_{-k}). \end{aligned} \tag{4.17}$$

Notice that the right-hand side of the above inequality is smaller than

$$-N^{com}_{-k} g\left(\frac{\kappa+\bar{b}}{N^{com}_{-k}}\right)\left(\frac{\kappa+\bar{b}}{N^{com}_{-k}}\right) - \tau(N^{com}_{-k})$$

by Assumption 2 and that

$$-N^k g\left(\frac{\kappa+\bar{b}+\epsilon}{N^k}\right)\left(\frac{\kappa+\bar{b}+\epsilon}{N^k}\right) - \tau(N^k) > -N^{com}_{-k} g\left(\frac{\kappa+\bar{b}}{N^{com}_{-k}}\right)\left(\frac{\kappa+\bar{b}}{N^{com}_{-k}}\right) - \tau(N^{com}_{-k})$$

by our assumption toward a contradiction and the definition of N^{**} as the unique minimizer of $Ng\left(\frac{\kappa+\bar{b}}{N}\right)\left(\frac{\kappa+\bar{b}}{N}\right) + \tau(N)$. Consequently, Inequality (4.17) holds for sufficiently small values of ϵ; a contradiction. Hence, we conclude that $N^{com}_k \leq \lceil N^{**} \rceil$ in any equilibrium.

3. STEP: Suppose that both firms offer the contract

$$\boldsymbol{c}^{bon}(M=1, N^{com}) = \left(\bar{b}, \frac{\kappa+\bar{b}}{N^{com}}, \ldots, \frac{\kappa+\bar{b}}{N^{com}}\right),$$

where N^{com} is chosen as to minimize $Ng\left(\frac{\kappa+\bar{b}}{N}\right)\left(\frac{\kappa+\bar{b}}{N}\right)+\tau(N)$; that is, $N^{com} \in \{\lfloor N^{**} \rfloor, \lceil N^{**} \rceil\}$. By STEP 3, no firm has an incentive to increase the number of regular payments, which by the way implies that $\underline{N} \leq \lceil N^{**} \rceil$. In addition, notice that the regular payments of firm k would determine the focus-weights if it decides to decrease the number of regular payments in a way that allows for non-negative profits. But then, by the definition of N^{com}, decreasing the number of regular payments cannot increase focus-weighted utility and yield non-negative profits at the same time. Hence, no firm has an incentive to decrease the number of regular payments and therefore no incentive to deviate, which was to be proven.

4. STEP: Suppose that both firms offer the contract

$$\boldsymbol{c}^{bon}_k(M=1, N^{com}_k) = \left(\bar{b}, \frac{\kappa+\bar{b}}{N^{com}_k}, \ldots, \frac{\kappa+\bar{b}}{N^{com}_k}\right),$$

where $N^{com}_k \in \{2, \ldots, \lfloor N^{**} \rfloor\}$. First, suppose that both firms choose the same number of regular payments, that is, $N_1^{com} = N_2^{com} = N^{com}$. Since $N^{com} \leq \lfloor N^{**} \rfloor$, by same argument as in STEP 3, firms do not have an incentive to decrease the number of regular payments. In addition, firms do not have an incentive to increase the number of payments if and only if

$$\begin{aligned} &N^{com} g\left(\frac{\kappa+\bar{b}}{N^{com}}\right)\left(\frac{\kappa+\bar{b}}{N^{com}}\right) - \tau(N^{com}) \\ &\qquad > N^{com} g\left(\frac{\kappa+\bar{b}}{N^{com}}\right)\left(\frac{\kappa+\bar{b}}{N^{com}+1}\right) + g\left(\frac{\kappa+\bar{b}}{N^{com}+1}\right)\left(\frac{\kappa+\bar{b}}{N^{com}+1}\right) - \tau(N^{com}+1), \end{aligned}$$

which holds if and only if (4.16) holds at $N = N^{com}$.

Second, notice that

$$\underbrace{\frac{\partial}{\partial N}\frac{\kappa+\bar{b}}{N+1}\left[g\left(\frac{\kappa+\bar{b}}{N}\right) - g\left(\frac{\kappa+\bar{b}}{N+1}\right)\right]}_{<0 \text{ by A.3}} - \underbrace{\frac{\partial}{\partial N}\Big[\tau(N+1) - \tau(N)\Big]}_{<0 \text{ as } \tau''>0} < 0,$$

which in turn implies that (4.16) is more likely to hold for larger values of N.

Third, let $N_1^{com} \neq N_2^{com}$. If $N_k^{com} \leq N_{-k}^{com} - 2$, firm k could profitbaly deviate to the contract

$$\tilde{\boldsymbol{c}}_k^{bon}(M=1, N_k^{com}+1) = \left(\bar{b}, \frac{\kappa+\bar{b}+\epsilon}{N_k^{com}+1}, \ldots, \frac{\kappa+\bar{b}+\epsilon}{N_k^{com}+1}\right)$$

for some sufficiently small $\epsilon > 0$, as $N_k^{com} \leq \lfloor N^{**} \rfloor - 1$ and as firm k's regular payments would fully determine the focus-weights. If $N_k^{com} = N_{-k}^{com} - 1$, then

$$\begin{aligned} &N_k^{com} g\left(\frac{\kappa+\bar{b}}{N_k^{com}}\right)\left(\frac{\kappa+\bar{b}}{N_k^{com}}\right) - \tau(N_k^{com}) \\ &\qquad = N_k^{com} g\left(\frac{\kappa+\bar{b}}{N_k^{com}}\right)\left(\frac{\kappa+\bar{b}}{N_k^{com}+1}\right) + g\left(\frac{\kappa+\bar{b}}{N_k^{com}+1}\right)\left(\frac{\kappa+\bar{b}}{N_k^{com}+1}\right) - \tau(N_k^{com}+1), \end{aligned}$$

has to hold, as consumers have to be indifferent between both contracts in equilibrium. But then (4.16) holds at $N = N_k^{com}$ and as it is more likely to hold for larger values of N it also holds at $N = N_{-k}^{com}$. This completes the proof. □

The preceding proposition shows that the competitive equilibrium has the same qualitative properties as before, namely, firms offer a single, maximum bonus payment and the regular payments are of equal size. The only difference compared to our baseline model is that there can exist multiple equilibria that differ in the number of non-zero regular payments. It is easy to see, however, that this multiplicity vanishes for sufficiently convex transaction costs. Consequently, as long as the transaction cost function is sufficiently convex, also our result on the comparison of monopolistic and competitive outcomes remains qualitatively the same.

Corollary 3. *If transaction costs are sufficiently convex (i.e., if $\tau''(N)$ is sufficiently large for any N), there is a unique (symmetric) competitive equilibrium. In this equilibrium all firms choose the same number of regular payments as a monopolist would do. If in addition the consumers' valuation for the product is sufficiently high, the contractual inefficiencies are strictly lower in a monopolistic than in a competitive market.*

Proof. As τ becomes more convex, the right-hand side of (4.16) becomes smaller for small values of N and larger for large values of N. Hence, $\underline{N}$ becomes larger as τ becomes more convex and eventually only one equilibrium candidate survives. In addition, as τ becomes more convex, both N^* and N^{**} become less sensitive to the level of the regular payments, so that for sufficiently convex transaction costs $N^{mon} = N^{com}$. $\square$

Declaration of Contribution

Hereby I, Mats Köster, declare that the chapter "Attention-Driven Demand for Bonus Contracts" is co-authored by Markus Dertwinkel-Kalt and Florian Peiseler. It has been published in the European Economic Review (Dertwinkel-Kalt *et al.*, 2019b). All authors contributed equally to the chapter.

Signature of coauthor 1 (Markus Dertwinkel-Kalt): M. Dertwinkel-Kalt

Signature of coauthor 2 (Florian Peiseler): ______________________

Conclusion

My dissertation is part of a broader research agenda in which I try to shed some light on the mechanisms underlying various behavioral puzzles in economics — such as the Allais paradoxes and the simultaneous demand for gambling and insurance (Chapter 2) or the demand effects of "shrouding" surcharges (Dertwinkel-Kalt *et al.*, 2020b) —, finance — such as the disposition effect (Dertwinkel-Kalt *et al.*, 2020a) or under-diversification (Chapter 3) —, and business economics — such as behavior observed in the newsvendor problem (Dertwinkel-Kalt and Köster, 2017). Combining rigorous theoretical models of individual behavior with experimental tests of the behavioral implications in lab and field settings, this line of work contributes to our understanding of how attention shapes economic decisions. In a second line of work, I make use of these insights on behavior to improve our understanding of how firms in different market environments respond to consumers being susceptible to salience effects or being inattentive more generally (Chapter 4; Dertwinkel-Kalt and Köster, 2020a; Heidhues, Köster and Kőszegi, 2020).

While this dissertation provides new answers to a few old questions (e.g. how to rationalize the Allais paradoxes), it also raises new ones for which we do not have a satisfactory answer yet. The questions I am currently most excited about include:

- What makes a choice problem complex?
- Does salience mitigate the "costs" associated with making complex decisions?

Building on the basic frameworks laid out in Chapters 1 and 3, I plan to develop a fully-fledged model of the (cognitive) costs associated with making complex decisions, which takes seriously the role that salient cues play in "helping" consumers reach a decision. Subsequently, having established a better understanding of the interaction of complexity and salience, I want to analyze how firms — offering (intentionally) complex products — can manipulate consumption behavior by providing consumers with salient cues that push them in a certain direction, and how regulators — acting in the interest of consumers — may best respond to these practices. Understanding the determinants of choice complexity and its strategic implications may indeed inform policy design across various domains.

Bibliography

Abeler, J., Huffman, D. and Raymond, C. (2020). Bounded rationality, complexity and optimal incentives. Mimeo, unpublished.

— and Jäger, S. (2015). Complex tax incentives. *American Economic Journal: Economic Policy*, **7** (3), 1–28.

Ackert, L. F., Church, B. K. and Qi, L. (2015). An experimental examination of portfolio choice. *Review of Finance*, **20** (4), 1427–1447.

Ahmetoglu, G., Furnham, A. and Fagan, P. (2014). Pricing practices: A critical review of their effects on consumer perceptions and behaviour. *Journal of Retailing and Consumer Services*, **21** (5), 696–707.

Andersen, S., Harrison, G. W., Lau, M. I. and Rutström, E. E. (2014). Discounting behavior: A reconsideration. *European Economic Review*, **71**, 15–33.

Andreoni, J. and Sprenger, C. (2012). Estimating time preferences from convex budgets. *American Economic Review*, **102** (7), 3333–3356.

Augenblick, N., Niederle, M. and Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics*, **130** (3), 1067–1115.

— and Rabin, M. (2019). An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies*, **86** (3), 941–975.

Azevedo, E. M. and Gottlieb, D. (2012). Risk-neutral firms can extract unbounded profits from consumers with prospect theory preferences. *Journal of Economic Theory*, **147** (3), 1291–1299.

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

Balakrishnan, U., Haushofer, J. and Jakiela, P. (2020). How soon is now? evidence of present bias from convex time budget experiments. *Experimental Economics*, **23**, 294–321.

Bali, T. G., Cakici, N. and Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, **99** (2), 427–446.

Baltussen, G. and Post, T. (2011). Irrational diversification: An examination of the portfolio construction decision. *Journal of Financial and Quantitative Analysis*, **46** (5), 1463–1491.

Banovetz, J. and Oprea, R. (2020). Complexity and procedural choice. `https://drive.google.com/file/d/15wtpNeRnNSqFJZg6GUePnjh2UG5gmHFM/view`, unpublished.

Barberis, N. and Huang, M. (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review*, **98** (5), 2066–2100.

Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives*, **27** (1), 173–195.

Barseghyan, L., Molinari, F., O'Donoghue, T. and Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review*, **103** (6), 2499–2529.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, **30** (5), 961–981.

Benartzi, S. and Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review*, **91** (1), 79–98.

Berkhout, P., Hartog, J. and Webbink, D. (2010). Compensation for earnings risk under worker heterogeneity. *Southern Economic Journal*, **76** (3), 762–790.

Bernheim, B. D. and Sprenger, C. (2020). On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting. *Econometrica*, **88** (4), 1363–1409.

Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Comentarii Academiae Scientiarum Imperialis Petropolitanae*, **5** (2), 175–92 [translated by L. Sommer (1954) in *Econometrica*, 22, 23–36].

Birnbaum, M. H. (2005). A comparison of five models that predict violations of first-order stochastic dominance in risky decision making. *Journal of Risk and Uncertainty*, **31** (3), 263–287.

Bordalo, P., Coffman, K., Gennaioli, N. and Shleifer, A. (2016a). Stereotypes. *Quarterly Journal of Economics*, **131** (4), 1753–1794.

—, Gennaioli, N. and Shleifer, A. (2012). Salience theory of choice under risk. *Quarterly Journal of Economics*, **127** (3), 1243–1285.

—, — and — (2013a). Salience and asset prices. *American Economic Review, Papers & Proceedings*, **103** (3), 623–628.

—, — and — (2013b). Salience and consumer choice. *Journal of Political Economy*, **121** (5), 803–843.

—, — and — (2016b). Competition for attention. *Review of Economic Studies*, **83** (2), 481–513.

—, — and — (2020a). Memory, attention, and choice. *Quarterly Journal of Economics*, **135** (3), 1399–1442.

—, Tabellini, M. and Yang, D. (2020b). Stereotypes and politics. `https://www.dropbox.com/s/epn2m7nf7fugd2x/Bordalo\%20Tabellini\%20Yang_april2020_v2.pdf?dl=0`, unpublished.

Boyer, B., Mitton, T. and Vorkink, K. (2010). Expected idiosyncratic skewness. *Review of Financial Studies*, **23** (1), 169–202.

Breunig, C., Huck, S., Schmidt, T. and Weizsacker, G. (2019). The standard portfolio choice problem in germany. `https://www.wiwi.hu-berlin.de/en/Professorships/vwl/microeconomics/people/gweizsaecker/paper_jul2019.pdf`, unpublished.

Brocas, I., Carrillo, J., Giga, A. and Zapatero, F. (2016). Skewness seeking in a dynamic portfolio choice experiment. `https://www.cemfi.es/ftp/pdf/papers/wshop/skweness_march17.pdf`, unpublished.

Bruhin, A., Fehr-Duda, H. and Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica*, **78** (4), 1375–1412.

—, Manai, M. and Santos-Pinto, L. (2019). Risk and rationality: The relative importance of probability weighting and choice set dependence. `http://www.hec.unil.ch/lspinto/Papers%20&%20CV/PrWeighting_vs_ChoiceSetDependence.pdf`, unpublished.

Brunnermeier, M. K. and Parker, J. A. (2005). Optimal expectations. *American Economic Review*, **95** (4), 1092–1118.

Bushong, B., Rabin, M. and Schwartzstein, J. (forthcoming). A model of relative thinking. *Review of Economic Studies*.

Camerer, C. F. and Ho, T.-H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, **8** (2), 167–196.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, **1** (3), 195–228.

Caplin, A. and Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, **105** (7), 2183–2203.

Carlin, B. I. (2009). Strategic price complexity in retail financial markets. *Journal of Financial Economics*, **91** (3), 278–287.

Carlson, J. P. and Weathers, D. (2008). Examining differences in consumer reactions to partitioned prices with a variable number of price components. *Journal of Business Research*, **61** (7), 724–731.

Chen, J., Hong, H. and Stein, J. C. (2001). Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices. *Journal of Financial Economics*, **61** (3), 345–381.

Chernev, A., Böckenholt, U. and Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, **25** (2), 333–358.

Chetty, R., Looney, A. and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, **99** (4), 1145–77.

Choi, D., Lou, D. and Mukherjee, A. (2016). The effect of superstar firms on college major choice. `http://www.pbcsf.tsinghua.edu.cn/research/loudong/paper/W%20The%20Effect%20of%20Superstar%20Firms%20on.pdf`, unpublished.

Chung, S.-H. and Herrnstein, R. J. (1967). Choice and delay of reinforcement. *Journal of the Experimental Analysis of Behavior*, **10** (1), 67–74.

Chunhachinda, P., Dandapani, K., Hamid, S. and Prakash, A. J. (1997). Portfolio selection and skewness: Evidence from international stock markets. *Journal of Banking & Finance*, **21** (2), 143–167.

Cohen, J. D., Ericson, K. M., Laibson, D. and White, J. M. (2020). Measuring time preferences. *Journal of Economic Literature*, **58** (2), 299–347.

Conrad, J., Dittmar, R. F. and Ghysels, E. (2013). Ex ante skewness and expected stock returns. *Journal of Finance*, **68** (1), 85–124.

Crawford, V. P. and Iriberri, N. (2007). Fatal attraction: Salience, naivete, and sophistication in experimental" hide-and-seek" games. *American Economic Review*, **97** (5), 1731–1750.

de Clippel, G., Eliaz, K. and Rozen, K. (2014). Competing for consumer inattention. *Journal of Political Economy*, **122** (6), 1203–1234.

DellaVigna, S. and Malmendier, U. (2004). Contract design and self-control: Theory and evidence. *Quarterly Journal of Economics*, **119** (2), 353–402.

Dertwinkel-Kalt, M., Frey, J. and Köster, M. (2020a). Optimal stopping in a dynamic salience model. `https://www.cesifo.org/DocDL/cesifo1_wp8496.pdf`, unpublished.

—, Gerhardt, H., Riener, G., Schwerter, F. and Strang, L. (2019a). Concentration bias in intertemporal choice. `https://www.dropbox.com/s/dv20mcu0qkygmjz/Concentration_Bias_in_Intertemporal_Choice.pdf?dl=0`, unpublished.

— and Köster, M. (2015). Violations of first-order stochastic dominance as salience effects. *Journal of Behavioral and Experimental Economics*, **59**, 42–46.

— and Köster, M. (2017). Salient compromises in the newsvendor game. *Journal of Economic Behavior & Organization*, **141**, 301–315.

— and Köster, M. (2020a). Attention to online sales: The role of brand image concerns. *DICE Discussion Paper No. 335*.

— and Köster, M. (2020b). Salience and skewness preferences. *Journal of the European Economic Association*, **18** (5), 2057–2107.

— and Köster, M. (2020c). Salient cues and complexity. `https://drive.google.com/file/d/1l8XVfCMVlb1RUnGlJhv7V6VhzdJLO9g9/view`, unpublished.

—, Köster, M. and Peiseler, F. (2019b). Attention-driven demand for bonus contracts. *European Economic Review*, **115**, 1–24.

—, — and Sutter, M. (2020b). To buy or not to buy? Price salience in an online shopping field experiment. *European Economic Review*, **130**, Article 103593.

—, Lange, M., Köhler, K. and Wenzel, T. (2017). Demand shifts due to salience effects: Experimental evidence. *Journal of the European Economic Association*, **15** (3), 626–653.

— and Wenner, L. (2020). Testing fundamentals of salience theory. `http://sesp.bplaced.net/dertwinkel/wp-content/uploads/2020/04/DW_v13.pdf`, unpublished.

Diecidue, E. and Somasundaram, J. (2017). Regret theory: A new foundation. *Journal of Economic Theory*, **172**, 88–119.

Ebert, S. (2013). Moment characterization of higher-order risk preferences. *Theory and Decision*, **74** (2), 267–284.

— (2015). On skewed risks in economic models and experiments. *Journal of Economic Behavior & Organization*, **112**, 85–97.

— and Hilpert, C. (2019). Skewness preference and the popularity of technical analysis. *Journal of Banking and Finance*, **109**, Article 105675.

— and Karehnke, P. (2019). Skewness preferences in choice under risk. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3480519`, unpublished.

— and Strack, P. (2015). Until the bitter end: On prospect theory in a dynamic context. *American Economic Review*, **105** (4), 1618–1633.

— and — (2018). Never, ever getting started: On prospect theory without commitment. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765550`, unpublished.

— and van de Kuilen, G. (2016). Measuring multivariate risk preferences. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2637964`, Working Paper.

— and Wiesen, D. (2011). Testing for prudence and skewness seeking. *Management Science*, **57** (7), 1334–1349.

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

Ellis, A. and Masatlioglu, Y. (2020). Choice with endogenous categorization. `http://econweb.umd.edu/~masatlioglu/CTM.pdf`, unpublished.

Enke, B. and Graeber, T. (2020). Cognitive uncertainty. `https://benjamin-enke.com/pdf/Cognitive_uncertainty.pdf`, unpublished.

— and Zimmermann, F. (2019). Correlation neglect in belief formation. *Review of Economic Studies*, **86**, 313–332.

Eyster, E. and Weizsäcker, G. (2016). Correlation neglect in portfolio choice: Lab evidence. `https://www.wiwi.hu-berlin.de/en/Professorships/vwl/microeconomics/people/gweizsaecker/paper_gweizsaecker_ewmaintext.pdf`, unpublished.

Falk, A., Kosse, F. and Pinger, P. (2020). Re-revisiting the marshmallow test: A direct comparison of studies by Shoda, Mischel, and Peake (1990) and Watts, Duncan, and Quan (2018). *Psychological Science*, **31** (1), 100–104.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, **47** (2), 427–465.

Farrell, J. and Klemperer, P. (2007). Coordination and lock-in: Competition with switching costs and network effects. *Handbook of Industrial Organization*, **3**, 1967–2072.

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

Fishburn, P. C. (1970). *Utility theory for decision making*. New York: Wiley.

— (1989). Non-transitive measurable utility for decision under uncertainty. *Journal of Mathematical Economics*, **18** (2), 187–207.

— (1990a). Continuous nontransitive additive conjoint measurement. *Mathematical Social Sciences*, **20** (2), 165–193.

— (1990b). Skew symmetric additive utility with finite states. *Mathematical Social Sciences*, **19** (2), 103–115.

— (1991). Nontransitive additive conjoint measurement. *Journal of Mathematical Psychology*, **35** (1), 1–40.

Forrest, D., Simmons, R. and Chesters, N. (2002). Buying a dream: Alternative models of demand for lotto. *Economic Inquiry*, **40** (3), 485–496.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, **19**, 24–42.

Friedrichsen, J. (2018). Signals sell: Product lines when consumers differ both in taste for quality and image concern. `https://rationality-and-competition.de/wp-content/uploads/discussion_paper/70.pdf`, unpublished.

Frydman, C. and Mormann, M. M. (2018). The role of salience in choice under risk: An experimental investigation. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2778822`, unpublished.

— and Wang, B. (2020). The impact of salience on investor behavior: Evidence from a natural experiment. *Journal of Finance*, **75** (1), 229–276.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *Quarterly Journal of Economics*, **129** (4), 1661–1710.

— (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, vol. 2, pp. 261–343.

— and Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics*, **121** (2), 505–540.

—, —, Li, D., Li, H., Resnick, S. and de Vries, C. G. (2016). The impact of competition on prices with numerous firms. *Journal of Economic Theory*, **165**, 1–24.

Gagnon-Bartsch, T., Rabin, M. and Schwartzstein, J. (2020). Channeled attention and stable errors. `https://scholar.harvard.edu/files/gagnonbartsch/files/channeled_attention.pdf`, unpublished.

Garrett, T. A. and Sobel, R. S. (1999). Gamblers favor skewness, not risk: Further evidence from united states' lottery games. *Economics Letters*, **63** (1), 85–90.

Genest, W., Stauffer, W. R. and Schultz, W. (2016). Utility functions predict variance and skewness risk preferences in monkeys. *Proceedings of the National Academy of Sciences*, **113** (30), 8402–8407.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, **62**, 451–482.

Golec, J. and Tamarkin, M. (1998). Bettors love skewness, not risk, at the horse track. *Journal of Political Economy*, **106** (1), 205–225.

Gollier, C. (2020). Aversion to risk of regret and preference for positively skewed risks. *Economic Theory*, **70**, 913–941.

Gonzalez, R. and Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, **38** (1), 129–166.

Graham, J. R., Harvey, C. R. and Huang, H. (2009). Investor competence, trading frequency, and home bias. *Management Science*, **55** (7), 1094–1106.

Green, T. C. and Hwang, B.-H. (2012). Initial public offerings as lotteries: Skewness preference and first-day returns. *Management Science*, **58** (2), 432–444.

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

Grossman, P. J. and Eckel, C. C. (2015). Loving the long shot: Risk taking with skewed lotteries. *Journal of Risk and Uncertainty*, **51** (3), 195–217.

Grove, W. A., Jetter, M. and Papps, K. L. (2018). Career lotto: Labor supply in winner-take-all markets. *IZA DP No. 12012*.

Guiso, L., Sapienza, P. and Zingales, L. (2018). Time varying risk aversion. *Journal of Financial Economics*, **128** (3), 403–421.

Halevy, Y. (2007). Ellsberg revisited: An experimental study. *Econometrica*, **75** (2), 503–536.

— and Mayraz, G. (2020). Identifying rule-based rationality. `https://yoram-halevy.faculty.economics.utoronto.ca/wp-content/uploads/rules.pdf`, unpublished.

Handel, B. and Schwartzstein, J. (2018). Frictions or mental gaps: what's behind the information we (don't) use and when do we care? *Journal of Economic Perspectives*, **32** (1), 155–78.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, **103** (7), 2643–82.

Hartog, J. and Vijverberg, W. P. (2007). On compensation for risk aversion and skewness affection in wages. *Labour Economics*, **14** (6), 938–956.

Hartzmark, S., Hirshman, S. and Imas, A. (2020). Ownership, learning, and beliefs. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3465246`, unpublished.

Hastings, J. S., Madrian, B. C. and Skimmyhorn, W. L. (2013). Financial literacy, financial education, and economic outcomes. *Annual Review of Economics*, **5** (1), 347–373.

Haucap, J., Kollmann, D., Nöcker, T., Westerwelle, A. and Zimmer, D. (2013). *Neunzehntes Hauptgutachten der Monopolkommission 2010/2011. Stärkung des Wettbewerbs bei Handel und Dienstleistungen*. Nomos Verlagsgesellschaft, Baden-Baden.

Heidhues, P., Köster, M. and Kőszegi, B. (2020). Steering fallible consumers. `http://www.personal.ceu.hu/staff/Botond_Koszegi/steering.pdf`, unpublished.

— and Kőszegi, B. (2010). Exploiting naivete about self-control in the credit market. *American Economic Review*, **100** (5), 2279–2303.

— and Kőszegi, B. (2018). Behavioral industrial organization. *Handbook of Behavioral Economics: Applications and Foundations 1*, Chapter 6, 517–612.

—, Kőszegi, B. and Murooka, T. (2016). Inferior products and profitable deception. *Review of Economic Studies*, **84** (1), 323–356.

Herman, L. M. and Bahrick, H. P. (1966). Information encoding and decision time as variables in human choice behavior. *Journal of Experimental Psychology*, **71** (5), 718.

Hortaçsu, A., Madanizadeh, S. A. and Puller, S. L. (2017). Power to choose? An analysis of consumer inertia in the residential electricity market. *American Economic Journal: Economic Policy*, **9** (4), 192–226.

Huber, J., Payne, J. W. and Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, **9** (1), 90–98.

Ito, K. (2014). Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing. *American Economic Review*, **104** (2), 537–63.

Johnen, J. (2019). Automatic-renewal contracts with heterogeneous consumer inertia. *Journal of Economics and Management Strategy*, **28**, 765–786.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47** (2), 263–291.

Kalayci, K. and Serra-Garcia, M. (2016). Complexity and biases. *Experimental Economics*, **19** (1), 31–50.

Kallir, I. and Sonsino, D. (2009). The neglect of correlation in allocation decisions. *Southern Economic Journal*, **75** (4), 1045–1066.

Kőszegi, B. and Szeidl, A. (2013). A model of focusing in economic choice. *Quarterly Journal of Economics*, **128** (1), 53–104.

Klemperer, P. (1995). Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *Review of Economic Studies*, **62** (4), 515–539.

Kontek, K. (2016). A critical note on salience theory of choice under risk. *Economics Letters*, **149**, 168–171.

Krantz, D. H., Luce, R. D., Suppes, P. and Tversky, A. (1971). *Foundations of measurement: Additive and Polynomial Representations*, vol. 1. New York: Academic Press.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, **112**, 443–477.

LANZANI, G. (2020). Correlation made simple: Applications to salience and regret theory. http://economics.mit.edu/files/20177, unpublished.

LEVY, G. and RAZIN, R. (2015). Correlation neglect, voting behavior, and information aggregation. *American Economic Review*, **105** (4), 1634–45.

LI, X. and CAMERER, C. (2019). Using visual salience in empirical game theory. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3308886, unpublished.

LOEWENSTEIN, G. and PRELEC, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics*, **107** (2), 573–597.

LOOMES, G. and SUGDEN, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, **92** (368), 805–824.

— and — (1987). Some implications of a more general form of regret theory. *Journal of Economic Theory*, **41** (2), 270–287.

LUSARDI, A. and MITCHELL, O. (2011). *Financial Literacy and Planning: Implications for Retirement Wellbeing*. Oxford University Press.

— and MITCHELL, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, **52** (1), 5–44.

MAO, J. C. (1970). Survey of capital budgeting: Theory and practice. *Journal of Finance*, **25** (2), 349–360.

MARKOWITZ, H. (1952). Portfolio selection. *Journal of Finance*, **7** (1), 77–91.

MATĚJKA, F. and MCKAY, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, **105** (1), 272–98.

MENEZES, C. F., GEISS, C. and TRESSLER, J. (1980). Increasing downside risk. *American Economic Review*, **70** (5), 921–932.

MISCHEL, W. and EBBESEN, E. B. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology*, **16** (2), 329.

MITTON, T. and VORKINK, K. (2007). Equilibrium underdiversification and the preference for skewness. *Review of Financial Studies*, **20** (4), 1255–1288.

MORWITZ, V. G., GREENLEAF, E. A. and JOHNSON, E. J. (1998). Divide and prosper: consumers' reactions to partitioned prices. *Journal of Marketing Research*, **35** (4), 453–463.

Nunnari, S. and Zápal, J. (2020). A model of focusing in political choice. `http://www.salvatorenunnari.eu/nz_focusinpolitics.pdf`, unpublished.

O'Donoghue, T. and Sprenger, C. (2018). Reference-dependent preferences. *Handbook of Behavioral Economics: Applications and Foundations 1*, Chapter 1, 1–77.

Oprea, R. (2020). What makes a rule complex? *American Economic Review*, **110** (12), 3913–3951.

Prakash, A. J., Chang, C.-H. and Pactwa, T. E. (2003). Selecting a portfolio with skewness: Recent evidence from us, european, and latin american equity markets. *Journal of Banking & Finance*, **27** (7), 1375–1390.

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A. and Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (irt). *Journal of Behavioral Decision Making*, **29** (5), 453–469.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, **3** (4), 323–343.

Richard, S. F. (1975). Multivariate risk aversion, utility independence and separable utility functions. *Management Science*, **22** (1), 12–21.

Rieger, M. O. and Wang, M. (2006). Cumulative prospect theory and the St. Petersburg paradox. *Economic Theory*, **28** (3), 665–679.

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.

Samuelson, P. A. (1937). A note on measurement of utility. *Review of Economic Studies*, **4** (2), 155–161.

Schkade, D. A. and Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, **9** (5), 340–346.

Sepulveda, P., Usher, M., Davies, N., Benson, A., Ortoleva, P. and De Martino, B. (2020). Visual attention modulates the integration of goal-relevant evidence and not value. `https://www.biorxiv.org/content/10.1101/2020.04.14.031971v1`, unpublished.

Shafir, E., Simonson, I. and Tversky, A. (1993). Reason-based choice. *Cognition*, **49** (1-2), 11–36.

Simonson, I. and Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, **29** (3), 281–295.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, **50** (3), 665–690.

Soltani, A., De Martino, B. and Camerer, C. (2012). A range-normalization model of context-dependent choice: a new model and evidence. *PLoS Computational Biology*, **8** (7), e1002607.

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

Strack, P. and Viefers, P. (forthcoming). Too proud to stop: Regret in dynamic decisions. *Journal of the European Economic Association*.

Strait, C. E. and Hayden, B. Y. (2013). Preference patterns for skewed gambles in rhesus monkeys. *Biology Letters*, **9** (6), Article 20130902.

Sydnor, J. (2010). (Over)insuring modest risks. *American Economic Journal: Applied Economics*, **2** (4), 177–199.

Taubinsky, D. and Rees-Jones, A. (2018). Attention variation and welfare: theory and evidence from a tax salience experiment. *Review of Economic Studies*, **85** (4), 2462–2496.

Thakral, N. and Tô, L. T. (forthcoming). Daily labor supply and adaptive reference points. *American Economic Review*.

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

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124–1131.

— and — (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5** (4), 297–323.

— and Simonson, I. (1993). Context-dependent preferences. *Management Science*, **39** (10), 1179–1189.

Ungeheuer, M. and Weber, M. (forthcoming). The perception of dependence, investment decisions, and stock prices. *Journal of Finance*.

Valletti, T. and Veiga, A. (2020). Attention, recall and purchase: Experimental evidence on online news and advertising. https://www.cesifo.org/sites/default/files/events/2020/ed20_Valletti.pdf, unpublished.

van Rooij, M., Lusardi, A. and Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, **101** (2), 449–472.

Westfall, J., Van Boven, L., Chambers, J. R. and Judd, C. M. (2015). Perceiving political polarization in the united states: Party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspectives on Psychological Science*, **10** (2), 145–158.

Wilcox, N. T. (1993). Lottery choice: Incentives, complexity and decision time. *Economic Journal*, **103** (421), 1397–1417.

Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making*, **12**, 93–106.

Appendix

Heinrich Heine
Universität Düsseldorf

Eidesstattliche Versicherung

Ich, Herr M.Sc. Mats Köster, versichere an Eides statt, dass die vorliegende Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf" erstellt worden ist.

Düsseldorf, der 17. März 2021

M. Köster

Unterschrift