

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jeboOn correlated lotteries in economic applications [☆]Markus Dertwinkel-Kalt ^{a,*}, Sebastian Ebert ^b, Mats Köster ^c^a University of Münster & Max Planck Institute for Research on Collective Goods, Germany^b Heidelberg University, Germany^c Central European University, Austria

ARTICLE INFO

JEL classification:

D01
D81
C90

Keywords:

Binary lotteries
Correlation
Correlation neglect
Attention
Portfolio choice

ABSTRACT

Economic models and experiments frequently use lotteries with only a few outcomes. We study the correlation of such lotteries and discuss its relevance for economic applications. In particular, we fully characterize the joint distribution of two binary lotteries via their first three univariate moments and their correlation coefficient. As we illustrate alongside several examples, the resulting parametrization may be useful for economic modeling and experimental design.

1. Introduction

Numerous important economic problems involve multiple risks, and the corresponding outcomes depend on the interdependence of these risks. A portfolio, for instance, combines multiple assets, with the resulting return distribution depending on the comovement of asset returns. As another example, how informative a series of signals (e.g., a student's grades over time) is about a variable of interest (e.g., the student's skills) depends on the signals' correlation. This implies that a researcher, studying beliefs and economic behavior, not only needs to specify the marginal distributions of the involved risks, but has to model their interdependence as well.

At the same time, researchers want to keep their models and analyses tractable. Economic behavior can often be illustrated at the hand of lotteries with only a few outcomes, which capture many essentials of risk. Due to their simplicity, their use is also standard in experiments on choice under risk (Allais, 1953; Kahneman and Tversky, 1979) or information economics (Charness and Levin, 2005), for example. Oftentimes, even binary lotteries — lotteries with only two outcomes — are sufficient for modeling the conceptual effects of risk on economic behavior.

In this paper, we study the correlation of lotteries and discuss its relevance for economic applications. In particular, we illustrate the usefulness of correlated binary lotteries for economic modeling and experimental design. We first show that correlation, or more precisely, Pearson's correlation coefficient,

[☆] We thank the editor, an associated editor, and two anonymous referees for useful comments and suggestions that helped us in clarifying the contribution of the paper. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 803380) and the German Research Foundation (DFG grant No. 490725816).

* Corresponding author.

E-mail addresses: markus.dertwinkel-kalt@wiwi.uni-muenster.de (M. Dertwinkel-Kalt), sebastian.ebert@awi.uni-heidelberg.de (S. Ebert), kosterm@ceu.edu (M. Köster).

<https://doi.org/10.1016/j.jebo.2023.08.025>

Received 17 October 2022; Received in revised form 2 May 2023; Accepted 30 August 2023

Available online 27 September 2023

0167-2681/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

$$\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}, \tag{1}$$

describes the dependence of two lotteries X and Y exhaustively if and only if both lotteries are binary. We also prove equivalence to other dependence notions for the case of binary lotteries, and we show that these equivalences break down as soon as either X or Y (or both) has more than two outcomes. Although such results may hardly surprise and possibly be found in other contexts elsewhere, an accessible and self-contained treatment with an eye on economic applications seems worthwhile. The observation that correlation may fail at measuring dependence properly even if just one lottery is not binary motivates our subsequent focus on studying correlation effects on economic beliefs and behavior using binary lottery pairs. As a core result, we present a novel parametrization of the joint distribution of two binary lotteries that is given in terms of the lotteries’ first three univariate moments and their correlation coefficient. The direct relationship to statistical moments allows for convenient comparative statics results in terms of these moments as well as in terms of stochastic dominance, and it eases economic interpretations. Related results were stated for the univariate case and the case of Bernoulli variables, which have outcomes 0 and 1. Explicating the result for the economically relevant case of binary lottery pairs with arbitrary outcomes proves useful in a number of economic applications.

The simplicity and tractability of our parametrization of correlated binary lotteries makes it amenable for economic modeling and experiment design. Experimental evidence tells us, for instance, that a pre-specified correlation structure matters for choice under risk (e.g., Andreoni and Sprenger, 2012; Cheung, 2015; Bordalo et al., 2012), and individuals also reveal strong preferences over correlation structures when being allowed to implement one freely (e.g., Andersen et al., 2018). We show that our parametrization allows us to sort through these existing experimental paradigms in a unified and structured way. This exercise reveals distributional commonalities and differences of the risks used in these experiments, which is useful for understanding the precise aspects of preferences that are identified by the experimental tasks. Our parametrization may also prove useful in economic modeling. Just like in the univariate case (e.g., in models of precaution and prevention like Denuit et al. 2016, Heinzel and Peter 2021, or Courbage et al. 2022), it may allow for analytical results and simple interpretations that are difficult to obtain otherwise. We demonstrate this at the hand of four examples prominently featured in the literature: testing for “correlation neglect” in belief formation (e.g., Enke and Zimmermann, 2019), understanding the testable implications of “rational inattention” (e.g., Matějka and McKay, 2015; Dean and Neligh, 2019), studying “Bayesian persuasion” à la Kamenica and Gentzkow (2011), and identifying misperceptions of correlation in portfolio selection (e.g., Ungeheuer and Weber, 2021).

2. Correlation and dependence of two (binary) lotteries

We study the joint distribution of two lotteries (discrete random variables with finite support), X and Y . Let $m, n \in \mathbb{N}$ be the number of distinct outcomes of X and Y , respectively. Lottery X pays with probability $p_i \in (0, 1)$ an amount $x_i \in \mathbb{R}$ such that $\sum_{i=1}^m p_i = 1$, with the outcomes of X being ordered as $x_1 > x_2 > \dots > x_m$. Analogously, lottery Y pays with probability $q_j \in (0, 1)$ an amount $y_j \in \mathbb{R}$ such that $\sum_{j=1}^n q_j = 1$ and $y_1 > y_2 > \dots > y_n$. Our object of interest is the joint distribution of these two lotteries: $\omega_{ij} := \mathbb{P}[\{X = x_i\} \cap \{Y = y_j\}]$.

2.1. A parametrization of the joint distribution of two binary lotteries

In this subsection, let X and Y be binary lotteries. To save on notation, let $p := \mathbb{P}[\{X = x_1\}]$ and $q := \mathbb{P}[\{Y = y_1\}]$. Denote as $\mathbb{E}[X] =: E_X$, $\text{Var}(X) =: V_X$, and $\text{Skew}(X) =: S_X$ the mean, variance, and (central- and standardized) skewness of Lottery X . The first three moments of Y are defined analogously. As before, denote as $\rho(X, Y) =: \rho$ the lotteries’ correlation coefficient.

Our main result in this subsection states that we can fully describe the joint distribution of the binary lotteries X and Y via (a) their first three univariate moments and (b) their correlation coefficient. The result — essentially a combination of existing results discussed below — comes with a self-contained and constructive proof (see Appendix A).

Proposition 1 (Distribution of two binary lotteries). Consider the binary lotteries X and Y .

(a) The marginal distributions of X and Y can be expressed in terms of their first three moments:

$$x_1 = E_X + \sqrt{\frac{(1-p)}{p}} V_X \quad \text{and} \quad x_2 = E_X - \sqrt{\frac{p}{(1-p)}} V_X \quad \text{and} \quad p = \frac{1}{2} - \frac{S_X}{2\sqrt{4 + S_X^2}},$$

and

$$y_1 = E_Y + \sqrt{\frac{(1-q)}{q}} V_Y \quad \text{and} \quad y_2 = E_Y - \sqrt{\frac{q}{(1-q)}} V_Y \quad \text{and} \quad q = \frac{1}{2} - \frac{S_Y}{2\sqrt{4 + S_Y^2}}.$$

(b) Let $\kappa := \sqrt{p(1-p)q(1-q)}$. The joint distribution of X and Y can be parameterized as:

	$Y = y_1$	$Y = y_2$
$X = x_1$	$pq + \kappa\rho$	$p(1-q) - \kappa\rho$
$X = x_2$	$(1-p)q - \kappa\rho$	$(1-p)(1-q) + \kappa\rho$

The lotteries' correlation coefficient, $\rho = \rho(X, Y)$, lies within the so-called Fréchet bounds:

$$\rho \in \left[-\min \left\{ \frac{\kappa}{pq}, \frac{\kappa}{(1-p)(1-q)} \right\}, \min \left\{ \frac{\kappa}{(1-p)q}, \frac{\kappa}{p(1-q)} \right\} \right].$$

Part (a) of the proposition says that the marginal distribution of a binary lottery can be parameterized by its first three moments (see also Chiu, 2010), and it makes this parametrization explicit (as in Ebert, 2015, Proposition 1). As noticed by Chiu (2010), changing the k th moment of a binary lottery, while holding the other two fixed, shifts its marginal distribution in terms of k th-order stochastic dominance. More precisely, ceteris paribus (c.p.), an increase in the lottery's expected value improves its marginal distribution in terms of first-order stochastic dominance. Increasing the lottery's variance (c.p.), on the other hand, results in a mean-preserving spread (in the sense of Rothschild and Stiglitz, 1970) and, thereby, deteriorates the marginal distribution in terms of second-order stochastic dominance. Finally, increasing the lottery's skewness, while holding its expected value and variance fixed, decreases the lottery's downside risk (in the sense of Menezes et al., 1980) and improves the distribution in terms of third-order stochastic dominance. For general probability distributions such ceteris paribus arguments cannot be made (Brockett and Kahane, 1992), while binary lotteries allow for this.¹

Part (b) of the proposition parameterizes the probability mass function of the joint distribution of X and Y in terms of their correlation $\rho(X, Y)$. Mathematically equivalent parametrizations have been proposed by Loomes and Sugden (1987) for the case of $x_1 > y_1 > 0$, $x_2 = y_2 = 0$, and $p < q$ as well as by Joe (1997, p. 225) for the case of $x_1 = y_1 = 1$ and $x_2 = y_2 = 0$.²

A crucial property of the parametrization provided in Proposition 1 (b) is that the correlation coefficient is bounded, with the so-called Fréchet bounds depending on the lotteries' skewness. The relationship to the lotteries' skewness becomes clear through Part (a) of the proposition, which implies that a lottery's skewness is one-to-one to the marginal probability of its larger outcome. As an illustration, if both lotteries have identical skewness (i.e. $S_X = S_Y \equiv S_{id}$),

$$\rho(X, Y) \in \left[-\min \left\{ \frac{\sqrt{4 + S_{id}^2} + S_{id}}{\sqrt{4 + S_{id}^2} - S_{id}}, \frac{\sqrt{4 - S_{id}^2} - S_{id}}{\sqrt{4 + S_{id}^2} + S_{id}} \right\}, 1 \right] \subseteq [-1, 1]. \tag{2}$$

Whenever $S_{id} \neq 0$, the lower Fréchet bound restricts the admissible correlation values, and the bound gets tighter the more skewed the lotteries are in absolute terms. This insight generalizes to arbitrary pairs of binary lotteries: as the lotteries become more skewed in absolute terms, the set of feasible correlations shrinks to a strict(er) subset of $[-1, 1]$. As pointed out in Joe (1997, p. 225), this means that it is difficult to compare two pairs of binary lotteries via their correlation coefficient whenever the marginal distributions differ across these two pairs. This is less of an issue in many economic applications, however, where the researcher is interested in comparing pairs of binary lotteries with identical marginals, in the way described below.

In many applications that study the impact of correlation on economic behavior, the researcher starts out from a single pair of binary lotteries X and Y . The researcher then wishes to study how a marginal c.p. change in the lotteries' correlation — a change that keeps their marginal distributions unchanged — changes economic outcomes. (For example, in Section 3 we consider correlated assets X and Y , and study how (a change in) their correlation affects an individual's choice among portfolios consisting of these assets.) Proposition 1 describes how the four state probabilities ω_{ij} must be changed to identify c.p. effects of changes in correlation. To be clear, keeping the marginals unchanged, while changing the correlation, means that p , x_1 , x_2 , q , y_1 , and y_2 remain unchanged, which — by Part (a) of Proposition 1 — is equivalent to all univariate moments remaining unchanged. As Proposition 1 further highlights, the Fréchet bounds are such that for any pair of binary lotteries X and Y , there exists some $\epsilon > 0$, so that any correlation $\rho \in (-\epsilon, \epsilon)$ is feasible. Hence, regardless of how skewed the marginal distributions are, it is always possible to impose negative, zero, or positive correlation values. And, starting out from any such correlation value, it is also always possible to investigate the effect of a marginal decrease or marginal increase in correlation on economic outcomes.

To summarize, for a pair of binary lotteries X and Y , Proposition 1 describes how to make c.p. changes of any of the six univariate moments and the correlation coefficient. Going further, it shows that (away from the Fréchet bounds) the researcher can vary any univariate moment — including the skewness of the lotteries — independently of the imposed correlation. Driessen et al. (2020), for example, conduct such (c.p.) analyses, studying the price impact of volatility, skewness, and correlation within a tractable asset pricing model with probability weighting.

2.2. Special cases and examples

We illustrate our parametrization alongside existing experimental paradigms. By recovering these paradigms from our parametrization in a unified and structured way, we reveal distributional commonalities and differences of the risks used in the experiments. These are useful for understanding the precise aspects of preferences that are identified by the experimental tasks.

¹ Moreover, binary lotteries have the advantage that their skewness is unambiguously defined; for more complex lotteries, different definitions exist and may yield conflicting classifications. For example, consider the lottery that pays with 39% an amount of 0, with 60% an amount of 50, and with 1% an amount of 100. When defining skewness in the usual way, that is, as the standardized third moment of the probability distribution, this lottery is left skewed, even though it arguably looks right-skewed.

² Loomes and Sugden (1987) do not explicitly state the joint distribution in terms of the correlation coefficient. Joe (1997) uses the correlation coefficient of the lotteries, but states their joint distribution in a slightly different way: $\omega_{ii}(\rho) = \omega_{ii}(0)(1 + \frac{\kappa}{\omega_{ii}(0)}\rho)$ and $\omega_{ij}(\rho) = \omega_{ij}(0)(1 - \frac{\kappa}{\omega_{ij}(0)}\rho)$ for $i, j \in \{1, 2\}$ and $i \neq j$.

$X \backslash Y$	$100 - z$	0
z	0	$\frac{1}{2}$
0	$\frac{1}{2}$	0

$X \backslash Y$	$100 - z$	0
z	$\frac{1}{4}$	$\frac{1}{4}$
0	$\frac{1}{4}$	$\frac{1}{4}$

$X \backslash Y$	$100 - z$	0
z	$\frac{1}{2}$	0
0	0	$\frac{1}{2}$

Fig. 1. On the left, the lotteries have correlation $\rho = -1$ (as used by Miao and Zhong, 2015); in the middle, they have correlation $\rho = 0$ (as used by Andreoni and Sprenger, 2012, Cheung, 2015, and Miao and Zhong, 2015); on the right, they have correlation $\rho = 1$ (as used by Cheung, 2015, and Miao and Zhong, 2015). Subjects must choose in each case an allocation $z \in [0, 100]$.

We begin with the experiments of Andreoni and Sprenger (2012), Cheung (2015), and Miao and Zhong (2015), who study the separability of risk and time preferences. Subjects obtain a budget of 100 experimental currency units. They have to decide on an amount z to be allocated to a first period, with the remaining amount $100 - z$ being allocated to a second period. (For the ease of presentation, we ignore the possibility of a non-zero interest rate as in the original experiments.) For payment, subjects receive their first-period allocation z and their second-period allocation $100 - z$, each with probability $1/2$. What varies across the three experiments, however, is the correlation between the two payments, and it has a significant impact on choices.

Our parametrization makes this intertemporal correlation structure transparent. Let X be a binary lottery that pays with equal probability either z or zero, capturing payments in the first period of the experiments described above. Similarly, we model second-period payments via the binary lottery Y that pays with equal probability either $100 - z$ or zero. Inserting $p = 1/2 = q$ in Proposition 1, the feasible joint distributions of payments over time can be described as follows:

$$\omega_{11} = \frac{1}{4} + \frac{1}{4}\rho = \omega_{22} \quad \text{and} \quad \omega_{12} = \frac{1}{4} - \frac{1}{4}\rho = \omega_{21} \quad \text{for } \rho \in [-1, 1]. \tag{3}$$

In Andreoni and Sprenger (2012), the payments in the first and second period are independent of each other (i.e. $\rho = 0$), which yields the joint distribution shown in the middle of Fig. 1. When instead plugging $\rho = -1$ and $\rho = 1$ in Eq. (3), respectively, we obtain the joint distributions shown on the left and on the right in Fig. 1.³ Cheung (2015) and Miao and Zhong (2015) argue that behavior will be different across these three problems, and they confirm this hypothesis experimentally. As Proposition 1 makes clear, the three experimental designs exclusively differ in how payments are correlated over time, while keeping all univariate moments constant, and thus perfectly identify the impact of intertemporal correlation on the choice of z .

The lottery pairs in Fig. 1 are closely related to Richard’s (1975) seminal work on multivariate risk (or correlation) preferences. Richard theoretically studies choices of the following type: fix the value of z (to, say, $z = 50$) and offer the choice between the consumption streams on the left and on the right of Fig. 1. Rather than identifying choice behavior for a given correlation structure, these choices identify preferences over correlation structures between sooner and later payments. Andersen et al. (2018), Ebert and van de Kuilen (2016), and Rohde and Yu (2023) elicit exactly such choices in controlled lab experiments. Moreover, when X and Y describe payments to different people, such choices shed light on individuals’ social preferences. Rohde and Rohde (2015), for example, elicit choices between independent and positively correlated payments; like, between the middle and the right payment structures in Fig. 1, again with z being fixed. All of these experiments report significant evidence for preferences over correlation structures.

Proposition 1 highlights that, because both X and Y above are symmetric ($S_X = 0 = S_Y$), the Fréchet bounds are -1 and $+1$, respectively. Hence, these experimental paradigms (and only these paradigms) allow for implementing the full range of correlation coefficients, going from perfectly negative to perfectly positive correlation.

This is different when studying preferences toward skewed risks (as in the experiments of Ebert and Wiesen, 2011 or Dertwinkel-Kalt and Köster, 2020), because then the Fréchet bounds are binding. We illustrate this property of binary lotteries by recovering the experimental designs used in Ebert and Wiesen (2011) and Dertwinkel-Kalt and Köster (2020). In doing so, we further illustrate the usefulness of our parametrization in the study of context-dependent theories of choice under risk (e.g., Loomes and Sugden, 1982; Bordalo et al., 2012).

Let $X \equiv R$ be a right-skewed binary lottery and $Y \equiv L$ be a left-skewed binary lottery with equal mean and variance but skewness opposite in sign. Denote the skewness of R by S so that the skewness of L is $-S$. By Proposition 1, the correlation coefficient satisfies:

$$\rho(R, L) \in \left[-1, \frac{\sqrt{4 + S^2} - S}{\sqrt{4 + S^2} + S} \right] \subset [-1, 1]. \tag{4}$$

For $S = 2.7$, for example, $\rho \in [-1, 1/9]$. Fig. 2 illustrates three feasible joint distributions for this case, going from perfectly negative correlation (shown on the left) over independence (shown in the middle) to the maximal positive correlation (shown on the right).

Ebert and Wiesen (2011) study the choice between such pairs of left- and right-skewed lotteries under independence, as is the case for R and L in the middle of Fig. 2. Dertwinkel-Kalt and Köster (2020) argue, and confirm experimentally, that — even though only the outcome of one marginal distribution, R or L , is being paid — a different correlation leads to different choice behavior. They compare the choice between R and L under perfectly negative correlation (as on the left of Fig. 2) to the same choice under the maximal positive correlation (as on the right of Fig. 2). They find that subjects choose R more often under negative correlation, which is consistent with context-dependent theories of choice under risk such as regret theory (Loomes and Sugden, 1982) and salience

³ The three joint distributions shown in Fig. 1 coincide with those shown in Fig. 1 of Miao and Zhong (2015).

R\L	120	90
216	0	$\frac{1}{10}$
96	$\frac{9}{10}$	0

R\L	120	90
216	$\frac{9}{100}$	$\frac{1}{100}$
96	$\frac{81}{100}$	$\frac{9}{100}$

R\L	120	90
216	$\frac{1}{10}$	0
96	$\frac{8}{10}$	$\frac{1}{10}$

Fig. 2. Lotteries such as those used in the experiments by Ebert and Wiesen (2011) and Dertwinkel-Kalt and Köster (2020). Here, we assume $E_L = E_R = 108$ and $V_L = V_R = 1296$ as well as $S = 2.7$.

Table 1
Joint distribution of a binary lottery X and a non-binary lottery Y .

	$Y = y_1$	$Y = y_2$	$Y = y_3$
$X = x_1$	$\alpha(1 - \beta)p$	$(1 - \alpha)p$	$\alpha\beta p$
$X = x_2$	$p - \alpha(1 - \beta)p$	αp	$1 - 2p - \alpha\beta p$

theory (Bordalo et al., 2012), but inconsistent with expected utility theory or cumulative prospect theory (Tversky and Kahneman, 1992).

2.3. Moving beyond binary lotteries: an impossibility result

We now relax the assumption of binary lotteries, and — to begin with — consider the following example: Lottery X is a binary lottery that pays x_1 with probability $p \in (0, 1/3)$ and x_2 with probability $1 - p$. Lottery Y is non-binary, however, as it pays y_1 and y_2 with probability p and y_3 with probability $1 - 2p$. It is easily verified that the joint distribution of X and Y satisfies

$$\omega_{11} = \omega_{22} + \omega_{23} - (1 - 2p), \omega_{12} = p - \omega_{22}, \omega_{13} = 1 - 2p - \omega_{23}, \text{ and } \omega_{21} = 1 - p - (\omega_{22} + \omega_{23}),$$

with $\omega_{22} \in [0, p]$ and $\omega_{23} \in [0, 1 - 2p]$, so that $\omega_{22} + \omega_{23} \geq 1 - 2p$. We can describe the distribution of X and Y via two parameters $\alpha, \beta \in [0, 1]$ by setting $\omega_{22} = \alpha p$ and $\omega_{23} = 1 - 2p - \alpha\beta p$ (Table 1).

The main take-away from this example is that with non-binary lotteries a single parameter — like the correlation of X and Y — may not be enough to fully describe their joint distribution. The following proposition generalizes this observation: the joint distribution of X and Y can be fully parameterized via their correlation coefficient if and only if both lotteries are binary.

Proposition 2. *The following two statements are equivalent:*

- (a) X and Y are both binary lotteries.
- (b) The correlation of X and Y pins down their joint distribution.

2.4. Correlation and invariance to increasing transformations

As argued in Epstein and Tanny (1980, p. 23), “A proper notion of [...] interdependence should be ‘ordinal,’ or invariant to increasing transformations.”⁴ Joe (1997, p. 38) includes this invariance notion in the list of desirable properties for a dependence measure. In general, however, the correlation coefficient does not satisfy this property (e.g., Meyer and Strulovici, 2012, p. 1465). The following result shows that the correlation coefficient is invariant to increasing transformations (Statement (a)) if and only if both marginal distributions are binary (Statement (b)).

Proposition 3. *Consider two pairs of lotteries, (X, Y) and (X', Y') , with $X' \stackrel{d}{=} X$ and $Y' \stackrel{d}{=} Y$. The following two statements are equivalent:*

- (a) For all nondecreasing functions $r_1, r_2 : \mathbb{R} \rightarrow \mathbb{R}$, the following implication holds:

$$\rho(X, Y) \geq \rho(X', Y') \implies \rho(r_1(X), r_2(Y)) \geq \rho(r_1(X'), r_2(Y')).$$

- (b) X and Y are both binary lotteries.

Another example for which Statement (a) holds is when X and Y are both normally distributed (Levy and Paroush, 1974, p. 140). Proposition 3 says that, for pairs of discrete random variables with identical marginals, invariance holds if and only if both marginals are binary.

Next we show that, for pairs of binary lotteries with identical marginals, the correlation coefficient also satisfies the other desirable properties of a dependence measure stated in Joe (1997, p. 38). To prove this, we relate the correlation coefficient to

⁴ Intuitively, invariance to increasing transformations implies that a dependence measure is unaffected by a monotonic relabeling of possible outcomes; see, for instance, Meyer and Strulovici (2012, p. 1465) for an explicit illustration of this property and its desirability.

another, stronger notion of dependence — namely, *concordance* as introduced in Tchen (1980) and Epstein and Tanny (1980) — that is known to satisfy the aforementioned properties (e.g., Meyer and Strulovici, 2012).

Definition 1 (Concordance). Consider two pairs of lotteries, (X, Y) and (X', Y') , with $X' \stackrel{d}{=} X$ and $Y' \stackrel{d}{=} Y$. Suppose that the joint distribution of X' and Y' is obtained from that of X and Y as follows: for $x_i > x_k$ and $y_j > y_l$, decrease the probability of the states (x_i, y_j) and (x_k, y_l) by some $\epsilon > 0$ and simultaneously increase the probability of the states (x_i, y_l) and (x_k, y_j) by the same amount. We then say that the lotteries X and Y are more concordant than the lotteries X' and Y' .

It is easily verified (and well known) that, if X and Y are more concordant than X' and Y' , then $\rho(X, Y) - \rho(X', Y') \propto \epsilon(x_i - x_k)(y_j - y_l) > 0$ for some $x_i > x_k$ and $y_j > y_l$ and $\epsilon > 0$. While it is also well known that the converse is generally not true (e.g., Epstein and Tanny, 1980, Theorem 7), the following proposition establishes this converse for pairs of binary lotteries.

Proposition 4. Consider pairs of binary lotteries, (X, Y) and (X', Y') , with $X' \stackrel{d}{=} X$ and $Y' \stackrel{d}{=} Y$. The following are equivalent: (a) X and Y are more concordant than X' and Y' . (b) $\rho(X, Y) > \rho(X', Y')$.

As alluded to before, combining this result with Theorem 1 in Meyer and Strulovici (2012) implies that ranking the dependence of two pairs of binary lotteries, (X, Y) and (X', Y') , with identical marginals via correlation is not only equivalent to ranking it via concordance but also via other meaningful dependence concepts (as discussed in Meyer and Strulovici, 2012).

2.5. Summary

The joint distribution of two binary lotteries is conveniently parameterized using their first three univariate moments and their correlation coefficient (Proposition 1). This parametrization allows us to recover and sort through several existing experimental paradigms in a unified and structured way, and — as we demonstrate in the next section — is useful for economic modeling and developing new experimental designs. An important aspect to keep in mind is that the range of admissible correlation values is a strict subset of $[-1, 1]$ whenever one of the lotteries is skewed. Still, larger correlation meaningfully indicates greater dependence for pairs of binary lotteries with identical marginals (Propositions 3 and 4). For lotteries with more than two outcomes, correlation does not determine the joint distribution of the lotteries (Proposition 2) and fails to meet some desirable properties of a dependence measure (Propositions 3 and 4).

3. Economic applications

Our parametrization of correlated binary lotteries may prove useful in economic modeling and developing new experimental designs. We demonstrate this at the hand of four examples prominently featured in the literature: testing for, and quantifying, “correlation neglect” in belief formation (e.g., Enke and Zimmermann, 2019), understanding the testable implications of “rational inattention” (e.g., Matějka and McKay, 2015; Caplin and Dean, 2015; Dean and Neligh, 2019), studying “Bayesian persuasion” à la Kamenica and Gentzkow (2011), and identifying misperceptions of correlation in portfolio selection (e.g., Ungeheuer and Weber, 2021).

Learning from correlated signals Binary lotteries, in general, and our parametrization of their joint distribution, in particular, allow us to quantify *correlation neglect* in belief updating. Existing experimental studies, relying on more complicated distributions (e.g., Enke and Zimmermann, 2019), document that subjects do not properly account for the interdependence of the signals they see. With our parametrization at hand, one could extend their experimental design to elicit which correlation — if not the actual one — is the one perceived by subjects.

Just like Enke and Zimmermann (2019), consider a subject who wants to estimate the mean of some random variable X . Suppose that X follows a Bernoulli distribution with a success probability $p \in (0, 1)$, so that the task boils down to estimating exactly this success probability. The subject’s prior is that p is drawn from a uniform distribution over $(0, 1)$. She observes two signals $X_1 \stackrel{d}{=} X$ and $X_2 \stackrel{d}{=} X$ with true correlation $\rho = \rho(X_1, X_2) \in [0, 1]$ that is communicated to the subject.⁵ The experimenter is interested in the (mis)perceived correlation $\hat{\rho} \in [0, 1]$ that is consistent with the subject’s estimate \hat{p} of the posterior mean $\mathbb{E}[X | \{X_1 = x_1\} \cap \{X_2 = x_2\}]$.⁶

Using Proposition 1, we obtain a simple closed-form expression for the Bayesian posterior mean. For the sake of concreteness, let us focus on the case of observing two successes in a row (i.e. $X_1 = 1$ and $X_2 = 1$), in which case the Bayesian posterior mean is given by:⁷

⁵ An experimenter could, for instance, easily explain to subjects — in plain English — that the two signals are independent or perfectly correlated.

⁶ Because X_1 and X_2 have the same distribution and, thus, also the same skewness, for any $p \in (0, 1)$, the upper Fréchet bound is 1 (and the lower Fréchet bound is strictly negative). Hence, holding a prior about p with support $(0, 1)$ is consistent with an estimate $\hat{\rho} \in [0, 1]$ of the signals’ correlation.

⁷ Denote as $f(\cdot)$ the prior and as $f(\cdot | \{X_1 = 1\} \cap \{X_2 = 1\})$ the posterior density of $p \in (0, 1)$. By Bayes’ rule,

$$f(p | \{X_1 = 1\} \cap \{X_2 = 1\}) = \frac{\mathbb{P}[\{X_1 = 1\} \cap \{X_2 = 1\} | p] f(p)}{\mathbb{P}[\{X_1 = 1\} \cap \{X_2 = 1\}]} = \frac{p^2 + \rho(1-p)\rho}{\int_0^1 q^2 + q(1-q)\rho \, dq} = \frac{p^2 + \rho(1-p)\rho}{1/3 + \rho/6},$$

$$\mathbb{E}[X|\{X_1 = 1\} \cap \{X_2 = 1\}] = \frac{3}{4} \left(\frac{1 + \rho/3}{1 + \rho/2} \right). \tag{5}$$

The less correlated the two signals are the more informative the second signal is about the success probability. A misspecified subject, however, might over- or underestimate the informativeness of this signal. Assuming that the subject is a Bayesian, given her reported estimate \hat{p} , we can back out the subject’s perceived correlation $\hat{\rho}$ that is consistent with this estimate, as:

$$\hat{\rho}(\hat{p}) = \frac{4\hat{p} - 3}{1 - 2\hat{p}}. \tag{6}$$

As an example, suppose that in a task in which the signals are perfectly positively correlated the subject reports $\hat{p} = 3/4$. From Eq. (5), the correct Bayesian estimate would be equal to $\mathbb{E}[X|\{X_1 = 1\} \cap \{X_2 = 1\}] = 2/3$. From Eq. (6), it follows that the subject’s estimate is consistent with a perceived correlation of $\hat{\rho} = 0$; that is, the subject’s estimate is consistent with complete correlation neglect. In general, Eq. (6) allows for a precise quantification of correlation neglect with the only necessary input being the response to a single and easily stated question.

Eq. (6) identifies a subject’s perceived correlation under the assumption that she updates according to Bayes’ rule. As a matter of fact, however, people often update their beliefs in ways that are inconsistent with Bayes’ rule (see Benjamin, 2019, for a survey). Our parametrization allows us to infer (mis)perceptions of correlation also in this case. Just like in Eq. (5), once we have specified an updating rule, we can (possibly numerically) calculate the implied posterior expectation of X upon observing the signals X_1 and X_2 as a function of their correlation ρ . And by a comparison similar to that in Eq. (6), we can then back out the perceived correlation $\hat{\rho}$ that is consistent with a subject’s guess \hat{p} under the respective updating rule.

As a specific example, we can easily do so for Grether’s (1980) model of the representativeness heuristic. For general specifications of this model, we can back out a subject’s perceived correlation numerically. And for empirically relevant parameters, we even obtain a closed-form representation of the posterior expectation of X as a function of the signal correlation ρ .⁸ This, in turn, allows us to easily determine a subject’s perceived correlation $\hat{\rho}$. Going further, our parametrization makes transparent how certain forms of non-Bayesian updating can be confused with Bayesian updating under misperceptions of the correlation among signals.⁹

Rational inattention Our parametrization is not only useful for studying how people update their beliefs in the light of new information or, more formally, upon receiving a *given* signal. It can also help us to analyze, and test, how people (optimally) acquire information to start with. Information acquisition is often formalized as “choosing” one of several available signals about a variable of interest. In doing so, individuals face the trade-off that more informative signals are also more costly. We can think of this process as the individual *literally* acquiring and observing a signal (as in consulting a doctor for a diagnosis) that is the more costly the more precise it is (e.g., when having to wait for an appointment with a more skilled doctor). In other situations, however, information acquisition happens more subconsciously, and may not be observable to an outsider (e.g., an experimenter). As an illustration, we re-visit an experimental test of models on *rational inattention* by Dean and Neligh (2019), and we show that our parametrization allows us to derive the optimal information acquisition and predicted behavior in closed form.

As depicted in Fig. 3, Dean and Neligh (2019) show their subjects a picture of 100 balls — some of which are blue and some of which are red —, and ask them to guess whether there are more blue or more red balls presented on the screen. If the subject’s guess is correct, she is paid \$ x . If her guess is wrong, she gets nothing. Before observing the picture of 100 balls, subjects are told that there are two possible states of the world: in state B (lue) there are 51 blue and 49 red balls on the screen, while in state R (ed) there are 49 blue and 51 red balls on the screen. Subjects are further told that state B occurs with probability $q \geq \frac{1}{2}$ while state R occurs with

where the second equality follows from Proposition 1 and $f(\rho) = 1$. Thus, because $x_1 = 1$ and $x_0 = 0$, we have

$$\mathbb{E}[X|\{X_1 = 1\} \cap \{X_2 = 1\}] = \mathbb{E}[p|\{X_1 = 1\} \cap \{X_2 = 1\}] = \int_0^1 p \left(\frac{p^2 + p(1-p)\rho}{1/3 + \rho/6} \right) dp = \frac{1/4 + \rho/12}{1/3 + \rho/6}.$$

⁸ Grether (1980) provides experimental evidence on people overreacting to signals, and develops a quasi-Bayesian framework to account for this evidence. Again, we denote as $f(\cdot)$ the subject’s prior density of $p \in (0, 1)$. For parameters $\beta_1 > \beta_2 \geq 0$, Grether (1980) assumes that the posterior density of $p \in (0, 1)$ upon observing $X_1 = 1$ and $X_2 = 1$ is given by

$$\tilde{f}(p|\{X_1 = 1\} \cap \{X_2 = 1\}) = \frac{\mathbb{P}[\{X_1 = 1\} \cap \{X_2 = 1\}|p]^{\beta_1} f(p)^{\beta_2}}{\int_0^1 \mathbb{P}[\{X_1 = 1\} \cap \{X_2 = 1\}|q]^{\beta_1} f(q)^{\beta_2} dq} = \frac{(p^2 + p(1-p)\rho)^{\beta_1}}{\int_0^1 (q^2 + q(1-q)\rho)^{\beta_1} dq},$$

where the second equality follows from Proposition 1 and $f(\rho) = 1$. For a given β_1 , we can calculate the implied posterior expectation of X . For $\beta_1 = 2$, which is consistent with the empirical results in Grether (1980), we obtain

$$\tilde{\mathbb{E}}[X|\{X_1 = 1\} \cap \{X_2 = 1\}] = \frac{10 + 4\rho + \rho^2}{12 + 6\rho + 2\rho^2}.$$

Suppose, for example, that the subject reports $\hat{p} = 10/12$ in a task with perfectly positively correlated signals. From the above we can then infer that the subject’s perceived correlation is $\hat{\rho} = 0$.

⁹ Consider again a subject who follows Grether’s (1980) model of the representativeness heuristic with parameter $\beta_1 = 2$. If the two signals X_1 and X_2 are perfectly positively correlated, the subject estimates the mean of X as $3/4$. The subject, therefore, behaves as if she was Bayesian with complete correlation neglect.



Fig. 3. Screenshot of the experimental task in Dean and Neligh (2019). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

the remaining probability $1 - q$.¹⁰ Thus, through the lens of our parametrization, the subject’s *prior* knowledge can be described by a binary lottery Y with outcomes $y_1 = B$ and $y_2 = R$.¹¹

We are interested in how subjects come up with their guesses upon observing the picture of blue and red balls. Given the incentive structure, a subject (who likes more money over less) will guess the state to be B if and only if she thinks that — after looking at the picture — it is weakly more likely that there are more blue than red balls on the screen. In principle, if she takes the time to count the number of blue balls, the subject can resolve all the uncertainty and receive $\$x$ with probability one. If she is less attentive, however, the subject is still somewhat uncertain after looking at the picture. For example, a subject might count the blue balls in the left half of the picture and extrapolate this to the overall number of blue balls. Or a subject might decide to not pay any attention at all and simply rely on her prior knowledge that state B is weakly more likely to occur. How much information a subject acquires (by being more or less attentive) depends on (a) the benefit of being attentive $\$x$, (b) her prior knowledge about the state q , and (c) the (cognitive) costs of being attentive. For cost functions often used in the literature (as discussed below), our parametrization allows us to characterize optimal information acquisition and to derive the resulting probability of a subject guessing B , as a function of the problem parameters in closed form. This makes comparative statics easy.

As noted earlier, the subject’s attention choice is usually modeled as her acquiring an informative signal X about the variable of interest Y . The optimal signal trades off the cost (as introduced below) and benefits of being attentive. In this kind of application, the signal X should not be interpreted literally, however; it is simply a modeling device to represent the cognitive processing of the information on the screen.¹² As we will see below, in the context of this experiment the optimal signal X is again a binary lottery. Hence, we can use our parametrization to describe the joint distribution of X and Y , thereby fully characterizing the optimal signal. Without loss of generality, we assume that X and Y are (weakly) positively correlated, so that we can label $x_1 = y_1 = B$ and $x_2 = y_2 = R$.¹³ Since the parameters x_1, x_2, y_1, y_2 and $\mathbb{P}[\{Y = B\}] = q$ are fixed, finding the optimal signal X corresponds to finding its marginal distribution, $\mathbb{P}[\{X = B\}] = p^*$, and its correlation with Y , $\rho(X, Y) = \rho^*$.

¹⁰ As an example, the experimenter could create 100 different pictures of blue and red balls, with $100q$ pictures having 51 blue and 49 red balls and $100(1 - q)$ pictures having 49 blue and 51 red balls. The experimenter explains this procedure to the subjects and tells them that for every decision one of these 100 pictures is drawn at random.

¹¹ This is with slight abuse of the previously used notation, because the outcomes of Y can no longer be ordered.

¹² As shown in Morris and Strack (2019), however, our results below can be directly applied also to settings, with the same incentive structure, where the agent literally observes a sequence of signals at decreasing marginal costs, and decides when to stop acquiring more information. Precisely, they show that static information acquisition with entropy-cost (that we define below) is identical to that in a sequential learning problem (as introduced in Wald, 1945) with a binary state and a constant flow cost that is increasing in the variance of the posterior belief.

¹³ This is doing nothing more than labeling $x_1 = B$ ($x_2 = R$, respectively) as the signal realization that is more likely to occur in state B (state R , respectively) and thus makes the subject guessing B (R , respectively) more likely.

Upon observing the signal X , the subject’s posterior belief can be described by the conditional distribution $Y|X$. Given the incentive structure, the subject then guesses the state to be B if and only if $\mathbb{P}[\{Y = B\}|X] \geq \frac{1}{2}$. When the signal is chosen optimally, the subject maximizes her expected utility by following the signal: it is optimal for the subject to guess B if $x_1 = B$ is realized and to guess R if $x_2 = R$ is realized.¹⁴ Hence, by deriving the probability p^* of observing a signal indicating that the subject should guess B , we also obtain our object of interest: the subject’s probability of guessing the state to be B , which is observable to the experimenter.

Following the large literature on rational inattention (see Maćkowiak et al., 2023, for a recent overview), we model the subject’s cost of being attentive via the reduction in entropy of Y due to observing the signal X . Precisely, the cost of observing X is given by $\lambda I(X; Y)$, where $I(X; Y)$ is the mutual information of X and Y ,¹⁵ and where $\lambda > 0$ is a cost parameter that reflects the difficulty of the task and/or a subject’s ability. When normalizing utility over money such that $u(0) = 0$, the optimal signal structure satisfies (see Matějka and McKay, 2015):

$$\frac{\mathbb{P}[Y = B|X = B]}{\mathbb{P}[Y = B|X = R]} = \exp(u(x)/\lambda) = \frac{\mathbb{P}[Y = R|X = R]}{\mathbb{P}[Y = R|X = B]}, \tag{7}$$

that is, the log-likelihood ratio of the state being $Y \in \{B, R\}$ (conditional on the subject’s guess) is equal to the benefit of being correct, $u(x)$, divided by the cost of being attentive, λ .

Our parametrization allows us to derive the optimal signal (and guesses) in closed form.¹⁶ Plugging the conditional probabilities from Proposition 1 into Eq. (7), we obtain two equations,

$$\frac{q + \frac{\kappa}{p}\rho}{q - \frac{\kappa}{1-p}\rho} = \frac{1 - q + \frac{\kappa}{1-p}\rho}{1 - q - \frac{\kappa}{p}\rho} \quad \text{and} \quad \frac{q + \frac{\kappa}{p}\rho}{q - \frac{\kappa}{1-p}\rho} = \exp(u(x)/\lambda), \tag{8}$$

which we can solve for the two unknowns, p and ρ . Re-arranging the first equation tells us that the optimal signal satisfies:

$$\rho^* = \frac{1 - 2q}{\sqrt{q(1 - q)}} \frac{\sqrt{p^*(q, x, \lambda)(1 - p^*(q, x, \lambda))}}{1 - 2p^*(q, x, \lambda)}.$$

Plugging this into the second equation in (8), and taking into account that $p^* \in [\frac{1}{2}, 1]$, we get

$$p^*(q, x, \lambda) = \begin{cases} q \frac{\exp(u(x)/\lambda)+1}{\exp(u(x)/\lambda)-1} - \frac{1}{\exp(u(x)/\lambda)-1} & \text{if } \lambda < u(x)/\ln \frac{q}{1-q}, \\ 1 & \text{if } \lambda \geq u(x)/\ln \frac{q}{1-q}. \end{cases} \tag{9}$$

If being attentive is sufficiently costly (i.e. if $\lambda \geq u(x)/\ln \frac{q}{1-q}$), the subject finds it optimal not to pay any attention to the problem: $\rho^* = 0$ and $p^* = 1$ (because, according to her prior, state B is more likely in this case). Otherwise the subject pays attention and sometimes guesses the state to be R instead: $\rho^* > 0$ and $p^* \in (\frac{1}{2}, 1)$. Eq. (9) now tells us exactly how the subject’s probability of guessing the state to be B varies with the problem parameters.

Having a closed-form solution for a subject’s predicted behavior is useful for illustrating the model’s predictions and interpreting experimental tests thereof (e.g. Dean and Neligh, 2019). It may also be useful for researchers who want to integrate rational inattention into broader economic models (e.g., to study how pricing responds to consumers being inattentive).

Bayesian persuasion Going beyond information acquisition, we can use our parametrization to study how a decision-maker wants to “present” information to persuade others (à la Kamenica and Gentzkow, 2011). In their leading example, Kamenica and Gentzkow consider a prosecutor (or sender) who tries to convince a judge (or receiver) that a defendant is guilty. The prosecutor receives a utility of one if the judge convicts the defendant and a utility of zero otherwise. The judge, in contrast, receives a utility of one if he convicts a guilty defendant or acquits an innocent one and a utility of zero otherwise. The prosecutor and judge share the common prior that the defendant is guilty with probability $q \in (0, \frac{1}{2})$. The judge makes a decision that maximizes her expected utility and thus, under this prior, acquits the defendant 100% of the time. At the other extreme, if the prosecutor conducts an investigation that is “fully revealing” (in the sense of perfectly identifying the defendant’s guilt or innocence), the judge will convict the defendant 100% of the time. Can the prosecutor come up with a line of arguments (modeled as a signal) that convinces a Bayesian judge¹⁷ to convict the defendant more than 100% of the time? Kamenica and Gentzkow (2011) have shown that the answer is yes.

¹⁴ Because state B is weakly more likely to begin with, for any (even non-optimal) signal X it is optimal to guess the state to be B if $x_1 = B$ is realized. If the signal is not chosen optimally, however, it might be better to guess B even when observing $x_2 = R$; namely if and only if

$$\frac{\mathbb{P}[Y = R|X = R]}{\mathbb{P}[Y = B|X = R]} = \frac{1 - q + \frac{\kappa}{1-p}\rho}{q - \frac{\kappa}{1-p}\rho} < 1 \quad \text{or, equivalently,} \quad q > \frac{1}{2} + \frac{\kappa}{1-p}\rho.$$

But, because being attentive is costly, in this case the subject could do better by not paying any attention at all.

¹⁵ See Appendix B for the precise definition of mutual information as well as some results on how the mutual information of two binary lotteries depends on their correlation and skewness.

¹⁶ In general, it is hard (or impossible) to characterize the optimal signal structure. A well-known exception is the case of linear-quadratic utility and a normal prior (see Maćkowiak et al., 2023).

¹⁷ Because the judge is Bayesian, her posterior belief must be *Bayes plausible*.

Kamenica and Gentzkow (2011) approach the problem of persuading a Bayesian agent in a geometric way. While their approach of “concavification” makes the problem tractable in general, our parametrization of binary lotteries allows us to set up, and solve, simple examples like the one above as standard constrained optimization problems. This can be useful for researchers trying to incorporate persuasion (about a binary state) into broader economic models.

To illustrate, we will solve the example above. Kamenica and Gentzkow (2011) argue that, if $q = 3/10$, the prosecutor can make the judge convict the defendant 60% of the time, although only 30% of the defendants are guilty. This insight generalizes to an arbitrary prior $q \in (0, \frac{1}{2})$, for which the prosecutor can obtain a conviction with probability $2q$. And, with our parametrization of binary lotteries at hand, we can provide a simple and elementary proof for this claim.

As in the previous application, we can formalize the judge’s prior knowledge via a binary lottery Y , with outcomes $y_1 = G$ (guilty) and $y_2 = I$ (nnocent). Kamenica and Gentzkow (2011) formalize persuasion as the prosecutor constructing a signal X that makes the judge update her belief to $Y|X$. By Proposition 1 in Kamenica and Gentzkow (2011), we can restrict attention to a binary signal X . Again it is without loss of generality to impose a (weakly) positive correlation between X and Y , so that we denote the outcomes of X as $x_1 = C$ (onvict) and $x_2 = A$ (cquit).

The judge convicts the defendant if and only if $\mathbb{P}\{Y = G|X\} \geq \frac{1}{2}$, and because $q \in (0, \frac{1}{2})$, the judge never convicts the defendant when the realized signal is $x_2 = A$. Hence, the prosecutor wants to maximize the probability $p = \mathbb{P}\{X = C\}$ subject to $\mathbb{P}\{Y = G|X = C\} \geq \frac{1}{2}$ and Bayes plausibility (here coming in the form of a constraint on the correlation coefficient). Because a fully revealing signal yields $p = q$, we know that the optimal signal satisfies $p^* \geq q$. Using our parametrization, we can thus write the prosecutor’s optimization problem as follows:

$$\max_{p, \rho} \underbrace{p}_{= \mathbb{P}\{X=C\}} \quad \text{subject to} \quad \underbrace{q + \sqrt{\frac{1-p}{p}} \sqrt{q(1-q)}\rho}_{= \mathbb{P}\{Y=G|X=C\}} \geq \frac{1}{2} \quad \text{and} \quad \rho \leq \underbrace{\sqrt{\frac{1-p}{p}} \sqrt{\frac{q}{1-q}}}_{\text{upper Fréchet bound}}.$$

Because $\mathbb{P}\{Y = G|X = C\}$ monotonically decreases in p and monotonically increases in ρ , the solution, (p^*, ρ^*) , to the above problem satisfies both constraints with equality. Thus,

$$\rho^*(X, Y) = \sqrt{\frac{1-p^*}{p^*}} \sqrt{\frac{q}{1-q}}.$$

Plugging this into the first constraint and re-arranging gives $p^*(q) = 2q$.

We can easily generalize the example by allowing for a “biased” judge, and still solve it in the exact same way as described above. Suppose, for instance, that the judge convicts the defendant if and only if $\mathbb{P}\{Y = G|X\} \geq \frac{1}{2} + b$ for some $b \in (0, \frac{1}{2})$. With our parametrization at hand, we conclude right away that the prosecutor can make the judge convict with probability

$$p^*(q, b) = \frac{2q}{1 + 2b}.$$

The above expression makes it easy to study how prior knowledge and bias interact in shaping the effectiveness of persuasion. This kind of trade-off is not unique to the example of a prosecutor trying to convince a judge, but arises across various economic domains (e.g., a manager or politician trying to convince (biased) shareholders or citizens to vote in a particular way).

Portfolio selection Finally, we show how our parametrization can be useful for identifying individuals’ potential misperceptions of — or misreactions to — correlation in portfolio choice, with minimal assumptions on preferences. How a subject’s portfolio choice changes in response to a c.p. change in the correlation of the underlying assets depends on her perception of that change as well as on her risk preferences. Hence, a researcher — who wants to identify a misperception of correlation — must impose assumptions on the subject’s preferences (e.g., that she is an EUT maximizer with a particular utility function), and the more restrictive these assumptions are, the easier it is to identify such misperceptions.

Existing experimental studies try to identify (mis)perceptions of correlation in selecting one of several portfolios. Ungeheuer and Weber (2021), for instance, assume that subjects are EUT maximizers who exhibit constant relative risk aversion. Under this assumption, they conclude that subjects neglect changes in correlation: their subjects behave as if the underlying assets were correlated less than they actually are. Eyster and Weizsäcker (2016) test for (mis)perceptions of correlation under weaker assumptions on preferences. They construct portfolio selection problems which feature the same set of available portfolio outcome distributions, but that differ in the underlying assets. According to any model that describes behavior based on portfolio outcomes, subjects should choose portfolios that result in the same portfolio outcome distributions across problems. The problems are set up in such a way, however, that an EUT maximizer who treats correlated assets as if being independent changes her behavior across the two problems. This experimental design comes close to an ideal test of correlation neglect in portfolio selection, with the one remaining caveat that Eyster and Weizsäcker (2016) do not only change the assets’ correlation, but also their marginal distributions across problems. Hence, although they rule out a preference-based explanation, Eyster and Weizsäcker (2016) cannot conclusively classify the behavioral change as a misreaction to correlation only. It is therefore natural to ask whether we can identify misreactions to correlation while making minimal assumptions on preferences and fixing the marginal distributions of the underlying assets.

Under the weak assumption that preferences over money are monotonic with respect to first-order stochastic dominance (FOSD), our parametrization allows us to identify misreactions to correlation in selecting a portfolio. Consider the assets in Table 2 as a concrete example. Asset X not only first-order stochastically dominates Asset Y , but also any portfolio $\alpha X + (1 - \alpha)Y$ with $\alpha \leq 25/33$.

Table 2
Joint distribution of two symmetric assets related via first-order stochastic dominance.

	Y = 100	Y = 40
X = 106	$\frac{1}{4} + \frac{1}{4}\rho$	$\frac{1}{4} - \frac{1}{4}\rho$
X = 90	$\frac{1}{4} - \frac{1}{4}\rho$	$\frac{1}{4} + \frac{1}{4}\rho$

Merely assuming that preferences over money are monotonic with respect to FOSD, a subject should choose X over any portfolio with $\alpha \leq 25/33$, no matter the correlation of the underlying assets. As a consequence, any change in behavior due to a change in correlation cannot be preference-based, but must be considered a mistake. At the same time, the dominance relationship is much easier to spot if the assets are perfectly positively correlated, in which case Asset X dominates any portfolio with $\alpha \leq 25/33$ *state-by-state*, so that one might expect subjects to make more mistakes when the assets are correlated less.¹⁸ And because the marginal distributions are fixed these mistakes must be attributed to the change in correlation.¹⁹

4. Concluding remarks

While we focus in this paper on describing the joint distribution of two binary lotteries, there are other classes of random variables, with a continuous distribution, for which the correlation coefficient is a proper dependence measure. As discussed in Levy and Paroush (1974) or Epstein and Tanny (1980), for instance, the dependence of two normally distributed random variables can be fully described with their correlation coefficient. Binary lotteries, however, come with two potential advantages from an economist’s perspective. First, their simplicity buys tractability in modeling and allows for easy communication to experimental subjects. Second, binary lotteries are flexible regarding manipulations of correlation as well as of the marginals, allowing researchers to independently vary expected value, variance, skewness — and correlation.

Declaration of competing interest

All three authors state that there are no interests to declare.

Data availability

No data was used for the research described in the article.

Appendix A. Proofs

Proof of Proposition 1. PART (a): Proposition 1 in Ebert (2015).

PART (b): We assume throughout that $q \geq p$. (The other case is analogous.) The joint distribution of X and Y is then constrained through the following system of linear equations:

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \omega_{11} \\ \omega_{12} \\ \omega_{21} \\ \omega_{22} \end{pmatrix} = \begin{pmatrix} p \\ q \\ 1-p \\ 1-q \end{pmatrix}. \tag{10}$$

We distinguish three cases: (i) $q \geq \frac{1}{2}$ and $p \geq 1 - q$, (ii) $q \geq \frac{1}{2}$ and $p < 1 - q$, and (iii) $q < \frac{1}{2}$.

Case (i): Let $q \geq \frac{1}{2}$ and $p \geq 1 - q$. Since these assumptions, together with $q \geq p$, imply $1 - q \leq \min\{q, p, 1 - p\}$, the solutions to the linear system in Eq. (10) satisfy:

$$\begin{pmatrix} \omega_{11} \\ \omega_{12} \\ \omega_{21} \\ \omega_{22} \end{pmatrix} = \begin{pmatrix} p - (1 - q) + \omega_{22} \\ 1 - q - \omega_{22} \\ 1 - p - \omega_{22} \\ \omega_{22} \end{pmatrix} \quad \text{with} \quad 0 \leq \omega_{22} \leq 1 - q. \tag{11}$$

Since $\text{Cov}(X, Y) = (x_1 - x_2)(y_1 - y_2)(\omega_{22} - (1 - q)(1 - p))$ and $\text{Var}(X) = p(1 - p)(x_1 - x_2)^2$ and $\text{Var}(Y) = q(1 - q)(y_1 - y_2)^2$, the correlation coefficient of X and Y equals

¹⁸ Comparing perfectly positive to perfectly negative correlation, the researcher can even fix the number of states.
¹⁹ Using the following straightforward observation, it is easy to construct further examples of this type: for symmetric assets X and Y with (i) $E_X > E_Y$ and (ii) $V_X < V_Y$ and (iii) $\rho(X, Y) \in [0, 1)$, X first-order stochastically dominates $\alpha X + (1 - \alpha)Y$ if and only if $E_X - E_Y > \max\{\frac{1+\alpha}{1-\alpha}\sqrt{V_X} - \sqrt{V_Y}, \sqrt{V_Y} - \sqrt{V_X}\}$.

$$\rho(X, Y) = \frac{\omega_{22} - (1 - q)(1 - p)}{\sqrt{(1 - p)(1 - q)pq}} = \frac{\omega_{22} - (1 - q)(1 - p)}{\kappa}, \tag{12}$$

which is strictly increasing in ω_{22} . Because $0 \leq \omega_{22} \leq 1 - q$, it follows that

$$\rho(X, Y) \in \left[-\sqrt{\frac{(1 - q)(1 - p)}{pq}}, \sqrt{\frac{(1 - q)p}{q(1 - p)}} \right] = \left[-\frac{\kappa}{qp}, \frac{\kappa}{q(1 - p)} \right].$$

Solving Eq. (12) for ω_{22} yields the identity $\omega_{22} = (1 - q)(1 - p) + \kappa\rho$. Plugging this expression into Eq. (11), we end up with the joint distribution of X and Y as stated in the proposition.

Case (ii) and (iii): Suppose that either $q \geq \frac{1}{2}$ and $p < 1 - q$ or $q \leq \frac{1}{2}$. In each of these cases, we have $p < \min\{q, 1 - q, 1 - p\}$. The proof goes along the same lines as that of the first case, with the one exception that now ω_{11} is the free state probability, which takes values in $[0, p]$. □

Proof of Proposition 2. The implication “(a) \Rightarrow (b)” follows from Proposition 1 (b).

“(b) \Rightarrow (a)”: The marginal distribution of X must satisfy $m \geq 2$ equations — one equation for each of its outcomes x_i — of the form $\sum_{j=1}^n \omega_{ij} = p_i$. Analogously, the marginal distribution of Y must satisfy $n \geq 2$ equations of the form $\sum_{i=1}^m \omega_{ij} = q_j$. Overall, we thus have $m + n$ equations that define the joint distribution over $m \cdot n$ states of the world. Since, by definition, $\sum_{i=1}^m \sum_{j=1}^n \omega_{ij} = 1 = \sum_{i=1}^m p_i$ and $\sum_{j=1}^n \sum_{i=1}^m \omega_{ij} = 1 = \sum_{j=1}^n q_j$, we have at most $m + n - 1$ linearly independent equations, however. As a consequence, we have at least $m \cdot n - (m + n) + 1$ “free” state probabilities that can take a range of values. This implies that we need at least $m \cdot n - (m + n) + 1$ independent parameters to describe the lotteries’ joint distribution exhaustively. A single parameter — such as the lotteries’ correlation — is sufficient to describe the joint distribution *only* if $m \cdot n = m + n$, which is indeed the case if and only if $m = n = 2$. □

Proof of Proposition 3. “(b) \Rightarrow (a)”: By Proposition 1, because the marginals are fixed and $\rho > \rho'$,

$$\begin{aligned} & \rho(r_1(X), r_2(Y)) > \rho(r_1(X'), r_2(Y')) \\ \Leftrightarrow & \mathbb{E}[r_1(X)r_2(Y)] > \mathbb{E}[r_1(X')r_2(Y')] \\ \Leftrightarrow & \kappa(\rho - \rho') (r_1(x_1)r_2(y_1) + r_1(x_2)r_2(y_2)) > \kappa(\rho' - \rho) (r_1(x_1)r_2(y_2) + r_1(x_2)r_2(y_1)) \\ \Leftrightarrow & r_1(x_1) (r_2(y_1) - r_2(y_2)) > r_1(x_2) (r_2(y_1) - r_2(y_2)), \end{aligned}$$

which is true for all r_1 and r_2 . (The claim follows from the necessities of the equivalences.)

“(a) \Rightarrow (b)”: Fix X and Y with a joint distribution $\{\omega_{ij}\}_{i \leq m, j \leq n}$. Without loss of generality, we can assume that Y is non-binary (i.e. $n \geq 3$). Now we construct X' and Y' such that $X \stackrel{d}{=} X'$ and $Y \stackrel{d}{=} Y'$, with a joint distribution that is derived from $\{\omega_{ij}\}_{i \leq m, j \leq n}$ as follows: Decrease ω_{mn} and $\omega_{m(n-2)}$ by $\epsilon/2$ for some $\epsilon > 0$, and increase $\omega_{(m-1)n}$ and $\omega_{(m-1)(n-2)}$ by the same amount. To preserve the marginals, we further increase $\omega_{m(n-1)}$ by ϵ and decrease $\omega_{(m-1)(n-1)}$ by ϵ .

Because $X \stackrel{d}{=} X'$ and $Y \stackrel{d}{=} Y'$, we have $\rho(X, Y) > \rho(X', Y')$ if and only if $\mathbb{E}[XY] > \mathbb{E}[X'Y']$. By construction, it follows that $\mathbb{E}[XY] - \mathbb{E}[X'Y'] = \frac{\epsilon}{2}(x_{m-1} - x_m)(y_{n-1} - y_{n-2} + y_{n-1} - y_n)$. We distinguish two cases. First, if $\rho(X, Y) \geq \rho(X', Y')$, we consider the nondecreasing functions $r_1(z) := z$ as well as $r_2(z) := y_n$ if $z \leq y_{n-1}$ and $r_2(z) := z$ otherwise. Then,

$$\mathbb{E}[r_1(X)r_2(Y)] - \mathbb{E}[r_1(X')r_2(Y')] = \frac{\epsilon}{2}(x_{m-1} - x_m)(y_n - y_{n-2}) < 0,$$

and, therefore, $\rho(r_1(X), r_2(Y)) < \rho(r_1(X'), r_2(Y'))$. Second, if $\rho(X, Y) < \rho(X', Y')$, we consider the functions $r_1(z) := z$ as well as $r_2(z) := y_{n-1}$ if $z \geq y_{n-1}$ and $r_2(z) := z$ otherwise. Then,

$$\mathbb{E}[r_1(X)r_2(Y)] - \mathbb{E}[r_1(X')r_2(Y')] = \frac{\epsilon}{2}(x_{m-1} - x_m)(y_{n-1} - y_n) > 0,$$

and, therefore, $\rho(r_1(X), r_2(Y)) > \rho(r_1(X'), r_2(Y'))$. This proves the claim. □

Proof of Proposition 4. We only prove that (b) implies (a). Fix binary lotteries X and Y , and consider an increase in their correlation by $\epsilon > 0$. By Proposition 1, the probabilities of states (x_1, y_1) and (x_2, y_2) increase by $\epsilon\kappa > 0$, while the probabilities of states (x_1, y_2) and (x_2, y_1) decrease by $\epsilon\kappa$. Thus, if $\rho(X, Y) > \rho(X', Y')$, X and Y are *more concordant* than X' and Y' . □

Appendix B. Correlation and mutual information

As we have seen in Proposition 4, when fixing X and Y , the correlation coefficient $\rho(X, Y)$ ranks the feasible joint distributions in the same way as does the concordance ordering. Thus, by Theorem 1 in Meyer and Strulovici (2012), also other prominent bivariate dependence orderings — greater weak association, the supermodular ordering, the convex-modular ordering, and the dispersion ordering — imply the exact same ranking.

A different way of thinking about the dependence of two random variables is in terms of their *mutual information* (Cover and Thomas, 1999). Mutual information — a concept from information theory — is a measure of how much information a lottery X

contains about another lottery Y and vice versa. Economists have used this concept to model the costs of information (e.g. Sims, 2003; Matějka and McKay, 2015; Maćkowiak et al., 2023), assuming that signals with more mutual information are more costly. Formally, the mutual information of X and Y equals the reduction in entropy of X when conditioning on the realization of Y .

Definition 2 (Mutual information). The mutual information of X and Y is given by

$$I(X; Y) := \sum_{i=1}^m \sum_{j=1}^n \omega_{ij} \log_2 \frac{\omega_{ij}}{p_i q_j}.$$

While the correlation coefficient distinguishes between positive and negative dependence, mutual information does not have such a directional component. In fact, the mutual information of X and Y is non-negative, and equal to zero if and only if $\rho(X, Y) = 0$ (e.g. Cover and Thomas, 1999). This already implies that, even when fixing the marginals of X and Y , ranking the feasible joint distributions in terms of their correlation is *not* the same as ranking them in terms of their mutual information. One might expect though that ranking the feasible joint distributions in terms of their *absolute* correlation, $|\rho(X, Y)|$, is identical to ranking them in terms of their mutual information. But also this is true if and only if at least one of the two lotteries is symmetric. Otherwise, depending on whether X and Y are skewed in the same or the opposite direction, either positive or negative correlation carries more mutual information.

Proposition 5. Fix any binary lotteries X and Y with upside probabilities $p \in (0, 1)$ and $q \in (0, 1)$.

(a) Let $p = \frac{1}{2}$ or $q = \frac{1}{2}$. Then, for any $\epsilon \in (0, \min\{p/(1-p), (1-p)/p, q/(1-q), (1-q)/q\})$, we have

$$I(X; Y) \Big|_{\rho(X, Y) = \epsilon} = I(X; Y) \Big|_{\rho(X, Y) = -\epsilon}.$$

(b) Let $p \neq \frac{1}{2}$ and $\text{sgn}(q - 1/2) = \text{sgn}(p - 1/2)$. For any $\epsilon \in (0, \min\{\kappa/pq, \kappa/(1-p)(1-q)\})$, we have

$$I(X; Y) \Big|_{\rho(X, Y) = \epsilon} < I(X; Y) \Big|_{\rho(X, Y) = -\epsilon}.$$

(c) Let $p \neq \frac{1}{2}$ and $\text{sgn}(q - 1/2) \neq \text{sgn}(p - 1/2)$. For any $\epsilon \in (0, \min\{\kappa/(1-p)q, \kappa/p(1-q)\})$, we have

$$I(X; Y) \Big|_{\rho(X, Y) = \epsilon} > I(X; Y) \Big|_{\rho(X, Y) = -\epsilon}.$$

Proof. PRELIMINARIES. Let $\Psi(x) := (1 + x) \log_2(1 + x)$ with $\Psi'(x) = 1 + \log_2(x)$. Then,

$$I(X; Y) = pq\Psi\left(\sqrt{\pi(1-p)\pi(1-q)\rho}\right) + p(1-q)\Psi\left(-\sqrt{\pi(1-p)\pi(q)\rho}\right) \\ + (1-p)q\Psi\left(-\sqrt{\pi(p)\pi(1-q)\rho}\right) + (1-p)(1-q)\Psi\left(\sqrt{\pi(p)\pi(q)\rho}\right).$$

Taking the derivative with respect to ρ and dividing it by $\sqrt{p(1-p)q(1-q)}$, we obtain

$$\frac{\partial}{\partial \rho} I(X; Y) \propto \log_2 \left(\overbrace{1 + \sqrt{\pi(1-p)\pi(1-q)\rho}}^{>0 \text{ if and only if } \rho > 0} \right) - \log_2 \left(\overbrace{1 - \sqrt{\pi(1-p)\pi(q)\rho}}^{>0 \text{ if and only if } \rho < 0} \right) \\ - \log_2 \left(\overbrace{1 - \sqrt{\pi(p)\pi(1-q)\rho}}^{>0 \text{ if and only if } \rho < 0} \right) + \log_2 \left(\overbrace{1 + \sqrt{\pi(p)\pi(q)\rho}}^{>0 \text{ if and only if } \rho > 0} \right).$$

Hence, $\frac{\partial}{\partial \rho} I(X; Y) > 0$ if and only if $\rho > 0$.

PART (a). Without loss of generality, we can assume $p = \frac{1}{2}$. Then,

$$I(X; Y) = q \left[\Psi\left(\sqrt{\pi(1-q)\rho}\right) + \Psi\left(-\sqrt{\pi(1-q)\rho}\right) \right] + (1-q) \left[\Psi\left(-\sqrt{\pi(q)\rho}\right) + \Psi\left(\sqrt{\pi(q)\rho}\right) \right].$$

Hence, for any $\epsilon > 0$, $I(X; Y) \Big|_{\rho(X, Y) = \epsilon} = I(X; Y) \Big|_{\rho(X, Y) = -\epsilon}$.

PART (b). Let $\epsilon > 0$. In addition, we assume $p \geq q > 1/2$, so that

$$\pi(p)\pi(q) > \pi(p)\pi(1-q) \geq \pi(q)\pi(1-p) > \pi(1-p)\pi(1-q).$$

The case with $p \leq q < 1/2$ is analogous. Next, we define $\Lambda(x) := \log_2(1 + x) + \log_2(1 - x)$, which is a strictly decreasing and strictly concave function with $\Lambda'(x) < 0$ and $\Lambda''(x) < 0$.

Table 3
Conditional distribution of X given Y .

	$X Y = y_1$	$X Y = y_2$
x_1	$p + \frac{\kappa}{q}\rho$	$p - \frac{\kappa}{1-q}\rho$
x_2	$(1-p) - \frac{\kappa}{q}\rho$	$(1-p) + \frac{\kappa}{1-q}\rho$

Building on the proof of Part (a), we observe that

$$\begin{aligned}
 & \left| \frac{\partial}{\partial \rho} I(X; Y) \Big|_{\rho(X, Y) = \epsilon} \right| - \left| \frac{\partial}{\partial \rho} I(X; Y) \Big|_{\rho(X, Y) = -\epsilon} \right| \\
 & \propto \Lambda \left(\sqrt{\pi(1-p)\pi(1-q)\epsilon} \right) + \Lambda \left(\sqrt{\pi(p)\pi(q)\epsilon} \right) \\
 & \quad - \Lambda \left(\sqrt{\pi(1-p)\pi(q)\epsilon} \right) - \Lambda \left(\sqrt{\pi(p)\pi(1-q)\epsilon} \right) \\
 & = \int_{\frac{\sqrt{\pi(1-p)\pi(1-q)\epsilon}}{\sqrt{\pi(1-p)\pi(q)\epsilon}}}^{\sqrt{\pi(1-p)\pi(1-q)\epsilon}} \Lambda'(z) dz - \int_{\frac{\sqrt{\pi(p)\pi(1-q)\epsilon}}{\sqrt{\pi(p)\pi(q)\epsilon}}}^{\sqrt{\pi(p)\pi(1-q)\epsilon}} \Lambda'(z) dz \\
 & < \epsilon \sqrt{\pi(1-p)} \left[\sqrt{\pi(1-q)} - \sqrt{\pi(q)} \right] \Lambda' \left(\sqrt{\pi(1-p)\pi(q)\epsilon} \right) \\
 & \quad - \epsilon \sqrt{\pi(p)} \left[\sqrt{\pi(1-q)} - \sqrt{\pi(q)} \right] \Lambda' \left(\sqrt{\pi(p)\pi(1-q)\epsilon} \right) \\
 & < \epsilon \sqrt{\pi(p)} \underbrace{\left[\sqrt{\pi(1-q)} - \sqrt{\pi(q)} \right]}_{<0 \text{ since } q > 1/2} \underbrace{\left[\Lambda' \left(\sqrt{\pi(1-p)\pi(q)\epsilon} \right) - \Lambda' \left(\sqrt{\pi(p)\pi(1-q)\epsilon} \right) \right]}_{\geq 0 \text{ since } \pi(p)\pi(1-q) \geq \pi(q)\pi(1-p) \text{ and } \Lambda''(\cdot) < 0} \\
 & \leq 0,
 \end{aligned} \tag{13}$$

where the first inequality follows from the fact that $p \geq q > 1/2$ and Λ is decreasing and concave, the second one follows from $p > 1/2$, and the third one again by the concavity of Λ .

Using the Eq. (13), we then conclude that

$$\begin{aligned}
 & I(X; Y) \Big|_{\rho(X, Y) = \epsilon} - I(X; Y) \Big|_{\rho(X, Y) = -\epsilon} \\
 & = \int_0^\epsilon \frac{\partial}{\partial \rho} I(X; Y) \Big|_{\rho(X, Y) = z} dz + \int_{-\epsilon}^0 \frac{\partial}{\partial \rho} I(X; Y) \Big|_{\rho(X, Y) = z} dz \\
 & = \int_0^\epsilon \frac{\partial}{\partial \rho} I(X; Y) \Big|_{\rho(X, Y) = z} dz - \int_0^\epsilon (-1) \frac{\partial}{\partial \rho} I(X; Y) \Big|_{\rho(X, Y) = -z} dz \\
 & = \int_0^\epsilon \left| \frac{\partial}{\partial \rho} I(X; Y) \Big|_{\rho(X, Y) = z} \right| - \left| \frac{\partial}{\partial \rho} I(X; Y) \Big|_{\rho(X, Y) = -z} \right| dz < 0,
 \end{aligned}$$

where the first equality follows from adding and subtracting $I(X; Y) \Big|_{\rho(X, Y) = 0}$, the second equality follows by a change of variable in the second integral, the third equality follows from Part (a), and the inequality from the inequality derived in (13).

PART (c). The proof goes along the same lines as in Part (b). \square

We observe that, if X and Y are skewed in the same direction, positive correlation carries less mutual information than the same degree of negative correlation. This explains why the optimal signal in the rational inattention model with entropy costs (as discussed in Section 3) has the same skewness as the variable of interest. A positively correlated signal with the same skewness is equally precise, but less costly. If X and Y are skewed in opposite directions, on the other hand, positive correlation carries more mutual information than negative correlation.

To get the intuition, consider the latter case. Here, positive correlation makes the less likely “signal” more informative — i.e. shifting the conditional probabilities away from one-half — and the more likely “signal” less informative. Negative correlation has the exact opposite effect. Since, by Theorem 2.7.4 in Cover and Thomas (1999), mutual information is convex in the conditional distribution of say X given Y , making a signal more (less) informative tends to increase (decrease) mutual information. Moreover, as depicted in Table 3, when conditioning on the less likely signal the same change in correlation induces a larger shift in conditional probabilities. Combining this with the preceding considerations yields Part (c) of the above proposition.

References

- Allais, M., 1953. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica* 21 (4), 503–546.
- Andersen, S., Harrison, G., Lau, M., Rutström, E., 2018. Multiattribute utility theory, intertemporal utility, and correlation aversion. *Int. Econ. Rev.* 59 (2), 537–555.
- Andreoni, J., Sprenger, C., 2012. Risk preferences are not time preferences. *Am. Econ. Rev.* 102 (7), 3357–3376.
- Benjamin, D.J., 2019. Errors in probabilistic reasoning and judgment biases. In: *Handbook of Behavioral Economics: Applications and Foundations* 1, vol. 2, pp. 69–186.
- Bordalo, P., Gennaioli, N., Shleifer, A., 2012. Salience theory of choice under risk. *Q. J. Econ.* 127 (3), 1243–1285.
- Brockett, P.L., Kahane, Y., 1992. Risk, return, skewness and preference. *Manag. Sci.* 38 (6), 851–866.
- Caplin, A., Dean, M., 2015. Revealed preference, rational inattention, and costly information acquisition. *Am. Econ. Rev.* 105 (7), 2183–2203.
- Charness, G., Levin, D., 2005. When optimal choices feel wrong: a laboratory study of Bayesian updating, complexity, and affect. *Am. Econ. Rev.* 95 (4), 1300–1309.
- Cheung, S.L., 2015. Risk preferences are not time preferences: on the elicitation of time preference under conditions of risk: comment. *Am. Econ. Rev.* 105 (7), 2242–2260.
- Chiu, W.H., 2010. Skewness preference, risk taking and expected utility maximisation. *Geneva Risk Insur. Rev.* 35 (2), 108–129.
- Courbage, C., Peter, R., Rey, B., 2022. Incentives and welfare effects of correlated returns. *J. Risk Insur.* 89 (1), 5–34.
- Cover, T.M., Thomas, J.A., 1999. *Elements of Information Theory*. John Wiley & Sons.
- Dean, M., Neligh, N.L., 2019. **Experimental tests of rational inattention**. http://www.columbia.edu/~md3405/Working_Paper_21.pdf.
- Denuit, M.M., Eeckhoudt, L., Liu, L., Meyer, J., 2016. Tradeoffs for downside risk averse decision makers and the self-protection decision. *Geneva Risk Insur. Rev.* 41 (1), 19–47.
- Dertwinkel-Kalt, M., Köster, M., 2020. Salience and skewness preferences. *J. Eur. Econ. Assoc.* 18 (5), 2057–2107.
- Driessen, J., Ebert, S., Köster, J., 2020. **PI-CAPM: an Asset Pricing Model with Probability Weighting**. Available at SSRN.
- Ebert, S., 2015. On skewed risks in economic models and experiments. *J. Econ. Behav. Organ.* 112, 85–97.
- Ebert, S., van de Kuilen, G., 2016. Measuring multivariate risk preferences. Available at SSRN.
- Ebert, S., Wiesen, D., 2011. Testing for prudence and skewness seeking. *Manag. Sci.* 57 (7), 1334–1349.
- Enke, B., Zimmermann, F., 2019. Correlation neglect in belief formation. *Rev. Econ. Stud.* 86 (1), 313–332.
- Epstein, L.G., Tanny, S.M., 1980. Increasing generalized correlation: a definition and some economic consequences. *Can. J. Econ.* 13 (1), 16–34.
- Eyster, E., Weizsäcker, G., 2016. Correlation neglect in portfolio choice: Lab evidence. Available at SSRN.
- Grether, D.M., 1980. Bayes rule as a descriptive model: the representativeness heuristic. *Q. J. Econ.* 95 (3), 537–557.
- Heinzel, C., Peter, R., 2021. **Precautionary Motives with Multiple Instruments**. Available at SSRN.
- Joe, H., 1997. *Multivariate Models and Dependence Concepts*. Chapman & Hall.
- Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47 (2), 263–291.
- Kamenica, E., Gentzkow, M., 2011. Bayesian persuasion. *Am. Econ. Rev.* 101 (6), 2590–2615.
- Levy, H., Paroush, J., 1974. Toward multivariate efficiency criteria. *J. Econ. Theory* 7 (2), 129–142.
- Loomes, G., Sugden, R., 1982. Regret theory: an alternative theory of rational choice under uncertainty. *Econ. J.* 92 (368), 805–824.
- Loomes, G., Sugden, R., 1987. Testing for regret and disappointment in choice under uncertainty. *Econ. J.* 97 (388a), 118–129.
- Mačkowiak, B., Matějka, F., Wiederholt, M., 2023. Rational inattention: a review. *J. Econ. Lit.* 61 (1), 226–273.
- Matějka, F., McKay, A., 2015. Rational inattention to discrete choices: a new foundation for the multinomial logit model. *Am. Econ. Rev.* 105 (1), 272–298.
- Menezes, C.F., Geiss, C., Tressler, J., 1980. Increasing downside risk. *Am. Econ. Rev.* 70 (5), 921–932.
- Meyer, M., Strulovici, B., 2012. Increasing interdependence of multivariate distributions. *J. Econ. Theory* 147 (4), 1460–1489.
- Miao, B., Zhong, S., 2015. Risk preferences are not time preferences: on the elicitation of time preference under conditions of risk: comment. *Am. Econ. Rev.* 105 (7), 2272–2286.
- Morris, S., Strack, P., 2019. **The Wald problem and the relation of sequential sampling and ex-ante information costs**. Available at SSRN.
- Richard, S.F., 1975. Multivariate risk aversion, utility independence and separable utility functions. *Manag. Sci.* 22 (1), 12–21.
- Rohde, I., Rohde, K., 2015. Managing social risks—tradeoffs between risks and inequalities. *J. Risk Uncertain.* 51 (2), 103–124.
- Rohde, K., Yu, X., 2023. **Intertemporal correlation aversion—a model-free measurement**. *Manag. Sci.* <https://doi.org/10.1287/mnsc.2023.4863>.
- Rothschild, M.D., Stiglitz, J.E., 1970. Increasing risk: I. A definition. *J. Econ. Theory* 2 (3), 225–243.
- Sims, C.A., 2003. Implications of rational inattention. *J. Monet. Econ.* 50 (3), 665–690.
- Tchen, A.H., 1980. Inequalities for distributions with given marginals. *Ann. Probab.* 8 (4), 814–827.
- Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5 (4), 297–323.
- Ungeheuer, M., Weber, M., 2021. The perception of dependence, investment decisions, and stock prices. *J. Finance* 76 (2), 797–844.
- Wald, A., 1945. Sequential tests of statistical hypotheses. *Ann. Math. Stat.* 16 (2), 117–186.