

The cognitive basis for decision making under risk and uncertainty: research programs & controversies

Edited by

Samuel Shye, Riccardo Viale and Shabnam Mousavi

Published in

Frontiers in Psychology



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-6108-9
DOI 10.3389/978-2-8325-6108-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

The cognitive basis for decision making under risk and uncertainty: research programs & controversies

Topic editors

Samuel Shye — Hebrew University of Jerusalem, Israel

Riccardo Viale — University of Milano-Bicocca, Italy

Shabnam Mousavi — Max Planck Institute for Human Development, Germany

Citation

Shye, S., Viale, R., Mousavi, S., eds. (2025). *The cognitive basis for decision making under risk and uncertainty: research programs & controversies*.

Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6108-9

Table of contents

- 04 **Editorial: The cognitive basis for decision making under risk and uncertainty: research programs & controversies**
Samuel Shye and Riccardo Viale
- 07 **Toward an attentional turn in research on risky choice**
Veronika Zilker and Thorsten Pachur
- 13 **Rebiasing: Managing automatic biases over time**
Aleksey Korniychuk and Eric Luis Uhlmann
- 26 **The “beauty premium” effect of voice attractiveness of long speech sounds in outcome-evaluation event-related potentials in a trust game**
Junchen Shang and Zhihui Liu
- 35 **Humans as intuitive classifiers**
Ido Erev and Ailie Marx
- 47 **Exploring the determinants of reinvestment decisions: Sense of personal responsibility, preferences, and loss framing**
Johannes T. Doerflinger, Torsten Martiny-Huenger and Peter M. Gollwitzer
- 64 **Risk attitude and belief updating: theory and experiment**
Evelyn Y. H. Huang and Benson Tsz Kin Leung
- 76 **How general is the natural frequency effect? The case of joint probabilities**
Nathalie Stegmüller, Karin Binder and Stefan Krauss
- 93 **Stochastic heuristics for decisions under risk and uncertainty**
Leonidas Spiliopoulos and Ralph Hertwig
- 104 **Uncertainty about paternity: a study on deliberate ignorance**
Gerd Gigerenzer and Rocio Garcia-Retamero
- 114 **Gambling on others’ health: risky pro-social decision-making in the era of COVID-19**
Leyla Loued-Khenissi and Corrado Corradi-Dell’Acqua



OPEN ACCESS

EDITED AND REVIEWED BY
Snehlata Jaswal,
Sikkim University, India

*CORRESPONDENCE
Samuel Shye
✉ samuel.shye@mail.huji.ac.il

RECEIVED 16 December 2024
ACCEPTED 13 February 2025
PUBLISHED 27 February 2025

CITATION
Shye S and Viale R (2025) Editorial: The
cognitive basis for decision making under risk
and uncertainty: research programs &
controversies. *Front. Psychol.* 16:1546461.
doi: 10.3389/fpsyg.2025.1546461

COPYRIGHT
© 2025 Shye and Viale. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC
BY\)](#). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Editorial: The cognitive basis for decision making under risk and uncertainty: research programs & controversies

Samuel Shye^{1,2*} and Riccardo Viale³

¹Department of Philosophy, Economics & Political Science (PEP), Hebrew University of Jerusalem, Jerusalem, Israel, ²Department of Psychology, Hebrew University of Jerusalem, Jerusalem, Israel, ³Department of Economics, University of Milano-Bicocca, Milan, Italy

KEYWORDS

expected utility theory (EUT), adaptive heuristics, ecological bounded rationality, challenge theory of decision under risk, dual system theory, enactive problem solving

Editorial on the Research Topic

The cognitive basis for decision making under risk and uncertainty: research programs & controversies

This volume showcases alternative research strategies in decision-making under risk and presents thought-provoking decision problems. The dominant approach in this research domain, rooted in Expected Utility Theory (EUT), emphasizes identifying functions that account for deviations from EUT, typically overlooking the cognitive processes involved. This limits its explanatory power and offers little guidance for improving decision-making. Nonetheless, the insights and terminology introduced by [Kahneman and Tversky \(1979\)](#) and [Tversky and Kahneman \(1992\)](#), who advanced this approach, remain influential, as reflected throughout this Research Topic.

An alternative approach, developed by [Simon \(1981, 1982\)*](#) and expanded by [Gigerenzer et al. \(1999\)](#) and [Gigerenzer and Selten \(2001\)](#), focuses on simple adaptive heuristics and ecological bounded rationality. It aims to improve decision theory by identifying cognitive processes that allow satisfactory choices when perfect optimization is not possible. Moreover, it focuses on the features of the environment, which are often characterized by uncertainty, complexity, and ambiguity. This volume contributes to this perspective by exploring attention allocation in heuristic decisions (Chapter 1), proposing a free parameter for error adjustment in heuristic choices (Chapter 3), examining the role of heuristics in biased choices (Chapter 4), and identifying heuristic elements in intuitive decision-making (Chapter 2).

A novel approach, Challenge Theory (CT; [Shye and Haber, 2020a,b](#)), integrates elements from the two approaches mentioned above. CT conceptualizes cognitive decision-making as two sequential thought processes: the heuristic (System 1), which reacts to probabilities and defines the default option, and the deliberate (System 2), which reevaluates the default and may opt for the “bold” alternative. Thus, in loss contexts, CT redefines the risky option as the “default” and the safe option as “bold,” providing a streamlined, one-parameter explanation for the psychological effects—the certainty effect, Allais Paradox, reflection, overweighting of low probabilities,

* When we began planning this project, Shabnam Mousavi was one of our co-editors. Her intelligence, creativity, intellectual independence, and profound knowledge would have been invaluable to the success of this endeavor. Tragically, Shabnam passed away before we could benefit from her contributions. We have strived to honor her vision in shaping this issue, which is dedicated to her memory.

and loss aversion—identified by Kahneman and Tversky as deviations from EUT. Initial experiments suggest that CT outperforms traditional economic models. Key elements of CT are echoed in various chapters of this Research Topic, such as the two-system approach (Chapter 4), the possible prominence of probabilities over outcomes (Chapter 2), and heuristics as a starting point in a sequential cognitive decision process (Chapter 1).

A more radical approach connected to ecological bounded rationality is introduced by Viale (2024) and Viale et al. (2023a), who integrate Simon and Newell's (1971) problem-solving framework into the emerging research paradigm of embodied cognition (Viale et al., 2023b). Simon (1986) emphasizes the centrality of problem-solving, distinguishing it from decision-making, which he considers a subsequent phase. According to Simon, the essence of rationality lies in the ability to adapt, with adaptation relying more on external environmental interactions than on internal cognition. Behavior aligns with external objectives, revealing systemic constraints on adaptation. Simon (1981) highlights the critical role of environmental feedback in shaping actions and narrowing the problem space—the set of potential situations to explore for solutions. In the context of embodied cognition, the problem space represents solutions enabled by environmental affordances (Viale, 2024). This perspective of enactive problem-solving bypasses the analytic phase of decision-making, reducing reliance on symbolic representation and focusing on iterative, action-driven feedback processes.

Below are brief descriptions of each chapter in this volume:

1. To enhance the explanatory power of decision theory, Zilker and Pachur advocate shifting the focus toward how imbalances in attention allocation, rather than distorted risk perceptions, shape decision-making. This approach offers deeper insights into how preferences are formed and holds promise for refining current heuristic models of risky decision-making.

2. Erev and Marx argue that the mainstream assumption of separating judgment from decision-making leads to oversensitivity to rare events. Additionally, the belief that providing a full description of incentives replaces judgment and past experiences overlooks the significant role past experiences play in decision-making. They propose that decision processes are more akin to machine learning classification, where patterns are recognized, rather than the traditional two-stage model of judgment and utility calculation.

3. Spiliopoulos and Hertwig propose that heuristic models, traditionally deterministic and parameter-free, can be enhanced by incorporating an error mechanism to account for stochastic choice. This modification introduces only a single free parameter while preserving the core cognitive processes of the original models. They explore different error mechanisms and examine how this adjustment influences comparisons between heuristics and more complex, parameter-rich models.

4. Korniychuk and Uhlmann model how automatic preferences influence decision-making during problem-solving through trial and error. They show that biases are beneficial early on but detrimental later. Timely “rebiasing”—reversing initial preferences—can lead to superior outcomes. This approach offers a strategic alternative to correcting biases, suggesting that organizations can improve performance by changing key decision-makers rather than eliminating biases entirely.

5. Gigerenzer and Garcia-Retamero find that men's widespread reluctance to take DNA tests to determine biological fatherhood is empirically linked to risk aversion. They conclude that this reluctance stems from anticipated regret: men fear potential embarrassment, if non-paternity is discovered; or potential strain on their relationship, if paternity is confirmed.

6. Loued-Khenissi and Corradi-Dell'Acqua investigated people's choices between two treatment options for serious diseases: a sure but mild improvement (sure option) or a riskier cure with a given probability of success (risky option). Results revealed a general preference for the riskier option, regardless of whether the recipient was oneself, a loved one, or a stranger. However, this preference diminished as the severity of the disease increased.

7. Two common errors in sequential investment decisions are escalation of commitment—persisting with a failing course of action—and prematurely abandoning a successful one. Doerflinger et al., using an incentivized task, identified three key determinants of escalation: personal responsibility, preference for initial investments, and loss framing. Notably, personal responsibility worsened decision quality, as participants were more likely to reinvest when accountable for prior decisions.

8. Huang and Leung examine how risk aversion influences belief updating, showing that stronger risk aversion leads to more conservative actions and reduces the value of new information. With self-relevant information (e.g., IQ), greater risk aversion leads to more belief change, while with self-irrelevant information, it leads to less belief change. Experimental results support this theory, with implications for persuasion, advertising, and political campaigns.

9. Shang and Liu explored how voice attractiveness influences cooperative behavior in economic games, with voices presented for 2,040 ms. Participants were more likely to invest in partners with attractive voices, confirming the “beauty premium” effect. They also invested more in male partners. Event-related potential (ERP) analysis showed that attractive voices reduced negative feelings after losses, suggesting that voice attractiveness weakens frustration and enhances cooperative behavior during feedback evaluations.

10. Stegmüller et al. examine how natural frequencies, known to aid Bayesian reasoning, perform in scenarios involving joint probabilities of binary events. Using a $2 \times 5 \times 2$ design, they explored different information formats and visualization types. Surprisingly, natural frequencies did not show the same advantage for joint probabilities as in typical Bayesian tasks. The format effect interacted with visualization types, with natural frequencies aiding understanding in some cases (like a double tree) but not in others (like a 2×2 table).

Conclusion

In conclusion, while cognitive-psychological approaches provide a more appropriate framework for understanding human decision-making under risk than Expected Utility Theory (EUT) or its derivatives, Aumann's (2019) thesis remains relevant: people generally make decisions that align with EUT. Indeed, people follow behavioral rules of thumb, which have evolved because they generally promote human goals, that is, accord

with EUT. Deviations from EUT typically occur in rare or contrived scenarios that are not subject to evolutionary pressures. As Tversky wryly observed while developing Prospect Theory: “Despite deviations from EUT, humans have managed rather well” (personal communication, Tversky, 1975). Thus, while cognitive-psychological processes undeniably shape decision-making under risk, their outcomes align sufficiently with EUT to ensure human survival.

Author contributions

SS: Conceptualization, Project administration, Writing – original draft, Writing – review & editing. RV: Conceptualization, Writing – review & editing.

References

- Aumann, R. J. (2019). A synthesis of behavioural and mainstream economics. *Nat. Human Behav.* 3, 666–670. doi: 10.1038/s41562-019-0617-3
- Gigerenzer, G., and Selten, R. (2001). *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press.
- Gigerenzer, G., Todd, P. M., and the, A. B. C., Research Group (1999). *Simple Heuristics that Make us Smart*. New York, NY: Oxford University Press.
- Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47, 263–291.
- Shye, S., and Haber, I. (2020a). Risk as challenge: a dual system stochastic model for binary choice behavior. *Appl. Econ. Finan.* 7:4714. doi: 10.11114/aef.v7i2.4714
- Shye, S., and Haber, I. (2020b). Challenge theory: the structure and measurement of risky binary choice behavior. *Appl. Econ. Finan.* 7:4845. doi: 10.11114/aef.v7i4.4845
- Simon, H. A. (1981). *The Sciences of the Artificial*. Cambridge, MA: MIT Press.
- Simon, H. A. (1982). *Models of Bounded Rationality, Volume 1: Economic Analysis and Public Policy; Volume 2: Behavioural Economics and Business Organization*. Cambridge, MA: MIT Press.
- Simon, H. A. (1986). *Decision Making and Problem Solving*. Washington, DC: National Academy Press.
- Simon, H. A., and Newell, A. (1971). Human problem solving: the state of the theory in 1970. *Am. Psychol.* 26, 145–159. doi: 10.1037/h0030806
- Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncert.* 5, 297–323.
- Viale, R. (2024). “Enactive problem solving: an alternative to the limits of decision making,” in Companion to Herbert Simon, eds. G. Gigerenzer, S. Mousavi, and R. Viale (Cheltenham: Cheltenham: Edward Elgar Publishing).
- Viale, R., Gallagher, S., and Gallese, V. (2023a). Bounded rationality, enactive problem solving, and the neuroscience of social interaction. *Front. Psychol.* 14:1152866. doi: 10.3389/fpsyg.2023.1152866
- Viale, R., Gallagher, S., and Gallese, V. (2023b). *Embodied Bounded Rationality*. Lausanne: Frontiers.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



OPEN ACCESS

EDITED BY

Samuel Shye,
Hebrew University of Jerusalem, Israel

REVIEWED BY

Jan B. Engelmann,
University of Amsterdam, Netherlands

*CORRESPONDENCE

Veronika Zilker
zilker@mpib-berlin.mpg.de

SPECIALTY SECTION

This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

RECEIVED 25 May 2022

ACCEPTED 08 August 2022

PUBLISHED 06 September 2022

CITATION

Zilker V and Pachur T (2022) Toward an
attentional turn in research on risky
choice. *Front. Psychol.* 13:953008.
doi: 10.3389/fpsyg.2022.953008

COPYRIGHT

© 2022 Zilker and Pachur. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Toward an attentional turn in research on risky choice

Veronika Zilker ^{1,2*} and Thorsten Pachur ^{1,2}

¹Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany,

²TUM School of Management, Technical University of Munich, Munich, Germany

For a long time, the dominant approach to studying decision making under risk has been to use psychoeconomic functions to account for how behavior deviates from the normative prescriptions of expected value maximization. While this neo-Bernoullian tradition has advanced the field in various ways—such as identifying seminal phenomena of risky choice (e.g., Allais paradox, fourfold pattern)—it contains a major shortcoming: Psychoeconomic curves are mute with regard to the cognitive mechanisms underlying risky choice. This neglect of the mechanisms both limits the explanatory value of neo-Bernoullian models and fails to provide guidance for designing effective interventions to improve decision making. Here we showcase a recent “attentional turn” in research on risk choice that elaborates how deviations from normative prescriptions can result from imbalances in attention allocation (rather than distortions in the representation or processing of probability and outcome information) and that thus promises to overcome the challenges of the neo-Bernoullian tradition. We argue that a comprehensive understanding of preference formation in risky choice must provide an account on a mechanistic level, and we delineate directions in which existing theories that rely on attentional processes may be extended to achieve this objective.

KEYWORDS

decision making under risk, risky choice, theory development, attention, cumulative prospect theory (CPT)

1. Introduction

One of the longest standing puzzles in the decision sciences is how people evaluate and choose between risky options, whose consequences cannot be predicted with certainty. Researchers have tried to address this question primarily by identifying how people deviate from the predictions of a normative economic model, then modifying the model to make it descriptively more valid. For instance, in the St. Petersburg paradox, most people are willing to pay only a moderate amount of money to play a coin toss game that involves risk but has an infinite expected value. Yet expected value (EV) theory would predict that people should be willing to pay a large sum to play the game. To account for this violation of EV theory, expected utility (EU) theory modified EV theory by introducing a concave utility function that assumes diminishing returns by transforming the objective outcomes of options into subjective utilities before weighting them by their probabilities (von Neumann and Morgenstern, 1944; Bernoulli, 1954).

However, human behavior also violates EU theory. For instance, EU theory cannot accommodate the fourfold pattern of risk attitudes (Tversky and Fox, 1995)—the phenomenon that people are risk averse (risk seeking) for high probability gains (losses) and low probability losses (gains). Consequently, prospect theory (PT; Kahneman and Tversky, 1979) and its extension, cumulative prospect theory (CPT; Tversky and Kahneman, 1992), were introduced. CPT modifies EU theory such that the subjective values of outcomes are no longer assumed to be weighted by their objective probabilities; instead, an inverse S-shaped probability weighting function overweights low-probability events and underweights medium- to high-probability events. The shape of the probability weighting function is characterized by its curvature (indicating the decision maker's probability sensitivity) and its elevation (indicating the decision maker's optimism/pessimism). Moreover, probability weighting in CPT depends on an event's rank among all possible events. In CPT, EU theory's utility function is replaced by a value function with a reference point; the function is concave (convex) for gains (losses) and steeper in the loss than in the gain domain, implementing the assumption of loss aversion. We refer to models such as EU theory and CPT—which use psychoeconomic functions (i.e., nonlinear utility and probability weighting functions) to account for violations of EV maximization—as *neo-Bernoullian* models. The approach of identifying behavioral deviations from EV and EU theory and capturing them in psychoeconomic functions has been highly influential, both in behavioral economics and in psychology, and has helped identify and characterize several key regularities of risky choice (see also Birnbaum, 2008).

The neo-Bernoullian modeling tradition, however, has an important downside: Its main concern is to account for choice behavior, with only little interest in the underlying psychological processes (Friedman and Savage, 1948). This imposes limitations. For instance, without understanding the cognitive processes underlying people's decisions it is impossible to design effective interventions to mitigate deviations from benchmarks such as EU or EV maximization (Weber and Johnson, 2009; Payne and Venkatraman, 2010). To illustrate, someone who ignores probabilities and someone who fails to understand probabilities would deviate from choices predicted based on an objective treatment of probabilities in similar ways but require a different intervention to rectify these deviations. Moreover, without a theoretical account of the cognitive processes it is difficult to specify at which stage and how psychological variables (e.g., affect; Lerner and Keltner, 2001; Pachur et al., 2014; Suter et al., 2016) or features of the choice context (e.g., whether people learn about the options via description or experience; Hertwig and Erev, 2009), can modulate risky choice behavior. For instance, characterizing patterns in description-based and experience-based choice with psychoeconomic functions indicates notable differences in probability weighting between these two learning modes

(Wulff et al., 2018). Understanding the mechanisms responsible for this gap in choice is difficult without a model that also explains the underlying information processing.

In this article we review recent progress toward overcoming the neglect of mechanisms that characterizes neo-Bernoullian models. Decades ago, Simon (1978) highlighted the key role attention plays in understanding decision making, but only recently have attempts to integrate attentional mechanisms in computational process models of decision making started to emerge. Here we review work showcasing that attentional processes can provide relatively simple yet powerful mechanistic explanations for longstanding puzzles in research on risky choice, and we delineate how these process-level insights can be linked to characteristic distortions in psychoeconomic functions. We discuss remaining questions and argue that attentional processes should be integrated more comprehensively in theories of risky choice. We begin by outlining how research on riskless choice has started to uncover how attentional processes modulate preference formation.

2. Preference construction and attention

In contrast to research on risky choice, a more process-oriented approach has been more readily adopted in research on riskless choice, such as choosing among food items (Shimojo et al., 2003; Krajbich et al., 2010). Measures of information search (e.g., eye tracking) have revealed some striking regularities—for instance, people tend to increasingly look at the item they ultimately choose over the time course of choice (the gaze cascade; Shimojo et al., 2003), and people are more likely to choose an item if they look at it longer than at the alternative (e.g., Krajbich et al., 2010; Zilker, 2022). These phenomena suggest a tight coupling between attention and preference formation.

Cognitive processes that might lead to these phenomena were formalized in computational models, which—unlike neo-Bernoullian theories—operate on the level of cognitive processing. An influential model of this type, the attentional drift diffusion model (aDDM; Krajbich et al., 2010), posits the decision process as an accumulation of evidence on the options over time and assumes that the accumulation of evidence toward an option is amplified (relative to the unattended option) whenever this option is attended to. The aDDM accounts for the gaze cascade and the attention–preference link described above. Investigations spanning various domains of decision making—including choices between monetary and food items, and even social decision making—have revealed that this attentional mechanism also seems to be at work in risky choice (Smith and Krajbich, 2018). Nevertheless, research on the aDDM has not addressed how attentional processes in risky choice might relate to the systematic deviations from EV and EU maximization

carved out by the neo-Bernoullian tradition. Recent work has started to bridge this gap between research traditions (thereby also contributing to theory integration; e.g., Gigerenzer, 2017; Pachur, 2020).

3. Linking psychoeconomic constructs in risky choice to imbalances in attention

Neo-Bernoullian models rely on psychoeconomic functions (e.g., nonlinear value and weighting functions) to account for deviations from EV and EU maximization. Recent work has uncovered that variability in predecisional attention allocation—measured, e.g., using eye tracking—may explain how the characteristic shapes of these functions come about. Using the process-tracing tool Mouselab (Payne et al., 1993), Pachur et al. (2018) measured how long participants inspected information on the attributes of risky options before making a choice. They modeled participants' choices with CPT and related the resulting parameter estimates to the attentional measures. The estimated value functions were more strongly curved for participants who inspected outcome information for a shorter time than for participants who inspected outcome information for a longer time. Furthermore, the estimated probability weighting functions were more strongly curved for participants who inspected probability information for a shorter time than for participants who inspected probability information for a longer time. These findings indicate that attention allocation to specific attributes in risky choice may modulate how severely people deviate from EV and EU maximization.

In addition, there are theory-driven analyses of how attentional mechanisms might relate to CPT's constructs. We (Zilker and Pachur, 2021) linked the mechanism proposed in the aDDM to nonlinear probability weighting. Specifically, we used the aDDM to simulate choices for binary choice problems and varied the strength and direction of attentional biases to one of the options in the choice problem. The simulated choices were modeled with CPT. The choice patterns arising from attentional biases in information search were reflected in highly systematic differences in the shape of the estimated probability weighting functions. For instance, when attention was biased to the safe option in a choice between a safe and a risky option, the resulting probability weighting functions were less elevated and more strongly curved—indicating a stronger overweighting of certainty—than when attention was biased to the risky option.

These results point to a process-level, mechanistic explanation for choice patterns that are commonly described with CPT's probability weighting function. Notably, the aDDM gives rise to these distortions in probability weighting merely by assuming attentional biases during evidence accumulation, without applying any nonlinear distortion to the options' outcomes or probabilities. This highlights that choice patterns

captured by distorted psychoeconomic functions can arise even when the attributes of choice problems are processed and evaluated in a non-distorted manner.

The analyses presented in Zilker and Pachur (2021) reveal how deviations from maximization might be linked to biases in attention allocation across options. To test whether such a link holds empirically, we reanalyzed a large pool of data from previous process-tracing studies. Indeed, attentional imbalances between options during predecisional information search were associated with specific distortions in probability weighting (Zilker and Pachur, 2021).

Using a similar approach but a different class of cognitive process models, heuristics, Pachur et al. (2017) analyzed how imbalances in attention allocation across attributes in risky choice might be related to the shape of psychoeconomic functions. For instance, some heuristics (e.g., minimax and maximax) focus on outcome information only and ignore probabilities; other heuristics (e.g., the least-likely heuristic) consider both outcome and probability information. By modeling choices predicted by heuristics with CPT, the authors showed that the different attentional policies implied by various heuristics are linked with specific distortions in the shape of CPT's psychoeconomic functions.

The insights obtained by these analyses help alleviate the neglect of mechanisms in neo-Bernoullian theories. They also suggest novel, process-based explanations for the impact of contextual variables (e.g., learning about options via description or experience) on psychoeconomic functions. For instance, people might show systematically different attentional biases depending on whether they learned about the options from description or experience, which might explain the description-experience gap in terms of probability weighting. Likewise, other psychological variables known to modulate risky decision making (e.g., affect) might operate by modulating the attentional process (e.g., Fehr-Duda et al., 2011). Moreover, the analyses point toward novel ways of designing interventions that might render probability weighting more objective (i.e., linear). Specifically, if preferences deviating from linear weighting can be attributed to attentional biases, a greater adherence to objective weighting might be achieved by manipulations (e.g., attentional cues) that lead to a more balanced allocation of attention across attributes and options.

In addition to the attention-based approaches to modeling risky choice originating in cognitive psychology, recent research in behavioral economics has also contributed to elaborating this link. For instance, Smith et al. (2019) propose a random utility approach for estimating the impact of attention on preferences. Similarly, Engelmann et al. (2021) integrate prominent economic theories—salience theory and rational inattention (Sims, 2003; Bordalo et al., 2012)—to disentangle bottom-up and top-down effects of attention on economic decisions. Overall, the quest for a unified theory of the attentional roots of decision making under risk will thus be

an interdisciplinary one, involving both cognitive psychology, behavioral economics, and neuroscience. Moreover, it is pertinent to note that attention is not the only cognitive process that can modulate decision making under risk. Memory, executive functions, and learning processes may shape risky decisions as well.

4. Discussion

We have reviewed empirical and theoretical work revealing how attentional processes can explain the risky choice phenomena that shaped the development of neo-Bernoullian theories. Although these analyses have enriched the understanding of how patterns in attention allocation relate to the shapes of psychoeconomic functions, existing theoretical accounts of the attentional process in risky decision making are incomplete. Perhaps most importantly, relatively little is known about how attentional biases in risky choice come about in the first place. Ideally, a comprehensive theoretical account on the level of cognitive processing should be able to predict (a) how attention is initially allocated to the different attributes in a given choice problem, (b) how attention allocation and potential biases therein unfold during evidence accumulation, (c) how a preference is formed based on the sampled information about the options' outcomes and probabilities, and how this process is modulated by attention, and ultimately, (d) how a choice is made. We next outline possible starting points for theorizing about attention allocation in risky choice.

4.1. Attention guided by accumulated evidence

Although the aDDM is a simple and elegant tool for explaining how imbalances in attention allocation can lead to deviations from EV or EU maximization, it does not predict attention allocation itself. How might imbalances in attention allocation come about? One possibility is that such imbalances emerge dynamically during evidence accumulation. If information search is guided by the amount of evidence accumulated for an option at a given time, attention allocation might reflect differences between the options in the evidence accumulated thus far (e.g., Gluth et al., 2020; Callaway et al., 2021; Glickman et al., 2022). The observation that options sometimes tend to capture attention proportional to their value (Anderson et al., 2011; Le Pelley et al., 2016; Gluth et al., 2018) seems consistent with this possibility. However, unless accompanied by further factors that shape attention allocation, value-based attention may not be able to explain systematic deviations from EV or EU maximization—the more valuable option would almost necessarily end up being attended to more and would thus be more likely to be chosen.

4.2. Attention guided by features of the choice problem

Unbalanced attention allocation might also be driven by specific features of the options—leading to regularities such as the Allais paradox and the fourfold pattern. Among such stimulus features might be the size and salience of stimuli (Orquin et al., 2021). Although in standard paradigms font size and display features are usually kept constant (but see Weber and Kirsner, 1997), the options might still differ in visual properties. For instance, when people choose between a safe option and a risky option—where the latter usually consists of more information than the former—the options differ in complexity (Zilker et al., 2020). These differences in complexity may lead to differences in attention allocation (Orquin and Loose, 2013), which in turn might affect evidence accumulation and choice.

Even when the options do not differ in visual complexity, differences in their riskiness might still modulate attention. For instance, it has been proposed that options with higher variance are associated with more extensive internal sampling (Johnson and Busemeyer, 2005). To the extent that external attention also reflects internal sampling processes, this might lead to attentional imbalances between options. Recent formal models posit that people may predominantly direct their attention to the option whose outcome distribution is more variable, thus reducing uncertainty about its value (Callaway et al., 2021; Jang et al., 2021). Consistently, in paradigms in which people learn about the options based on free sampling from the options' outcome distributions, they tend to draw more samples from the option with higher variance (Lejarraga et al., 2012; Pachur and Scheibehenne, 2012).

Further, attention seems to be sensitive to the magnitude of the attribute values. For instance, larger outcomes or probabilities tend to receive more attention than smaller ones (e.g., Fiedler and Glöckner, 2012). According to decision field theory (Busemeyer and Townsend, 1993), an outcome should receive more attention the more likely it is to occur (see also Bhatia, 2014), but empirical tests of this prediction have yielded mixed results (Glöckner and Herbold, 2011; Stewart et al., 2016).

4.3. Strategic determinants of attention

Strategic factors might also guide attention allocation. Heuristic strategies describe how choice processes can be simplified, often by focusing on specific attributes (Thorngate, 1980; Payne et al., 1993). For instance, the minimax heuristic (Savage, 1954) makes decisions by comparing the least attractive outcome of each option; the maximax heuristic (Thorngate, 1980) compares the most attractive outcomes. Both heuristics ignore probabilities. Their attentional policies imply different

risk attitudes, with minimax leading to risk-averse and maximax to risk-seeking choices, reflected in distinct shapes of psychoeconomic functions (Pachur et al., 2017).

Other heuristics (e.g., the priority heuristic, the lexicographic heuristic; Payne et al., 1993; Brandstätter et al., 2006) predict a sequential inspection of attributes and that search is stopped as soon as the options differ on a given attribute. Heuristics might thus serve as a starting point for predicting patterns in attention allocation, and exploring how heuristic strategies are selected depending on the structure of the choice problem might illuminate how attentional policies vary across trials (Lieder and Griffiths, 2017; Mohnert et al., 2019).

5. Conclusion

For decades, a key approach to developing descriptive models for decision making under risk has been to modify a normative model, EV theory, by introducing transformations that distort the attributes of risky options (i.e., outcomes and probabilities) such that the predicted decisions match the observed ones. The psychological processes underlying the observed decisions were neither modeled nor measured. In this article we described an alternative approach. In process-level theories, deviations from EV maximization are not modeled by distorting outcome or probability information; instead, they are explained by a directly measurable aspect of cognitive processing: attention allocation. The development toward more cognitively grounded, attentional explanations of deviations from EV or EU maximization can be viewed as an “attentional turn” in research on risky choice. We argued that attention-based theoretical accounts of risky choice should not only predict how attention allocation shapes the accumulation of evidence, but also how patterns in attention allocation arise (for an example of how this might be achieved, see Johnson and Busemeyer, 2016).

As Herbert Simon noted, “attention is a major scarce resource” and people “cannot afford to attend to information simply because it is there” (Simon, 1978, p. 11). As researchers, “we must give an account not only of substantive rationality—the extent to which appropriate courses of action are chosen—but also procedural rationality—the effectiveness, in light of human cognitive powers and limitations, of the

procedures used to choose actions” (Simon, 1978, p. 9). With the attentional turn we have outlined, research in risky choice might finally take on Simon’s call and account for the intricate interplay between attention and preference.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

VZ and TP: conceptualization, writing, and editing. Both authors contributed to the article and approved the submitted version.

Funding

The Max Planck Society provided funding for the Article Processing Fees for this article.

Acknowledgments

The authors thank Deb Ain for editing the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Anderson, B. A., Laurent, P. A., and Yantis, S. (2011). Value-driven attentional capture. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10367–10371. doi: 10.1073/pnas.1104047108
- Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica* 22, 23–36. doi: 10.2307/1909829
- Bhatia, S. (2014). Sequential sampling and paradoxes of risky choice. *Psychon. Bull. Rev.* 21, 1095–1111. doi: 10.3758/s13423-014-0650-1
- Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychol. Rev.* 115, 463–501. doi: 10.1037/0033-295X.115.2.463

- Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *Q. J. Econ.* 127, 1243–1285. doi: 10.1093/qje/qjs018
- Brandstätter, E., Gigerenzer, G., and Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychol. Rev.* 113, 409–432. doi: 10.1037/0033-295X.113.2.409
- Busemeyer, J. R., and Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychol. Rev.* 100, 432–459. doi: 10.1037/0033-295X.100.3.432
- Callaway, F., Rangel, A., and Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLoS Comput. Biol.* 17, e1008863. doi: 10.1371/journal.pcbi.1008863
- Engelmann, J., Hirmas, A., and van der Wee, J. J. (2021). Top down or bottom up? Disentangling the channels of attention in risky choice. *Tinbergen Inst. Discuss.* 27, 52. doi: 10.2139/ssrn.3834381
- Fehr-Duda, H., Epper, T., Bruhin, A., and Schubert, R. (2011). Risk and rationality: The effects of mood and decision rules on probability weighting. *J. Econ. Behav. Organ.* 78, 14–24. doi: 10.1016/j.jebo.2010.12.004
- Fiedler, S., and Glöckner, A. (2012). The dynamics of decision making in risky choice: An eye-tracking analysis. *Front. Psychol.* 3, 335. doi: 10.3389/fpsyg.2012.00335
- Friedman, M., and Savage, L. J. (1948). The utility analysis of choices involving risk. *J. Polit. Econ.* 56, 279–304. doi: 10.1086/256692
- Gigerenzer, G. (2017). A theory integration program. *Decision* 4, 133–145. doi: 10.1037/dec0000082
- Glickman, M., Moran, R., and Usher, M. (2022). Evidence integration and decision confidence are modulated by stimulus consistency. *Nat. Hum. Behav.* 6, 988–999. doi: 10.1038/s41562-022-01318-6
- Glöckner, A., and Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *J. Behav. Decis. Mak.* 24, 71–98. doi: 10.1002/bdm.684
- Gluth, S., Kern, N., Kortmann, M., and Vitali, C. L. (2020). Value-based attention but not divisive normalization influences decisions with multiple alternatives. *Nat. Hum. Behav.* 4, 634–645. doi: 10.1038/s41562-020-0822-0
- Gluth, S., Spektor, M. S., and Rieskamp, J. (2018). Value-based attentional capture affects multi-alternative decision making. *Elife* 7, e39659. doi: 10.7554/eLife.39659
- Hertwig, R., and Erev, I. (2009). The description-experience gap in risky choice. *Trends Cogn. Sci.* 13, 517–523. doi: 10.1016/j.tics.2009.09.004
- Jang, A. I., Sharma, R., and Drugowitsch, J. (2021). Optimal policy for attention-modulated decisions explains human fixation behavior. *Elife* 10, e63436. doi: 10.7554/eLife.63436
- Johnson, J. G., and Busemeyer, J. R. (2005). A dynamic, stochastic, computational model of preference reversal phenomena. *Psychol. Rev.* 112, 841–861. doi: 10.1037/0033-295X.112.4.841
- Johnson, J. G., and Busemeyer, J. R. (2016). A computational model of the attention process in risky choice. *Decision* 3, 254–280. doi: 10.1037/dec0000050
- Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47, 263–292. doi: 10.2307/1914185
- Krajovich, I., Armel, C., and Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.* 13, 1292–1298. doi: 10.1038/nm.2635
- Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., and Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychol. Bull.* 142, 1111–1140. doi: 10.1037/bul0000064
- Lejarraga, T., Hertwig, R., and Gonzalez, C. (2012). How choice ecology influences search in decisions from experience. *Cognition* 124, 334–342. doi: 10.1016/j.cognition.2012.06.002
- Lerner, J. S., and Keltner, D. (2001). Fear, anger, and risk. *J. Pers. Soc. Psychol.* 81, 146–159. doi: 10.1037/0022-3514.81.1.146
- Lieder, F., and Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychol. Rev.* 124, 762–794. doi: 10.1037/rev0000075
- Mohnert, F., Pachur, T., and Lieder, F. (2019). “What’s in the adaptive toolbox and how do people choose from it? Rational models of strategy selection in risky choice,” in *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, eds A. K. Goel, C. M. Seifert, and C. Freksa (Montreal, QB: Cognitive Science Society), 2378–2384.
- Orquin, J. L., Lahm, E. S., and Stojić, H. (2021). The visual environment and attention in decision making. *Psychol. Bull.* 147, 597–617. doi: 10.1037/bul0000328
- Orquin, J. L., and Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychol.* 144, 190–206. doi: 10.1016/j.actpsy.2013.06.003
- Pachur, T. (2020). “Mapping heuristics and prospect theory: A study of theory integration,” in *Routledge Handbook of Bounded Rationality*, ed R. Viale (Cambridge, MA: Routledge), 324–337.
- Pachur, T., Hertwig, R., and Wolkewitz, R. (2014). The affect gap in risky choice: Affect-rich outcomes attenuate attention to probability information. *Decision* 1, 64–78. doi: 10.1037/dec0000006
- Pachur, T., and Scheibehenne, B. (2012). Constructing preference from experience: The endorsement effect reflected in external information search. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 1108–1116. doi: 10.1037/a0027637
- Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O., and Hertwig, R. (2018). Prospect theory reflects selective allocation of attention. *J. Exp. Psychol. Gen.* 147, 147–169. doi: 10.1037/xge0000406
- Pachur, T., Suter, R. S., and Hertwig, R. (2017). How the twain can meet: Prospect theory and models of heuristics in risky choice. *Cogn. Psychol.* 93:44–73. doi: 10.1016/j.cogpsych.2017.01.001
- Payne, J. W., Bettman, J. R., and Johnson, E. J. (1993). *The Adaptive Decision Maker*. New York, NY: Cambridge University Press.
- Payne, J. W., and Venkatraman, V. (2010). “Opening the black box: Conclusions to a handbook of process tracing methods for decision research,” in *A Handbook of Process Tracing Methods for Decision Research*, eds M. Schulte-Mecklenbeck, A. Kühberger, and R. Ranyard (New York, NY: Psychology Press), 223–249.
- Savage, L. J. (1954). *The Foundations of Statistics*. New York, NY: John Wiley & Sons.
- Shimojo, S., Simion, C., Shimojo, E., and Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nat. Neurosci.* 6, 1317–1322. doi: 10.1038/nm1150
- Simon, H. A. (1978). Rationality as process and as product of thought. *Am. Econ. Rev.* 68, 1–16.
- Sims, C. A. (2003). Implications of rational inattention. *J. Monet. Econ.* 50, 665–690. doi: 10.1016/S0304-3932(03)00029-1
- Smith, S. M., and Krajovich, I. (2018). Attention and choice across domains. *J. Exp. Psychol. Gen.* 147, 1810–1826. doi: 10.1037/xge0000482
- Smith, S. M., Krajovich, I., and Webb, R. (2019). Estimating the dynamic role of attention via random utility. *J. Econ. Sci. Assoc.* 5, 97–111. doi: 10.1007/s40881-019-00062-4
- Stewart, N., Hermens, F., and Matthews, W. J. (2016). Eye movements in risky choice. *J. Behav. Decis. Mak.* 29, 116–136. doi: 10.1002/bdm.1854
- Suter, R. S., Pachur, T., and Hertwig, R. (2016). How affect shapes risky choice: Distorted probability weighting versus probability neglect. *J. Behav. Decis. Mak.* 29, 437–449. doi: 10.1002/bdm.1888
- Thorngate, W. (1980). Efficient decision heuristics. *Behav. Sci.* 25, 219–225. doi: 10.1002/bs.3830250306
- Tversky, A., and Fox, C. R. (1995). Weighing risk and uncertainty. *Psychol. Rev.* 102, 269–283. doi: 10.1037/0033-295X.102.2.269
- Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertain.* 5, 297–323. doi: 10.1007/BF00122574
- von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
- Weber, E., and Kirsner, B. (1997). Reasons for rank-dependent utility evaluation. *J. Risk Uncertain.* 14, 41–61. doi: 10.1023/A:1007769703493
- Weber, E. U., and Johnson, E. J. (2009). Mindful judgment and decision making. *Annu. Rev. Psychol.* 60, 53–85. doi: 10.1146/annurev.psych.60.110707.163633
- Wulff, D. U., Mergenthaler-Canseco, M., and Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychol. Bull.* 144, 140–176. doi: 10.1037/bul0000115
- Zilker, V. (2022). Stronger attentional biases can be linked to higher reward rate in preferential choice. *Cognition*. 225, 105095. doi: 10.1016/j.cognition.2022.105095
- Zilker, V., Hertwig, R., and Pachur, T. (2020). Age differences in risk attitude are shaped by option complexity. *J. Exp. Psychol. Gen.* 149, 1644–1683. doi: 10.1037/xge0000741
- Zilker, V., and Pachur, T. (2021). Nonlinear probability weighting can reflect attentional biases in sequential sampling. *Psychol. Re.* doi: 10.1037/rev0000304. [Epub ahead of print].



OPEN ACCESS

EDITED BY
Riccardo Viale,
University of Milano-Bicocca, Italy

REVIEWED BY
Ian Belton,
Middlesex University, United Kingdom
Dilek Onkal,
Northumbria University,
United Kingdom

*CORRESPONDENCE
Aleksey Korniychuk
ak.si@cbs.dk

SPECIALTY SECTION
This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

RECEIVED 06 April 2022
ACCEPTED 29 August 2022
PUBLISHED 29 September 2022

CITATION
Korniychuk A and Uhlmann EL (2022)
Rebiasing: Managing automatic biases
over time.
Front. Psychol. 13:914174.
doi: 10.3389/fpsyg.2022.914174

COPYRIGHT
© 2022 Korniychuk and Uhlmann. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided
the original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Rebiasing: Managing automatic biases over time

Aleksey Korniychuk^{1*} and Eric Luis Uhlmann²

¹Copenhagen Business School, Strategy and Innovation, Frederiksberg, Denmark, ²INSEAD, Organisational Behaviour Area, Singapore, Singapore

Automatic preferences can influence a decision maker's choice before any relevant or meaningful information is available. We account for this element of human cognition in a computational model of problem solving that involves active trial and error and show that automatic biases are not just a beneficial or detrimental property: they are a tool that, if properly managed over time, can give rise to superior performance. In particular, automatic preferences are beneficial early on and detrimental at later stages. What is more, additional value can be generated by a timely *rebiasing*, i.e., a calculated reversal of the initial automatic preference. Remarkably, rebiasing can dominate not only debiasing (i.e., eliminating the bias) but also continuously unbiased decision making. This research contributes to the debate on the adaptiveness of automatic and intuitive biases, which has centered primarily on one-shot controlled laboratory experiments, by simulating outcomes across extended time spans. We also illustrate the value of the novel intervention of adopting the opposite automatic preference—something organizations can readily achieve by changing key decision makers—as opposed to attempting to correct for or simply accepting the ubiquity of such biases.

KEYWORDS

automatic evaluations, automatic preferences, biases, adaptiveness, intuition, debiasing

Introduction

Decision making in organizations is prone to the effects of intuitive thinking, most notably biases (Khatiri and Ng, 2000; Kahneman, 2003; Miller and Ireland, 2005). Existing work in the organizational sciences and social-cognitive psychology often focuses on debiasing interventions, in other words strategies to remove automatic biases from organizational choices (Schwenk, 1986; Wilson and Brekke, 1994; Wilson et al., 2000; Winter et al., 2007; Christensen and Knudsen, 2010). However, we show that dynamically rebiasing—that is, reversing biases by periodically adopting the opposite automatic preference—can be a strictly dominant strategy. To do so, we extend the standard model of boundedly rational search with a first principle of biased decision-making—namely, the presence of spontaneous, intuitive thinking.

Social-cognitive psychology has highlighted the layered nature of the human mind, where decision making involves the functioning of both controlled

(System 2) and automatic (System 1) processes (Simon, 1990; Sloman, 1996; Stanovich and West, 2000; Newell and Simon, 2007; Evans, 2008; Evans and Stanovich, 2013). The former is the kind of thought process that comes with an effort: it is deliberate, slow, and self-aware. The latter, conversely, is the kind of thinking that we can only barely control or shape logically: it is fast, associative, and effortless (Stanovich and West, 2000). This intuitive component represents an important element of human judgment. Even in organizations, decision makers routinely call on their intuitions or “gut feelings” when making both day-to-day and long term strategic choices (Khatri and Ng, 2000; Miller and Ireland, 2005). But the effect of intuitive thinking on organizational choices is not always positive and indeed can be detrimental (Kahneman, 2003; Inbar et al., 2010). This has to do with the fact that a key aspect of effortless information processing is our ability or propensity to make automatic evaluations before perceiving complete or even meaningful information (Zajonc, 1980; Wilson and Brekke, 1994; Duckworth et al., 2002; Kahneman, 2003; Volz and von Cramon, 2006). Naturally, such reliance on arbitrary, immediately observable stimuli often results in biases, or deviations from what would be deemed appropriate by the more logical rules of System 2 (Kahneman, 2003).

Biased judgments are commonplace and have been documented in a wide spectrum of settings (e.g., Kramer et al., 1993; Stone, 1994; Nickerson, 1998; Raghurir and Valenzuela, 2006; Scott and Brown, 2006). However, despite their definitional conflict with the rule of logic in observable outcomes, beyond the scope of a single choice, biases may be beneficial (Arkes, 1991; Marshall et al., 2013). Cognitive processes of System 1 generate responses so efficiently that the organisms possessing them can have evolutionary advantages (Gigerenzer and Todd, 1999). Similarly, such responses may reflect the properties of the environments in which our intelligence has evolved (e.g., Haselton and Nettle, 2006; Johnson and Fowler, 2011). If a certain behavioral response confers propagation or survival advantages, it is more likely to be prevalent in the population long-term (Haselton and Nettle, 2006). Consequently, the positive effects of our less controlled cognitive processes and corresponding biases may only emerge over a sequence of choices and would not be captured in single-session experiments in laboratory settings.

Guided by this premise, we conjecture that positive or negative effects of cognitive manipulations (such as eliminating or altering biases) should likewise manifest themselves over a sequence of adaptive choices. Accordingly, we design a computational model of adaptive sequential trial and error that incorporates the first principles of human thinking and thus allows for a study of temporal effects of System 1 biases as well as interventions to eliminate or alter them.

We find that the consequences of biased judgments are indeed time-variant. System 1 automatic evaluations offer short-term benefits that will tend to propagate in dynamic

environments that remain stable only for a limited time. However, these benefits quickly disappear, causing profound long-term harm. The reason for the observed pattern is that automatic evaluations constrain the space of options for trial and error (e.g., pick only green, no red), thereby suppressing experimentation. Further analysis of this effect reveals that manipulations of biases can offer advantages in settings with more available time. However, contrary to what may be expected, it is not debiasing (or eliminating the bias) that betters both biased and unbiased decision making, it is rebiasing (or reversing the bias). To be effective, rebiasing must take place at a calculated moment in time. An advantage, therefore, may come not from eliminating biases but from effectively managing them. Unlike individuals, organizations can in principle reverse their biases by appointing different decision makers to key roles such as top leadership positions.

Theoretical background

Consider the following problem. A decision maker is faced with a set of options, each with a different payoff or score. These can represent monetary outcomes such as profit, or different measures of performance, for example, product quality, cost, or customer satisfaction. The goal is to discover options with greater scores (see, for example, Simon, 1955).

For a flawless intelligence, a problem like this is trivial. An omnipotent mind would immediately select the best option. Assuming that there are no information processing constraints, the number of possibilities is finite, and there are no impediments to choice, such behavior is rational. Indeed, in some situations, this kind of intelligent choice is a good proxy of that of humans. Think, for example, about choosing the biggest apple on a plate. The color, size, and shape are all directly observable and the choosing of the most appealing apple is not a problem. Given comprehensible information about all options, we simply pick the best one. However, the situation changes when we cannot process the entire set of possibilities or face noisy signals. Finding the biggest apple in a loaded trailer will already reveal the limits of our capacities.

In the middle of the last century, Herbert Simon postulated that in problems like the one above, human rationality is bounded (Simon, 1955, 1956). Instead of optimizing over the entire space of possibilities, we search and satisfice. That is, we sequentially generate and try new options until we find one that meets all essential criteria or as long as our outcomes are below aspirations (Simon, 1955; Levinthal and March, 1981; Lant, 1992). In other words, boundedly rational decision makers continuously search for better options. This model of decision making represents the kind of “behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms” (Simon, 1955, p. 99).

However, while certainly compatible with a limited intelligence, including that of a human, the Simonian representation of problem solving is not specifically human (or more broadly, biological). In particular, it omits biases that are typical of human cognition (see [Fiori, 2011](#)). The existing literature identifies a wide spectrum of intuitive biases or spontaneous “response[s] because of mental processing that is unconscious or uncontrollable” ([Wilson and Brekke, 1994](#), p. 117). These biases systematically contaminate decision making, often without the person’s awareness of their influence. Indeed, such blindness to the rationale behind one’s own choices reflects the complexity of human thought ([Nisbett and Wilson, 1977](#); [Greenwald and Banaji, 1995](#); [Haidt, 2001](#); [Kahneman et al., 2011](#)).

Extensive research in psychology indicates that human cognition involves the simultaneous functioning of two systems ([Slooman, 1996](#); [Kahneman, 2003](#)). One system (System 1) is spontaneous, intuitive, uncontrolled, and fast—this system is based on the law of association. The other system (System 2) is deliberate, effortful and relatively slow—this system can be said to rely on the law of logic ([Stanovich and West, 2000](#)). However, the responses of these systems to exogenous stimuli do not always align. In situations in which System 1 dominates System 2 (e.g., limited time, high cognitive load, or when the choice is closer to perception than to deliberate assessment), the decision maker’s judgment is especially likely to deviate from the rules of logic ([Fazio, 2001](#)). Although there are exceptions, such as expert intuition trained in repetitive and predictable settings—think about chess ([Kahneman and Klein, 2009](#))—in real-world situations automatic evaluations will not always be “reasonable by the cooler criteria of reflective reasoning. In other words, the preferences of System 1 are not necessarily consistent with preferences of System 2” ([Kahneman, 2003](#), p. 1463). This inconsistency can take multiple forms but fundamentally it reduces to an arbitrary preference for a certain, immediately observable or perceivable attribute of options ([Zajonc, 1980](#); [Fazio et al., 1986](#); [Fazio, 2001](#); [Duckworth et al., 2002](#); [Slovic et al., 2002](#)).

Such preferences form as a part of automatic evaluations that do not require conscious reasoning and occur even when the stimuli are novel ([Zajonc, 1980](#); [Fazio et al., 1986](#); [Greenwald and Banaji, 1995](#); [Fazio, 2001](#); [Duckworth et al., 2002](#)). While these affective responses are variegated ([Hutchinson and Gigerenzer, 2005](#)), in the context of choice, they fundamentally reduce to a form of heuristic that accepts or rejects based on a certain immediately perceivable attribute of options. That is, “pick A, if A is” more readily accessible, more representative of a category, implies lesser losses, etc.

To the extent that this immediately observable attribute is uncorrelated with the target criterion (i.e., the performance score, quality, cost, etc.), the ultimate choice will be subject to biases. Importantly, the presence of these biases is not uniform over all stages of the decision-making processes.

Specifically, the greater the involvement of System 1, the more liable to biases the choice is. This happens because intuitive judgments originate “between the automatic parallel operations of perception and the controlled serial operations of reasoning” ([Kahneman and Frederick, 2002](#), p. 50). Somewhere between perception and more deliberate processes of reasoning, a human-like intelligence will have a quick, spontaneous evaluative response that may direct the ultimate choice ([Zajonc, 1980](#); [Kahneman, 2003](#)).

Existing experimental studies have shown that biases appear in a wide variety of trivial choices ([Tversky and Kahneman, 1974](#)). A natural consequence is that biases permeate human and by extension organizational decision making. This, in turn, can hold implications for organizational performance. Accordingly, scholars have analyzed the role of biases from various organizational perspectives, from their effects on strategic decision making ([Schwenk, 1984, 1986](#); [Lyles and Thomas, 1988](#); [Reitzig and Sorenson, 2013](#)) to their implications for organizational adaptation ([Denrell and March, 2001](#)). However, in this stream of work, biases have been essentially equated with some form of evaluation imperfections and thus no different from systematic errors in deliberate decisions. The automatic, spontaneous nature of the underlying cognitive processes remains largely unintegrated with boundedly rational problem solving at the individual or organizational levels. This omission limits our understanding of how organizations can leverage the idiosyncrasies of human decision making.

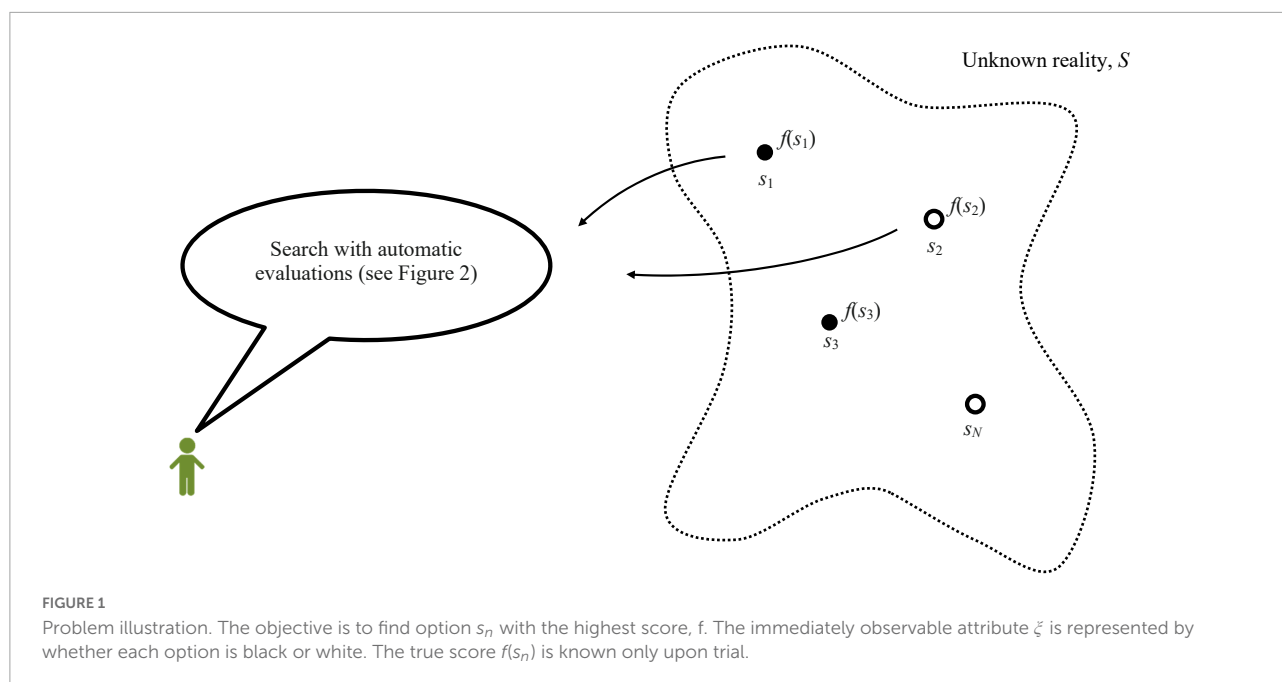
In the following section, we develop a parsimonious model of boundedly rational problem solving with unreasoned automatic evaluations (i.e., automatic biases). We then use this model to illustrate the temporal consequences of intervening to eliminate or change biases. Our work specifically assesses the effectiveness of two basic strategies that organizations can use to manipulate biases: de-biasing, or entirely eliminating a bias, and re-biasing, or adopting the exact opposite automatic preference, as well as their optimal timing.

Model setup and analyses

Our model has two basic elements: (i) an unknown reality with N options, (ii) a process of search that proxies problem solving by a boundedly rational intelligence with automatic evaluations. [Figure 1](#) illustrates these elements.

Unknown reality

Reality is represented by a set of options, S , where each option s_n has two attributes. For a trivial example, consider a bucket of exotic fruits. Let’s call them *karamzamsas*. The first attribute, ξ , is an immediately perceivable property, e.g., size, color, smell, etc. of a *karamzamsa*. We assume this attribute to



take on one of two values, 0 or 1, i.e., $\zeta \sim U\{0, 1\}$. The second attribute, f , represents the true value of the option, e.g., taste, nutritional content, etc. Without loss of generality, we assume that this value is distributed normally, i.e., $f(s_n) \sim N(0, 1)$. The true value of each option is observable only upon trial. That is, to know how a *karamzamsa* tastes, we need to take a bite.

Search with automatic evaluations

Consistent with the first principles of bounded rationality, our agents sequentially generate and try new options. However, we consider that although able to try only a single option at a time, agents can perceive multiple possibilities simultaneously. This is a key distinctive element of our conceptualization: at every moment in time, agents simultaneously perceive multiple options, but can try or experience only a single one. Continuing our example with a bucket of *karamzamsas*, consider that these exotic fruits are small and we can hold several of them in one hand. So we grab a handful and then drop all but the one we want to taste. For a more practical analogy, think about serial entrepreneurs or startups that come up with various business ideas but implement only a single one at a time. For an analogy that closely maps onto the underlying assumptions, think about the many choices organizational executives make on a daily basis: appointing the right subordinates, selecting suppliers, discontinuing products, etc.¹ In many ways, these decisions

are logically equivalent to exotic fruits: there is a multitude of them and their value, like that of *karamzamsas*, becomes fully identified only upon trial.

With this basic setup, we can understand the effect of biases that come with automatic evaluations. Unbiased agents will automatically select a random option. Think about a person who has never tried any fruit. This person will not be able to tell *karamzamsas* apart: a green *karamzamsa* looks just as good as a red one. On the contrary, a person who is fond of red apples, may automatically select red *karamzamsas*. Green *karamzamsas* are, of course, as good as red *karamzamsas*. But the person who likes red apples will tend to pick red *karamzamsas*. This is the logic of a biased agent, an agent with automatic evaluations who exhibits systematic preferences for an irrelevant immediately observable attribute of options. Although in the case of *karamzamsas*, such a bias will likely quickly disappear as the agent learns about the true taste of these wonderful fruits, many real-world biases are hard to eradicate even given the agent's full awareness (Wilson and Brekke, 1994). Such persistent biases in our automatic evaluations will interplay with our problem solving long-term.

Similar to Jung et al. (2021) we illustrate the logic of the search process with an algorithm. However, our algorithm

described as a collection of policies. States that differ by few policies are close to each other, whereas states that differ by many policies are distant. Naturally, correlation of performance tends to be higher for those states that are closer to each other and lower for those states that are far apart. On such a landscape, organizations tend to search within an immediate vicinity of the current state (see Simon, 1956; Levinthal, 1997). Our results are robust to such local adaptation on rugged performance landscapes simulated by means of the NK model (Kauffman and Levin, 1987; Kauffman, 1993; Rivkin, 2000).

¹ Combinations of these and similar decisions can be seen as locales on a rugged performance landscape (e.g., Levinthal, 1997; Rivkin, 2000). The idea in this line of work is simple: every (organizational) state is

does not have a defined stopping point. This implies that the agents continuously adjust their aspirations and continue searching for better solutions. **Figure 2** illustrates this algorithm and the distinction between the two categorical extremes, biased and unbiased search, in stricter terms. Unbiased search approximates problem solving of a bounded intelligence that has no automatic evaluations. Biased search is a proxy for a human-like intelligence that exhibits automatic evaluations. If the search is biased, the agents will effectively reject options based on the irrelevant criterion ξ every time they simultaneously perceive an option they prefer.

The logic of the algorithm is as follows. Generate or perceive several options. If one of these options dominates other options in terms of the immediately observable criterion ξ , select this option for thorough consideration and trial. If the selected option has been tried before, disregard it and restart the process of search. If the selected option has not been tried before, try it and observe its performance. We measure performance as the value $f(s_t)$ of the currently accepted option. If the performance improves, i.e., if $f(s_t) > f(s_{t-1})$, where t indicates the moment in time, accept this option, i.e., $f(s_t)$, as a new *status quo*. If the performance declines, i.e., if $f(s_t) < f(s_{t-1})$, continue to the next period and when it starts remember to return to the *status quo*, or the best option discovered thus far, i.e., $f(s_{t-1})$.

With this algorithm, we run a simulation model. In particular, we create a random set S of 100 options,² and assume that the agents sample options from this set with replacement. In every period, an agent generates two random alternatives from set S , picks one of the two generated options following the biased or unbiased process and then either tries this option or moves to the next period (see **Figure 2**). Our observations are averaged over at least 10^6 simulations. This amount of simulations ensures that the reported patterns are stable and reproduce with near certainty. Simulations were coded in Code:Blocks 16.01 in C++ programming language following C++ 11 ISO standard. The complete data and code are posted on the Open Science Framework at https://osf.io/sypn2/?view_only=1b00c0d2dc964bafadf10215bfca4743.

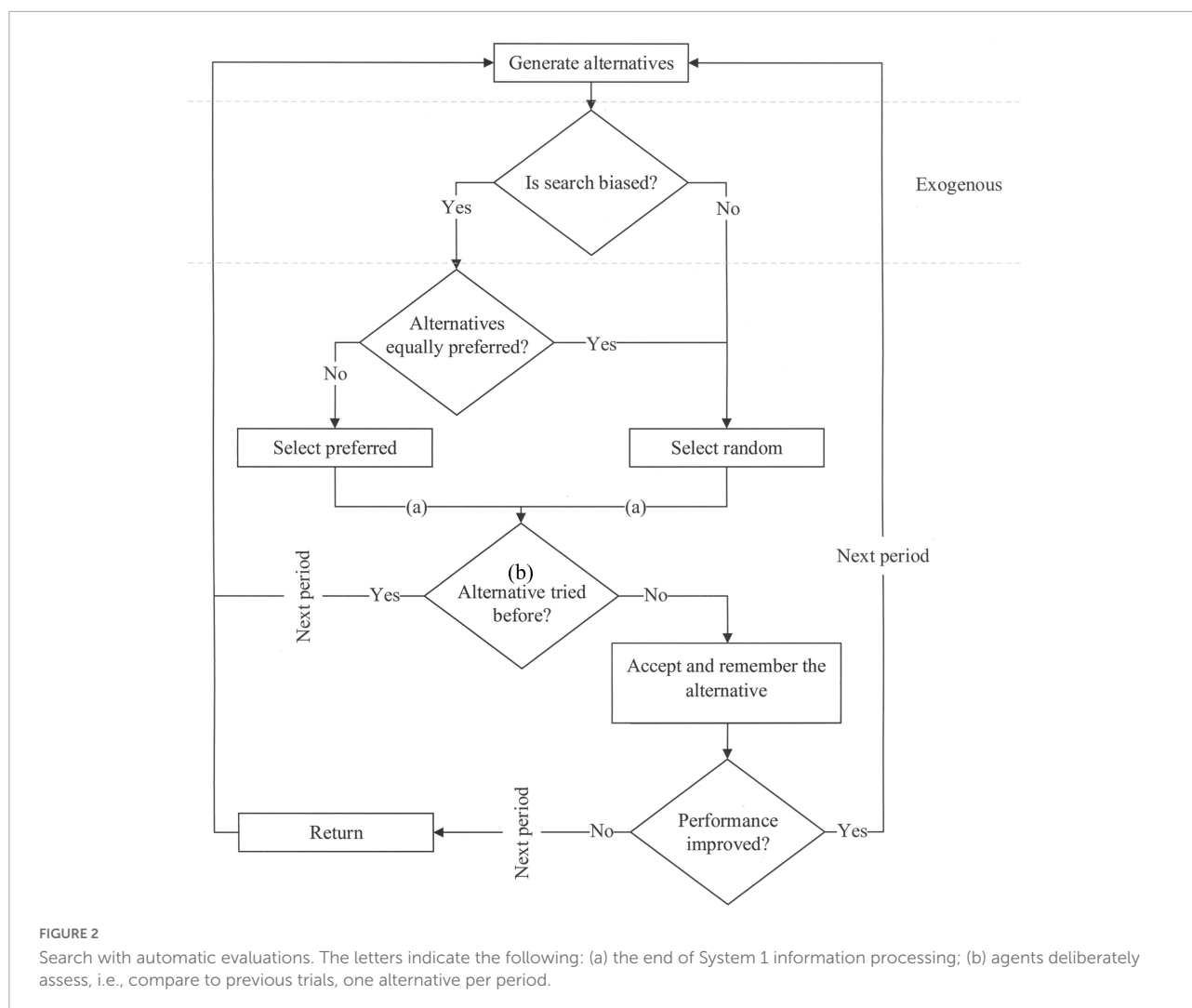
Before we proceed to our observations, let us make some important clarifications and caveats. First, the process, where the tried option can be sampled repeatedly, proxies a situation with a multiplicity of similar choices that have the same performance. To see what this means in the context of organizational decision making, consider, for example, a situation where a company from the capital region of Denmark unsuccessfully expands to the rest of the country. If establishing operations in Aalborg was not successful then probably (for the sake of argument, consider that these two

cities are sufficiently similar along the dimensions relevant for the organizational offer) it will also fail in Odense. Then, if after a failure in Aalborg, decision makers come up with the idea of starting operations in Odense, they will effectively have generated the same option again. This, of course, is only a hypothetical illustrative example. Possibilities vary (e.g., smaller cities in Denmark like Roskilde or Ringsted may turn out to represent a different option). The logic of the model is, of course, agnostic to the exact criterion. Sampling with replacement captures only the idea that some similar options have the same performance and can be intuitively generated or perceived separately.

Second, given the example above, a careful reader may wonder whether it is appropriate to compare an expansion to Aalborg in, for example, 2010 with an expansion to Odense in say 2035. Probably not. In fact, it may be equally unjustified to compare Aalborg in 2010 and Aalborg in 2035. The social, environmental, market, and even political conditions may be completely unlike. For this reason, time is a critical variable in our analysis because we compare performance in solving a given problem. The problem, of course, remains the same as long as the set of options S is constant. A meaningful change in the composition of this set, however, will essentially mean that the agents start solving another problem and the clock should start anew. Evolution of the problem, i.e., a gradual change in the composition of the set S , is another possibility. In the interest of clarity, we leave these issues beyond the scope of the present study and focus on the temporal effects of automatic biases when solving a given problem. That is, our agents search a fixed set of possibilities S and we observe their performance over time, i.e., the number of sequential choices made.

Finally, as any analytical tool, our model has boundary conditions. Our analysis captures a specific task environment designed to reflect the essential basics of many decision making situations. Although properties of this task environment are arguably general and sufficient for the following effects to hold in some other contexts of interest, the characteristics and complexities of specific real-world situations may differ and the model does not necessarily bear on them. These properties of the model can be summarized as follows: each option is characterized by two variables, one of which is directly observable and the other requires at least partial testing; decision makers are biased with respect to the observable variable but have no bias with respect to the unobservable variable of interest; the bias with respect to the observable variable materializes before any testing of the observable variable can be performed; and the two variables do not correlate with each other. The more overlapping features between the real situation and the simulated one, the more the simulation is relevant. The core code for our analyses is publicly posted, and we encourage the scientific community to explore alternative parameters more closely aligned with their specific decision making environments of interest.

² Recall that $f(s_n) \sim N(0, 1)$.



The basic effect

Figure 3 shows the relative effect of biased search. Positive (negative) values indicate that at the given moment in time, the biased agent has an advantage (disadvantage) over the unbiased agent. The value of zero means that biased and unbiased agents tend to have exactly the same performance.

An immediate observation is that the effect of automatic evaluations is time-variant. System 1 biases are beneficial in the short-term and yet harmful in the long run. Note that the model timings have no direct correspondence to real-world time. The model time is measured in terms of the number of steps or decisions made or, equivalently, the number of options considered for trial. A few steps (decisions) into the process of search, automatic evaluations can generate better performance by up to ~ 0.12 scores or 27% of the absolute performance of unbiased agents. Note that the magnitude of the advantage in terms of percentage peaks earlier. Early in the process of search, the absolute performance is relatively low and thus,

every additional score represents a greater portion. Consider that 65 steps into the process of search, the benefit of biased search equals 0.1192 scores or 11.4% of 1.045 scores gained at that point by the unbiased agent. On the contrary, 5 steps into the process of search, the benefit of biased search is only 0.008163 scores. But in percentage terms, this represents 27.21% of 0.03 scores gained by the unbiased agent at that time. This advantage, however, is relatively short-lived. Already 187 steps into the process of search, biases become detrimental. Although the magnitude of this effect does not exceed 2.7%, it continues (albeit monotonically declining) until the problem is solved, at which point biased and unbiased agents find the best alternative and their performances converge.

The mechanism

To understand the reasons for the observed pattern, consider what happens as the agents search the set of possibilities

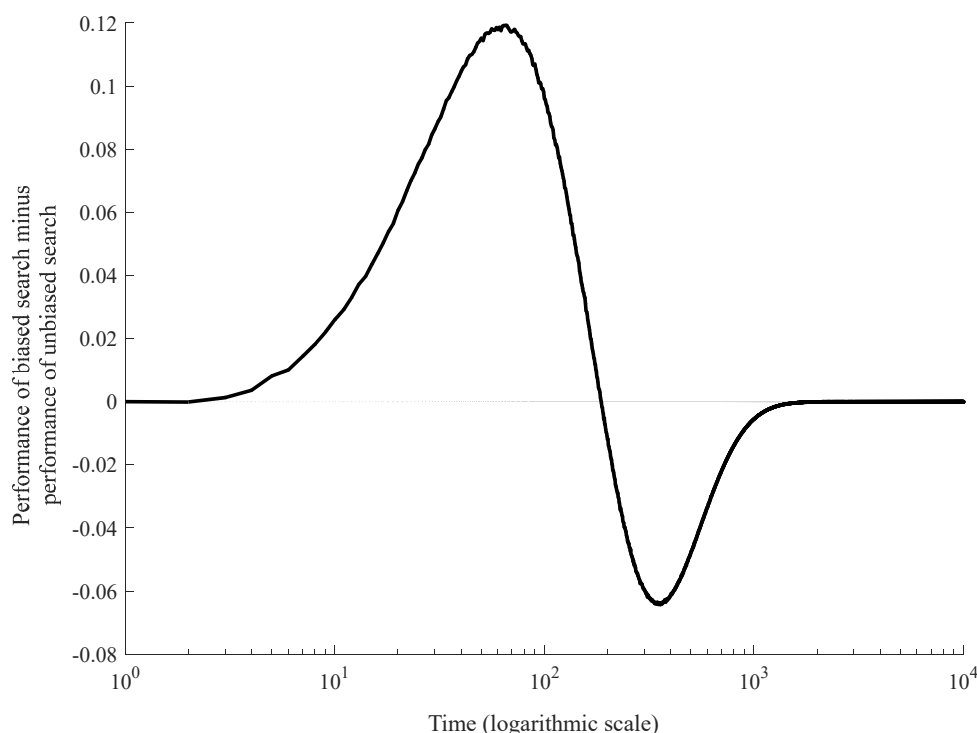


FIGURE 3
Performance of biased search relative to unbiased search.

S. Every time the agents try a new option, their expected performance is 0. Recall that since $f(s_n) \sim N(0, 1)$, $E[f(s_n)] = 0$. The difference between their *status quo* and the expected performance is essentially the implicit cost of experimentation. As long as their performance is greater than 0, every time they try a new option, their performance will fall until they return to the *status quo*. However, sometimes it will rise and their new *status quo* will improve measurably. This is how the agents learn, i.e., increase their accumulated knowledge about the problem.

Accordingly, the effect in Figure 3 is a product of two processes (see Figure 4). First, automatic evaluations direct agents to the options they prefer (i.e., are biased toward). As a result, a biased agent learns less, i.e., accumulated knowledge is lower, because it repeatedly draws from the same subset of possibilities. In contrast, an unbiased decision maker does not rely on automatic evaluations and therefore faces lower redundancies in learning.

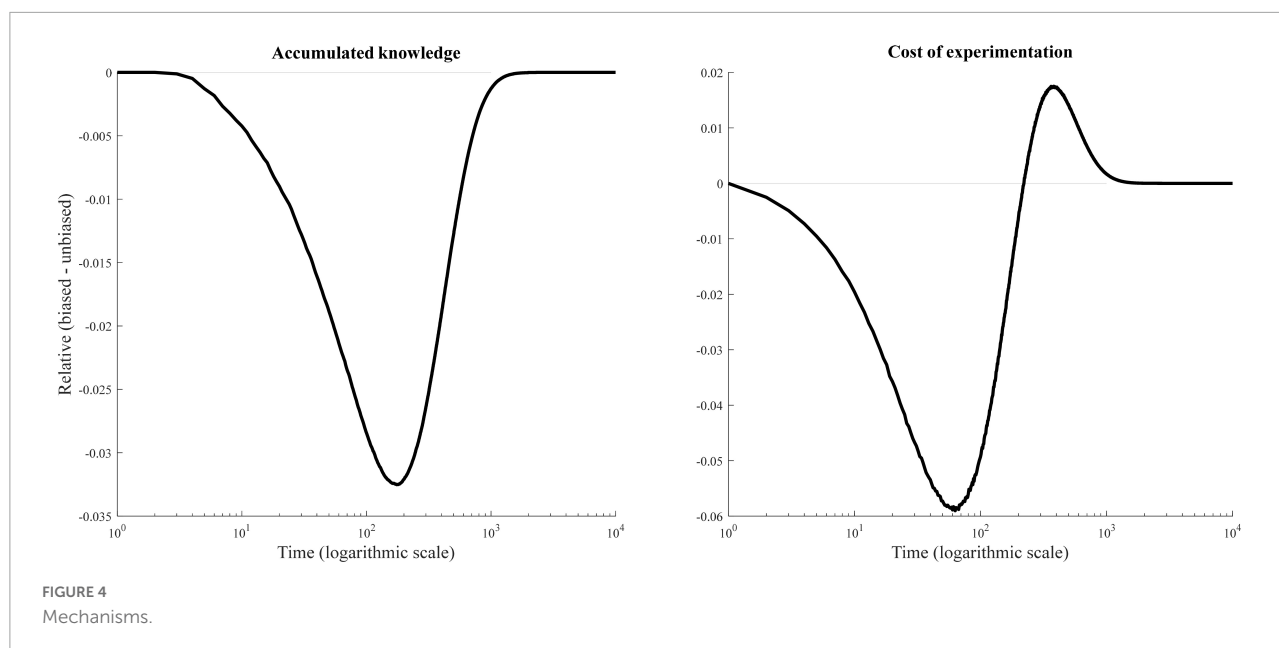
However, there is a second process. Learning about the problem requires experimentation, and experimentation is costly. Automatic evaluations make it less likely that the agents try new options and thereby regulate the excess of experimentation in the initial phase of problem solving. Early in the process of search, there is little knowledge about the set of possibilities S, which means that there are plenty of unknown options, each of which has an expected performance of 0. The probability of trying new options is very high during

this time. Automatic evaluations reduce this probability and thereby increase the value from stability. Over time, this value declines as the agents learn about the problem. Past experience with a given option helps resolve uncertainty about its potential: agents know that such an option is inferior to their *status quo* and therefore need not try it.

The curves in Figure 4 illustrate the dynamics of accumulated knowledge and the implicit cost of experimentation in relative terms, where zero means that there is no difference between biased and unbiased agents. The left panel shows the dynamics of accumulated knowledge. We measure accumulated knowledge as the score of the best option known to the agent. The right panel shows the cost of experimentation. We measure the cost of experimentation as the probability of trying a new option.

Rebiased and debiased search

In our analyses above, we assumed that biases remain constant during the entire process of search. While this is often the case, biases need not persist unchanged. Automatic evaluations exhibit high degrees of variability across people, such that different individuals can have idiosyncratic and atypical biases (Fazio et al., 1986; Baron, 2000). This variability may be used to change biases without altering the encoded



memory or association. Teams, organizations, and societies can replace key decision makers with others who are less biased or hold different biases. Case studies highlight instances in which companies have changed management teams and completely reversed their previous management practice orientations (see for example, Maddux et al., 2014). At the individual level, various psychological techniques, such as framing, may activate different automatic associations and thus elicit different automatic preferences or biases within the same person (Kühberger, 1998; Chong and Druckman, 2007). Scholars in psychology as well as industry practitioners have discussed an array of techniques that can abate the effect of biases, or debias, decision making (see Kahneman et al., 2011). Similarly, the literature in management has shown that organizations have structural means to manipulate and attempt to reduce bias in organizational decision making (see Christensen and Knudsen, 2010).

Accordingly, we examine temporal implications of two interventions or manipulations of bias: rebiasing (changing the bias to its opposite), and debiasing (eliminating the bias entirely). We operationalize rebiasing as adopting the exact opposite of the initial bias, i.e., pick red instead of green, when previously the automatic preferences was green over red. Debiasing means the agent no longer relies on any irrelevant signal. Consider our example with the exotic fruit *karamzamsa* and suppose that this fruit comes in two colors: red and green. As before, both green and red *karamzamsas* are equally tasty. Then, if our decision maker prefers red apples, this decision maker will likely favor red *karamzamsas*. Rebiasing in this case would be to now have a decision maker who prefers green apples. By analogy, debiasing would mean having a decision maker who equally prefers red and green apples. We are agnostic

as to the exact levers that organizations or collectives use to manipulate biases—whether they involve replacement of the key decision makers or implementation of other management practices—and focus solely on the outcomes of such strategic interventions. Our starting condition is that of the biased firm and its performance dynamics. Subsequently, we examine the temporal implications of rebiasing and debiasing.

Figure 5 shows the effects of these manipulations. The curves show relative performance of debiased and rebiased search (cf. Figure 3). The value of zero indicates that the difference between unbiased and debiased or rebiased agents is nil.

Contrary to what might be expected, debiasing does not result in simple convergence with unbiased search. Immediately after debiasing, there is a sharp decline in performance (see Figure 5). This happens because the set of options that used to be intuitively discarded remains comparatively unknown. So, when the bias disappears, the likelihood of trying new options goes up, which in turn increases the cost of experimentation. However, since a large portion of the possibilities are already encoded in the agent's memory, an increase in experimentation does not provide a commensurate improvement in the best-known state. As the agents gradually discover superior options, this initial shock of debiasing fades out and the performance of the debiased search ultimately converges to that of the continuously unbiased search.

In contrast, rebiasing leads to a second-order advantage. That is, after an initial drop in performance, rebiasing produces a temporary, but significant improvement in performance. A greater focus on the underexplored subset of the possibilities allows for a speeded accumulation of knowledge, which soon approaches that of the continuously unbiased search. As this

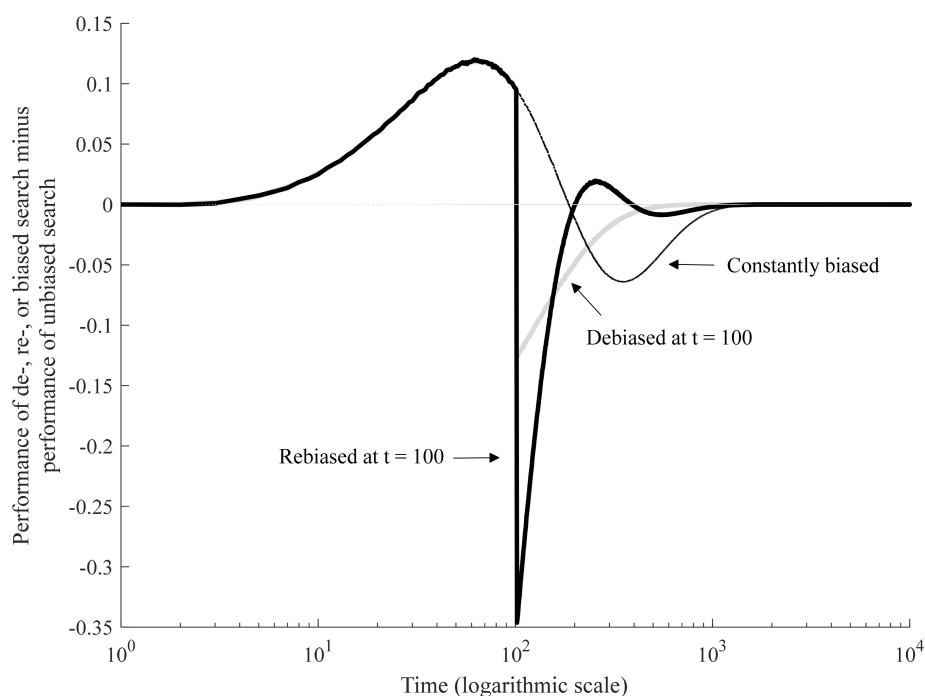


FIGURE 5
Rebiased, debiased, and constantly biased search compared to unbiased search.

happens, the implicit relative cost of experimentation declines and the agent takes advantage of the new bias. We call this effect a second-order advantage because it builds on the asymmetries in knowledge accumulation that were generated in the course of exercising the initial automatic bias.

The optimal timing of rebiasing

Significant declines in relative performance may naturally cause the species and by extension their behaviors to go extinct, or the company to become bankrupt. However, if the challenge of survival is taken out of the picture, the net effect of volatility is not clear. In particular, short-term losses can be seen as a form of investment for delayed gains. With this in mind, we compare the levels of cumulative scores of various behaviors (biased, unbiased, debiased, and rebased search) over different time spans. Note that there is no real-world time in the model. Therefore, as a proxy of actual time we take the count of search iterations or steps. In other words, one iteration of generating and evaluating a pair of alternatives corresponds to one unit on the time scale.

The curves in Figure 6 plot the relative cumulative performance of a given manipulation of biases. The value of zero indicates that the average accumulated performance of the unbiased and rebased or debiased agents are equal. For example, a point on the solid black line (left panel) that

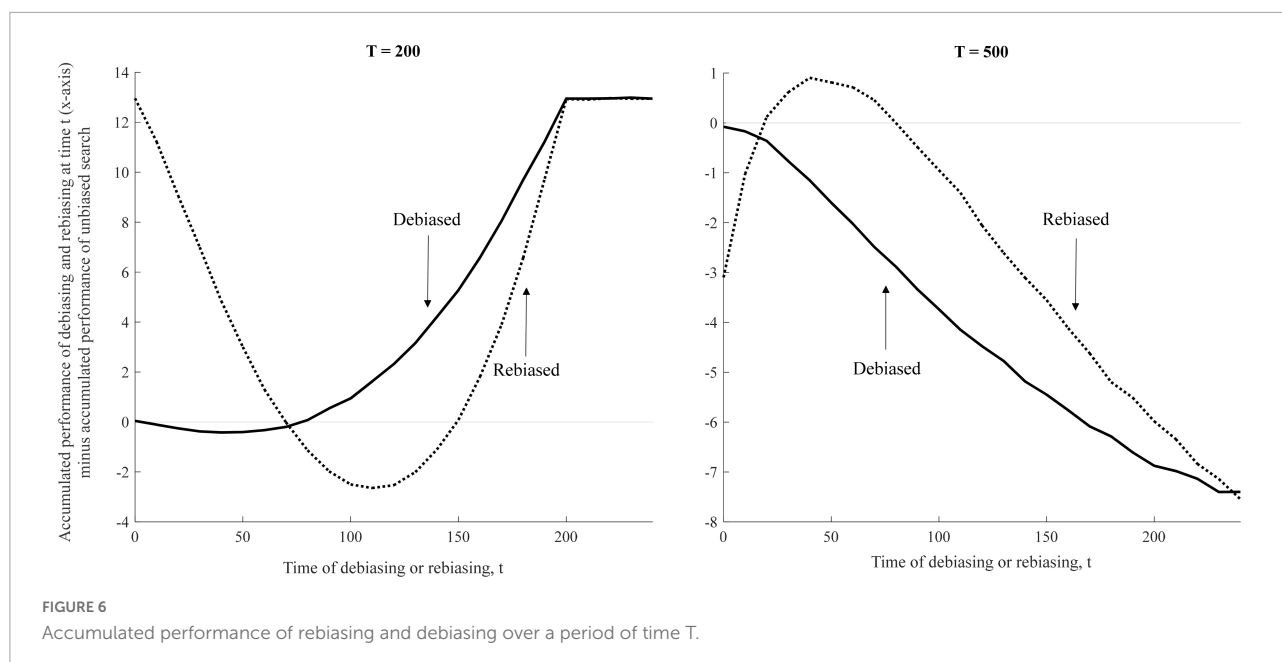
coordinates approximately (50, 2.5) means that rebiasing at $t = 50$ in a setting with significant time pressure leads to the overall gain of approximately 2.5 performance scores over the entire period ($T = 200$).

Figure 6 shows that rebiasing (and not debiasing) can be a superior intervention. With short or moderate time spans in a given setting ($T = 500$), agents benefit from periodically changing their biases. In other words, if human decision makers have a sufficiently limited time to solve a certain recombination problem, i.e., if they have relatively few trial attempts, rebased search may be their optimal form of behavior.

Strikingly, although debiasing occasionally outperforms rebiasing, it is never the dominant approach. Debiasing is always dominated either by continuously unbiased or by rebased search. When it comes to recombination problems that involve active trial and errors, organizations should not seek to debias their decision makers. In fact, they may want to do the exact opposite and seek to rebias organizational decisions. This observation, unique to the present research, has important implications for how we manage human biases that originate in our less deliberate cognitive processes.

Discussion

System 1 automatic evaluations are endemic to human mental functioning, and as some have argued may contribute



to our intelligence. Yet because of them, our specific judgments are often deeply biased. Arbitrary signals activate our automatic preferences and make us gravitate toward some options even before we know how good or bad they truly are. This tendency may undermine the quality of any single choice. At the same time, it is so fast and effortless that over populations of choices it may prove to be useful and adaptive (e.g., Gigerenzer and Goldstein, 1996; Gigerenzer and Todd, 1999; Bernardo and Welch, 2001; Johnson and Fowler, 2011). Drawing on this prior work, we find that biases improve decision maker's performance over a sequence of choices. As we illustrate, System 1 biases serve as a cognitive tool regulating excess experimentation, producing substantial benefits. Strikingly, this benefit of bias occurs even when there is no correlation between the variable of interest and the bias-generating variable. Automatic biases should be even more useful, and return value for longer, when they map closely onto environmental regularities (Gigerenzer and Todd, 1999).

In and of itself, this effect parallels other evolutionary advantages. But when paired with our present-day self-awareness and psychological toolkit, it offers the possibility of uncovering value beyond that of survival. Changing a bias, including debiasing, comes with a major short-term penalty: there is an immediate and profound decline in expected performance. However, the immediate disadvantage of changing biases are outweighed by the long-run benefits. Contrary to what might be anticipated, we find that organizations can most benefit by periodically reversing the biases of their decision makers. In complex settings with limited available time, a dominant strategy can be to rebias, in other words to strategically shift the overall decision making bias to its precise opposite. This provides a novel perspective on managing biases as previous

work in experimental settings has focused almost exclusively on debiasing: in other words the reduction, correction, and elimination of bias (e.g., Wilson and Brekke, 1994; Wilson et al., 2000). The present analyses identify rebiasing as an unconsidered but highly effective strategy for organizations. The benefits of rebiasing, however, emerge only if decision makers reverse their biases at a calculated moment in time, when the benefits of the initial automatic preference are no longer materializing.

Time is an essential variable in our analyses. First, we use time to show that biases in solving recombination problems that involve active trial and error are not uniformly negative or positive. In complex environments full of uncertainty, acting on automatic preferences is associated with short-term gains in performance and yet long-term costs. In addition, time can underlie an important variance in how effectively organizations manage biases. We show that biases should be managed, and time is a critical component in the effectiveness of this process. The optimal strategy may be to first leverage initial biases, and then engage in a timely rebiasing, adopting the exact opposite automatic preference. Our work thus answers calls to explore the role of intuition and affect in decision making over time (see George and Dane, 2016). *Via* the computational experiments used in the present research, we can point to the plausibility of phenomena that would be otherwise difficult to observe empirically (e.g., Epstein, 1999; Gray et al., 2014; Jung et al., 2021; Schaller and Muthukrishna, 2021).

Although, we cannot say if the observed differences will translate into meaningful effects in the real world—this requires empirical measurement—within the modeled universe, the effects are not as small as they might seem. Indeed, the gain

of biased search is ~ 0.119 , which is around 11%. Further, with regards to performance in highly competitive environments, even small differences can prove crucial. Seemingly minor discrepancies in outcomes accumulate over time (Hardy et al., 2022) and may provide key advantages over rivals, especially in winner take all competition formats. Consider a rivalry between two firms, in which company A achieving a certain market share will drive company B out of the market entirely and vice versa. In such a scenario, real-world differences far less than 11% could prove decisive.

A further important caveat concerns how the model time translates into the real-world time and whether such a translation is plausible. In other words, what is the meaning of 10, 100, or 1,000 search iterations in real-world settings? At this point, we cannot answer this question directly. But we can claim that a thousand iterations, or even more, may be well within many real-world time horizons over which performance plays out. To see this, consider the many decisions organizations make on a daily basis, i.e., decisions regarding personal remuneration, monetary and non-monetary rewards, product size, packaging, pricing, etc. All of these decisions seem to solve various problems and many of them take little to no time. At the same time, there is a combination of choices that will result in superior performance. Assuming that each possible combination of choices represents a single alternative in the model, by making day-to-day decisions, organizations effectively select different options. This means that a few years of routine organizational decision making can be realistically analogous to a thousand search iterations in the model. This, however, is only speculative at this point. Further empirical analyses of decision frequency in ecological contexts are needed to understand how the model time translates into the real-world time as well how organizations can use this to rebias productively.

Although judicious timing is clearly critical, another practical question is how feasible it is to debias or rebias decisions. Numerous experimental interventions have been developed in an effort to achieve unbiased or at least less biased decisions, with decidedly mixed success (Wilson and Brekke, 1994; Kahneman, 2003; Kahneman et al., 2011). Some interventions do attempt to push decision makers in the opposing direction, such as the consider-the-opposite strategy (Lord et al., 1984), or exhibiting pictures of widely admired Black Americans to reduce implicit prejudice (Dasgupta and Greenwald, 2001). However, the underlying goal is typically to shift decision makers toward neutrality, in other words to debias rather than rebias. For instance, Dasgupta and Greenwald (2001) presented White American research participants with photographs of Dr. Martin Luther King Jr. in the hopes of reducing their implicit preference for White over Black, not to create a bias against Whites. With regard to rebiasing at the individual level, there is the possibility of using framing to activate alternative automatic preferences (e.g., directly opposed

values both endorsed by the same person, such as group loyalty vs. merit; Haidt, 2001; Chong and Druckman, 2007). A more pragmatic and sustainable option, readily available to most organizations, is to switch the key decision makers to persons already known to hold the opposite automatic inclinations. For example, an organization that senses it is no longer reaping the benefits of its initial automatic preferences and needs to re-bias might change their leadership team to executives with directly contrary automatic biases. Re-biasing, however, would not be advisable in cases where the initial bias maps closely on to environmental regularities, as often happens in the natural world (e.g., wild animals relying on predictive cues to identify predators and prey in their natural habitat). Yet, in the turbulent environments faced by many contemporary organizations, well-timed reversals in leadership approach could prove advantageous.

Consider an example of a football team. From the perspective of the coach, choosing the right players is a standard problem that requires trial and error. While searching for an efficient solution to this problem, the coach may automatically discard some options. For example, the coach may intuitively reject those alternatives that do not favor players with whom the coach has friendly relationships. However, should this coach be removed after a time, her or his successor is likely to already hold or shortly form a different pattern of liking and disliking toward the players. A change of the key decision maker, therefore, represents a basic instrument that can lead to a change in the automatic evaluations, or rebiasing, at the organizational level.

Our model indicates that the success of a debiasing or rebiasing intervention is contingent on intervening at the correct moment. But how can an individual or organization determine when that moment is, or in other words, where they are currently situated in the performance curve? We conjecture that an organization can leverage its traditional performance indicators to get a sense its performance has dropped substantially and is on a downward trajectory from earlier time periods relative to peers. If so, this suggests they could now benefit from a change in automatic decision tendencies at the top. Our results highlight to an organization that is underperforming relative to its comparative performance in the past, and decides they need a significant change, that rebiasing may benefit them more than debiasing.

Previous work has pointed to the possibly positive and adaptive role of biases (e.g., Gigerenzer and Todd, 1999; Johnson and Fowler, 2011). Building on this idea, we use simulations to capture the temporal dimension long under-recognized in the experimental literature. By doing so, we analyze the lifecycles of biases and demonstrate that time is an important factor in managing them. Notably, our longitudinal pattern is distinct, but also non-contradictory, to what scholars studying fast and frugal heuristics have previously theorized. Specifically, they suggest biases that lead to errors in one-shot laboratory experiments can be adaptive in the long

term in complex naturalistic environments. In contrast, our simulations capture situations in which biases are beneficial in the short term but hurt performance in the long term—unless the decision making agent rebiasing itself at an opportune moment. Although this argument is substantially different, it does not contradict the existing theories. Like Gigerenzer and colleagues, we argue that biases can be adaptive over multiple choices. However, we further suggest that this effect is non-monotone and may reverse over time. Organizations—unlike individuals—possess instruments to calibrate and manipulate biases, such as changing decision-making processes, redesigning organizational structures, or simply replacing key decision makers entirely (Christensen and Knudsen, 2010). That is, organizations have structural and contextual means to alter the effective biasedness of their decisions, and therefore can proactively and profitably manage their effects.

Data availability statement

The complete data and code are posted on the Open Science Framework at https://osf.io/sypn2/?view_only=1b00c0d2dc964bafadf10215bfca4743.

References

- Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychol. Bull.* 110, 486–498. doi: 10.1037/0033-2909.110.3.486
- Baron, J. (2000). *Thinking and deciding*. Cambridge: Cambridge University Press.
- Bernardo, A. E., and Welch, I. (2001). On the evolution of overconfidence and entrepreneurs. *J. Econ. Manage. Strategy* 10, 301–330. doi: 10.1162/105864001316907964
- Chong, D., and Druckman, J. N. (2007). Framing theory. *Annu. Rev. Polit. Sci.* 10, 103–126. doi: 10.1146/annurev.polisci.10.072805.103054
- Christensen, M., and Knudsen, T. (2010). Design of decision-making organizations. *Manage. Sci.* 56, 71–89. doi: 10.1287/mnsc.1090.1096
- Dasgupta, N., and Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *J. Pers. Soc. Psychol.* 81, 800–814. doi: 10.1037/0022-3514.81.5.800
- Denrell, J., and March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organ. Sci.* 12, 523–538. doi: 10.1287/orsc.12.5.523.10092
- Duckworth, K. L., Bargh, J. A., Garcia, M., and Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychol. Sci.* 13, 513–519. doi: 10.1111/1467-9280.00490
- Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity* 4, 41–60. doi: 10.1002/(SICI)1099-0526(199905/06)4:5<41::AID-CPLX9>3.0.CO;2-F
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.* 59, 255–278. doi: 10.1146/annurev.psych.59.103006.093629
- Evans, J. S. B., and Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspect. Psychol. Sci.* 8, 223–241. doi: 10.1177/1745691612460685
- Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cogn. Emot.* 15, 115–141. doi: 10.1080/02699930125908
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., and Kardes, F. R. (1986). On the automatic activation of attitudes. *J. Pers. Soc. Psychol.* 50, 229–238. doi: 10.1037/0022-3514.50.2.229
- Fiori, S. (2011). Forms of bounded rationality: The reception and redefinition of Herbert A. Simon's perspective. *Rev. Polit. Econ.* 23, 587–612. doi: 10.1080/09538259.2011.611624
- George, J. M., and Dane, E. (2016). Affect, emotion, and decision making. *Organ. Behav. Hum. Decis. Process.* 136, 47–55. doi: 10.1016/j.obhdp.2016.06.004
- Gigerenzer, G., and Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychol. Rev.* 103, 650–669. doi: 10.1037/0033-295X.103.4.650
- Gigerenzer, G., and Todd, P. M. (1999). “Fast and frugal heuristics: The adaptive toolbox,” in *Simple heuristics that make us smart*, eds G. Gigerenzer, P. M. Todd, and The ABC Research Group (Oxford: Oxford University Press), 3–34.
- Gray, K., Rand, D. G., Ert, E., Lewis, K., Hershman, S., and Norton, M. I. (2014). The emergence of “us and them” in 80 lines of code: Modeling group genesis in homogeneous populations. *Psychol. Sci.* 25, 982–990. doi: 10.1177/0956797614521816
- Greenwald, A. G., and Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychol. Rev.* 102, 4–27. doi: 10.1037/0033-295X.102.1.4
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychol. Rev.* 108, 814–834. doi: 10.1037/0033-295X.108.4.814
- Hardy, J. H. III, Tey, K. S., Cyrus-Lai, W., Martell, R. F., Olstad, A., and Uhlmann, E. L. (2022). Bias in context: Small biases in hiring evaluations have big consequences. *J. Manage.* 48, 657–692. doi: 10.1177/0149206320982654
- Haselton, M. G., and Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Pers. Soc. Psychol. Rev.* 10, 47–66. doi: 10.1207/s15327957pspr1001_3

Author contributions

AK and EU ideated the project and wrote the manuscript. AK conceptualized the model and performed the analyses. Both authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Hutchinson, J. M., and Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behav. Process.* 69, 97–124. doi: 10.1016/j.beproc.2005.02.019
- Inbar, Y., Cone, J., and Gilovich, T. (2010). People's intuitions about intuitive insight and intuitive choice. *J. Pers. Soc. Psychol.* 99, 232–247. doi: 10.1037/a0020215
- Johnson, D. D., and Fowler, J. H. (2011). The evolution of overconfidence. *Nature* 477, 317–320. doi: 10.1038/nature10384
- Jung, J., Bramson, A., Crano, W. D., Page, S. E., and Miller, J. H. (2021). Cultural drift, indirect minority influence, network structure, and their impacts on cultural change and diversity. *Am. Psychol.* 76, 1039–1053. doi: 10.1037/amp0000844
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *Am. Psychol.* 58, 697–720. doi: 10.1037/0003-066X.58.9.697
- Kahneman, D., and Frederick, S. (2002). "Representativeness revisited: Attribute substitution in intuitive judgment," in *Heuristics and biases: The psychology of intuitive judgment*, eds T. Gilovich, D. Griffin, and D. Kahneman (Cambridge: Cambridge University Press), 49–81. doi: 10.1017/CBO9780511808098.004
- Kahneman, D., and Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *Am. Psychol.* 64, 515–526. doi: 10.1037/a0016755
- Kahneman, D., Lovallo, D., and Sibony, O. (2011). Before you make that big decision. *Harv. Bus. Rev.* 89, 50–60.
- Kauffman, S., and Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* 128, 11–45. doi: 10.1016/S0022-5193(87)80029-2
- Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. New York, NY: Oxford University Press. doi: 10.1007/978-94-015-8054-0_8
- Khatri, N., and Ng, H. A. (2000). The role of intuition in strategic decision making. *Hum. Relat.* 53, 57–86. doi: 10.1177/0018726700531004
- Kramer, R. M., Newton, E., and Pommerenke, P. L. (1993). Self-enhancement biases and negotiator judgment: Effects of self-esteem and mood. *Organ. Behav. Hum. Decis. Process.* 56, 110–133. doi: 10.1006/obhd.1993.1047
- Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organ. Behav. Hum. Decis. Process.* 75, 23–55. doi: 10.1006/obhd.1998.2781
- Lant, T. K. (1992). Aspiration level adaptation: An empirical exploration. *Manage. Sci.* 38, 623–644. doi: 10.1287/mnsc.38.5.623
- Levinthal, D., and March, J. G. (1981). A model of adaptive organizational search. *J. Econ. Behav. Organ.* 2, 307–333. doi: 10.1016/0167-2681(81)90012-3
- Levinthal, D. A. (1997). Adaptation on rugged landscapes. *Manage. Sci.* 43, 934–950. doi: 10.1287/mnsc.43.7.934
- Lord, C. G., Lepper, M. R., and Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *J. Pers. Soc. Psychol.* 47, 1231–1243. doi: 10.1037/0022-3514.47.6.1231
- Lyles, M. A., and Thomas, H. (1988). Strategic problem formulation: Biases and assumptions embedded in alternative decision-making models. *J. Manage. Stud.* 25, 131–145. doi: 10.1111/j.1467-6486.1988.tb00028.x
- Maddux, W. W., Williams, E., Swaab, R., and Betania, T. (2014). *Ricardo Semler: A revolutionary model of leadership*. Case study. Boston, MA: Harvard Business Publishing.
- Marshall, J. A., Trimmer, P. C., Houston, A. I., and McNamara, J. M. (2013). On evolutionary explanations of cognitive biases. *Trends Ecol. Evol.* 28, 469–473. doi: 10.1016/j.tree.2013.05.013
- Miller, C. C., and Ireland, R. D. (2005). Intuition in strategic decision making: Friend or foe in the fast-paced 21st century? *Acad. Manage. Perspect.* 19, 19–30. doi: 10.5465/ame.2005.15841948
- Newell, A., and Simon, H. A. (2007). "Computer science as empirical inquiry: Symbols and search," in *Proceedings of the ACM turing award lectures* (New York, NY: Association for Computing Machinery). 113–126. doi: 10.1145/1283920.1283930
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* 2, 175–220. doi: 10.1037/1089-2680.2.2.175
- Nisbett, R. E., and Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychol. Rev.* 84, 231–259. doi: 10.1037/0033-295X.84.3.231
- Raghubir, P., and Valenzuela, A. (2006). Center-of-inattention: Position biases in decision-making. *Organ. Behav. Hum. Decis. Process.* 99, 66–80. doi: 10.1016/j.obhdp.2005.06.001
- Reitzig, M., and Sorenson, O. (2013). Biases in the selection stage of bottom-up strategy formulation. *Strateg. Manage. J.* 34, 782–799. doi: 10.1002/smj.2047
- Rivkin, J. W. (2000). Imitation of complex strategies. *Manage. Sci.* 46, 824–844. doi: 10.1287/mnsc.46.6.824.11940
- Schaller, M., and Muthukrishna, M. (2021). Modeling cultural change: Computational models of interpersonal influence dynamics can yield new insights about how cultures change, which cultures change more rapidly than others, and why. *Am. Psychol.* 76, 1027–1038. doi: 10.1037/amp0000797
- Schwenk, C. H. (1986). Information, cognitive biases, and commitment to a course of action. *Acad. Manage. Rev.* 11, 298–310. doi: 10.5465/amr.1986.4283106
- Schwenk, C. R. (1984). Cognitive simplification processes in strategic decision-making. *Strateg. Manage. J.* 5, 111–128. doi: 10.1002/smj.4250050203
- Scott, K. A., and Brown, D. J. (2006). Female first, leader second? Gender bias in the encoding of leadership behavior. *Organ. Behav. Hum. Decis. Process.* 101, 230–242. doi: 10.1016/j.obhdp.2006.06.002
- Simon, H. A. (1955). A behavioral model of rational choice. *Q. J. Econ.* 69, 99–118. doi: 10.2307/1884852
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychol. Rev.* 63, 129–138. doi: 10.1037/h0042769
- Simon, H. A. (1990). Invariants of human behavior. *Annu. Rev. Psychol.* 41, 1–20. doi: 10.1146/annurev.ps.41.020190.000245
- Slooman, S. A. (1996). The empirical case for two systems of reasoning. *Psychol. Bull.* 119, 3–22. doi: 10.1037/0033-2909.119.1.3
- Slovic, P., Finucane, M., Peters, E., and MacGregor, D. G. (2002). Rational actors or rational fools: Implications of the affect heuristic for behavioral economics. *J. Soc. Econ.* 31, 329–342. doi: 10.1016/S1053-5357(02)00174-9
- Stanovich, K. E., and West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behav. Brain Sci.* 23, 645–665. doi: 10.1017/S0140525X00003435
- Stone, D. N. (1994). Overconfidence in initial self-efficacy judgments: Effects on decision processes and performance. *Organ. Behav. Hum. Decis. Process.* 59, 452–474. doi: 10.1006/obhd.1994.1069
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124
- Volz, K. G., and von Cramon, D. Y. (2006). What neuroscience can tell about intuitive processes in the context of perceptual discovery. *J. Cogn. Neurosci.* 18, 2077–2087. doi: 10.1162/jocn.2006.18.12.2077
- Wilson, T. D., and Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychol. Bull.* 116, 117–142. doi: 10.1037/0033-2909.116.1.117
- Wilson, T. D., Lindsey, S., and Schooler, T. Y. (2000). A model of dual attitudes. *Psychol. Rev.* 107, 101–126. doi: 10.1037/0033-295X.107.1.101
- Winter, S. G., Cattani, G., and Dorsch, A. (2007). The value of moderate obsession: Insights from a new model of organizational search. *Organ. Sci.* 18, 403–419. doi: 10.1287/orsc.1070.0273
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *Am. Psychol.* 35, 151–175. doi: 10.1037/0003-066X.35.2.151



OPEN ACCESS

EDITED BY
Riccardo Viale,
University of Milano-Bicocca, Italy

REVIEWED BY
Sihua Xu,
Shanghai International Studies
University, China
Rui Shi,
Yanshan University, China
Yue Qi,
Renmin University of China, China

*CORRESPONDENCE
Junchen Shang
junchen_20081@163.com

SPECIALTY SECTION
This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

RECEIVED 08 August 2022
ACCEPTED 22 September 2022
PUBLISHED 13 October 2022

CITATION
Shang J and Liu Z (2022) The “beauty premium” effect of voice attractiveness of long speech sounds in outcome-evaluation event-related potentials in a trust game.
Front. Psychol. 13:1010457.
doi: 10.3389/fpsyg.2022.1010457

COPYRIGHT
© 2022 Shang and Liu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The “beauty premium” effect of voice attractiveness of long speech sounds in outcome-evaluation event-related potentials in a trust game

Junchen Shang^{1*} and Zhihui Liu²

¹Department of Medical Humanities, School of Humanities, Southeast University, Nanjing, China,

²School of Psychology, Liaoning Normal University, Dalian, China

Previous research suggested that people with attractive voices had an advantage in economic games, even if the voices were only presented for 400 ms. The present study investigated the influence of voice attractiveness on the cooperative trust behavior with longer exposure times to the voices. Event-related potentials (ERPs) were recorded during the feedback outcome evaluation. Participants heard a voice of the partner for 2,040 ms and decided whether to invest to the partner for a possibility to gain more money. The results showed that participants made more invest choices to the attractive partners, replicating the “beauty premium” effect of the attractive voices. Moreover, participants were more likely to invest to male partners. The ERP analysis for the outcome showed that the difference waves of feedback-related negativity (FRN) amplitude were smaller in the attractive voice condition than in the unattractive voice condition, suggesting that the rewarding effect of attractive voices weakened the frustrating feelings of the loss. In sum, the present study confirms that attractive voices with longer presentation durations facilitate cooperative behavior and modulate the processing of feedback evaluations.

KEYWORDS

voice attractiveness, duration, trust game, cooperative behavior, the “beauty premium” effect

Introduction

With the development of Internet communication technology, people use voice more often for communication with social network software, which can improve work efficiency. If you want to contact someone, you can hear the person's voice first not only by a telephone call but also by mobile apps. Face-to-face communication is eliminated, making social communication more convenient. The social signals conveyed by voices have important impact on daily life. Many mobile apps such as navigation software have begun to use voice for human-computer interaction. Human voice is applied more and more often in the media. The "sound industry" is booming, such as the Chinese TV show "The Sound," audiobooks, and radio apps. In literature, there are also depictions of personality traits conveyed by human voice. In "A Dream of the Red Mansions," one of the most famous literature in China, Wang Xifeng's hearty laughter and unfettered speech showed her shrewd, strong, pungent, and vicious traits even if she did not show up. Voice attractiveness is the extent to which the voice of the speaker can induce a positive and pleasant emotional experience and attract other people. Zuckerman and Driver (1989) revealed the "what sounds beautiful is good" stereotype, such that voice attractiveness influenced impressions of personality (Zuckerman et al., 1990). A recent study (Wu et al., 2021) confirmed that voice attractiveness was related to the speaker's personality, such as capability and approachability dimensions in Chinese culture. Voice attractiveness also plays an important role in evolution because it correlates with traits reflecting hormone levels and health (Groyecka et al., 2017). In addition to behavioral research, some studies provided evidence for the neural underpinnings of voice attractiveness processing with functional magnetic resonance imaging (fMRI) and event-related potentials (ERPs). When participants passively listened to voices in a pure tone detection task, the activities in the higher level auditory cortex and inferior prefrontal regions were correlated with voice attractiveness (Bestelmeyer et al., 2012). Moreover, compared to happiness and age judgments of voices, voice attractiveness judgments activated the bilateral inferior parietal cortex, and the dorsomedial prefrontal cortex extending into the perigenual anterior cingulate cortex (Hensel et al., 2015). An ERP study (Zhang et al., 2020) reported that attractive voices elicited larger N1, smaller P2, and larger P3 and late positive component (LPC) amplitudes than unattractive voices in an attractiveness rating task. Since attractive faces also evoked larger LPC than unattractive faces (Ma and Hu, 2015; Ma et al., 2015, 2017), and dorsomedial prefrontal cortex was also involved in facial attractiveness processing (Hensel et al., 2015), voice attractiveness may have a reward effect as facial attractiveness (Shang and Liu, 2022).

Research revealed that people who own attractive voices had some advantages in economic activities. A recent study (Shang et al., 2021) suggested that males' voice attractiveness affected

responders' fairness considerations during the ultimatum game even though the voices were only presented for 400 ms. More offers were accepted from proposers who had attractive voices in a two-person ultimatum game. Moreover, voice attractiveness of the third player also influenced decision-making in a three-person ultimatum game which included a proposer, a responder, and a powerless third player. Participants (responders) accepted more offers if the third player had an attractive voice even though the offer was unfair for them but fair for the third player. The above findings confirmed that voice attractiveness induced the "beauty premium" effect. Shang and Liu (2022) further explored the influence of vocal attractiveness on cooperative behavior in a trust game using similar voice stimuli as Shang et al. (2021). The participants made more invest choices to the partners with attractive voices. However, vocal attractiveness did not impact the feedback-related negativity (FRN) related to the outcome which is an important component in the economic decision-making.

Despite previous research showing the influence of voice attractiveness on decision-making in ultimatum game and trust game (Shang et al., 2021; Shang and Liu, 2022), the exposure time to the voices in these studies is only 400 ms. Research showed that increased exposure time resulted in more differentiated trait inferences of an unfamiliar face although people can form impressions (such as attractiveness) even after a 100-ms exposure time (Willis and Todorov, 2006). Moreover, the judgment of facial attractiveness correlates with the judgment of voice attractiveness (Saxton et al., 2009; Hughes and Miller, 2016). It is possible that "beauty premium" effect would be different between long speech voices and short speech voices. In addition, voices last for much longer time in daily life. Some studies (Krumpholz et al., 2021, 2022) suggested that it took around 1 s for the stable judgment for voice attractiveness. This duration is much longer than the exposure time to the voices in decision-making research (Shang et al., 2021; Shang and Liu, 2022). However, it is unclear whether decision-making toward voices with exposure time longer than 1 s would be different with previous studies (Shang et al., 2021; Shang and Liu, 2022).

Especially, the FRN amplitude, which represents the brain sensitivity to the failure of decision-making, has been well documented (e.g., Xu et al., 2020a,b). It is considered as the brain response to positive and negative outcomes, such as gain and loss (Gehring and Willoughby, 2002; Yeung and Sanfey, 2004; Martín, 2012; Ma and Hu, 2015; Ma et al., 2017). For example, unfair offers elicited larger FRN than fair offers when the partner's face was unattractive in the Ultimatum Game, whereas there was no difference in attractive-face condition (Ma et al., 2015, 2017). It is possible that voice attractiveness would induce the same effect in trust game. The reason that Shang and Liu (2022) did not find the influence of vocal attractiveness on FRN may be because of the short duration of voices. Therefore, the present study investigated the influence of voice attractiveness

of long speech voices (lasted for 2,040 ms) on the investment behavior and the neural underpinnings for outcome evaluation in a trust game (Shang and Liu, 2022) using ERPs. The present study predicted that participants would invest more money to the attractive partners. It is also predicted that the FRN effect in the attractive voice condition would be different with that in the unattractive voice condition.

Materials and methods

Participants

Using G* Power v. 3.1.9.6 (Faul et al., 2007), the sample size was determined based on the sample size of previous ERP research about voice attractiveness and trust game (Shang and Liu, 2022). Given the power of a statistical test of 0.95, and the effect size of 0.25, 64 students (33 female participants, $M_{age} = 20.94$ years, $SD = 2.79$ years) at Liaoning Normal University participated in this study. The participants all had normal or corrected to normal vision and normal hearing. All participants were physically healthy and had no neurological damage. Each participant was paid a certain money reward after the experiment. For this research, we obtained approval from the ethics committee of Liaoning Normal University and written informed consent from each participant before the experiment.

Design and materials

This experiment employed a within-subject design with voice attractiveness (attractive vs. unattractive) and voice gender (female participant vs. male participant) as within-subject factors. The ratio of investment and the ERP amplitudes (FRN) were the dependent variables.

To be comparable with Shang and Liu (2022), the neutral vowels were used. Voice stimuli were chosen from Ferdenzi et al. (2015). There were 111 voice samples (61 female voices, 50 male voices, $M_{age} = 22.9$ years, $SD = 4.3$ years). Each voice sample included three neutral vowel syllables (/i/, /a/, /ou/). The duration of all voice recordings is adjusted to 2,040 ms, by using Praat software v.5.3.85. The sound intensity is adjusted to 70 dB. Fifty-eight participants (18 male participants, $M_{age} = 21.60$ years, $SD = 2.43$ years) who did not take part in the ERP experiment were asked to rate the attractiveness of the voices on a seven-point Likert scale (from 1 = “very unattractive” to 7 = “very attractive”).

According to the mean rating value of each voice across the 58 participants, we chose 30 female voices (15 most attractive voices and 15 most unattractive voices) and 30 male voices (15 most attractive voices and 15 most unattractive voices) for use as partners in the trust game for the ERP experiment. The attractiveness ratings of the four categories

of voices were compared using a two-way ANOVA. The voice attractiveness was significantly different ($F_{(1,56)} = 362.55$, $p < 0.001$, $\eta_p^2 = 0.87$, 95%CI [0.79, 0.90]). The main effect of voice gender was not significant [$F_{(1,56)} = 0.40$, $p = 0.531$]. The interaction between voice attractiveness and voice gender was not significant [$F_{(1,56)} = 2.15$, $p = 0.148$]. The attractive female voices ($M = 5.15$, $SD = 0.31$) were rated as more attractive than unattractive female voices ($M = 2.81$, $SD = 0.49$). The attractive male voices ($M = 4.91$, $SD = 0.38$) were rated as more attractive than unattractive male voices ($M = 2.91$, $SD = 0.55$).

The acoustic parameters of attractive voices and unattractive voices were calculated using Praat software and were compared using paired *t*-tests (as shown in Tables 1, 2). Previous research suggested that lower-pitched male voices are more attractive than higher-pitched male voices (Collins, 2000; Feinberg et al., 2005; Jones et al., 2010; Re et al., 2012). Also, higher-pitched female voices are more attractive than lower-pitched female voices (Feinberg et al., 2008; Zheng et al., 2020). Almost consistent with prior studies, differences in voice attractiveness for the present experiment were accompanied by differences in acoustic parameters. F0 of unattractive male voices was higher than attractive male voices. Moreover, f3 of attractive male voices was higher than unattractive male voices. F0 and f4 of unattractive female voices were lower than attractive female voices. In attractive and unattractive condition, F0, f3, f4, Df, Pf, and HNR of female voices were higher than male voices. In attractive condition, the jitter of female voices was lower than male voices. In addition, the shimmer was lower in female voices.

Procedure

Participants comfortably completed the experiment individually in a sound-attenuated lab. A chin rest was used to eliminate head movements. The voices were presented binaurally over Sennheiser headphones. Before the experiment, we adjusted the loudness for each participant for the comfortableness. The instructions and measurements were controlled by E-prime version 2.

This experiment employed the same trust game in Shang and Liu (2022), except for the voice samples and duration of voices (see Figure 1). We clarify the procedure succinctly. In the beginning, there were eight practice trials containing the voices which were not shown in the formal experiment. First, participants got ¥20 to play the game. They were asked to decide whether to invest to a “real” partner (who was actually fictional and represented by an attractive or an unattractive voice) in each trial for a chance to earn the real monetary remuneration as the final rewards they gained in the game. In each trial, a central fixation cross was first shown for 1,000 ms. Then, a voice of the partner was presented for 2,040 ms. Afterward, two sentences “invest ¥0.5” and

TABLE 1 Means (and standard deviations) and acoustic differences between attractive and unattractive voices.

	Female voices					Male voices				
	Attractive voices	Unattractive voices	<i>t</i>	<i>p</i>	Cohen's <i>d</i>	Attractive voices	Unattractive voices	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
F0	252.51 (18.69)	229.72 (31.44)	2.41	0.023	0.88	132.65 (15.50)	151.02 (25.77)	-2.37	0.025	-0.86
f1	747.77 (58.11)	701.19 (97.84)	1.59	0.124	0.58	712.56 (107.99)	653.57 (78.97)	1.71	0.099	0.62
f2	1,781.70 (79.23)	1,756.65 (105.50)	0.74	0.468	0.27	1,783.02 (121.58)	1,727.10 (78.07)	1.50	0.145	0.55
f3	3,055.14 (97.62)	3,032.83 (92.25)	0.64	0.525	0.24	2,982.77 (90.17)	2,900.92 (101.09)	2.34	0.027	0.86
f4	4,192.94 (73.43)	4,078.56 (130.68)	2.96	0.006	1.08	3,981.34 (171.73)	3,945.07 (119.19)	0.67	0.507	0.25
Df	1,148.39 (33.06)	1,125.79 (31.98)	1.90	0.067	0.70	1,089.60 (40.10)	1,097.17 (32.55)	-0.57	0.575	-0.21
Pf	0.54 (0.44)	0.12 (0.83)	1.75	0.091	0.64	-0.05 (0.92)	-0.60 (0.63)	1.91	0.067	0.70
Jitter	1.94 (0.60)	1.87 (0.79)	0.28	0.782	0.10	2.55 (0.82)	2.37 (0.68)	0.66	0.516	0.24
Shimmer	7.47 (1.33)	8.41 (2.48)	-1.29	0.207	-0.47	10.82 (3.47)	10.84 (2.20)	-0.02	0.981	-0.01
HNR	14.05 (2.28)	13.87 (3.38)	0.18	0.862	0.06	10.00 (2.03)	11.07 (1.46)	-1.66	0.109	-0.61

F0, fundamental frequency in Hz; f1–f4, formant frequencies in Hz; Df, formant dispersion in Hz; Pf, formant position; jitter, variation of pitch in μ s; Shimmer, variation of energy in dB; HNR, harmonic-to-noise ratio in dB.

“keep ¥0.5” appeared on the screen. Half of the participants pressed the “F” key once they decided to invest and pressed the “J” key once they decided to keep ¥0.5. For the other half of participants, the response keys were counterbalanced. If the participants submitted the choice, the final decision would be shown for 1,000 ms. If the participant chose to invest, a blank screen was shown for 600–1,000 ms, and the partner would receive ¥2. The partner would either pay ¥1 to the participant or keep all of the rewards. Then, the feedback from the partner was presented for 1,000 ms. If participants refuse to invest, the current amount of money would not be changed. Finally, participants were asked to press the space key to start the next trial. Each voice was repeated eight times. Half times the voice was accompanied by gains, while the other half times it was accompanied by losses. The ERP experiment contained 480 trials presented in a pseudorandom order, whereas the participants were not told about the regularity.

Event-related potential recording and analysis

The electroencephalography (EEG) was continuously recorded from 64 scalp sites arranged on an elastic cap (Brain Products, GmbH, Germany). The sampling rate was 500 Hz. The ground electrode was on the cephalic (forehead) location. The vertical and horizontal electrooculogram (EOG) were recorded with two electrodes which were placed below and on the right side of the right eye. All electrode impedances were kept below 5 k Ω . The EEG signals were re-referenced offline to the average of the left and right mastoids. First, the EOG artifacts were corrected. Then, digital bandpass filtering was employed between 0.01 and 30 Hz. We applied an independent component analysis algorithm to correct EOG. Epochs, which contained EOG artifacts and amplifier clipping artifacts, were excluded before averaging. Other recording artifacts were also excluded when the EEG amplitudes exceeded $\pm 80 \mu$ V. The ERPs were extracted and segmented with time-locked signal averaging by adopting the time window initiated at -200 ms and stopped at 1,000 ms relative to the feedback stimuli onset.

The average amplitude of FRN differential waves (280–310 ms) was measured to investigate the ERP waves evoked by feedback stimuli in “investment” trials. Based on the methodology in previous research (Holroyd and Krigolson, 2007; Chen et al., 2012), the difference waves of FRN amplitude were calculated by subtracting the average amplitude of the gain ERP wave from the average amplitude of the loss ERP wave. Five electrode sites (Fz, FCz, Cz, CPz, and Pz) were selected. The FRN difference waves were separately calculated for two conditions: attractive voice-related FRN difference wave and unattractive voice-related

TABLE 2 Means (and standard deviations) and acoustic differences between female and male voices.

	Attractive voices					Unattractive voices				
	Female voices	Male voices	<i>t</i>	<i>p</i>	Cohen's <i>d</i>	Female voices	Male voices	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
F0	252.51 (18.69)	132.65 (15.50)	19.12	< 0.001	6.98	229.72 (31.44)	151.02 (25.77)	7.50	< 0.001	2.74
f1	747.77 (58.11)	712.56 (107.99)	1.11	0.276	0.41	701.19 (97.84)	653.57 (78.97)	1.47	0.154	0.54
f2	1,781.70 (79.23)	1,783.02 (121.58)	−0.04	0.972	−0.01	1,756.65 (105.50)	1,727.10 (78.07)	0.87	0.391	0.32
f3	3,055.14 (97.62)	2,982.77 (90.17)	2.11	0.044	0.77	3,032.83 (92.25)	2,900.92 (101.09)	3.73	< 0.001	1.36
f4	4,192.94 (73.43)	3,981.34 (171.73)	4.39	< 0.001	1.60	4,078.56 (130.68)	3,945.07 (119.19)	2.92	0.007	1.07
Df	1,148.39 (33.06)	1,089.60 (40.10)	4.38	< 0.001	1.60	1,125.79 (31.98)	1,097.17 (32.55)	2.43	0.022	0.89
Pf	0.54 (0.44)	−0.05 (0.92)	2.26	0.032	0.83	0.12 (0.83)	−0.60 (0.63)	2.67	0.012	0.98
Jitter	1.94 (0.60)	2.55 (0.82)	−2.32	0.028	−0.85	1.87 (0.79)	2.37 (0.68)	−1.87	0.072	−0.68
Shimmer	7.47 (1.33)	10.82 (3.47)	−3.50	0.002	−1.28	8.41 (2.48)	10.84 (2.20)	−2.85	0.008	−1.04
HNR	14.05 (2.28)	10.00 (2.03)	5.15	< 0.001	1.88	13.87 (3.38)	11.07 (1.46)	2.95	0.006	1.08

F0, fundamental frequency in Hz; f1–f4, formant frequencies in Hz; Df, formant dispersion in Hz; Pf, formant position; Jitter, variation of pitch in μ s; Shimmer, variation of energy in dB; HNR, harmonic-to-noise ratio in dB.

FRN difference wave. We did not examine the voice gender effect since there were not enough artifact-free trials in each condition [we used criteria of Shang and Liu (2022) that at least 30 valid trials per condition]. Six participants were excluded, and there were 58 valid participants (32 female participants) in the analysis of FRN differential waves. A two-way repeated measures ANOVA was conducted on FRN differential amplitudes including voice attractiveness (attractive vs. unattractive) and electrode sites (Fz, FCz, Cz, CPz, and Pz) as within-subject factors. We adopted Greenhouse–Geisser corrections when the results violated the spherical assumption. All multiple comparisons were Bonferroni-corrected.

Results

Behavioral results

We calculated the percentage of average percentage of invest choices in attractive and unattractive voice conditions, respectively, as the ratio of investment. We conducted a 2 (voice attractiveness: attractive vs. unattractive) \times 2 (voice gender: female participant vs. male participant) repeated measures ANOVA on the ratio of investment.

This test yielded a significant effect of voice attractiveness ($F_{(1,63)} = 63.47$, $p < 0.001$, $\eta_p^2 = 0.50$, 95%CI [0.32, 0.62]). Participants were more willing to invest to attractive partners ($M = 0.66$, $SD = 0.14$) than unattractive partners ($M = 0.55$, $SD = 0.15$). There was also a significant effect of voice gender ($F_{(1,63)} = 9.62$, $p = 0.003$, $\eta_p^2 = 0.13$, 95%CI [0.02, 0.29]), indicating that participants were more likely to cooperate with male partners ($M = 0.63$, $SD = 0.13$) than female partners ($M = 0.58$, $SD = 0.16$). The interaction between voice attractiveness

and voice gender was not significant [$F_{(1,63)} = 0.15$, $p = 0.698$].

Event-related potential results: The feedback-related negativity difference wave (280–310 ms)

The results showed a significant effect of voice attractiveness ($F_{(1,57)} = 4.33$, $p = 0.042$, $\eta_p^2 = 0.07$, 95%CI [0.00, 0.22]). Specifically, a larger difference wave of FRN amplitude was elicited by unattractive voices than attractive voices. The main effect of electrode sites was significant ($F_{(1.39,79.25)} = 13.97$, $p < 0.001$, $\eta_p^2 = 0.20$, 95%CI [0.06, 0.34]). *Post-hoc* comparisons showed that a smaller FRN difference wave was elicited in parietal region than the other regions ($ps < 0.003$). A larger FRN difference wave was elicited in fronto-central and central regions rather than central-parietal region ($ps \leq 0.006$). The interaction between voice attractiveness and electrode sites was not significant [$F_{(1.67,94.88)} = 0.50$, $p = 0.576$] (Figure 2).

Discussion

The current research investigated the effect of voice attractiveness on investments in a trust game when the duration of partner's voice was 2,040 ms. The behavioral responses suggested that participants made more invest choices in the attractive partner condition than in the unattractive partner condition. The finding was in line with prior studies which revealed the “beauty premium” effect of voice attractiveness in the ultimatum games and a trust game using short voices which lasted for 400 ms (Shang et al., 2021; Shang and Liu, 2022). The result was also similar with the “beauty premium”

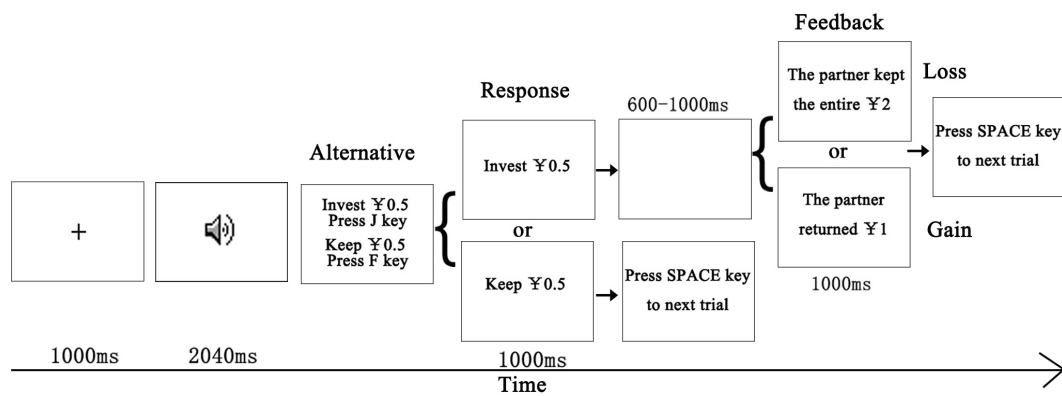


FIGURE 1

Trial procedure of the trust game task. A central fixation cross was presented followed by a partner's voice presented for 2,040 ms. The participant was asked to decide whether to invest ¥0.5 or refuse to invest ¥0.5. The decision would be presented for 1,000 ms. If the participant decided to invest, the partner would gain ¥2. Then, the partner either gave ¥1 back to the participant or kept the entire ¥2. If the participant refused to invest ¥0.5, the amount of money would remain unchanged.

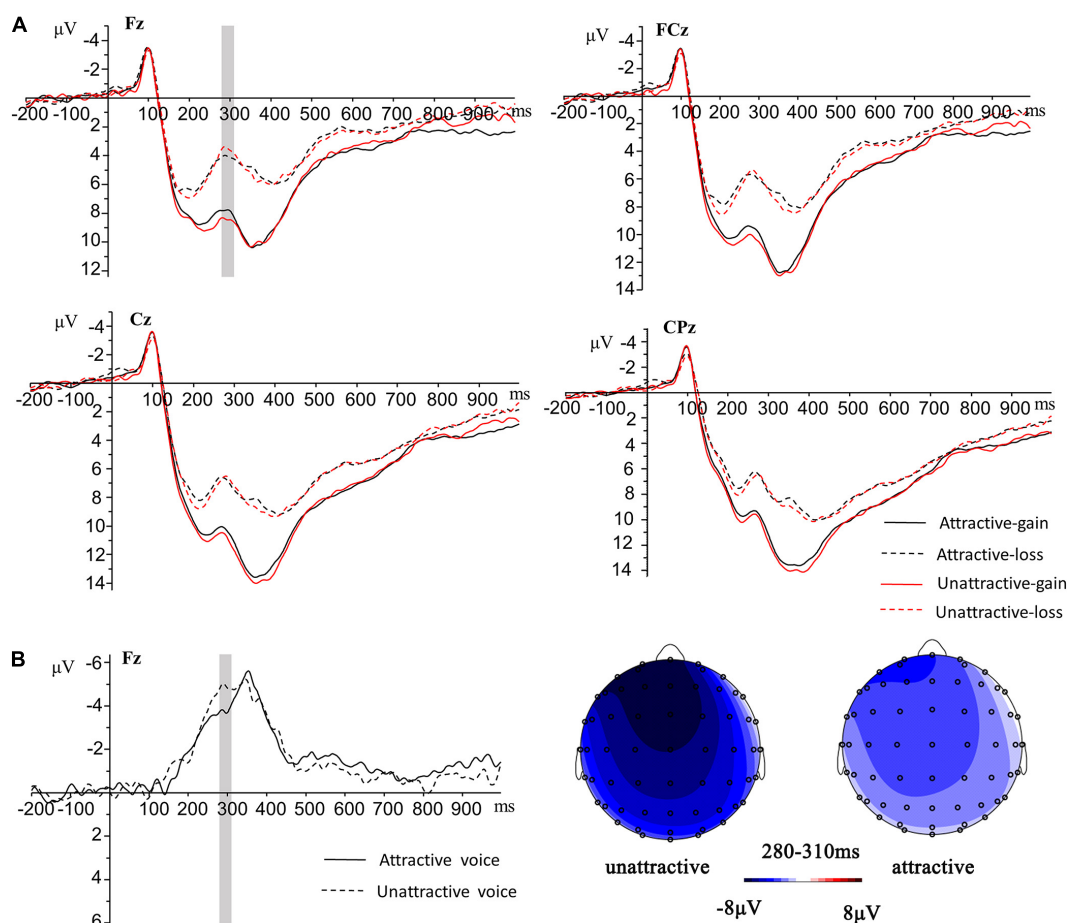


FIGURE 2

(A) Grand average event-related potentials (ERPs) induced by feedback of gain and loss at four representative electrodes in the attractive and unattractive voice conditions. (B) Topography of scalp distribution and differential waves generated by gain and loss in the attractive and unattractive voice conditions.

effect of facial attractiveness in decision-making (e.g., Ma et al., 2015, 2017). Hensel et al. (2015) suggested that the dorsomedial prefrontal cortex which was activated by voice attractiveness also played an important role in processing of facial attractiveness, indicating that the brain regions related to voice attractiveness overlapped with those activated by facial attractiveness. The present study confirmed this by behavioral findings.

We also analyzed participants' decisions toward male partners and female partners. Specifically, participants made more cooperation choices (investments) to male partners compared to female partners. These results were inconsistent with previous research (Shang and Liu, 2022) which reported that people made more investments to female partners compared to male partners when the voices were unattractive. A possible explanation might be that objective and subjective judgments of voice traits can be different between short speech sounds and longer speech sounds (Pisanski and Feinberg, 2018; Krumpholz et al., 2022). Shang and Liu (2022) used short vowels, which may convey a first impression (Krumpholz et al., 2022). It is possible that gender influenced first impression of voice attractiveness. In the current study, each voice consisted of three vowels and lasted longer than 1 s. This duration enabled stable voice judgment (Krumpholz et al., 2021). Thus, the discrepancy of gender effect may be attributed to the exposure time of voices. Furthermore, the findings of present research supported previous research on facial attractiveness and decision-making, reporting that people allocated more money to male partners in an ultimatum game (Solnick and Schweitzer, 1999). This study further indicates that decision-making may be influenced by the gender of the partner with whom we interact even though only a voice was presented.

In addition, we analyzed participants' brain activities in outcome feedback evaluations. We assumed that Shang and Liu (2022) did not yield the FRN effect because of the short exposure times to voices. Consistent with our hypothesis, the present research showed that the different waves of FRN amplitude (loss ERP minus gain ERP elicited by the feedback) in the unattractive voice condition were larger than in the attractive voice condition. This may be interpreted as a reward effect of voice attractiveness, which could affect participants' fairness considerations and reduce their negative emotion toward loss even though the attractive partners did not return the reward. The findings also supported the beauty premium effect that participants may show more prosocial behaviors to partners with attractive voices (Shang et al., 2021). A similar FRN effect was reported by previous research about facial attractiveness and decision-making (Ma et al., 2017), such that unfair offers elicited larger FRN than fair offers in the unattractive face condition during an ultimatum game. Research showed that increased exposure time to a face may boost confidence in

impressions (Willis and Todorov, 2006). The discrepancy of FRN observed in the present study compared with Shang and Liu (2022) might also be interpreted as a boosted confidence in voice attractiveness judgments after a longer exposure time, since impressions of voice attractiveness correlated with impressions of facial attractiveness (Saxton et al., 2009; Hughes and Miller, 2016). Again, the present study provides more evidence for the beauty premium of voice attractiveness of longer speech sounds in a social economic game.

There were two limitations in the current study. First, the attractiveness ratings of voices were from 58 participants (18 male participants) and were mainly based on female participants. Although the gender of participants was approximately balanced in the ERP experiment, the effect of attractiveness may be influenced by the biased ratings in the pretest selection. Second, we used long neutral vowel stimuli to rule out irrelevant variables, such as semantic meaning (Ferdenzi et al., 2013). However, the vowel sounds were not representative in everyday life and less ecologically. The raters' evaluations of vowels may be different with words and speech in a real-life situation (Ferdenzi et al., 2013; Pisanski and Feinberg, 2018). Future research should test the beauty premium effect of voices using real speech sounds in natural social conditions.

Conclusion

The present study suggested that both voice attractiveness and gender influenced investments in a trust game. Attractive voices facilitated cooperative behaviors, demonstrating the "beauty premium" effect. Participants were more likely to cooperate with male partners. Regarding the evaluation of feedback, larger FRN effects were observed in the unattractive voice condition than in the attractive voice condition, suggesting that the level of reward expectation may be higher in the unattractive partner condition.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board of the Liaoning Normal University, China. The participants provided their written informed consent to participate in this study.

Author contributions

JS designed the experiment. ZL prepared the materials and performed the experiments. Both authors analyzed the data and wrote the manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (31400869), the Fundamental Research Funds for the Central Universities (2242022S20009), and the research funds for ideological and political education in postgraduate courses of Southeast University (yjgkcsz2229).

References

- Bestelmeyer, P. E. G., Latinus, M., Bruckert, L., Rouger, J., Crabbe, F., and Belin, P. (2012). Implicitly perceived vocal attractiveness modulates prefrontal cortex activity. *Cereb. Cortex* 22, 1263–1270. doi: 10.1093/cercor/bhr204
- Chen, J., Zhong, J., Zhang, Y., Li, P., Zhang, A., Tan, Q. B., et al. (2012). Electrophysiological correlates of processing facial attractiveness and its influence on cooperative behavior. *Neurosci. Lett.* 517, 65–70. doi: 10.1016/j.neulet.2012.02.082
- Collins, S. (2000). Men's voices and women's choices. *Anim. Behav.* 60, 773–780. doi: 10.1006/anbe.2000.1523
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/bf03193146
- Feinberg, D. R., DeBruine, L. M., Jones, B. C., and Perrett, D. I. (2008). The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Perception* 37, 615–623. doi: 10.1068/p5514
- Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., and Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Anim. Behav.* 69, 561–568. doi: 10.1016/j.anbehav.2004.06.012
- Ferdenzi, C., Delplanque, S., Mehu-Blantar, I., Cabral, K. M. D. P., Felicio, M. D., and Sander, D. (2015). The geneva faces and voices (GEFAV) database. *Behav. Res. Methods* 47, 1110–1121. doi: 10.3758/s13428-014-0545-0
- Ferdenzi, C., Patel, S., Mehu-Blantar, I., Khidasheli, M., Sander, D., and Delplanque, S. (2013). Voice attractiveness: Influence of stimulus duration and type. *Behav. Res. Methods* 45, 405–413. doi: 10.3758/S13428-012-0275-0
- Gehring, W. J., and Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295, 2279–2282. doi: 10.1126/science.1066893
- Groyeck, A., Pisanski, K., Sorokowska, A., Havlíček, J., Karwowski, M., Puts, D., et al. (2017). Attractiveness is multimodal: Beauty is also in the nose and ear of the beholder. *Front. Psychol.* 18:778. doi: 10.3389/fpsyg.2017.00778
- Hensel, L., Bzdok, D., Müller, V. I., Zilles, K., and Eickhoff, S. B. (2015). Neural correlates of explicit social judgments on vocal stimuli. *Cereb. Cortex* 25, 1152–1162. doi: 10.1093/cercor/bht307
- Holroyd, C. B., and Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology* 44, 913–917. doi: 10.1111/j.1469-8986.2007.00561
- Hughes, S. M., and Miller, N. E. (2016). What sounds beautiful looks beautiful stereotype: The matching of attractiveness of voices and faces. *J. Soc. Personal Relat.* 33, 984–996. doi: 10.1177/0265407515612445
- Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., and Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Anim. Behav.* 79, 57–62. doi: 10.1016/j.anbehav.2009.10.003
- Krumholz, C., Quigley, C., Ameen, K., Reuter, C., Fusani, L., and Leder, H. (2022). The effects of pitch manipulation on male ratings of female speakers and their voices. *Front. Psychol.* 13:911854. doi: 10.3389/fpsyg.2022.911854
- Krumholz, C., Quigley, C., Little, A., Zäske, R., and Riebel, J. K. (2021). Multimodal signalling of attractiveness. *Proc. Annu. Meet. Cogn. Sci. Soc.* 43, 31–32.
- Ma, Q., and Hu, Y. (2015). Beauty matters: Social preferences in a three-person ultimatum game. *PLoS One* 10:e0125806. doi: 10.1371/journal.pone.0125806
- Ma, Q., Hu, Y., Jiang, S., and Meng, L. (2015). The undermining effect of facial attractiveness on brain responses to fairness in the Ultimatum Game: An ERP study. *Front. Neurosci.* 9:77. doi: 10.3389/fnins.2015.00077
- Ma, Q., Qian, D., Hu, L., and Wang, L. (2017). Hello handsome! Male's facial attractiveness gives rise to female's fairness bias in ultimatum game scenarios—an ERP study. *PLoS One* 12:e0180459. doi: 10.1371/journal.pone.0180459
- Martin, R. S. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Front. Hum. Neurosci.* 6:304. doi: 10.3389/fnhum.2012.00304
- Pisanski, K., and Feinberg, D. R. (2018). "Vocal attractiveness," in *The oxford handbook of voice perception*, eds S. Frühholz and P. Belin (Oxford: Oxford University Press), 607–626.
- Re, D. E., O'Connor, J. J. M., Bennett, P. J., and Feinberg, D. R. (2012). Preferences for Very low and very high voice pitch in humans. *PLoS One* 7:e32719. doi: 10.1371/journal.pone.0032719
- Saxton, T. K., Burriss, R. P., Murray, A. K., Rowland, H. M., and Roberts, S. C. (2009). Face, body and speech cues independently predict judgments of attractiveness. *J. Evol. Psychol.* 7, 23–35. doi: 10.1556/JEP.7.2009.1.4
- Shang, J., and Liu, Z. (2022). Vocal attractiveness matters: Social preferences in cooperative behavior. *Front. Psychol.* 13:877530. doi: 10.3389/fpsyg.2022.877530
- Shang, J., Liu, Z., Wang, X., Chi, Z., and Li, W. (2021). Influence of vocal attractiveness on decision-making in a two-person ultimatum game and a three-person ultimatum game. *Adv. Psychol. Sci.* 29, 1402–1409. doi: 10.3724/SP.J.1042.2021.01402
- Solnick, S. J., and Schweitzer, M. E. (1999). The influence of physical attractiveness and gender on ultimatum game decisions. *Organ. Behav. Hum. Decis. Process.* 79, 199–215. doi: 10.1006/obhd.1999.2843
- Willis, J., and Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* 17, 592–598. doi: 10.1111/j.1467-9280.2006.01750.x
- Wu, Q., Liu, Y., Li, D., Leng, H., Iqbal, Z., and Jiang, Z. (2021). Understanding one's character through the voice: Dimensions of personality perception from

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Chinese greeting word “Ni Hao”. *J. Soc. Psychol.* 161, 653–663. doi: 10.1080/00224545.2020.1856026

Xu, S., Wang, H., and Wang, C. (2020a). Paying out one versus paying out all trials and the decrease in behavioral and brain activity in the balloon analogue risk task. *Psychophysiology* 57:e13510. doi: 10.1111/psyp.13510

Xu, S., Wang, M., Liu, Q., Wang, C., and Zhang, C. (2020b). Exploring the valence-framing effect: Gain frame enhances behavioral and brain sensitivity to the failure of decision-making under uncertainty. *Int. J. Psychophysiol.* 153, 166–172. doi: 10.1016/j.ijpsycho.2020.05.006

Yeung, N., and Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.* 24, 6258–6264. doi: 10.1523/JNEUROSCI.4537-03.2004

Zhang, H., Liu, M., Li, W., and Sommer, W. (2020). Human voice attractiveness processing: Electro physiological evidence. *Biol. Psychol.* 150:1078. doi: 10.1016/j.biopsycho.2019.107827

Zheng, Y., Compton, B. J., Heyman, G. D., and Jiang, Z. (2020). Vocal attractiveness and voluntarily pitch-shifted voices. *Evol. Hum. Behav.* 41, 170–175. doi: 10.1016/j.evolhumbehav.2020.01.002

Zuckerman, M., and Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *J. Nonverb. Behav.* 13, 67–82. doi: 10.1007/BF00990791

Zuckerman, M., Hodgins, H., and Miyake, K. (1990). The vocal attractiveness stereotype: Replication and elaboration. *J. Nonverb. Behav.* 14, 97–112. doi: 10.1007/BF01670437



OPEN ACCESS

EDITED BY

Samuel Shye,
Hebrew University of Jerusalem, Israel

REVIEWED BY

Ninja Katja Horr,
Brain Intelligence Neuro-Technology
Ltd., China
Tomás Lejarraga,
University of the Balearic Islands, Spain
Amos Schurr,
Ben-Gurion
University of the Negev, Israel

*CORRESPONDENCE

Ido Erev
✉ erev@tx.technion.ac.il

SPECIALTY SECTION

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

RECEIVED 11 September 2022

ACCEPTED 30 November 2022

PUBLISHED 12 January 2023

CITATION

Erev I and Marx A (2023) Humans as
intuitive classifiers.
Front. Psychol. 13:1041737.
doi: 10.3389/fpsyg.2022.1041737

COPYRIGHT

© 2023 Erev and Marx. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Humans as intuitive classifiers

Ido Erev^{1*} and Ailie Marx²

¹Faculty of Data and Decisions Sciences, Technion Israel Institute of Technology, Haifa, Israel,

²Department of Computer Science, Technion Israel Institute of Technology, Haifa, Israel

Mainstream decision research rests on two implicit working assumptions, inspired by subjective expected utility theory. The first assumes that the underlying processes can be separated into judgment and decision-making stages without affecting their outcomes. The second assumes that in properly run experiments, the presentation of a complete description of the incentive structure replaces the judgment stage (and eliminates the impact of past experiences that can only affect judgment). While these working assumptions seem reasonable and harmless, the current paper suggests that they impair the derivation of useful predictions. The negative effect of the separation assumption is clarified by the predicted impact of rare events. Studies that separate judgment from decision making document oversensitivity to rare events, but without the separation people exhibit the opposite bias. The negative effects of the assumed impact of description include masking the large and predictable effect of past experiences on the way people use descriptions. We propose that the cognitive processes that underlie decision making are more similar to machine learning classification algorithms than to a two-stage probability judgment and utility weighting process. Our analysis suggests that clear insights can be obtained even when the number of feasible classes is very large, and the effort to list the rules that best describe behavior in each class is of limited value.

KEYWORDS

J/DM separation paradox, description-experience gap, wavy recency effect, underweighting of rare events, the RUB assumption

Introduction

Classical studies of human decision making (Allais, 1953; Tversky and Kahneman, 1974; Kahneman and Tversky, 1979) use Savage's (1954) Subjective Expected Utility (SEU) theory as a benchmark. The most influential experimental studies focus on deviations from this benchmark, and the leading descriptive models focus on additions to this benchmark theory that explain the results. This research relies on two implicit working assumptions that facilitate the formulation of clear testable predictions from Savage's theory. The first implies that the underlying processes can be separated into two distinct stages: Judgment and Decision-Making (Edwards, 1954). Under this "J/DM separation" assumption (Erev and Plonsky, 2022), the decision makers first form beliefs concerning the payoff distributions of the feasible actions, and then use these beliefs (often referred to as judgements) to make decisions. The second assumption is that the participants in properly run experiments Read, Understand and Believe (RUB) the instructions (Erev, 2020).

While Savage's theory has lost popularity, the two working assumptions that were introduced to facilitate evaluation of this theory still underlie mainstream decision research. The current paper describes some of the negative impacts of this "working assumptions inertia," and highlights the potential benefit of relaxing these assumptions. Under the proposed relaxation, the cognitive processes that underlie decision making resemble machine learning classification algorithms.

J/DM separation: The assumption and the paradox

Savage (1954) showed that under a reasonable set of axioms (which generalizes the set used by von Neumann and Morgenstern, 1947 to support Expected Utility Theory), people behave "as-if" they form beliefs concerning the payoff distributions associated with all the feasible actions, and select the action that maximizes personal (subjective) expected utility given these beliefs. To illustrate the potential generality of this theory, Savage describes the preparation of an omelet. Specifically, he considers the decision made after breaking five good eggs into a bowl, and when considering the option of adding a sixth egg. It is easy to see that even this trivial decision is affected by personal beliefs: the belief concerning the probability that the egg is rotten. In addition, the omelet example clarifies the term "as-if" in Savage's analysis: our experience with the preparation of omelets suggests that it is possible to behave "as-if" we hold beliefs without explicitly considering these beliefs.

As noted above, behavioral decision research focuses on a sequential interpretation of Savage's theory. Specifically, the "as-if" part is replaced with the assumption that the underlying process can be separated into two stages: Explicit belief formation that involves probability judgment, and decision making. The leading studies of belief formation focus on human judgment; they examine how people estimate the probabilities of different events based on their past experiences. The top panel in Figure 1 presents one example from Rapoport et al.'s (1990) replication of Phillips and Edwards' (1966) classical study of revision of opinion. This study focuses on the way people form beliefs (judge probabilities) based on observable past experiences (the observed draws of red or white balls). The most influential studies of decision-making focus on "decisions under risk," and explore the way people decide when they are presented with a description of the payoff distributions (and do not have to judge probabilities based on past experience). The middle panel in Figure 1 presents one example from Erev et al.'s (2017) replication of Kahneman and Tversky's (1979) classical analysis of decisions under risk.

Although separating studies of judgment and of decision making is consistent with a feasible cognitive interpretation of SEU theory, the results presented by Barron and Erev (2003; lower

panel of Figure 1) suggest that it can lead to incorrect conclusions. The clearest demonstration of the shortcoming of the J/DM separation comes from studies of the impact of rare (low probability) events. Studies of judgment highlight robust overestimation of the probability of rare events (Phillips and Edwards, 1966; Erev et al., 1994), and studies of decisions under risk document overweighing of low probability outcomes (Kahneman and Tversky, 1979), thus, it is natural to conclude that oversensitivity to rare events is a general tendency (Fox and Tversky, 1998). In sharp contrast to this natural conclusion, Barron and Erev find that in tasks where judgment and decision making are not separated and people decide based on past experiences (as in Savage's omelet example), their behavior reflects underweighing of rare events. That is, separately both judgment and decision making reflect oversensitivity to rare events, but without the experimental separation these processes often lead to the opposite bias. Erev and Plonsky (2022) refer to this puzzle as the *J/DM separation paradox*.

The mere-presentation explanation

The difference between the middle and lower panels in Figure 1 is known as the description-experience gap (Hertwig and Erev, 2009): It implies higher sensitivity to rare events in decisions from description (middle panel) than in decisions from experience (lower panel). Erev et al. (2008a) show that part of this gap can be explained as a reflection of a mere-presentation effect: The rare outcomes receive more weight when they are explicitly presented (in the middle panel, but not in the lower panel). Erev and Plonsky (2022) note that the mere-presentation effect can also explain why the deviations from the rational model in judgment from experience (upper panel in Figure 1) are more similar to decisions from description than to decisions from experience. The results suggest that the mere-presentation of the rare events increases their weighting, in both judgment and decision tasks.

The overestimation of the probability of the less likely events in the top panel of Figure 1 can also be explained as the impact of response errors given the bounded response scale (see Erev et al., 1994); since the response scale is bounded between 0 and 1, response errors (e.g., some random responses) are expected to move the mean response toward 0.5. In agreement with this explanation, studies of judgment from experience in tasks in which the bias implied by random responses is minimized (like judgment of the mean of a series of observations, Spencer, 1961) reveal smaller biases (Peterson and Beach, 1967; Lejarraga and Hertwig, 2021). Yet, controlling the impact of response errors does not eliminate the indication of the mere presentation effect in judgment tasks. An indication of the impact of mere presentation that cannot be explained by response error is presented by Fischhoff et al. (1978). In one of the conditions they examined, the participants were asked to judge the

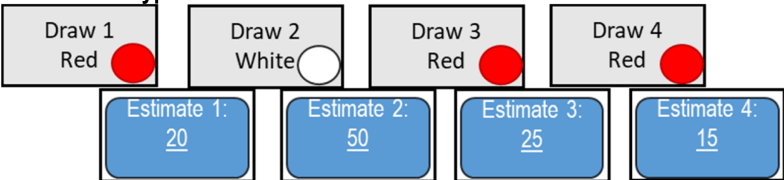

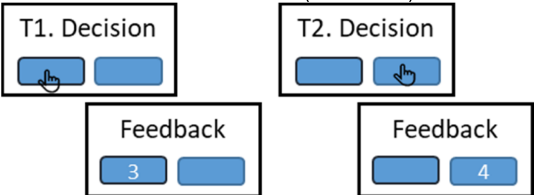
Typical experimental tasks	Typical results				
<p>Judgment: Two urns (Rapoport et al., 1990; following Phillips & Edwards, 1966)</p> <p>Instructions: Urn A holds 30 red balls and 70 white balls. Urn B holds 70 red balls and 30 white balls. One of the two urns was randomly selected. The experimenter will sample (with replacement) balls from that urn. After observing each ball, you will be asked to estimate the probability that the selected urn is A.</p> <p>Screens in a typical trial</p> 	<p>After experiencing a sequence with three red balls and one white ball:</p> <table border="1"> <tr> <th>Objective probability (under Bayes rule)</th><th>Mean judgment</th></tr> <tr> <td>0.06</td><td>0.24</td></tr> </table> <p>That is, the probability of the rare event (urn A) is overestimated</p>	Objective probability (under Bayes rule)	Mean judgment	0.06	0.24
Objective probability (under Bayes rule)	Mean judgment				
0.06	0.24				
<p>Decisions under risk (from description) (Erev et al., 2017; following Kahneman & Tversky, 1979)</p> <p>Instructions: Choose between the following two options:</p> 	<p>The maximization rate (rate of right button choice) was only 41%. This and similar results are captured, in prospect theory, with the assertion that the low-probability outcome (the payoff 0) is over-weighted.</p>				
<p>J/DM without separation. Decisions from experience with partial feedback (Barron & Erev, 2003).</p> <p>Instructions: The current experiment includes many trials. Your task, in each trial, is to click on one of the two keys presented on the screen. Each click will be followed by the presentation of that key's payoff. Your payoff for the trial is the payoff of the selected key.</p> <p>Screens in the first two trials (T1 and T2):</p> 	<p>When one option provided "3 with certainty," and the alternative "4 in 80% of the trials, 0 otherwise" the choice rate for the risky, EV maximizing, option was 65%. In contrast, when the risky option provided "32 in 10% of the trials, 0 otherwise," the maximization rate was only 30%. This, and similar results suggest insufficient sensitivity to the rare events.</p>				

FIGURE 1
Examples of studies of judgement and decision making with and without the J/DM separation.

probability that the reason for the observation that a "a car will not start," is "fuel system defective." The mere-presentation of a list of possible fuel system problems increased the mean estimate from 0.15 to 0.23.

Another indication of the descriptive value of the mere-presentation effect comes from studies that compare implicit and explicit perceptual decisions. One example (from Erev et al., 2008b) is presented in Figure 2. Condition Memory requires an implicit judgment of the probability that the central stimulus is the letter "B" rather than the number "13." In Condition Memory and Decision, the participants were explicitly asked to decide if the central stimulus is "B" or "13" in addition to being asked to memorize the list. This explicit request includes a presentation of the possibility that the list of letters includes a number. The results reveal that it increased the proportion of participants that remember "13" from 12 to 44%.

The RUB assumption and the impact of experience

The predictions of SEU theory depend on the information the decision maker uses to form beliefs and decide. Almost any behavior can be consistent with SEU theory given certain assumptions concerning the information the decision maker uses. Thus, it is impossible to test this theory without the addition of auxiliary assumptions regarding that information. The common additions rely on the working assumption that the participants in experimental studies Read, Understand and Believe (RUB) the information provided by the experimenter.

Careful experimenters focus on conditions that facilitate the descriptive value of the RUB assumption, and ensure that rational individuals who RUB the information provided by the

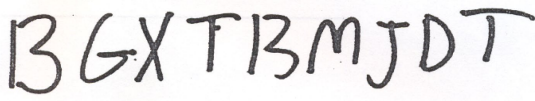
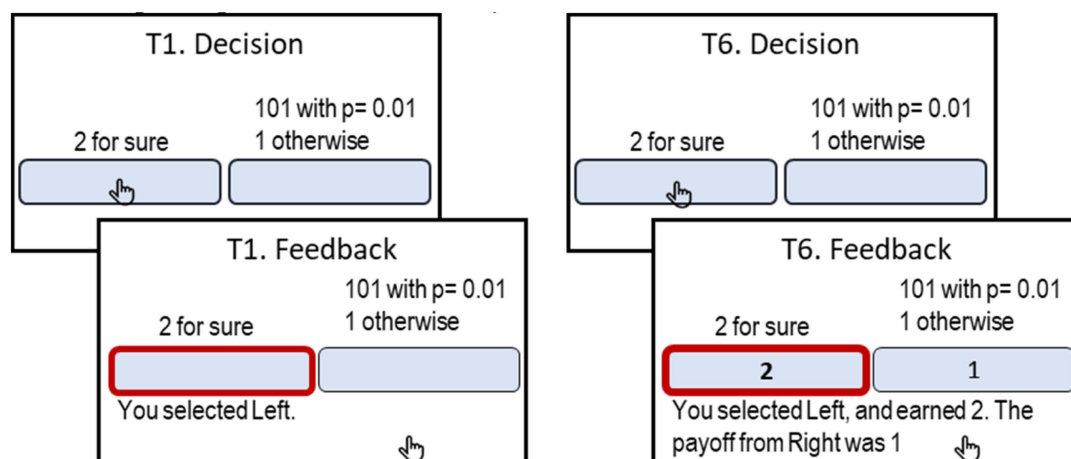
The list of stimuli	Main results
	<p>When asked to memorize the list, only 12% interpreted the central stimuli as the number 13.</p> <p>With mere presentation (of the event “number in the list”), when asked if the central stimulus is “B” or “13,” 44% answered “13.”</p>

FIGURE 2
The list of stimuli used by Erev et al. (2008b).



Main results: Before receiving feedback (Trials 1 to 5) most participants (55%) preferred the risky gamble. Feedback reduced this rate to 41%.

FIGURE 3
The screens in Trials 1 and 6 in one of the conditions studied by Erev et al. (2017), when the participant chose the left key.

experimenter will not be motivated to use other sources of information. For example, careful experimenters use easy to understand instructions, exclude participants that fail attention tests, and avoid running experiments that involve deception. Under these conditions, the RUB assumption implies that the availability of the description of the incentive structure replaces the judgment stage, and determines the information used by the decision makers. However, experimental studies question the success of this effort. For example, in the studies conducted by Erev et al. (2017, and see Figure 3), each of the participants were presented with 30 choice tasks for 25 trials (and were paid for one, randomly selected, of the 750 choices). The participants were first presented with a description of the payoff distributions, and after the 5th trial, received feedback after each choice. The results reveal that the availability of feedback affected the choice rate even when it did not add information concerning the

incentive structure. For example, consider the choice task presented in Figure 3, where the participants are asked to select between “2 with certainty” and “1% chance to win 101, 1 otherwise,” 25 times, and told that they will be paid for one randomly selected choice. Erev et al. found that in most cases (55%) the participants chose the risky prospect in the first five trials, but after receiving feedback the choice rate of this prospect dropped to 41%.¹

1 Erev et al. also show that the impact of experience cannot be explained by assuming that it only improves understanding of the incentive structure. Their results reveal that experience can increase violations of stochastic dominance. Specifically, when the correlation between the payoffs of the two prospects was negative, experience reduced the choice rate of “50% to win 9, 0 otherwise” over “50% to win 6, 0 otherwise.”

Three direct costs of the J/DM separation and RUB assumptions

In order to clarify the potential negative effects of the tendency to rely on the J/DM separation and the RUB assumptions, and ignore the shortcomings of these assumptions summarized above, we chose to highlight three direct costs of this “working assumptions inertia.”

Incorrect implementation of basic research results

One of the clearest direct costs of the reliance on the J/DM separation and RUB assumptions is overgeneralization of the results of studies of one-shot decisions under risk (like the middle panel in [Figure 1](#)). This research demonstrates overweighting of low probability outcomes. For example, 83% of the participants in [Kahneman and Tversky's \(1979\)](#) study preferred a loss of 5 with certainty over a 1/1000 chance to lose 5,000. Natural generalization of this finding suggests that the best way to avoid crime involves the use of severe punishments, even if the increase in severity implies lower probability of enforcement. While this prediction seems reasonable under the assumption that people overestimate and overweight rare costs, empirical research shows that using gentle punishments with high probability tend to be more effective ([Erev et al., 2010c](#); [Teodorescu et al., 2021](#)). For example, Erev et al. found that asking proctors in college exams to delay the preparation of a map of the students seating (that can be used to detect cheating and justify harsh punishments), and focus on moving students that appear to look around to the first row (a punishment that implies a loss of time of about a minute), reduces cheating.

Another example involves the effort to use lotteries to facilitate COVID-19 vaccination. The use of lotteries is predicted to be effective if people overweight rare rewards, but the effort to use this method to facilitate vaccination was not successful (see [Gandhi et al., 2021](#)). In contrast, the use of Green Pass policies that impose gentle punishments on individual that delay vaccination (the requirement to perform time consuming tests to allow entering public areas) appears to be more effective ([Mills and Rüttenauer, 2022](#)).

Suboptimal design of field experiments

In theory, the risk of overgeneralizing basic research can be addressed by running field experiments than compare alternative generalizations. This method is often used by applied behavioral economists that study nudge-based intervention ([Thaler and Sunstein, 2008](#)). However, most of these studies focus on the initial reaction to the intervention (see [Beshears and Kosowsky, 2020](#)). While this solution is likely to hold if experience does not affect choice behavior, as

expected in many settings under the RUB assumption, it might lead to incorrect conclusions if this working assumption does not hold.

Oversimplification and exaggeration of the impact of the choice environment

One of the contributors to the popularity of the J/DM separation and the RUB assumptions is the fact that they facilitate the simplification of complex decision problems. Yet, in some settings these assumptions simplify the problems too much. One demonstration of the cost of oversimplification is provided by the leading explanations of deviations from maximization in natural settings. Consider risk attitude in financial decisions: The observation that many investors prefer bonds over riskier stocks that provide higher average returns suggests risk aversion ([Mehra and Prescott, 1985](#)). In contrast, the observation that investors prefer individual stocks over safer index funds suggests risk-seeking ([Statman, 2004](#)). The leading explanations of these contradictories rest on the J/DM separation assumption, and ignore the impact of experience. They imply that the contradictory preferences reflect two distinct biases: Loss aversion in decisions under risk ([Benartzi and Thaler, 1995](#)), and overconfidence in probability judgment ([Odean, 1998](#)). These explanations suggest that the relative importance of the two biases is a function of the choice environment: Loss aversion is more important when investors choose between stock and bonds ([Benartzi and Thaler, 1995](#)), and overconfidence is more important when the investors select between stocks and index funds ([Odean, 1998](#)).

Recent research demonstrates that when the impact of experience is considered, the apparent contradiction can be explained without assuming two distinct biases and sensitivity to the choice environment. Specifically, under the assumption that people rely on past experiences, the tendency to select the riskier prospects is highly sensitive to the correlation between the different options. A tendency to avoid the risky options is expected when the differences between the payoffs of these options and the payoff from the safe choice are positively correlated (as in the case of a choice between different stocks and a safe bond), and a tendency to prefer the riskier options is expected these differences are negatively correlated (as in the case of a choice between stocks and index funds, see [Ben Zion et al., 2010](#)). [Figure 4](#) presents an experiment (from [Erev et al., 2023](#)) that tests and clarifies this prediction. In each of the 100 trials of this experiment the participants were asked to choose between an option that maintained the safe status quo (Option C, “0 for sure”), and two risky options with similar expected return. In the condition summarized in the top left panel, the two risky prospects were *negatively* correlated, and, the choice rate of the status quo was only 12%. In the condition summarized in the bottom left, the two risky prospects were *positively* correlated, and, the choice rate of the status quo was higher (34%) than the choice rate of the more attractive medium risk option (Option B, choice rate of 14%).

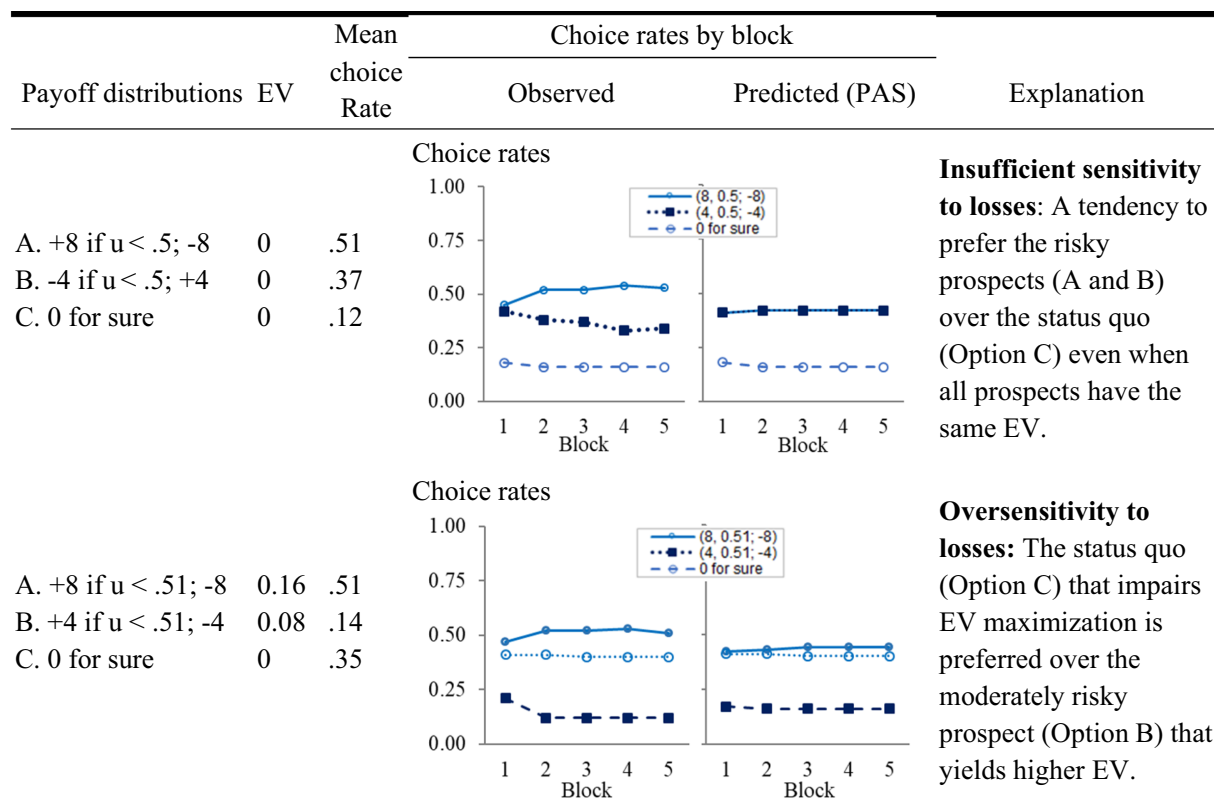


FIGURE 4

The impact of experience on sensitivity to losses (from Erev et al., 2023). The experiment used a variant of the experimental paradigm described in the lower panel of Figure 1. It included 100 trials, and the participants were presented with the payoff from all options after each choice. The left-hand column presents the incentive structure, u is a random draw from the range 0 to 1 [that is, from $u(0,1)$]. The right-hand choice rate graphs present the prediction of the PAS model, described below.

The reliance on small samples assumption, and the intuitive classifier explanation

Previous research that compares alternative explanations of the results exemplified in Figures 3, 4 highlights the advantage of models assuming the people tend to rely on small samples of past experience. Models that share this assumption won four choice prediction competitions (Erev et al., 2010a,b, 2017; Plonsky et al., 2019). The right side of Figure 4 demonstrates how a 2-parameter model of this type captures the contradictory sensitivity to losses described above. The model, referred to as Partially Attentive Sampler (PAS, Erev et al., 2023), assumes that after gaining experience each of the decisions of agent i in Task T is based on a sample of $\kappa_{i,T}$ past experiences (randomly drawn with replacement) with this task. The value of $\kappa_{i,T}$ is a free parameter. The agent selects the option with the highest average payoff in the sample, among the options it considers. At each trial the agent considers at least one option. The probability of considering each of the other options equals

$1 - \delta_{i,T}^{\left\lceil \frac{t-1}{\kappa_{i,T}} \right\rceil}$, where $\delta_{i,T}$ is another free parameter. The right-hand column in Figure 4 presents the prediction of this model for

Figure 4's tasks (when the distribution of parameters is estimated on a different set of tasks and different group of participants).

The wavy recency effect (a violation of the positive recency explanation)

The simplest explanations for the predictive value of models that assume reliance on small samples suggest that it reflects cognitive costs and limitations (see Hertwig and Pleskac, 2010). For example, it is possible that people overweight the easier to remember recent trials, or use a simple "win-stay-lose-shift" heuristic (Nowak and Sigmund, 1993). However, analysis of the sequential dependencies in the data rejects this simple explanation (Plonsky et al., 2015). The clearest evidence against the positive recency explanation comes from studies of decisions made between a safe prospect, and a binary risky prospect with a low probability extreme outcome. The results (see typical findings in Figure 5) reveal a wavy recency effect: The tendency to select the best reply to each occurrence of the rare and extreme outcomes is maximal 11 to 16 trials later. Moreover, the lowest best reply rate was observed 3 trials after the occurrence of the rare, extreme outcome.

The intuitive classifiers explanation

Plonsky et al. show that the wavy recency effect, and the descriptive value of the reliance on small samples hypothesis, can be explained with models that share two assumptions: (1) People try to select the option that led to the best outcomes in the most similar past experiences, and (2) The features used to judge similarity include the sequences of recent outcomes. These assumptions imply that the negative recency part of the wavy recency curve (the drop below 0 in [Figure 5C](#)) reflects the fact that the number of “similar past experiences” to decisions made immediately after a sequence that includes rare outcomes tends to be small. [Table 1](#) presents examples that clarify this assertion by focusing on the decision in Trial 64 of an experiment that studies the disaster problem of [Figure 5](#). It shows that if the payoff sequence immediately before Trial 64 includes a rare unattractive outcome (loss of -10), agents that select the option that led to the best outcome after a similar sequence are likely to rely on less than 5 past experiences, and are likely to underweight the rare events. Yet, if the sequence of last three recent payoffs does not include a loss, these agents rely on a larger sample (about 44 observations), and are not likely to underweight the rare events.

Plonsky et al. also demonstrate that when the environment is dynamic, judging similarity based on the sequence of recent outcomes can be highly adaptive. For example, consider the thought experiment described in [Figure 6](#). Intuition in this experiment favors a choice of Top in Trial 16. This behavior is implied by the assumption that similarity is determined based on the number of rare and extreme outcomes in the most recent 3 payoffs. And, under the assumption that the environment is dynamic (e.g., the payoffs are determined by the 4-state Markov chain described in [Figure 7](#)) it approximates the optimal strategy.

The assumption that people rely on similar past experiences can also explain the mere presentation effect. The mere presentation of a rare event (e.g., explicit description of the possibility of existence of a letter in a list of digits), under this account, changes the set of experiences that seem most similar to the current task. Specifically, it increases the probability of considering experiences with similar rare events. This account can also capture this initial tendency to overweight rare events in decisions from description (see [Marchiori et al., 2015](#)).

Notice that the current explanation, of the mere presentation effect and descriptive value of the reliance on small samples hypothesis, implies that the underlying processes resemble machine learning classification algorithms like Decision Tree ([Safavian and Landgrebe, 1991](#)), and Random Forest ([Breiman, 2001](#)). The basic idea behind these algorithms is the classification of the training data based on distinct features, assigning tasks to their appropriate classes, and deriving predictions based on past outcomes in these classes. For example, [Figure 8](#) presents a Decision Tree classification of [Figure 6](#)'s 15 observations based on the sign of the payoff from the risky choice in the last three trials (each as an individual feature). Trial 16 in this thought experiment is classified to the left most branch, and the implied decision is Top. While the popular machine learning tools were not designed to capture human cognition, their

success (for example, in controlling autonomous vehicles) suggests that it is possible that human cognitive processes were evolved to use the value of effective classifications, and people are “intuitive classifiers.”

It is important to emphasize that the intuitive classifiers explanation is not suggested here as a theory with testable predictions. Moreover, the intuitive classifiers explanation does not imply violations of SEU. Rather, it is an explanation of the observations described above. This explanation can be useful in two ways. First, it highlights the boundary conditions for the predictive value of the models we considered. For example, it implies that models like PAS that assume reliance on random samples of past experiences, and were found to provide good prediction of behavior in static settings, are not likely to provide useful prediction of behavior in dynamic settings (like [Figure 7](#) incentive structure). Second, it sheds light on the way in which these models can be extended.

The intuitive classifiers explanation (or view) is closely related to the assertion that behavior is selected by the contingencies of reinforcement ([Skinner, 1985](#), and see related ideas in [Nosofsky, 1984](#); [Gilboa and Schmeidler, 1995](#); [Gentner and Markman, 1997](#); [Dougherty et al., 1999](#); [Marchiori et al., 2015](#)). The current paper contributes to these analyses in two ways. First, the machine learning analogy highlights the possibility that the underlying processes use multiple classification methods, and it may not be possible to develop a simple model capturing people's response to the contingencies of reinforcements. Second, our analysis demonstrates that when it is difficult to correctly classify the current decision task (the contingencies of reinforcement are not clear) this process is likely to trigger behavior that appears to rely on randomly selected small samples of past experiences. This addition allows useful quantitative prediction of choice behavior in a wide set of situations.

The intuitive classifiers view can also be described as a generalization of the intuitive statistician assertion ([Peterson and Beach, 1967](#); [Gigerenzer and Murray, 1987](#); [Juslin et al., 2007](#)). Under the interpretation of the intuitive statistician assertion proposed by Gigerenzer and Murray, people tend to use cognitively efficient rules that approximate the outcomes of the more demanding computation required under traditional statistics. Thus, it assumes that the main deviations from maximization reflect cognitive limitations. The current generalization allows for the possibility of a second type of deviations from maximization: It addresses situations (like the ones considered here) in which the optimal choice rule is simple, but the decision makers cannot know it. In these situations, part of the deviation from maximization appears to reflect the use of cognitively inefficient similarity-based rules.

Relationship to the adaptive toolkit approach

The analysis presented by [Berg and Gigerenzer \(2010\)](#) suggests that the leading behavioral refinements of SEU (including

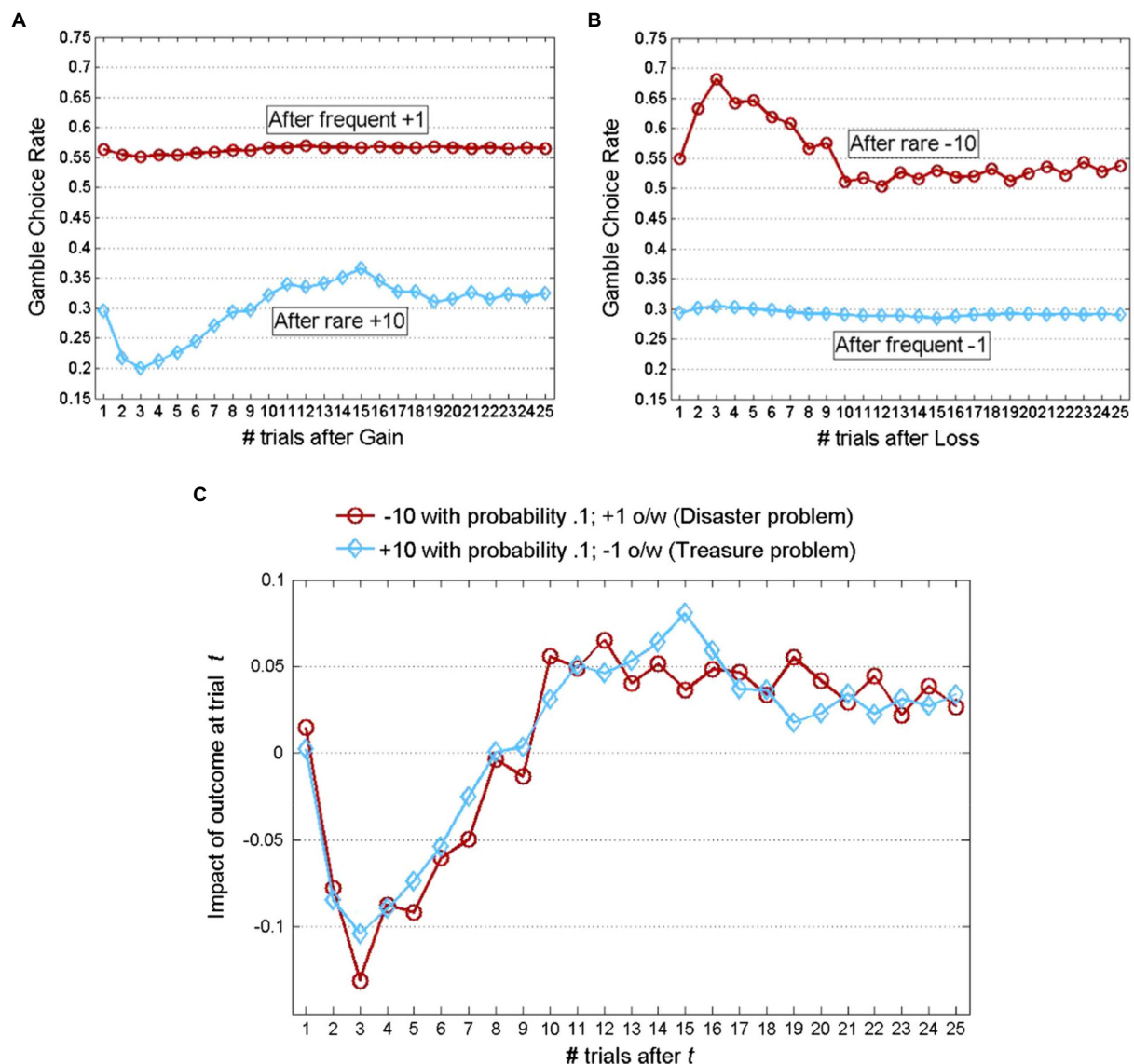


FIGURE 5

Demonstration of the wavy recency effect (adapted from Plonsky and Erev, 2017). Participants selected repeatedly for 100 trials between two unmarked buttons and received feedback concerning the payoff from both the chosen and the forgone option following each trial. One option generated a payoff of 0 with certainty while the other was a risky gamble detailed in the legend. (A) Exhibits the choice rates of the gamble contingent on the gamble providing a gain at trial t ; (B) exhibits the choice rates of the gamble contingent on the gamble providing a loss at trial t ; and (C) presents the difference between the corresponding plots in (A,B). Thus, the wavy curves in (C) reflect the impact of an outcome generated by the gamble at trial t on its choice rate in subsequent trials. Positive values (on the Y-axis) imply "positive recency" and negative values imply "negative recency." Data is averaged across 48 participants from Nevo and Erev (2012) and 80 participants from Teodorescu et al. (2013).

prospect theory, and other analyses that rest on the J/DM and RUB assumptions), are "as-if" models (like SEU itself); these models do not present a cognitively feasible description of the underlying cognitive processes. To advance toward better understanding of the underlying process, Berg and Gigerenzer (and see Gigerenzer and Selten, 2001) propose an adaptive toolkit (or toolbox) approach. This approach assumes that people use different "fast and frugal" cognitive tools (heuristics) in different settings (Gigerenzer and Todd, 1999). Thus, to understand choice behavior, it is necessary to map of the contextual variables that impact behavior by determining the boundaries of the different

areas in the map, and discover the heuristic people use in each area.

The current intuitive classifiers view is similar to the adaptive toolkit approach in several ways, but there are also important differences between the two approaches. One important similarity involves the fact that both approaches assume that decision making starts with a classification process. The main difference involves the assumed number of classes. The adaptive toolkit (or toolbox) approach rests on the (implicit) optimistic assumption that the number of significant classes (distinct areas in the map) is relatively small. This implies that it is possible to map the space

TABLE 1 Demonstration of the implications of sequence-based similarity rules.

Trials since the last loss	The payoff from the risky option in the three trials before Trial 64			Expected number of similar past experiences in Trial 64	The probability that the average payoff from the risky option over the similar past experiences is positive (and the implied decision reflects underweighting of rare events)
	Trial 61	Trial 62	Trial 63		
More than 3	+1	+1	+1	44.00	0.495
3	−10	+1	+1	4.70	0.593
2	+1	−10	+1	4.79	0.591
1	+1	+1	−10	4.79	0.602

The table considers Trial 64 in the “disaster problem” of Figure 5 (“0 with certainty” or “10% to lose 10, gain of 1 otherwise”), assuming that similarity is determined by the three recent payoffs from the risky option. It shows that when the recent payoff sequence includes a rare event, the number of similar past experiences decreases, and the probability of underweighting of the rare event (choosing the risky option) increases.

(a) Task:

In each trial of the current study, you are asked to choose between “Top” and “Bottom”, and earn the payoff that appears on the selected key after your choice is made. The following table summarizes the environment results of the first 15 trials. What would you select in trial 16?

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Top	-1	-1	-1	+2	-1	-1	-1	+2	-1	-1	-1	+2	-1	-1	-1	
Bottom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

(b) Implications:

In trial 16, intuition favors “Top” despite the fact that the average payoff from “Top” over the 15 trials is negative (-0.4). This intuition suggests that when facing Trial 16, people tend to rely on the most similar previous (trial 16, like 4, 8, and 12, follows a sequence of three -1 outcomes). Thus, the choice is made based on only three past experiences.

FIGURE 6
A thought experiment.

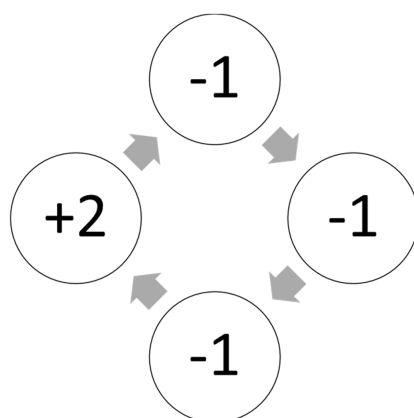


FIGURE 7
An example of a 4-state Markov chain that could determine the payoff from Top in Figure 6.

of decision tasks, and identify the heuristics that people tend to use in each area of the map. Partial support for this optimistic hypothesis is provided by studies demonstrating how specific “fast and frugal” heuristics can capture adaptive human behavior in specific settings. For example, the take-the-best heuristic (Gigerenzer and Goldstein, 1996) was found to facilitate performance in decisions based on multiple cues, and the priority heuristic (Brandstätter et al., 2006) was found to capture basic decisions from description. The current intuitive classifiers view is less optimistic. We believe that the number of classes that people consider can be extremely large, and it might not be possible to map them in a useful way. To address this possibility, we build on the premise that in many situations the impact of the multiple classifications can be predicted with simple approximations.

Part of our pessimism, concerning the predictive value of fast and frugal heuristics, reflects the outcomes of the choice prediction competitions conducted by Erev et al. (2017) and Plonsky et al. (2019). These competitions focused on decisions

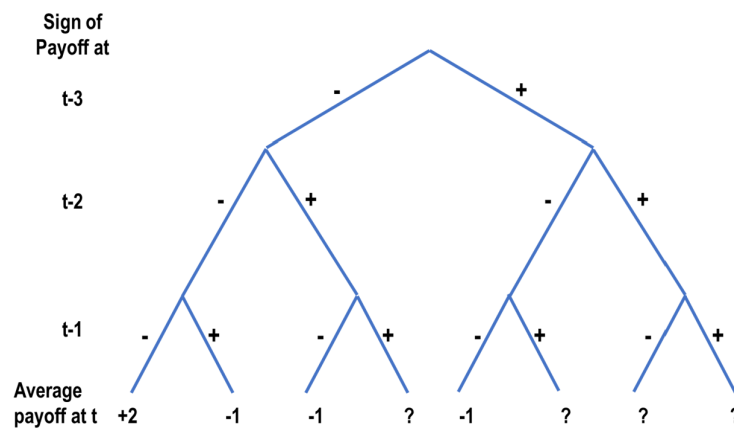


FIGURE 8

A decision tree analysis of the results in the first 15 trials of the thought experiment presented in Figure 6. The average payoff line presents the observed average payoff in each category. The question mark (?) implies that the training data do not include observation in the relevant branch.

from description (without and with feedback concerning the outcome of the previous choices, using the experimental paradigm describe in Figure 3). Under an optimistic interpretation of Brandstätter et al.'s (2006) results, this class of decision tasks is on the area of the map in which people are expected to use the priority heuristic. The results, did not support this prediction. Rather, the best models in the two completions can be described as quantifications of the intuitive classifiers explanation.

One demonstration of the potential of models that approximate the impact of a huge number of possible classifications, comes from the study of decisions from experience in static settings illustrated in Figure 4. As noted above, the choice rates in these experimental conditions can be captured with simple models that assume reliance on small samples, and this behavior can be the product of intuitive classification.

Summary

Research can be described as a hike through the land of assumptions in an attempt to find a hill with a good point of view on the lands of behaviors (Erev, 2020). Mainstream decision researchers tend to hike on a hill defined by the J/DM separation and RUB working assumptions. The view from this hill clarifies interesting deviations from specific rational models, but can also lead to incorrect conclusions. The current analysis highlights some of the shortcomings of the view from the J/DM separation and RUB hill, and the potential of exploring new areas in the land of assumptions.

The main cost of reliance on the J/DM separation assumption involves incorrect prediction of the impact of rare events. Studies that separate judgment and decision making suggest oversensitivity to rare events, while many natural decisions appear to reflect the opposite bias. This gap can be explained with the assertion that the separation requires an explicit presentation of the rare events that triggers a merge presentation effect. The costs of the RUB assumption include incorrect interpretation of short field experiments, and overestimation of the impact of the choice environment.

The potential of exploring other hills in the land of assumptions is clarified by the high predictive value of models that assume reliance on small samples of past experiences, and the observation that the success of these models can be explained by the assuming that humans are intuitive classifiers. While the intuitive classifiers view does not lead to testable predictions, our analysis suggests that exploring the possibility that people are intuitive classifiers can facilitate understanding and the derivation of models that provide useful predictions.

Ethics statement

Ethical review and approval was not required for the current study in accordance with the local legislation and institutional requirements. The studies which this study reviews were approved by the Technion IRB committee.

Author contributions

IE and AM thought about the basic idea together. IE wrote the first draft. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by a grant from the Israel Science Foundation (grant 861/22).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

References

- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica* 21, 503–546.
- Barron, G., and Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *J. Behav. Decis. Mak.* 16, 215–233. doi: 10.1002/bdm.443
- Benartzi, S., and Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *Q. J. Econ.* 110, 73–92. doi: 10.2307/2118511
- Ben Zion, U., Erev, I., Haruvy, E., and Shavit, T. (2010). Adaptive behavior leads to under-diversification. *J. Econ. Psychol.* 31, 985–995. doi: 10.1016/j.joep.2010.08.007
- Berg, N., and Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *Hist. Econ. Ideas* 18, 133–165. doi: 10.1400/140334
- Beshears, J., and Kosowsky, H. (2020). Nudging: Progress to date and future directions. *Organ. Behav. Hum. Decis. Process.* 161, 3–19. doi: 10.1016/j.obhdp.2020.09.001
- Brandstätter, E., Gigerenzer, G., and Hertwig, R. (2006). The priority heuristic: making choices without trade-offs. *Psychol. Rev.* 113:409. doi: 10.1037/0033-295X.113.2.409
- Breiman, L. (2001). Random forests. *Machine learning* 45, 5–32.
- Dougherty, M. R. P., Gettys, C. F., and Ogden, E. E. (1999). MINERVA-DM: a memory processes model for judgments of likelihood. *Psychol. Rev.* 106, 180–209. doi: 10.1037/0033-295X.106.1.180
- Edwards, W. (1954). The theory of decision making. *Psychol. Bull.* 51:380.
- Erev, I. (2020). Money makes the world go round, and basic research can help. *Judgment & Decision Making* 15, 304–310.
- Erev, I., Ert, E., and Roth, A. E. (2010a). A choice prediction competition for market entry games: An introduction. *Games* 1, 117–136. doi: 10.3390/g1020117
- Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., et al. (2010b). A choice prediction competition: Choices from experience and from description. *J. Behav. Decis. Mak.* 23, 15–47. doi: 10.1002/bdm.683
- Erev, I., Ert, E., Plonsky, O., Cohen, D., and Cohen, O. (2017). From anomalies to forecasts: toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychol. Rev.* 124, 369–409. doi: 10.1037/rev0000062
- Erev, I., Ert, E., Plonsky, O., and Roth, Y. (2023). Contradictory deviations from maximization: environment-specific biases, or reflections of basic properties of human learning?
- Erev, I., Glozman, I., and Hertwig, R. (2008a). What impacts the impact of rare events. *J. Risk Uncertain.* 36, 153–177. doi: 10.1007/s11166-008-9035-z
- Erev, I., Ingram, P., Raz, O., and Shany, D. (2010c). Continuous punishment and the potential of gentle rule enforcement. *Behav. Process.* 84, 366–371. doi: 10.1016/j.beproc.2010.01.008
- Erev, I., and Plonsky, O. (2022). *The J/DM separation paradox and the reliance on small samples hypothesis. To appear in sampling in judgment and decision making*, Fiedler, K., Juslin, P., and Denrell, J. (Eds.) Cambridge University Press.
- Erev, I., Shimonovich, D., Schurr, A., and Hertwig, R. (2008b). “Base rates: how to make the intuitive mind appreciate or neglect them” in *Intuition in judgment and decision making* (LEA, New York: Erlbaum), 135–148.
- Erev, I., Wallsten, T. S., and Budescu, D. V. (1994). Simultaneous over- and underconfidence: the role of error in judgment processes. *Psychol. Rev.* 101, 519–527.
- Fischhoff, B., Slovic, P., and Lichtenstein, S. (1978). Fault trees: sensitivity of estimated failure probabilities to problem representation. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 330–344.
- Fox, C. R., and Tversky, A. (1998). A belief-based account of decision under uncertainty. *Manag. Sci.* 44, 879–895.
- Gandhi, L., Milkman, K. L., Ellis, S., Graci, H., Gromet, D., Mobarak, R., et al. (2021). *An experiment evaluating the impact of large-scale High-Payoff Vaccine Regret Lotteries*.
- Gentner, D., and Markman, A. B. (1997). Structure mapping in analogy and similarity. *Am. Psychol.* 52:45.
- Gigerenzer, G., and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychol. Rev.* 103:650.
- Gigerenzer, G., and Murray, D. J. (1987). *Cognition as intuitive statistics* London: Psychology Press.
- Gigerenzer, G., and Selten, R. (2001). Rethinking rationality. *Bounded rationality: The adaptive toolbox*. 1:12.
- Gigerenzer, G., and Todd, P. M. (1999). “Fast and frugal heuristics: the adaptive toolbox” in *Simple heuristics that make us smart*. eds. G. Gigerenzer and P. Todd (Oxford University Press), 3–34.
- Gilboa, I., and Schmeidler, D. (1995). Case-based decision theory. *Q. J. Econ.* 110, 605–639.
- Hertwig, R., and Erev, I. (2009). The description–experience gap in risky choice. *Trends Cogn. Sci.* 13, 517–523. doi: 10.1016/j.tics.2009.09.004
- Hertwig, R., and Pleskac, T. J. (2010). Decisions from experience: why small samples? *Cognition* 115, 225–237. doi: 10.1016/j.cognition.2009.12.009
- Juslin, P., Winman, A., and Hansson, P. (2007). The naïve intuitive statistician: a naïve sampling model of intuitive confidence intervals. *Psychol. Rev.* 114:678. doi: 10.1037/0033-295X.114.3.678
- Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–292.
- Lejarraga, T., and Hertwig, R. (2021). How experimental methods shaped views on human competence and rationality. *Psychol. Bull.* 147:535. doi: 10.1037/bul0000324
- Marchiori, D., Di Guida, S., and Erev, I. (2015). Noisy retrieval models of over- and under-sensitivity to rare events. *Decision* 2, 82–106. doi: 10.1037/dec0000023
- Mehra, R., and Prescott, E. C. (1985). The equity premium: A puzzle. *J. Monet. Econ.* 15, 145–161. doi: 10.1016/0304-3932(85)90061-3
- Mills, M. C., and Rüttenauer, T. (2022). The effect of mandatory COVID-19 certificates on vaccine uptake: synthetic-control modelling of six countries. *Lancet Public Health* 7, e15–e22. doi: 10.1016/S2468-2667(21)00273-5
- Nevo, I., and Erev, I. (2012). On surprise, change, and the effect of recent outcomes. *Front. Psychol.* 3:24. doi: 10.3389/fpsyg.2012.00024
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *J. Exp. Psychol. Learn. Mem. Cogn.* 10:104.
- Nowak, M., and Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's dilemma game. *Nature* 364, 56–58.
- Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *J. Finance* 53, 1887–1934. doi: 10.1111/0022-1082.00078
- Peterson, C. R., and Beach, L. R. (1967). Man as an intuitive statistician. *Psychol. Bull.* 68:29.
- Phillips, L. D., and Edwards, W. (1966). Conservatism in a simple probability inference task. *J. Exp. Psychol.* 72, 346–354.
- Plonsky, O., Apel, R., Ert, E., Tennenholtz, M., Bourgin, D., Peterson, J. C., et al. (2019). Predicting human decisions with behavioral theories and machine learning. *ArXiv Preprint ArXiv:1904.06866*.
- Plonsky, O., and Erev, I. (2017). Learning in settings with partial feedback and the wavy recency effect of rare events. *Cogn. Psychol.* 93, 18–43. doi: 10.1016/j.cogpsych.2017.01.002
- Plonsky, O., Teodorescu, K., and Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychol. Rev.* 122, 621–647. doi: 10.1037/a0039413
- Rapoport, A., Wallsten, T. S., Erev, I., and Cohen, B. L. (1990). Revision of opinion with verbally and numerically expressed uncertainties. *Acta Psychol.* 74, 61–79.
- Safavian, S. R., and Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* 21, 660–674.
- Savage, L. J. (1954). *The foundations of statistics*. New York: John Wiley & Sons.
- Skinner, B. F. (1985). Cognitive science and behaviourism. *Br. J. Psychol.* 76, 291–301.

- Spencer, J. (1961). Estimating averages. *Ergonomics* 4, 317–328. doi: 10.1080/00140136108930533
- Statman, M. (2004). The diversification puzzle. *Financ. Anal. J.* 60, 44–53. doi: 10.2469/faj.v60.n4.2636
- Teodorescu, K., Amir, M., and Erev, I. (2013). The experience–description gap and the role of the inter decision interval. in *Progress in Brain Research*. eds. V. S. C. Pammi and N. Srinivasan (Amsterdam, The Netherlands) Vol. 202, 99–115.
- Teodorescu, K., Plonsky, O., Ayal, S., and Barkan, R. (2021). Frequency of enforcement is more important than the severity of punishment in reducing violation behaviors. *Proc. Natl. Acad. Sci.* 118:e2108507118. doi: 10.1073/pnas.2108507118
- Thaler, R. H., and Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin Group, New York.
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases: biases in judgments reveal some heuristics of thinking under uncertainty. *Science* 185, 1124–1131.
- von Neumann, J., and Morgenstern, O. (1947). *Theory of games and economic behavior*. 2nd Edn. (Princeton, NJ: Princeton University Press).



OPEN ACCESS

EDITED BY

Samuel Shye,
Hebrew University of Jerusalem,
Israel

REVIEWED BY

Eldad Yechiam,
Technion Israel Institute of Technology,
Israel
Hazik Mohamed,
Stellar Consulting Group,
Singapore

*CORRESPONDENCE

Johannes T. Doerflinger
✉ johannes.doerflinger@uni-konstanz.de

SPECIALTY SECTION

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

RECEIVED 22 August 2022

ACCEPTED 22 December 2022

PUBLISHED 12 January 2023

CITATION

Doerflinger JT, Martiny-Huenger T and
Gollwitzer PM (2023) Exploring the
determinants of reinvestment decisions:
Sense of personal responsibility,
preferences, and loss framing.
Front. Psychol. 13:1025181.
doi: 10.3389/fpsyg.2022.1025181

COPYRIGHT

© 2023 Doerflinger, Martiny-Huenger and
Gollwitzer. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Exploring the determinants of reinvestment decisions: Sense of personal responsibility, preferences, and loss framing

Johannes T. Doerflinger^{1*}, Torsten Martiny-Huenger² and
Peter M. Gollwitzer^{1,3}

¹Department of Psychology, University of Konstanz, Konstanz, Germany, ²Department of Psychology, UiT The Arctic University of Norway, Tromsø, Norway, ³Department of Psychology, New York University, New York, NY, United States

Two potentially costly errors are common in sequential investment decisions: sticking too long to a failing course of action (escalation of commitment), and abandoning a successful course of action prematurely. Past research has mostly focused on escalation of commitment, and identified three critical determinants: personal responsibility, preferences for prior decisions, and decision framing. We demonstrate in three studies using an incentivized poker inspired task that these determinants of escalation reliably lead decision makers to keep investing even when real money is on the line. We observed in Experiments 1, 2 and 3 that reinvestments were more likely when decision makers were personally responsible for prior decisions. This likelihood was also increased when the decision makers had indicated a preference for initial investments (Experiments 2 and 3), and when outcomes were framed in terms of losses as compared to gains (Experiment 3). Both types of decision errors – escalation of commitment and prematurely abandoning a course of action – could be traced to the same set of determinants. Being personally responsible for prior decisions, having a preference for the initial investment, and loss framing did increase escalation, whereas lacking personal responsibility, having no preference for the initial investment, and gain framing increased the likelihood of prematurely opting out. Finally, personal responsibility had a negative effect on decision quality, as decision-makers were still more likely to reinvest when they were personally responsible for prior decisions, than when prior decisions were assigned optimally by an algorithm (Experiments 2 and 3).

KEYWORDS

investment decisions, escalation of commitment, personal responsibility, framing, preferences, poker game, sunk cost

1. Introduction

When individuals repeatedly face the decision to further invest in or opt-out of a course of action, accurately using the available information is crucial to avoid two potential decision errors: first, abandoning the successful course of action too early and thus missing out on potential benefits, and second, persisting too long with a futile course of action (Drummond, 2014). The latter refers to one of the major branches of the sunk cost research (progress decisions; Roth et al., 2015) and is often labeled as escalation of commitment (EoC). The effect has been observed in a multitude of domains (e.g., personal, business, political, or gambling decisions; Sleesman et al., 2012), both on the individual and the group level (Sleesman et al., 2018). Research has identified factors enhancing EoC: among others personal responsibility for prior decisions (Staw, 1976), preferences for the initial decisions (Schulz-Hardt et al., 2009), and increased risk-seeking when dealing with losses (Soman, 2004).

Although there are experimental studies investigating EoC most of these relied on hypothetical scenarios (reviewed by Roth et al., 2015; Sleesman et al., 2018; see also Negrini et al., 2020), with some exceptions (e.g., Heath, 1995; Wong et al., 2006; Ronayne et al., 2021). More importantly, there are almost no studies investigating the responsibility/self-justification factor in experimental designs with real consequences. A study by Kirby and Davis (1998) is an exception. Kirby and Davis introduced a complex company setup in which participants solved anagrams using a specific strategy. In the escalation decision, participants could invest money into continuing with the (failing) strategy that they had previously chosen themselves or not. The results show an effect of responsibility; participants that made the first strategy decision invested more money into continuing with that strategy. There are some noteworthy details. The study involved a complex setup including an indirect relation between study performance and the real money to be gained, and the study setup relied on considerable false information given to participants. First, the real consequences were not directly contingent on the participants' decision. Instead, the money that participants ended up with as profit in their "company" were exchanged into raffle tickets. Thus, making more money in the study's game only increased the likelihood of winning the raffle. In addition, these supposedly real consequences were actually not implemented and all participants received the same amount of raffle tickets, a procedure that is not in line with the strict standards of behavioral economics research. Furthermore, the negative feedback that participants received regarding the initially chosen strategy was false information that was the same for all participants.

In sum, considering the prominence of the responsibility/self-justification factor in the literature on escalation of commitment and the limited experimental evidence for it, there is a need to validate this EoC determinant. Such validation does not only require study designs that focus on manipulating the critical factors but also on assessing actual task performance.

1.1. Determinants of escalation of commitment

A standard experimental task paradigm in EoC research (Staw, 1976) is confronting research participants with hypothetical investment scenarios, in which research participants are asked to take on the role of a CEO and then choose one of two investment alternatives. For example, participants are asked to select one of two projects in a company – the company could either develop consumer products or industrial products – in which they could hypothetically invest \$8 million. Participants then receive either positive or negative feedback, meaning that the project they had invested in has done well or poorly. Following this feedback, participants are asked to allocate a given amount of money (e.g., \$10 million) between the previously chosen and the non-chosen alternative. Reinvesting in the previously chosen option after negative feedback is regarded as a suboptimal strategy and labeled as EoC (e.g., Lipsey and Harbury, 1992). In such research, multiple psychological mechanisms and situational influences were found to affect EoC.

1.1.1. Personal responsibility and preferences for the initial decision

One of the most prominently studied determinants for EoC is personal responsibility: If decision-makers are personally responsible for initial investments and experience that their chosen course of action is failing, they are driven to stick with it as a form of ego-defense. For example, in hypothetical scenarios, participants are either asked to make the first investment decision or are told that this decision had been made by their predecessor. A common finding is that participants who made the first investment decision will later invest more money than participants who did not make this decision, even if the course of action invested in is failing (Staw, 1976; Bobocel and Meyer, 1994).

Schulz-Hardt et al. (2009) argued that decision-makers persist with a chosen course of action not because of a responsibility bias but because they originally had a higher preference for the chosen rather than the non-chosen option. As choices are to some extent guided by preferences, there is a greater fit between the choices and preferences in conditions in which decision-makers make the decision themselves compared to conditions in which the decision-makers do not make the decision themselves. Tasks in which only personal responsibility is manipulated thus confound personal responsibility with preferences. The authors found that preferences for the initial decision fully mediated the effects of manipulated personal responsibility on subsequent hypothetical reinvestment decisions (Study 1) or on sticking to a chosen strategy of task performance (Study 2).

Besides Schulz-Hardt et al. (2009), very few studies directly measured preferences for the initial investment decisions in an EoC task. A meta-analytic review by (Sleesman et al., 2012) found a small positive main effect of preferences for the initial decision on EoC. The authors concluded that this preference effect was not strong enough to fully account for responsibility effects. They also

categorized studies into those in which the decision-makers actually made the initial decision themselves versus those in which the initial decision was merely assigned to the participants. Sleesman et al. argued that studies requiring an actual initial decision should consist mostly of participants who prefer the chosen course of action, whereas studies in which the initial decision is assigned should consist of both participants with and without a high preference. Using this categorization of actually made versus adopting assigned decisions, the authors however found no difference in EoC. Whereas the authors conclude that this finding raises doubts regarding a preference-based explanation of EoC, it also raises questions regarding the responsibility explanation. Should one not expect to see more substantial personal responsibility effects in studies where participants are actually responsible instead of just being told that they are responsible even though they were not?

Taken together, evidence suggests that preferring the initially chosen option over the non-chosen option increases the likelihood of reinvesting in a chosen course of action. We expect that this will not just be the case in EoC situations – preference effects should also be a driver of continued investments, when reinvesting is the prudent course of action. However, it remains unclear whether preferences can fully account for personal responsibility effects. We therefore measured preferences and assessed whether there are personal responsibility effects beyond them.

1.1.2. Framing

EoC has also been interpreted as a consequence of negative decision framing (Thaler, 1980; Whyte, 1986; Arkes, 1991; Soman, 2004); decision makers are assumed to construe their previous investments after failure feedback in terms of losses. Prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) postulates that when people are dealing with gains, they are less sensitive to additional gains – high gain options with the risk for low (or no) gains are less attractive than risk-free moderate gain options. However, when people are dealing with losses, they are less sensitive to additional losses – no (or low) loss options with the risk for high losses are more attractive than risk-free moderate loss options. Typical EoC situations can be mapped onto the gain/loss framework described in prospect theory. When actors think about prior investments in terms of losses, additional (risky) investments may be perceived as an opportunity to avoid or recoup losses. Whether decision makers construe a situation in terms of either gains or losses can be manipulated by framing the outcomes accordingly (Tversky and Kahneman, 1981).

Experimental research on gain vs. loss framing and EoC is limited so far. Rutledge (Rutledge, 1995) examined EoC and framing effects in a modified investment task for small groups with a gain versus loss framing manipulation. Participants in their study were asked to assume the role of financial vice presidents of a fictitious company. They worked in groups of three and were asked to decide whether to make a reinvestment decision for a failing project. Personal responsibility was manipulated by telling participants that the initial investment decision was made because

they had recommended it themselves or because of the recommendation of another team. Consequences were presented for half the participants in terms of savings and the other half in terms of losses. The author observed personal responsibility and framing effects on EoC; responsibility effects were more pronounced in the loss frame than in the gain frame condition. This finding is in line with the prospect theory account of EoC, which predicts that loss framing should increase escalation of commitment. However, Schoorman et al. (1994) manipulated the gain/loss framing in hypothetical investment scenarios and observed framing effects only when little (vs. much) context information was given.

In sum, typical EoC tasks put participants into a situation that prospect theory would refer to as loss framing. Prospect theory and studies on EoC converge on predicting that participants are likely to take risks beyond what might be considered reasonable from a probability perspective. However, experimental evidence regarding gain/loss framing and reinvestments is limited so far and based solely on using hypothetical decision scenarios. Also, the question remains whether moving Rutledge's (1995) research conducted at the group level onto individual (non-hypothetical) decision-making results in comparable findings. If this is the case, then responsibility effects should be even more pronounced if outcomes are framed as losses. In any case, individuals should be more likely to invest in a loss than in a gain frame.

2. Present research

Prior research on EoC predominantly relied on hypothetical scenarios tasks in experimental studies, but the determinants are also supported by non-experimental studies examining investments on the organizational level [e.g., McNamara et al., 2002; Hsieh et al., 2015 review by Sleesman et al. (2018)]. We used a poker-game inspired computer task (VIP-Task; Doerflinger et al., 2017) that allowed us to conjointly manipulate the previously identified features (i.e., personal responsibility, loss/gain framing) and measure relevant variables for each decision (i.e., preferences) in a context with real consequences for participants.

Poker is a card game of chance and strategy, in which the players have to repeatedly decide whether to bet on their cards (i.e., invest further resources) or opt-out. Opting out means disregarding some still existing chances of winning and incurring a sure loss, but potentially avoiding throwing good money after bad. Between each bet, the chances of winning can change. Leaving strategic social interactions (e.g., bluffing) aside, poker players should only consider prospective gains and losses to arrive at the best outcome. Nonetheless, players regularly fall victim to EoC effects or miss out on good investments (Smith et al., 2009). The quality of poker decisions depends highly on the probability of success. The same is true for reinvestment decisions in many other real-life domains (e.g., in a business, there is the probability that competitors whom I did not yet know of enter the stage). Therefore, we see poker decisions as a suitable model environment for testing reinvestment decisions in realistic incentivized experiments.

The purpose of Experiment 1 is to demonstrate the personal responsibility effect on performing the VIP-Task. In Experiment 2, we tested whether personal responsibility effects still occur when initial preferences were controlled for; are personal responsibility effects independent of whether people prefer the initially chosen option or not? Experiment 3 was designed to test the effects of responsibility, preferences, and gain/loss framing of outcomes in parallel. The design of Experiments 2 and 3 allows evaluating the quality of those decisions where the participants chose to invest and those decisions where they chose to opt out based on the expected value of the decision. Experiments 2 and 3 include two benchmarks against which we compare participants' reinvestment decisions: decisions when the prior investment was made by an algorithm either (1) randomly or (2) optimally in line with expected-value principles. We obtained approval from the university's ethics committee for all of the studies reported in the present manuscript. We used the *simr* package for R (Green and MacLeod, 2016) to estimate the statistical power *via* simulations. The sample size and number of trials in all three experiments are sufficient to detect small within-participants effects ($OR = 1.4$) with a probability of $\beta - 1 > 0.95$ at the $\alpha = 0.05$ significance level in mixed effects logistic regressions with random effects for participants (Experiments 1, 2, and 3) and for trials (Experiment 1). Experiment 3 was preregistered, and the data for all three experiments and a pilot study for Experiment 3 are available at: https://osf.io/hdczr/?view_only=546aec80e6f7468685072d199a9d9821.

2.1. Experiment 1: Personal responsibility

Participants played multiple rounds of a poker inspired card game against a computer. To increase their payout, they had to repeatedly decide whether to keep investing into new cards or to quit a round. We tested responsibility effects on reinvestments by asking participants to make reinvestment decisions after having made prior investment decisions or having adopted prior investment decisions made by the computer.

2.1.1. Method

2.1.1.1. Participants and design

Fifty-one individuals (39 female) with a mean age of 23.0 (range 19 to 41, $SD = 4.9$) recruited at a German university participated. Personal responsibility was manipulated as a within-participants factor with two levels (personally responsible vs. assigned). A prior investment factor resulted from the repeated investment in a round where a given decision was preceded by one to four prior investments. Losing probability was calculated as a quasi-experimental predictor for each stage of the 100 decision trials. The dependent variable is the participant's decision to invest or opt-out in any given trial.

2.1.1.2. Procedure

The study was conducted as a laboratory experiment with each participant working alone and a maximum of 8 participants in any session. In each session, the participants first played the card game, and then demographic variables were assessed. Finally, the participants were debriefed, thanked, and paid 4 Euros and the performance-dependent bonus (potential range 0 to 7.80 Euros).

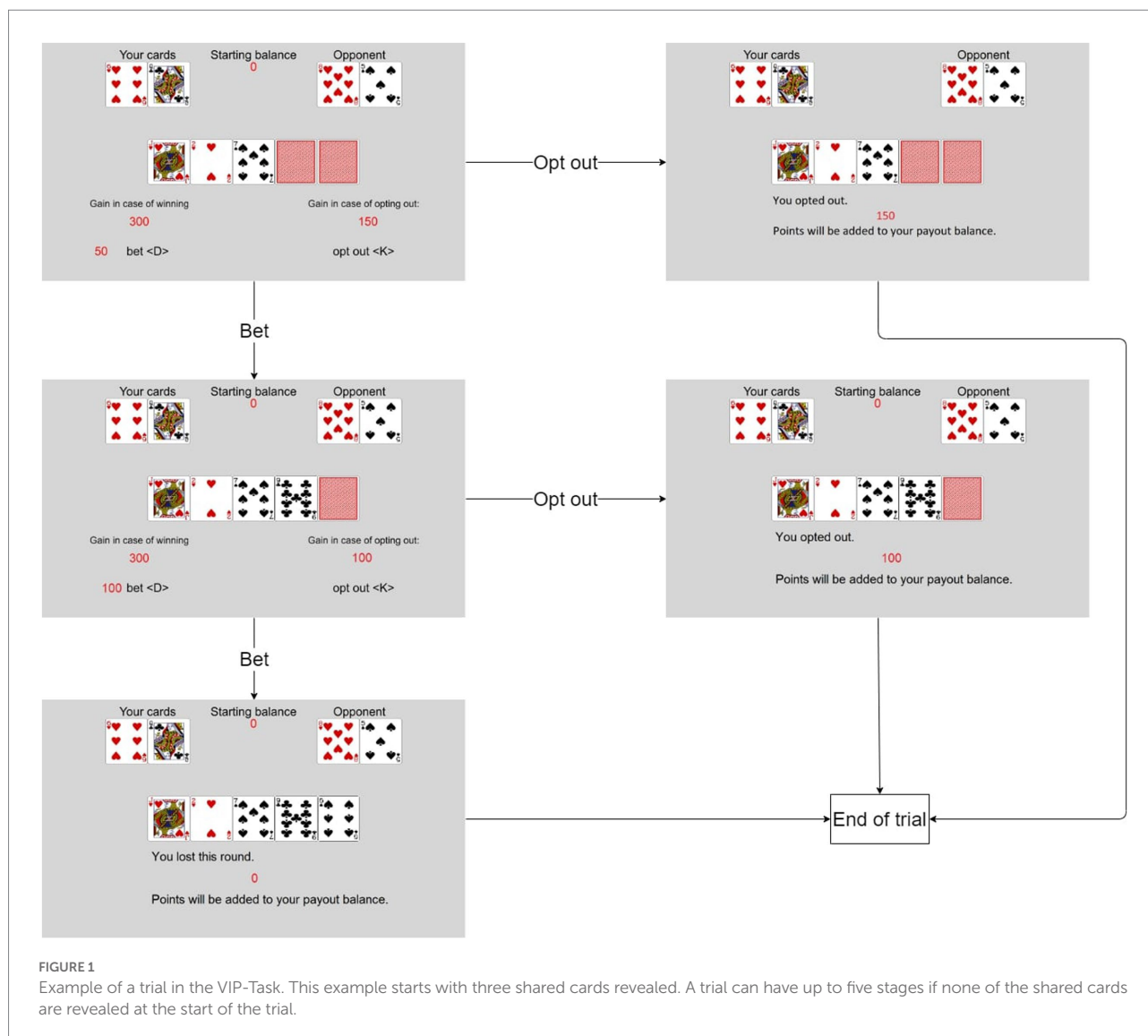
2.1.1.2.1. VIP-task

The VIP-Task was designed as a measure of incentivized sequential decisions in an uncertain and risky environment. It has previously been used to assess EoC (Doerflinger et al., 2017) and was implemented using PsychoPy (Peirce et al., 2019). The rules of the VIP-Task are based on the "Texas hold'em" variant of poker. In the VIP-Task, participants play against the computer (referred to as the opponent) and each trial contributes to the potential bonus. Trials are randomly generated for each participant. At the beginning of each trial, the participants had a fixed amount of points they could bet (i.e., invest) during the trial (see Figure 1 for an exemplary trial). Both the participant and the opponent have two individual cards in each trial (i.e., their hand). In addition, five shared cards can add to the value of both players' hands. All cards are randomly drawn from a list of standard poker cards. Usually, all shared cards are hidden at the beginning of each trial. Participants decide whether to invest further or to opt out.

If they decided to invest, one of the hidden shared cards was revealed. The cost of investing in a trial increased with each revealed card. If they decided to opt-out, the current trial ended, and the remaining points were added to the participants' payout. If participants invested until all shared cards are revealed, their cards were compared against their opponent's by using standard poker rules. The value of the best five cards out of a player's two individual cards and the five shared cards are compared to decide the winner. The five cards can be any combination of hand and shared cards. The points added to the payout depend on the outcome of this comparison: if the participant loses, no points are added; if the participant wins, twice the invested points are added; if the comparison ends in a tie, the invested points are added. After this comparison, the trial ends. At the end of each trial, a screen informs the participants of the results (i.e., win, lose, tie, opt out) and the points added to the payout.

Explicit probabilities were not shown to participants. We use the probability of losing (if the round is played until the end) as an independent variable ranging from 0 to 1. It is calculated based on the revealed cards. The probability of winning is complementary to it and using it produces the reversed pattern of results. Ties are rare and do not affect the overall results.

The participants were given a reference sheet explaining the standard poker rules. They could use the sheet throughout the experiment. After reading the rules of the game, the participants played three practice trials before the main task started. The task had 100 trials in total. Each trial had up to five stages in which the participants could invest up to 310 points, costing 10 points for



the first investment, 20 points for the second, 40 points for the third, 80 points for the fourth, and 160 points for the fifth.

2.1.1.2.2. Personal responsibility manipulation

In one half of the trials, the participants were personally responsible for each decision. In the other half, the trials started with some shared cards already revealed (between 1 and 4) and points invested accordingly. All trials were presented in random order. The trial stages after the participants had already made an investment decision themselves were coded as personally responsible. The trial stages directly following a computer-made investment decision were coded as assigned decisions, as the participants had no control over the invested points before their decisions. The first stage of a trial without revealed shared cards is not included because, at this stage, no prior investment had been made. Based on the assumption that when participants had made multiple prior investments, they shared more personal responsibility for any given situation, the stage of the trial

corresponding to the number of prior investments (between 1 and 4) was included as a further indicator of the degree of responsibility.

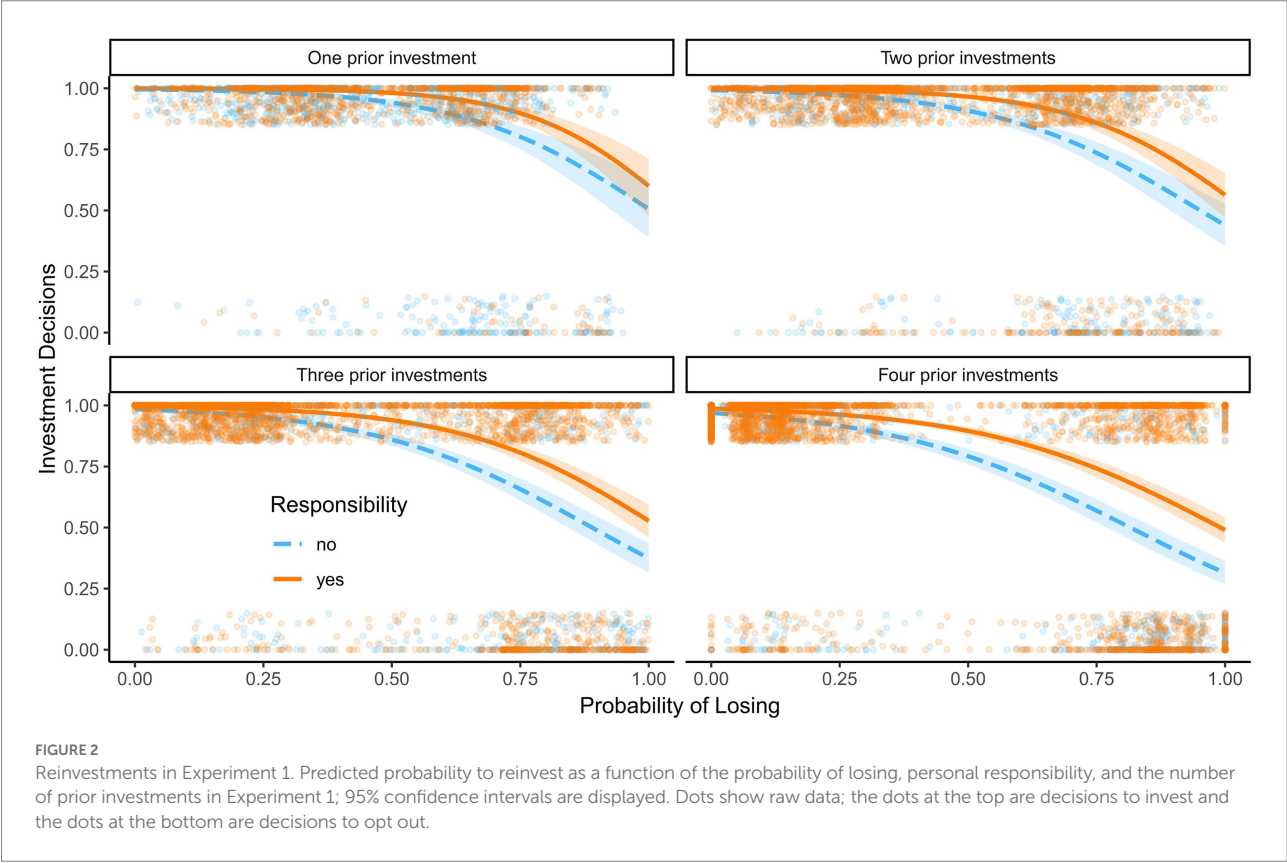
2.1.2. Results

A mixed effects logistic regression was used to predict the probability of investments. Independent variables were the probability of losing, personal responsibility, and the number of prior investments, as well as all their interaction terms. Random effects for participants and trial numbers were included. The model is summarized in Table 1. The main effect of the probability of losing, $z = -8.62$, $p < 0.001$, was significant, indicating a higher likelihood of reinvestments if the probability of losing was low. Personal responsibility, $z = 3.32$, $p < 0.001$, also had a significant main effect – participants were more likely to reinvest after decisions they were personally responsible for than after assigned decisions. Moreover, the number of prior investments, $z = -6.30$, $p < 0.001$, was a significant negative predictor of reinvestment, indicating that opting out was more prevalent in later than earlier

TABLE 1 Mixed effects logistic regression estimating the decision to bet in Experiment 1.

Variable	OR	B	SE B	z	p
Intercept	492.75	6.20	0.46	13.37	<0.001
Probability of losing	0.003	−5.90	0.68	−8.62	<0.001
Personal responsibility ^a	10.18	2.32	0.70	3.32	<0.001
Prior investments	0.50	−0.69	0.11	−6.30	<0.001
Probability of losing × Personal responsibility ^a	0.13	−2.07	1.03	−2.01	0.045
Probability of losing × Prior investments	1.52	0.42	0.16	2.58	0.010
Personal responsibility ^a × Prior investments	0.70	−0.36	0.15	−2.27	0.023
Probability of losing × Personal responsibility ^a × Prior investments	1.62	0.48	0.23	2.07	0.038
Random effects (s ²)	Participant: 0.35	Trial: <0.01			

^aAssigned = 0, responsible = 1.



stages of each trial. All two-way interaction effects were significant: the probability of losing and personal responsibility, $z = -2.01$, $p = 0.045$, the probability of losing and the number of prior investments, $z = 2.58$, $p = 0.010$, and personal responsibility and the number of prior investments, $z = -2.27$, $p = 0.023$. These effects were qualified by a significant three-way interaction effect between the probability of losing, personal responsibility, and the number of prior investments, $z = 2.07$, $p = 0.038$. For simple slope analyzes and Johnson Newman intervals see the supplemental materials.

The predicted probabilities are visualized in Figure 2. Participants were more likely to bet on a given hand at all levels of

probability of losing when they were personally responsible for prior decisions. The difference between decisions for which the participants were personally responsible and assigned decisions was larger at higher probabilities of losing. Furthermore, the responsibility effect was more pronounced with multiple prior investments.

2.1.3. Discussion

Participants made decisions in line with the rules of the task to increase their payout; they were more likely to reinvest if the probability of losing was low. Concerning personal responsibility,

we found that participants were indeed more likely to reinvest in a hand if they were personally responsible for prior decisions, especially when the probability of losing was high. Note that we compare self-chosen and (random) computer chosen precursor decisions on the same level of success/failure probability; despite expecting self-chosen participant decisions to be more aligned to probabilities than the random computer decisions in general, comparing responses at the same probability levels removes this difference in the quality of decisions.

The smaller effects at low losing-probability levels are most likely the result of a ceiling effect, as both decisions for which the participants were personally responsible and assigned decisions are close to 100% continue/invest decisions. Our results are in line with previous evidence (Sleesman et al., 2012) that personal responsibility increases EoC. We observed larger personal responsibility effects at later stages of a trial where more prior investments had been made, indicating that the degree of personal responsibility is positively related to reinvestments.

There is a limitation, however: defining the degree of personal responsibility in terms of the number of prior investments is confounded with outcome uncertainty. Negrini et al. (2020) observed no responsibility effects on reinvestments in an incentivized experimental study. However, in Negrini et al.'s study both the cost of the initial investment and the likelihood that investments would ultimately lead to success were unknown to the participants when the initial investment was made. This very high degree of uncertainty might have undermined potential responsibility effects, because the decision makers could not estimate the quality of the first investment.

In a series of multiple reinvestments, each investment can also be understood as costly information search to reduce uncertainty and the sampled information can inform search decisions (Cohen and Erev, 2021). The more prior investments have been made, the more shared cards are revealed in the VIP-Task. While the risk of losing can increase or decrease with each investment, the new shared cards reveal previously unavailable information, thus reducing uncertainty. Some decision-makers may seek to reduce uncertainty or avoid investments in more uncertain situations (Nau, 2006).

The design of Experiment 1 does not allow for a fair evaluation based on the expected value of the participants' decisions, because the expected value of decisions early in the sequence depends on the participants' future decisions. Thus, additional evidence is needed to evaluate, when exactly personal responsibility is beneficial (to avoid prematurely opting out) or detrimental (EoC) to decision quality. To exclude uncertainty avoidance as an alternative explanation, and to allow an evaluation of the expected value, in Experiments 2 and 3, we only focused on the last decision in the sequence and modified the Poker task accordingly. That is, for all critical decisions, the same number of cards was revealed. The level of uncertainty was thus held constant.

Finally, the assigned decision trials were randomly generated decision situations with prior investments made by the computer. These situations are equivalent to decision

situations following a sequence of arbitrary prior decisions. The participants were fully informed about the task procedure and the randomness of the assigned decision trials. Participants may have heuristically responded negatively to the random computer decisions. Thus, the personal responsibility effect could partially result from the negative connotation conveyed by knowing that the previous decision was made randomly. To account for the possibility of such a heuristic, we explicitly manipulated the quality of the assigned decisions in Experiment 2.

2.2. Experiment 2: Personal responsibility, preferences, and decision quality

To further test the personal responsibility hypothesis, we changed three aspects of the procedure: (1) We measured preferences to continue investing/opting out in the initial investment situation for each trial before the decision task. If we apply the arguments raised by Schulz-Hardt et al. (2009) to the results of Experiment 1, it seems possible that the responsibility effects we observed are due to the participants' preferences for the initial card combinations. To account for this alternative explanation and also to test whether responsibility effects do occur in parallel to preference effects, we measured the participants' preferences for each card combination that would be played (both for trials with and without personal responsibility). (2) In addition to assigned decision trials in which the initial decision was made randomly (equivalent to Experiment 1), we added assigned decision trials in which the initial decision was made optimally by the computer (based on the probability of losing). (3) The modified task also consisted of only one reinvestment decision per trial (at the last stage) after the initial investment decision. We thus avoided confounding uncertainty and responsibility. The expected value of each reinvestment decision made by the participants is now independent of future decisions, which allows us to analyze it as a dependent variable without taking into account the participants' potential future decisions. We used the expected value of reinvestments as a plausibility check. The assigned optimal decisions should lead to a higher expected value than the assigned random decisions and participants should be more likely to indicate a preference for initial investments when the probability of losing is low. With only one critical reinvestment decision in the sequence, the probability of losing at which reinvesting is prudent can also be easily determined based on the expected value as a benchmark for EoC vs. prematurely opting out. Because the second investment costs twice as much as the first and the payout in case of winning is double the investment, the expected value for reinvesting is higher than for opting out, if the probability of losing is less than $2/3$.

2.2.1. Method

2.2.1.1. Participants and design

Forty-nine participants (27 female) with a mean age of 23.7 (range 19 to 43, $SD = 4.2$) were recruited at a German university. Personal responsibility was manipulated within-participants as one of three trial types (personal responsibility: personally responsible vs. random assignment vs. optimal assignment). Preferences were measured as a dichotomous variable (preference vs. no-preference) for each trial before the card game started. Losing probability was calculated as a quasi-experimental predictor for each decision trial. The dependent variable is the participants' decision to invest or opt out in any given trial. The study was conducted in the laboratory with up to 8 participants per session. The participants first indicated their preferences for each trial, then they played the card game, and thereafter provided demographic information. Finally, they were debriefed, thanked, and paid 5 Euros and a bonus dependent on the card game (potential range: 0–8.2 Euros).

2.2.1.2. VIP-task

Standard poker rules were explained to the participants, and they were given a reference sheet that could be used during the experiment. In contrast to Experiment 1, all trials of the VIP-Task started with three of the five shared cards revealed. Therefore, in a standard trial, participants had to decide twice whether to invest or not – the first decision pertained to the initial investment and the second decision to the reinvestment. In each trial, 150 points could be invested. The first investment cost 50 points; the second investment cost 100 points. Decisions for the second investment are the dependent variable. Before the participants played the card game, we measured their initial preferences for each trial. The participants played 150 trials of the game, 50 trials for each within-participant condition.

2.2.1.2.1. Preference measure

All trials were generated at the beginning of the experiment and randomly mixed. Before the card game was played, the participants were confronted with the initial investment situation (with three revealed shared cards) and asked whether they would hypothetically invest or not – this was done for each of the 150 trials presented in random order. This procedure allows for assessing initial preferences (as a dichotomous measure: preference vs. no preference) for all trials, including non-responsible trials.

2.2.1.2.2. Personal responsibility manipulation

At the beginning of each trial, the participants saw their hand cards, the opponent's hand cards, and three revealed shared cards. This was the same configuration that they had rated before on the preference task. Personally responsible trials played out the same way as in the standard VIP-Task: the participants decided whether to invest in the shown cards or not. If they invested, the fourth shared card was revealed, and they had to decide whether to reinvest or opt out. Each trial started with 150 points available to

the participants to invest. If they opted out the first time around, the trial ended, and 150 points were added to their payout. If they opted out after having invested once, 100 points were added to their payout. If participants decided both times to invest in their cards, the fifth shared card was revealed, and the participants' cards were compared to the opponent's cards according to standard poker rules. Winning this comparison resulted in 300 points added to the participants' balance; losing the comparison resulted in 0 points added, and if the participant and the opponent tied, 150 points were added.

Personal responsibility and the quality of prior decisions were varied by introducing two computer "advisors" to the participant. In some rounds, the participants played from the beginning, in others, one of the computer advisors made the initial investment decision. The advisors either invested points to reveal a hidden shared card or opted out. The participants always had to make the reinvestment decisions. Rounds in which the advisors opted out did not count toward the payout and were not counted as trials played. The rules to determine the winner for the trials were the same, whether the participants made the first decision or one of the advisors. In the random assignment trials, the computer would invest with a probability of 50% and opt out with a probability of 50%, independent of the card values. In the optimal assignment trials, the computer would invest when the participants' chances of winning or a tie were better than the chances of losing. If the chances of losing were higher, the computer would opt out. If the advisors invested in the cards, a new shared card was revealed. This changed the chances of winning. See [Figure 3](#) for examples of the assigned decision trials. The 5th and last card decision was then always decided by the participant.

The procedure was thoroughly described to the participants on the screen before the card game task started. The random assignment advisor was named "Random," and the optimal assignment advisor was named "Maximize." The advisors' names were colored green versus blue (counterbalanced) to make it more intuitive to identify them. In the first stage of each random and optimal assignment trial, a message was displayed, indicating which advisor would make the first decision: "This time, the following computer advisor makes the first decision for you: Random/Maximize."

2.2.1.2.3. Expected value

Because the participants made only one reinvestment decision in each trial, the expected value of this decision could be calculated without further assumptions about future decisions. If participants opted out in the second decision, the expected value of that decision was 100 points. If they decided to invest, the expected value was calculated as $300 * (1 - p_{\text{losing}})$.

2.2.2. Results

2.2.2.1. Expected value

A mixed linear model (full summary in supplemental materials) was calculated to predict the expected point value of the

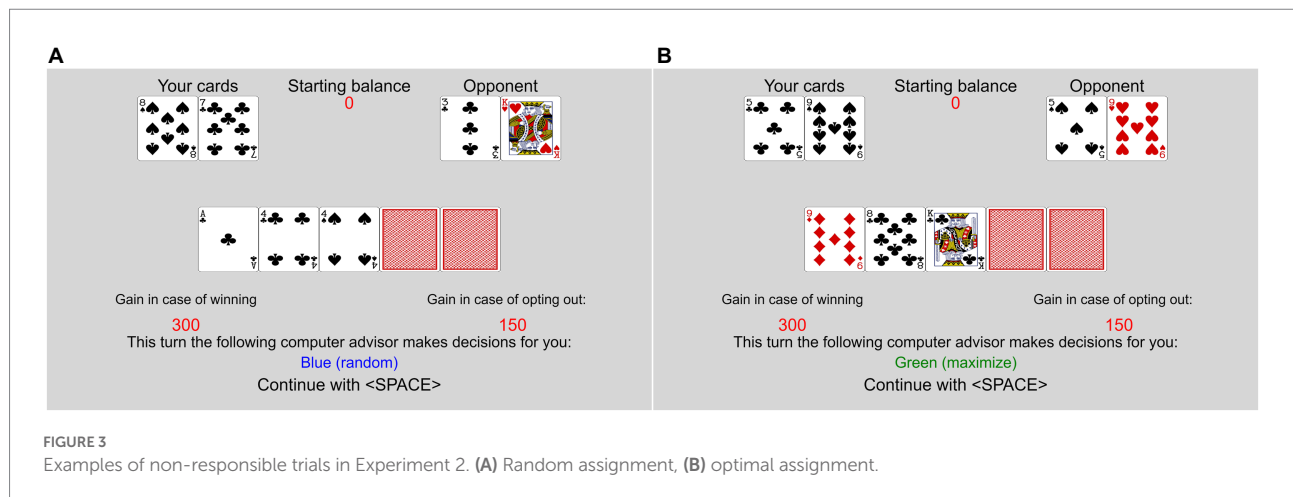


TABLE 2 Mixed effects logistic regression estimating the decision to bet in Experiment 2.

Variable	OR	B	SE B	z	p
Intercept	4.02	1.39	0.17	8.15	<0.001
Probability of losing	0.04	-3.18	0.23	-14.06	<0.001
Trial type optimal ^a	1.67	0.51	0.21	2.40	0.017
Trial type responsible ^a	5.87	1.77	0.37	4.73	<0.001
Preference ^b	3.46	1.24	0.21	5.87	<0.001
Probability of losing × Trial type optimal ^a	1.27	0.24	0.45	0.54	0.591
Probability of losing × Trial type responsible ^a	0.75	-0.29	0.51	-0.57	0.568
Probability of losing × Preference ^b	0.66	-0.42	0.33	-1.29	0.196
Trial type optimal ^a × Preference ^b	0.83	-0.19	0.29	-0.67	0.502
Trial type responsible ^a × Preference ^b	0.73	-0.32	0.48	-0.65	0.515
Probability of losing × Trial type optimal ^a × Preference ^b	0.68	-0.38	0.58	-0.65	0.514
Probability of losing × Trial type responsible ^a × Preference ^b	0.64	-0.44	0.69	-0.64	0.523
Random effects (s ²)	Participant: 0.25				

^aVariables are dummy coded, random assignment trials are coded 0 on both variables; ^bno-preference = 0, preference = 1.

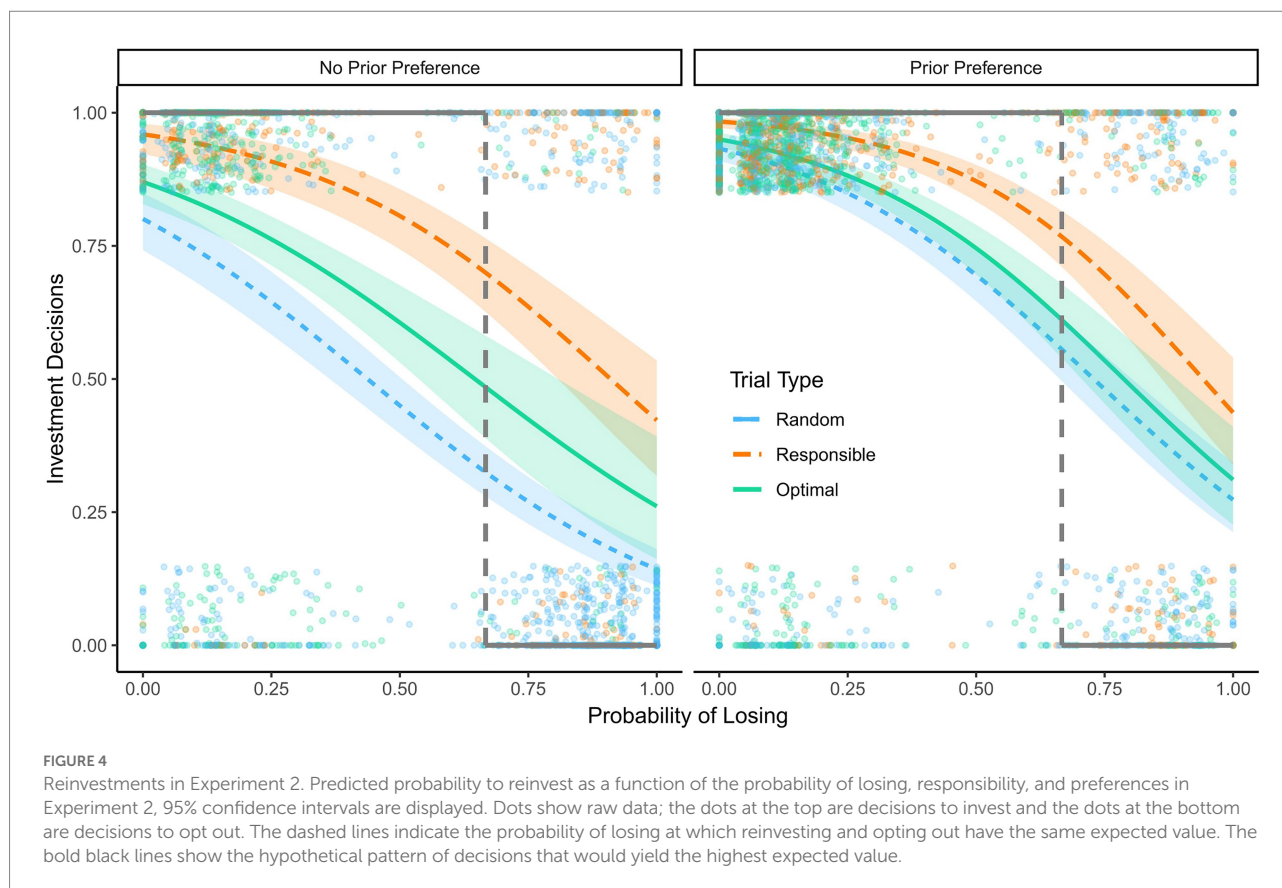
participants' decision in each trial based on personal responsibility, preferences, and the interaction terms with random effects for participants. The main effect of preferences was significant, $t(5066.7) = 18.92$, $p < 0.001$. Compared to randomly assigned trials, the expected value was significantly higher in personally responsible trials, $t(5065.2) = 8.29$, $p < 0.001$, and in optimal assignment trials, $t(5066.7) = 17.58$, $p < 0.001$. These main effects were qualified by significant interaction terms of preference and personally responsible trials, $t(5064.4) = -4.07$, $p < 0.001$, and preference and optimal assignment trials, $t(5070.4) = -7.97$, $p < 0.001$. The difference between personally responsible, randomly assigned and optimally assigned trials was smaller in trials for which the participants had indicated a preference to invest.

2.2.2.2. Reinvestment decision

We calculated a mixed effects logistic regression (summarized in Table 2) to predict investments based on the probability of losing, personal responsibility (dummy coded with random

assignment trials as the baseline), and initial preferences. Interaction terms of the independent variables and random effects for participants were included. The main effect of the probability of losing, $z = -14.06$, $p < 0.001$, was significant, indicating that participants were more likely to invest when the probability of losing was low. The dummy coded personal responsibility variables also showed significant main effects, indicating that participants were more likely to invest in the optimal assignment trials, $z = 2.40$, $p = 0.017$, and the personally responsible trials, $z = 7.73$, $p < 0.001$, than in the random assignment trials. None of the interaction terms were significant, $|z| < 1.30$, $ps > 0.196$.

A follow-up contrast test using a mixed effects logistic regression to compare only personally responsible trials versus optimal assignment trials with preferences as a control variable, and including random effects for participants, revealed in addition to the main effects of probability of losing, $z = -30.44$, $p < 0.001$, and preferences, $z = 10.78$, $p < 0.001$, a main effect of personal responsibility, indicating that participants were significantly more



likely to reinvest in personally responsible trials than optimal assignment trials, $z = 4.96$, $p < 0.001$.

The predicted probabilities are visualized in [Figure 4](#). Participants were least likely to invest in their cards in trials in which the random advisor had made the first decision. Investments were more likely in trials in which the optimal advisor had made the first decision. The highest rate of reinvestments was in trials in which the participants had made the first decision themselves (i.e., personally responsible trials). These differences were unaffected by the probability of losing and participants' initial preferences.

2.2.3. Discussion

The expected value of the participants' decisions was lowest in the random advisor trials. This is not surprising, as the initial decision quality of the random advisor is lower than that of both the participants and the optimal advisor. The expected value of the decisions in personally responsible trials was higher than in random assignment trials but still lower than in optimal assignment trials. Thus, the deviation induced by participants' personal responsibility did lower their expected outcome compared to the optimal advisor.

The observed main effect of the probability of losing on reinvestments demonstrates that the participants understood the task they had to perform. They were more likely to keep investing in good cards than in bad cards. Preferences had an incremental

effect on investments. Irrespective of the actual probability of losing, the participants were more likely to invest if they had indicated a preference for the initial decision, validating the preference measure and replicating prior evidence ([Schulz-Hardt et al., 2009](#)) that preferences are a factor in EoC. The effect of a preference for initial investments was present across all levels of the probability of losing – when the situation was unfavorable, an initial preference decreased decision quality (i.e., making EoC more likely), but when the situation was favorable, initial preferences increased decision quality (i.e., making prematurely opting out less likely).

We observed personal responsibility effects beyond measured preferences; responsibility remained a significant predictor even when controlling for preferences. Thus, we have some indication that preferences need to be considered but may not fully account for responsibility effects in reinvestment decisions. Personal responsibility made participants generally more likely to reinvest. At high probabilities of losing this effect resulted in decreased decision quality and at low probabilities of losing it increased decision quality. The dashed line in [Figure 4](#) indicates the probability of losing at which reinvesting and opting out have the same expected value. To maximize the expected value, one should always invest in situations to the left of the line and always opt out in situations to the right (illustrated by the bold line in the figure). The confidence intervals for the participants'

investment decisions do not include the optimal strategy at any level of probability. The participants abandoned reinvestments too early, because the likelihood of reinvesting is below the optimal pattern for situations where reinvestment increases the expected value, even at a very low probability of losing. They also demonstrated EoC, because the likelihood of reinvesting was higher than the optimal pattern for situations where reinvestment decreased the expected value, even at a very high probability of losing.

Participants were also more likely to invest when the optimal advisor had made the initial decision rather than the random advisor. This pattern may be a reasonable heuristic based on the negative connotation of “random choices” in this context. However, the participants were even more likely to invest if they were personally responsible for the initial decision themselves compared to the optimal advisor. We found these effects controlling for preferences. As the participants knew that the optimal advisor maximized the expected value, more reinvestments after their own initial decisions over the optimal advisor’s is a deviation from a normative expected utility perspective.

2.3. Experiment 3: Personal responsibility and framing effects

According to prospect theory accounts of EoC (Thaler, 1980; Whyte, 1986; Soman, 2004), loss framing should increase escalation behavior. Decision makers might construe invested resources as potential losses and seek to minimize them by making further risky investments. This however constitutes a risky option: failure means that the second investment will also be lost, whereas success could mitigate the loss of the investments made. Accordingly, we hypothesized that participants would be more likely to invest if the task is presented in a loss frame than a gain frame. Using incentivized decisions with real feedback goes beyond the past research on framing effects on EoC, which relied on hypothetical investment scenarios. We found such framing effects in a pilot study that followed a procedure similar to Experiment 1; we manipulated whether the outcomes were presented as gains or losses (see supplemental material). Based on this pilot study, in Experiment 3, we combined the three determinants of EoC: personal responsibility, preferences, and gain vs. loss framing. We also included the numeracy scale (Lipkus et al., 2001) and the gambling and investing risk-taking propensity subscales of the DOSPERT scale (Weber et al., 2002) as individual difference measures. Numeracy might be beneficial to participants when judging the probability of losing, while a general disposition for risk-taking might make reinvestments more likely in the present task. The experiment was preregistered on [osf.org](https://osf.io/zg7xm/?view_only=b913923de9ce4dccbddd6929678a398c): https://osf.io/zg7xm/?view_only=b913923de9ce4dccbddd6929678a398c.

2.3.1. Method

2.3.1.1. Participants and design

Eighty participants (23 female) with a mean age of 23.3 (range 18–46, SD = 5.0) were recruited online via [prolific.org](https://www.prolific.org) (Palan and Schitter, 2018). The sample included participants from Europe, Asia, Northern America, and Middle America. The experiment had a 3-within (personal responsibility: personally responsible vs. random assignment vs. optimal assignment) by 2-between (framing: gain vs. loss) design, with preferences as an additional independent variable measured for each trial. Probability of losing was calculated for each trial as a quasi-experimental factor. The dependent variable is the participants’ decision to invest or opt out in any given trial.

2.3.1.2. Procedure

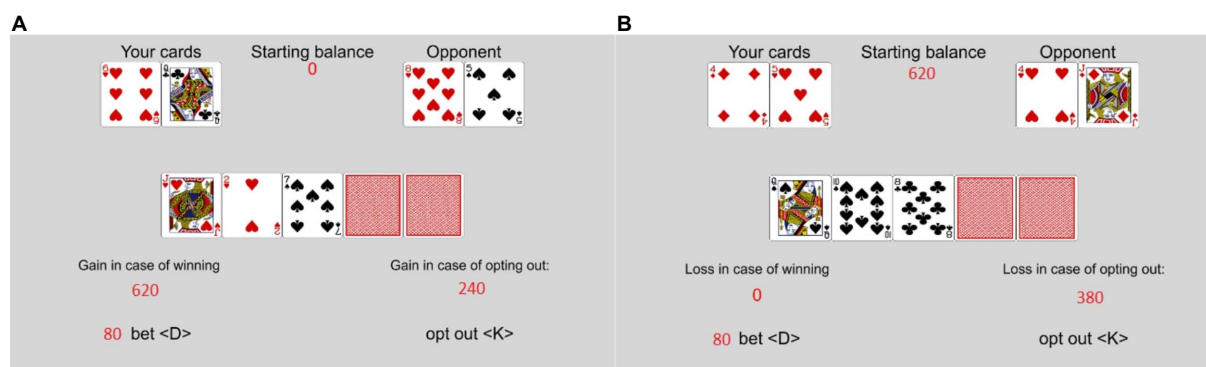
The experiment was conducted as an online experiment using the JavaScript functions of PsychoPy (Peirce et al., 2019) on the pavlovio.org platform. We asked participants first to fill out the numeracy scale (Lipkus et al., 2001) and the gambling and investing risk-taking propensity subscales of the DOSPERT scale (Weber et al., 2002). Then we measured preferences in the same way as in Experiment 2. Finally, the participants played the VIP-Task. They were paid 5 GBP and a bonus dependent on the card game (potential range: 0–5.4 GBP). Demographic information was obtained from [prolific.org](https://www.prolific.org).

2.3.1.2.1. VIP-task

The participants played 90 trials of the VIP-Task, which was structured in line with Experiment 2. Personal responsibility was manipulated the same way: in one third of the trials a random advisor made the first investment, in another third an optimal advisor made the first investment, and in the final third the participants made the first investment decision. The VIP-Task was presented with outcomes framed either in terms of gains or losses. Participants could display the Poker rules on the screen by pressing a key at any time during the experiment.

2.3.1.2.2. Framing

See Figure 5 for an example of trials in the gain and loss frame conditions. In the *gain frame* condition, participants were informed that they started the game with 0 points and could gain between 0 and 300 points each turn. The maximum gains were 27.000 points in total. In the *loss frame* condition, the participants were informed that they started the game with 27.000 points and could lose between 0 and 300 points each turn. The maximum losses were 27.000 points. The game played out the same way, and the bonus was calculated the same way in the two framing conditions. The only difference was in the presentation as gains versus losses. Five thousand points were equivalent to 1 GBP, to be paid rounded mathematically to one penny. In the gain frame condition, the optimal advisor was named “maximize,” in the loss frame condition it was named “minimize.” In each trial, the participants’ starting balance for the trial was displayed in the top

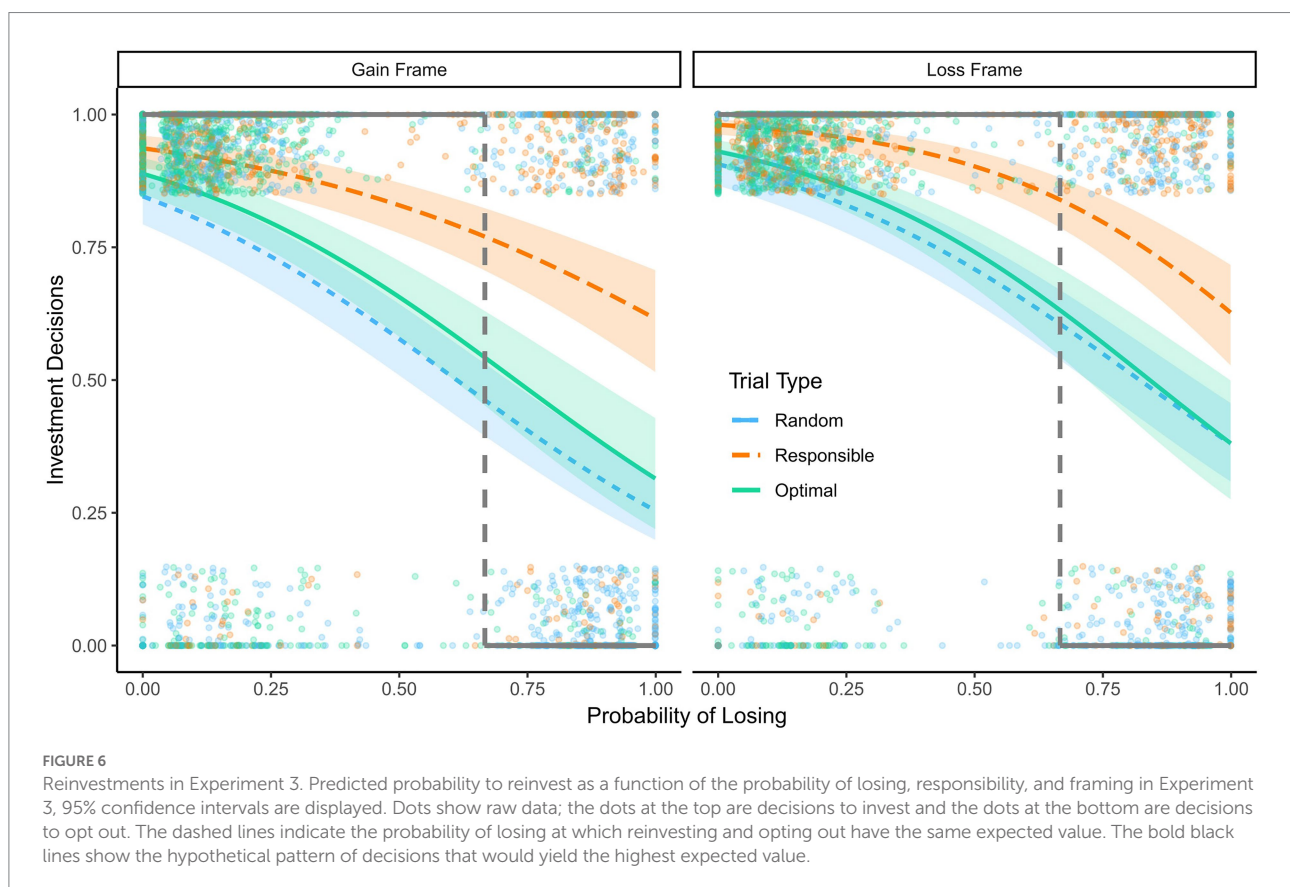


with random assignment trials as the baseline), initial preferences, and the interaction terms of these variables replicates the pattern of results found in Experiment 2. Because the addition of Experiment 3 is the framing manipulation, we focus here, as pre-registered, on analyzes including main effects and interaction terms of framing, treating preferences for prior investments as a covariate.

TABLE 3 Mixed effects logistic regression estimating the decision to bet in Experiment 3.

Variable	OR	B	SE B	z	p
Intercept	139.77	2.40	0.76	3.14	0.001
Probability of losing	0.06	-2.78	0.20	-14.11	<0.001
Trial type optimal ^a	1.45	0.37	0.18	2.07	0.039
Trial type responsible ^a	2.69	0.99	0.26	3.82	<0.001
Preference ^b	2.10	0.74	0.08	9.56	<0.001
Framing ^c	1.72	0.54	0.26	2.06	0.039
Numeracy	0.94	-0.06	0.07	-0.85	0.396
DOSPRT	0.94	-0.06	0.14	-0.42	0.672
Probability of losing × Trial type optimal ^a	0.93	-0.07	0.35	-0.22	0.828
Probability of losing × Trial type responsible ^a	1.77	0.57	0.37	1.54	0.125
Probability of losing × Framing ^c	0.99	-0.01	0.30	-0.04	0.967
Trial type optimal ^a × Framing ^c	0.99	-0.01	0.28	-0.03	0.978
Trial type responsible ^a × Framing ^c	1.95	0.67	0.44	1.52	0.130
Probability of losing × Trial type optimal ^a × Framing ^c	0.77	-0.26	0.51	-0.53	0.598
Probability of losing × Trial type responsible ^a × Framing ^c	0.31	-1.18	0.60	-1.99	0.047
Random effects (σ^2)	Participant: 0.25				

^aThe variables are dummy coded, random assignment trials are coded 0 on both variables; ^bno-preference = 0, preference = 1; ^cgain = 0, loss = 1.



random assignment vs. optimal assignment advisor) and between the assignment trials and the personally responsible trials in the loss frame condition compared to the gain frame condition. This

seems to be a consequence of the higher likelihood of participants in the loss frame condition to invest in the assignment trials, particularly when the probability of losing was high.

2.3.3. Discussion

2.3.3.1. Plausibility check

Framing had no significant effect on the expected value of the decisions. A likely reason for this null-effect is that participants were more likely to reinvest in the loss frame than the gain frame condition irrespective of the probability of losing. Across the probability range, the costs and benefits of the framing conditions canceled each other out. For the remaining effects the same pattern of results for preferences and personal responsibility as in Experiment 2 were observed. The expected value was highest in optimal assignment trials and lowest in random assignment trials, with personally responsible trials sitting in-between. Preferences were a positive predictor of expected value. The difference between trial types was smaller for trials that the participants preferred. This pattern shows that preferences were adequately measured and the trial types worked as intended.

Participants were more likely to invest in bad cards, when they were personally responsible for prior investments, than when they were not. They were also less likely to invest in good cards when dealing with assigned decisions (see [Figure 6](#)). Both of these trends observed on the level of decision-making lead to a suboptimal expected value of the reinvestment decisions, reflecting the two types of decision errors investigated in the present research.

Numeracy positively predicted the expected value, but risk taking did not. This is plausible because a risk main effect can yield better outcomes in some trials and worse ones in others. Being better able to accurately judge the probability of losing, which is related to numeracy, is likely beneficial to performing well on the task at hand – further indicating the task's validity. In contrast, a general preference for risky or safe options is neither an advantage nor a disadvantage.

2.3.3.2. Responsibility, decision quality, preferences, and framing

The present study replicated the central findings from Experiments 1 and 2. Thereby, we provide another replication of personal responsibility effects beyond preferences. Preference effects are observed parallel to framing effects. Framing effects were evident, as participants were more likely to bet in a loss than a gain frame. Besides these central findings, somewhat surprisingly, the personal responsibility-framing interaction is not in line with the findings reported by Rutledge ([Rutledge, 1995](#)). Personal responsibility effects were not enhanced in the loss-frame condition; actually, there was a smaller difference between the personal responsibility conditions in the loss framing condition. A possible explanation is that participants in the loss frame condition were more likely to bet in assignment trials than participants in the gain frame condition, especially when the probability of losing was high. In personally responsible trials, the participants were more likely to bet both in the gain as well as the loss frame condition compared to the assigned trials, but the probability to reinvest did not differ between framing conditions for the responsible trials. This led to a smaller difference between personally responsible and assignment trials in the loss condition than the gain condition, which resulted in the observed interaction effect.

The probability to reinvest for personally responsible trials was not diminished in the loss frame condition; instead, participants in the loss frame condition were more likely to reinvest in assigned trials than participants in the gain frame condition. This pattern may be driven by a ceiling effect for the personally responsible trials. Although the personally responsible decisions were not close to the actual ceiling of 100%, there is probably a limit to participants' mindlessly continuing at very high probabilities of losing. With a lower starting point for the assigned trials, there was more room to be pushed toward risky investing caused by loss framing. Also, there may generally be a limit to the degree of escalation (i.e., reinvestment at high losing probabilities) that can be expected for any given decision problem, and personal responsibility effects may be sufficient to push decision makers to that limit. Both construing the situation as a choice between losses and being personally responsible for prior decisions increase the probability of investments, but this does not mean that decision makers will completely disregard available information about the likelihood of success or failure.

3. General discussion

Decisions to continue with a previously chosen course of action or to quit can be critical for individuals (financially, personally) and even societies (e.g., when to continue or withdraw regulations to contain an ebbing pandemic). As such decisions can be highly consequential, it is important to understand psychological factors that can influence them. Different determinants of reinvestment decisions have been proposed in the literature on escalation of commitment (EoC) and some have been questioned (e.g., [Schulz-Hardt et al., 2009](#)). We argued that the determinants were often tested in hypothetical scenarios and by using bogus feedback. This is a critical limitation as anticipated responses do not necessarily line up with responses made in the actual situation ([Nordgren et al., 2009](#)). Other researchers agree with this assessment ([Roth et al., 2015](#); [Negrini et al., 2020](#)). Even more, a recent study ([Negrini et al., 2020](#)) showed that determinants of reinvestment (amount of prior investment) can differently affect hypothetical and financially consequential decisions. They observed that higher prior investments led to escalation of commitment in hypothetical scenarios, but a reverse effect in the incentivized task. The authors also varied responsibility for previous investments. They found no effect of responsibility on reinvestments in their incentivized task. These findings, which are seemingly inconsistent with the EoC literature, suggest that additional experiments using incentivized tasks are needed to probe whether responsibility effects occur when real money is on the line.

We use a behavioral decision task with real financial consequences ([Doerflinger et al., 2017](#)). We validated the previously proposed responsibility and framing factors, finding evidence for their effects even when controlling for alternative explanations (preferences). When comparing with objectively (i.e., probability based) optimal decisions, our studies indicate that the presence of these factors (responsibility or loss framing) are not only relevant

for continuing to invest beyond what is optimal, but that their absence can also lead to dropping out earlier than what is optimal.

3.1. Personal responsibility and preferences for the initial investment

Schulz-Hardt et al. (2009) observed in two studies that personal responsibility effects on EoC disappeared after statistically controlling for initial preferences. In our studies, preferences increased the probability of continuing investing, irrespective of success versus failure. However, preferences did not impact the influence of personal responsibility on reinvesting in the present studies. Instead, personal responsibility and preferences had an additive effect on the likelihood of reinvesting both in unfavorable situations (i.e., escalation of commitment) and favorable situations (i.e., avoiding prematurely opting out).

3.2. Personal responsibility and framing

According to the prospect theory account of EoC, framing in terms of gains or losses should influence the degree of escalation of commitment (Soman, 2004). In Experiment 3, the likelihood of reinvestments was lower when outcomes were framed as gains compared to losses. In line with the hypothesis that loss framing is a relevant factor in driving reinvestments, gain framing decreased reinvestments. In Experiment 3, where framing and personal responsibility were varied, framing moderated the effect of personal responsibility. A loss frame increased the probability of reinvesting for trials without personal responsibility for the initial decision; the difference between personally responsible and assignment trials was smaller in the loss frame condition than in the gain frame condition.

Based on Rutledge (Rutledge, 1995), we predicted that loss framing should have magnified responsibility effects as decision makers should be even more hesitant to lose something based on their own prior decision. However, we did not observe such an effect. This might be due to a ceiling effect of the already high probability of continuing to invest when participants were personally responsible so that there was less room for loss framing to drive this further. Beyond this unexpected effect, however, we can conclude that framing has an effect on reinvestments overall.

3.3. Decision quality

Escalation of commitment is often referred to as a decision bias, implying that it is not rational (Brockner, 1992) and therefore a suboptimal strategy. While EoC is unprofitable from an economic perspective, one may argue from a preference perspective that sticking to a chosen course of action is “rational” because it matches one’s preferences (Schulz-Hardt et al., 2009). We found evidence that personal responsibility increased reinvestments beyond what would be expected solely based on the preference for the initial decision.

Furthermore, in Experiments 2 and 3, we found that participants were more likely to keep investing if they were personally responsible for prior decisions compared to both assigned decision mechanisms – assigned decisions made randomly and systematically (i.e., following the expected value principle). This effect was observed in both experiments for trials where participants had indicated a prior preference and for trials where they had not. The pattern of results can thus not solely be attributed to overestimating the quality of one’s initial decision because participants were more likely to reinvest when they were personally responsible, even compared to trials for which they knew that the computer had made an optimal initial decision. Responsibility effects also held while controlling for preferences regarding the initial decisions.

The expected value of the participants’ decisions was lower when the initial decision was made by the participants themselves rather than made optimally by the computer. As the probability of investing was consistently higher in the personally responsible condition than in the optimal assignment condition at all probability levels, this difference in expected value is driven by participants’ tendency to bet too much on bad hands for which they were personally responsible – which can be expected as a result of personal responsibility effects in EoC.

From an expected-value perspective, the participants reinvested too often in bad cards if they were personally responsible for the initial decision, and they reinvested not often enough in good cards if the initial decision was not their responsibility. Participants were more likely to reinvest at a high probability of losing in responsible trials compared to the two assignment conditions (see Figures 4, 6). They were also less likely to reinvest at a low probability of losing in the assignment trials than in responsible decision trials. Too many bad reinvestments lower the expected value in the personally responsible condition, and too few good reinvestments lower the expected value in the two assignment conditions, compared to a hypothetical decision strategy that would maximize the expected value (the bold lines in Figures 4, 6). In conclusion, the sum of our results indicates that being personally responsible interfered with optimal subsequent decisions (even beyond preferences) in those situations where the probability of success was low. When the probability of success was high however, personal responsibility was beneficial as it increased the likelihood of (good) reinvestments. This reasoning can also be applied to loss framing – being in a loss frame increases the probability to reinvest. In situations with a high probability of success this can be advantageous, but in situations with a high probability of failure it can be disadvantageous.

3.4. Variable investment poker task

The incentivized poker-based task used in the present experiments has several advantages over standard EoC paradigms such as hypothetical investment scenarios (e.g., Staw, 1976; Garland and Newport, 1991; Schulz-Hardt et al., 2009; Feldman and Wong, 2018; Benschop et al., 2020; Lee et al., 2020), or paradigms using

deception (e.g., Strube and Lott, 1984; Schulz-Hardt et al., 2009). The rules of the task are fully and transparently described to the participants. Trials are generated randomly following these rules. This procedure created realistic decision problems, in some of which further investments yield a better outcome, while in others opting out results in higher payoffs. The participants' decisions can thus be analyzed in relation to normative decision theories (e.g., expected utility theory; Von Neumann and Morgenstern, 1944). From a procedural validity perspective, the participants in the VIP task know that their decisions are consequential compared to hypothetical scenarios where they make decisions while knowing that they merely pretend to be the CEO of a million-dollar company. Besides the issues with anticipated versus real decisions, outcomes of hypothetical scenarios may also be biased by participants not anticipating their own decisions but by what they anticipate what a CEO would or should do. Such problems are avoided in the task paradigm used in our present research.

Anecdotal feedback by our participants suggests that the task is highly engaging and holds the participants' interest and attention in the lab and online over even a large number of trials. The present task paradigm consists of multiple repeated trials, as opposed to one or two scenarios. This increases statistical power and allows within-participant manipulations of relevant factors. In the present line of studies, personal responsibility was manipulated within participants, while framing was manipulated between participants. But both variables could also be manipulated in a within-participants design using the VIP-Task. Materials to implement the task are available at https://osf.io/hdczr/?view_only=546aec80e6f7468685072d199a9d9821.

4. Conclusion and outlook

Being personally responsible for prior decisions and loss framing increased the likelihood that decision makers reinvested in an ongoing course of action. These effects were robust and occurred beyond preferences. In situations where the probability of failure was high, personal responsibility, preferences and loss framing decreased the expected value of the reinvestment decisions, but in situations where the probability of failure was low, these variables increased the expected value. A comparison to optimal assigned initial decisions indicates that decision quality suffered when decision-makers were personally responsible for prior investments; in particular, when they were personally responsible for prior decisions they were more likely to throw good money after bad.

The card-game task used in our experiments is a powerful tool for future research. For example, it could be adopted for investigating social decisions. In our studies, advisors were computer algorithms. This has the advantage that the decision rules for these advisors can be exactly determined. However, it will be interesting to investigate whether real human advice is treated differently. It would also be interesting to analyze cooperative or competitive decisions with a modified VIP-Task, in which participants' decisions affect the payout of other players.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

The studies involving human participants were reviewed and approved by Ethik-Kommision, University of Konstanz. The patients/participants provided their written informed consent to participate in this study.

Author contributions

JD created the experimental task, analyzed the data, and wrote the initial draft of the manuscript with input from TM-H and PG. TM-H and PG contributed additional written sections to the manuscript and provided guidance and editorial feedback. JD, TM-H, and PG involved in the conceptual development of the studies and the final preparation of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The authors gratefully acknowledge financial support from the Ausschuss für Forschungsfragen, University of Konstanz.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1025181/full#supplementary-material>

References

- Arkes, H. R. (1991). Costs and benefits of judgment errors: implications for Debiasing. *Psychol. Bull.* 110, 486–498. doi: 10.1037/0033-2909.110.3.486
- Benschop, N., Nuijten, A. L. P., Keil, M., Rohde, K. I. M., Lee, J. S., and Commandeur, H. R. (2020). Construal level theory and escalation of commitment. *Theor. Decis.* 91, 135–151. doi: 10.1007/s11238-020-09794-w
- Bobocel, D. R., and Meyer, J. P. (1994). Escalating commitment to a failing course of action: separating the roles of choice and justification. *J. Appl. Psychol.* 79, 360–363. doi: 10.1037/0021-9010.79.3.360
- Brockner, J. (1992). The escalation of commitment to a failing course of action: toward theoretical Progress. *Acad. Manag. Rev.* 17, 39–61. doi: 10.2307/258647
- Cohen, D., and Erev, I. (2021). Over and under commitment to a course of action in decisions from experience. *J. Exp. Psychol. Gen.* 150, 2455–2471. doi: 10.1037/xge0001066
- Doerflinger, J. T., Martiny-Huenger, T., and Gollwitzer, P. M. (2017). Planning to deliberate thoroughly: if-then planned deliberation increases the adjustment of decisions to newly available information. *J. Exp. Soc. Psychol.* 69, 1–12. doi: 10.1016/j.jesp.2016.10.006
- Drummond, H. (2014). Escalation of commitment: when to stay the course? *Acad. Manag. Perspect.* 28, 430–446. doi: 10.5465/amp.2013.0039
- Feldman, G., and Wong, K. F. E. (2018). When action-inaction framing leads to higher escalation of commitment: a new inaction-effect perspective on the sunk-cost fallacy. *Psychol. Sci.* 29, 537–548. doi: 10.1177/0956797617739368
- Garland, H., and Newport, S. (1991). Effects of absolute and relative sunk costs on the decision to persist with a course of action. *Organ. Behav. Hum. Decis. Process.* 48, 55–69. doi: 10.1016/0749-5978(91)90005-e
- Green, P., and MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol. Evol.* 7, 493–498. doi: 10.1111/2041-210X.12504
- Heath, C. (1995). Escalation and de-escalation of commitment in response to sunk costs: the role of budgeting in mental accounting. *Organ. Behav. Hum. Decis. Process.* 62, 38–54. doi: 10.1006/obhd.1995.1029
- Hsieh, K.-Y., Tsai, W., and Chen, M.-J. (2015). If they can do it, why not us? Competitors as reference points for justifying escalation of commitment. *Acad. Manag. J.* 58, 38–58. doi: 10.5465/amj.2011.0869
- Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291. doi: 10.2307/1914185
- Kirby, S. L., and Davis, M. A. (1998). A study of escalating commitment in principal-agent relationships: effects of monitoring and personal responsibility. *J. Appl. Psychol.* 83, 206–217. doi: 10.1037/0021-9010.83.2.206
- Lee, J. S., Keil, M., and Wong, K. F. E. (2020). When a growth mindset can backfire and cause escalation of commitment to a troubled information technology project. *Inf. Syst. J.* 31, 7–32. doi: 10.1111/isj.12287
- Lipkus, I. M., Samsa, G., and Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Med. Decis. Mak.* 21, 37–44. doi: 10.1177/0272989x0102100105
- Lipsey, R. G., and Harbury, C. (1992). *First Principles of Economics*. United States: Oxford University Press.
- McNamara, G., Moon, H., and Bromiley, P. (2002). Banking on commitment: intended and unintended consequences of an organization's attempt to attenuate escalation of commitment. *Acad. Manag. J.* 45, 443–452. doi: 10.5465/3069358
- Nau, R. F. (2006). Uncertainty aversion with second-order utilities and probabilities. *Manag. Sci.* 52, 136–145. doi: 10.1287/mnsc.1050.0469
- Negrini, M., Riedl, A. M., and Wibral, M. (2020). Still in search of the sunk cost bias. *SSRN Electron. J.* doi: 10.2139/ssrn.3706308
- Nordgren, L. F., Harreveld, F. V., and Pligt, J. V. D. (2009). The restraint bias: how the illusion of self-restraint promotes impulsive behavior. *Psychol. Sci.* 20, 1523–1528. doi: 10.1111/j.1467-9280.2009.02468.x
- Palan, S., and Schitter, C. (2018). Prolific. Ac—a subject pool for online experiments. *J. Behav. Exp. Financ.* 17, 22–27. doi: 10.1016/j.jbef.2017.12.004
- Pearce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* 51, 195–203. doi: 10.3758/s13428-13018-01193-y
- Ronayne, D., Sgroi, D., and Tuckwell, A. (2021). Evaluating the sunk cost effect. *J. Econ. Behav. Organ.* 186, 318–327. doi: 10.1016/j.jebo.2021.03.029
- Roth, S., Robbert, T., and Straus, L. (2015). On the sunk-cost effect in economic decision-making: a meta-analytic review. *Bus. Res.* 8, 99–138. doi: 10.1007/s40685-014-0014-8
- Rutledge, R. W. (1995). Escalation of commitment in groups and the moderating effects of information framing. *J. Appl. Bus. Res.* 11, 17–22. doi: 10.19030/jabr.v11i2.5870
- Schoorman, F. D., Mayer, R. C., Douglas, C. A., and Hetrick, C. T. (1994). Escalation of commitment and the framing effect: an empirical investigation. *J. Appl. Soc. Psychol.* 24, 509–528. doi: 10.1111/j.1559-1816.1994.tb00596.x
- Schulz-Hardt, S., Thurow-Kröning, B., and Frey, D. (2009). Preference-based escalation: a new interpretation for the responsibility effect in escalating commitment and entrapment. *Organ. Behav. Hum. Decis. Process.* 108, 175–186. doi: 10.1016/j.obhdp.2008.11.001
- Sleesman, D. J., Conlon, D. E., McNamara, G., and Miles, J. E. (2012). Cleaning up the big muddy: a meta-analytic review of the determinants of escalation of commitment. *Acad. Manag. J.* 55, 541–562. doi: 10.5465/amj.2010.0696
- Sleesman, D. J., Lennard, A. C., McNamara, G., and Conlon, D. E. (2018). Putting escalation of commitment in context: a multilevel review and analysis. *Acad. Manag. Ann.* 12, 178–207. doi: 10.5465/annals.2016.0046
- Smith, G., Levere, M., and Kurtzman, R. (2009). Poker player behavior after big wins and big losses. *Manag. Sci.* 55, 1547–1555. doi: 10.1287/mnsc.1090.1044
- Soman, D. (2004). “Framing, loss aversion, and mental accounting,” in *Blackwell Handbook of Judgment and Decision Making*. eds. D. J. Koehler and N. Harvey (Hoboken, New Jersey: Blackwell Publishing).
- Staw, B. M. (1976). Knee-deep in the big muddy: a study of escalating commitment to a chosen course of action. *Organ. Behav. Hum. Perform.* 16, 27–44. doi: 10.1016/0030-5073(76)90005-2
- Strube, M. J., and Lott, C. L. (1984). Time urgency and the type a behavior pattern: implications for time investment and psychological entrapment. *J. Res. Pers.* 18, 395–409. doi: 10.1016/0092-6566(84)90023-0
- Thaler, R. (1980). Toward a positive theory of consumer choice. *J. Econ. Behav. Organ.* 1, 39–60. doi: 10.1016/0167-2681(80)90051-7
- Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* 211, 453–458. doi: 10.1126/science.7455683
- Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5, 297–323. doi: 10.1007/BF00122574
- Von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton, New Jersey: Princeton University Press.
- Weber, E. U., Blais, A. R., and Betz, N. E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *J. Behav. Decis. Mak.* 15, 263–290. doi: 10.1002/bdm.414
- Whyte, G. (1986). Escalating commitment to a course of action: a reinterpretation. *Acad. Manag. Rev.* 11, 311–321. doi: 10.5465/AMR.1986.4283111
- Wong, K. F. E., Yik, M., and Kwong, J. Y. Y. (2006). Understanding the emotional aspects of escalation of commitment: the role of negative affect. *J. Appl. Psychol.* 91, 282–297. doi: 10.1037/0021-9010.91.2.282



OPEN ACCESS

EDITED BY

Samuel Shye,
Hebrew University of Jerusalem, Israel

REVIEWED BY

Christos Andreas Makridis,
Columbia University, United States
Richard S. John,
University of Southern California, United States

*CORRESPONDENCE

Benson Tsz Kin Leung
✉ btkleung@hkbu.edu.hk

[†]These authors share first authorship

RECEIVED 22 August 2023

ACCEPTED 30 November 2023

PUBLISHED 21 December 2023

CITATION

Huang EYH and Leung BTK (2023) Risk attitude and belief updating: theory and experiment. *Front. Psychol.* 14:1281296. doi: 10.3389/fpsyg.2023.1281296

COPYRIGHT

© 2023 Huang and Leung. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Risk attitude and belief updating: theory and experiment

Evelyn Y. H. Huang^{1†} and Benson Tsz Kin Leung^{2*†}

¹Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China, ²Hong Kong Baptist University, Kowloon, Hong Kong SAR, China

Despite the importance of risk attitude in decision-making, its role in belief updating has been overlooked. Using economic theory, we analyzed a dual-self equilibrium where an individual first updates her belief about an uncertain state and then takes an action to maximize her payoff. We showed that stronger risk aversion drives more conservative actions and thus decreases the instrumental value of information relative to the importance of belief-based utility. As a result, the relationship between risk attitude and belief updating depends on the nature of the belief-based utility. With self-relevant information, stronger risk aversion leads to more belief change, whereas with self-irrelevant information, stronger risk aversion leads to less belief change. Our experimental results concur with the theoretical predictions with two settings where subjects update their belief about their IQ and a randomly drawn number, respectively. We discuss implications on persuasion, advertisements, and political campaigns.

KEYWORDS

risk attitude, risk aversion, belief, learning, belief-based utility

1 Introduction

Research on risk attitude has received an abundance of attention across different disciplines including marketing, behavioral science, economic, and psychology (Weber et al., 2002; Wakebe et al., 2012). It affects individuals' financial decisions (Noussair et al., 2014; Oehler et al., 2018), career choices (Gaba and Kalra, 1999; Bonin et al., 2007; Jaeger et al., 2010; Argaw et al., 2017), medical decisions (Rosen et al., 2003; Arrieta et al., 2017; Massin et al., 2018), purchase and sales decisions (Okada, 2010; Shapiro, 2011; Jindal, 2015), etc. The existing research mainly focuses on the relationship between risk attitude and decision-making by assuming risk attitude is independent to belief updating, while there is scant knowledge about the relationship between risk attitude and belief updating.¹ However, in many situations with information transmission, it is important to understand how people update their beliefs with new information in order to determine their subsequent decisions. For example, to evaluate the impact of information campaigns, e.g., campaigns to convey the importance of stay-home policy during COVID-19 (Krpan et al., 2021), it is crucial to understand whether information could effectively influence people's belief, and if yes, to what extent.² This study aims to shed light on the role of risk aversion in belief updating and the underlying mechanism and discuss implications on persuasion, advertisements, and political campaigns.

¹ Ho et al. (2021) looks into the relationship between risk preference and preference of information acquisition. In contrast, we study how individuals with different risk preferences update their belief upon receiving the same piece of information.

² See Haaland et al. (2023) for a literature review of information provision experiments. In the conclusion, we discuss the implications of our results in the literature.

From a Bayesian perspective, risk attitude has no impact on belief updating. Given a piece of information, and the understanding of the underlying information structure, individuals have no incentive to distort their belief as it will otherwise lead to sub-optimal decision-making in future.³ Given the popularity of the Bayesian paradigm, the literature has instead focused on how different characteristics of information structures affect belief updating. To give a few examples, [Eil and Rao \(2011\)](#) find evidence of asymmetric updating toward good and bad news in self-relevant but not self-irrelevant context, while [Coutts \(2019\)](#) found no evidence of asymmetric updating across different contexts; [Alós-Ferrer and Garagnani \(2023\)](#) found that larger incentive leads to a more reinforcing belief updating and less Bayesian updating; [Coffman et al. \(2023\)](#) showed that individuals are more likely to update to reinforce stereotypes.

In contrast, this study intends to investigate how risk attitude affects belief updating. Contributing to the research program of decision-making under uncertainty, our results suggest that there is an inherent relationship between risk preference and belief formation, which calls for more future research. It also sheds light on the mechanism behind the heterogeneous belief-updating behavior across individuals (see for example, [Berlin and Dargnies, 2016](#); [Sinclair et al., 2020](#)), and could explain heterogeneous treatment effects in information provision experiments ([Haaland et al., 2023](#)). Moreover, it also has significant implications on persuasion, advertisement, politics, etc. First, belief updating behavior directly relates to consumers' susceptibility to being persuaded by advertisements. Our results hence speak to the empirical relationship between risk aversion and brand loyalty ([Matzler et al., 2008](#)) and between risk aversion and the effectiveness of advertisement ([Jeong and Kwon, 2012](#)). Second, our results also provide firms guidance on their advertisement strategy, depending on whether their target customers are more- or less-risk averse. Third, and similarly, our results also shed light on how politicians could target more- or less-risk-averse individuals more effectively in their political campaigns. Given the well-documented relationship between age and risk-aversion ([Albert and Duffy, 2012](#)), we also speak to the political divides between older and younger constituencies.

So, how would risk attitude affect belief updating? In this study, we first present an economic theory with the premise that individuals trade-off between the instrumental purpose and the non-instrumental (psychological) purposes of information. In a model of decision-making under uncertainty, the instrumental purpose of information refers to the need of improving decision-making: a more accurate belief enables the individual to better take into account available information and choose a better decision, e.g., to pick a better product or to vote for a better candidate. On the other hand, the non-instrumental purpose of information refers to the concept of belief-based utility such as motivated belief, a utility for reduced uncertainty, and updating cost (see [Loewenstein and Molnar, 2018](#) for a review). We analyze a dual-self equilibrium

where individuals first update their belief and afterwards take an action. Importantly, we show that individuals with stronger risk aversion choose more conservative actions and that diminishes the importance of the instrumental purpose of information relative to the non-instrumental purpose.⁴ As a result, more risk-averse individuals update their belief in a way that caters more to the non-instrumental purpose. In self-relevant settings, i.e., when the uncertainty is self-related, individuals have a higher demand for information ([Bargh, 1982](#); [Shapiro et al., 1997](#); [Symons and Johnson, 1997](#); [Gray et al., 2004](#); [Sui et al., 2006](#); [Turk et al., 2011](#)), the non-instrumental purpose of information resembles the utility for reduced uncertainty, thus more risk-averse individuals update their belief more. On the other hand, in a self-irrelevant setting, i.e., when the uncertainty is not self-related, there is less utility for reduced uncertainty, updating cost becomes more (relatively) important; thus, individuals with stronger risk aversion update their belief less.

We then test our theoretical prediction in an experiment with two settings, where subjects have to update their belief with self-relevant and self-irrelevant information, respectively. We find that upon receiving the same information, subjects with stronger risk aversion update more in the self-relevant setting and less in the self-irrelevant setting. It, therefore, confirms our theoretical predictions. We also report a significant relationship between demographics, such as gender and confidence, and belief updating in both self-relevant and self-irrelevant settings. Combined with existing literature on gender differences, we argue that our findings on demographics and belief updating support our theory and the trade-off between instrumental and non-instrumental value of information.

This study is organized as follows. We present the theoretical analysis in the next section. Afterward, we present the experimental design and the results in Sections 3 and 4. Section 5 discusses potential concerns of our study. Lastly, we conclude by discussing the implications of our results.

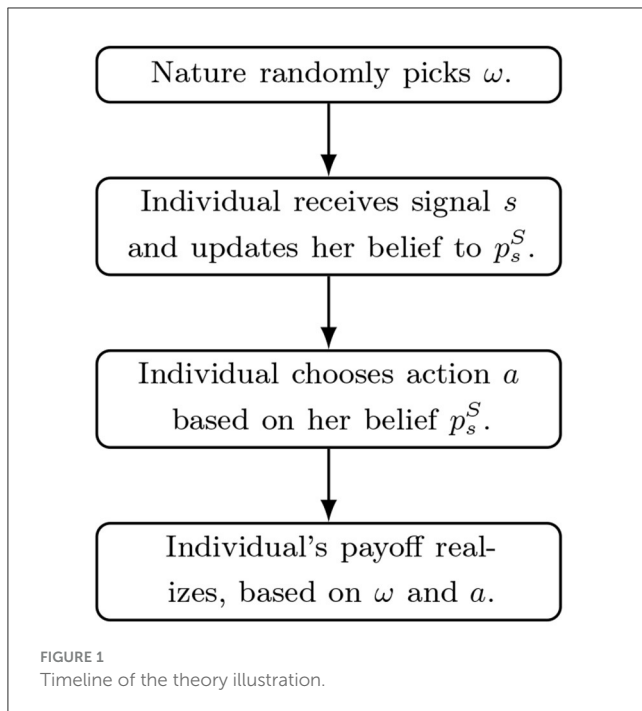
2 Theory illustration

In this section, we present the theoretical foundation that illustrates how risk aversion affects belief updating. It sheds light on the mechanism behind the relationship between risk aversion and belief updating and helps formulate our hypotheses. In particular, we show that individuals with stronger risk aversion take more conservative actions and thereby have more incentive to form belief catering to non-instrumental objectives instead of instrumental objectives.

Imagine an individual who tries to learn an unknown state of the world to improve her decision-making. For example, she learns whether her IQ is among the top half of society in order to plan her future career or evaluates the quality of a social media platform to decide how much time she spends on it or predicts the state of the economy in the coming year for her investment plan. The state of

³ This is true even when there are uncertainties over the information structure, or quality of the information. More specifically, uncertainty over the information structure could be incorporated to the information structure itself, just like compound lotteries can be reduced to simple lotteries.

⁴ As an extreme example, an infinitely risk averse individual always picks the safest action, thus information has no impact on her action, i.e., information has no instrumental value.



the world is denoted as ω , and for simplicity, ω equals either 0 or 1. We assume that the two states are a priori equally likely.

In what follows, we analyze a scenario where the individual first receives a piece of information and updates her belief, and afterwards chooses her action based on her belief. The updating rule and action rule is characterized by a dual-self equilibrium introduced in the next paragraph. The timeline is illustrated in Figure 1. In period 0, nature randomly picks ω , which equals 0 or 1 with equal probability. In period 1, the individual receives a signal $s \in S$, which induces a Bayesian posterior in which we denote as p_s^B . For simplicity, we assume $p_s^B \sim U[\underline{p}, \bar{p}]$, where $\bar{p} = 1 - \underline{p}$.⁵ Given the Bayesian belief p_s^B , the individual updates her belief to p_s^S according to a linear updating rule $p_s^S = (1 - \lambda)0.5 + \lambda p_s^B$.⁶ In period 2, the individual chooses an action a according to a linear action rule $a = (1 - \gamma)0.5 + \gamma p_s^S$ and receives a payoff $\pi^\omega(a) = 1 - (\omega - a)^2$ depending on the state of the world. To model risk aversion, we denote the utility function of the individual as $u(\pi) = \pi^{1-\theta}$ where $\theta \in (0, 1)$. A higher θ implies a stronger risk aversion.

The updating rule λ and the action rule γ are characterized as a dual-self Subgame Nash equilibrium, where the period-1 self first picks the updating rule λ and afterwards the period-2 self picks the action rule γ .⁷ The equilibrium solution is denoted as (λ^*, γ^*) . Given our linear formulation, the period-2 self picks γ to maximize

5 The Bayesian posterior is given by

$$p_s^B = \frac{\Pr(s | \omega = 1)}{\Pr(s | \omega = 1) + \Pr(s | \omega = 0)}.$$

Note that instead of specifying the signal distributions $\Pr(s | \omega = 1)$ and $\Pr(s | \omega = 0)$, we directly model the resulting Bayesian posterior as our primitive, akin to the approach in the information design literature (Bergemann and Morris, 2019).

6 This linear formulation combined with the quadratic loss function introduced later brings tractability, and also imposes less cognitive demand on the individual (Compte and Postlewaite, 2019).

her expected utility denoted as U_2 :

$$U_2(a) = \int_{0.5+\lambda(\underline{p}-0.5)}^{0.5+\lambda(\bar{p}-0.5)} [p_s^S u(\pi^1(a)) + (1 - p_s^S) u(\pi^0(a))] \frac{1}{\lambda(\bar{p} - \underline{p})} dp_s^S$$

On the other hand, for tractability of our analysis, we assume the period-1 self picks λ to solve the following minimization problem:

$$\min_{\lambda} \int_{\underline{p}}^{\bar{p}} [(a(p_s^S) - a(p_s^B))^2 + V^N(p_s^S)] \frac{1}{(\bar{p} - \underline{p})} dp_s^B. \quad (1)$$

Equation (1) captures and allows us to focus on the main building block of the model, i.e., the trade-off between the instrumental and non-instrumental value of belief.⁸

The first item of Equation (1) corresponds to the instrumental value of belief, which is the quadratic difference between the Bayesian action and the action chosen by the period-2 self. The closer the period-2 self's action is to the Bayesian action, the lower of the first item is. It thus represents the utility loss of taking actions that is away from the Bayesian optimal action, i.e., the instrumental value of belief. The second item of Equation (1) corresponds to the non-instrumental value of belief, which we provide a few examples below.⁹

1. Motivated belief: for example, $V^N(p_s^S) = w(1 - p_s^S)$. The individual gets higher utility if she believes state 1 is true.
2. Utility for reduced uncertainty: for example, $V^N(p_s^S) = -w(p_s^S - 0.5)^2$. The individual gets higher utility if she is confident about the state.
3. Updating cost: for example, $V^N(p_s^S) = w(p_s^S - 0.5)^2$. The individual incurs cost from updating her belief away from the prior.

Lastly, for ease of exposition, we assume that \underline{p} is small enough such that a and p_s^S are characterized by the first-order conditions and are inside $[0, 1]$. We proceed by backward induction and first characterize the action rule. The following proposition shows that individuals with stronger risk aversion are more conservative and choose action closer to 0.5.

Proposition 1. Denote the optimal action rule as $a^* = (1 - \gamma^*)0.5 + \gamma^* p_s^S$, γ^* decreases in θ .

7 See Bénabou and Tirole (2004), Brunnermeier and Parker (2005), and Wilson (2014) for examples of dual-self/multi-self models that study deviation from Bayesian updating.

8 An alternative formulation is to assume that the period-1 self maximizes a sum of expected utility and a belief-based utility, i.e.,

$$\max_{\lambda} \int_{\underline{p}}^{\bar{p}} [p_s^B u(\pi^1(a(p_s^S))) + (1 - p_s^B) u(\pi^0(a(p_s^S)))] - V^N(p_s^S) \frac{1}{(\bar{p} - \underline{p})} dp_s^B.$$

This, however, introduces an arbitrary scaling effect: as θ increases, i.e., level of risk aversion increases, $u^{1-\theta}$ also changes (increases when $u < 1$ and decreases when $u > 1$). Therefore, risk aversion affects the trade-off between instrumental and non-instrumental value in a way that arbitrarily depends on the magnitude of payoff and the functional form. Such scaling effect presents even in the extreme case where action does not depend on the belief. Equation (1) eliminates the scaling effect.

9 The non-instrumental purpose of belief corresponds to the belief-based utility, as discussed in Loewenstein and Molnar (2018).

The omitted proofs are shown in the [Appendix](#). The intuition of Proposition 1 is as follows: as the degree of risk aversion θ increases, the individual has more incentive to insure herself against the mistake she would have made, or put differently, balance the utility between the two states. As a result, she does not tailor her action to her belief as much and chooses action closer to 0.5.

Now, we are ready to characterize the optimal belief updating rule $p_s^S = (1 - \lambda^*)0.5 + \lambda^*p_s^B$. Given our linear formulation and γ^* , Equation (1) becomes

$$\min_{\lambda} (\gamma^*(1 - \lambda))^2 \text{Var}(p_s^B) + \int_{\underline{p}}^{\bar{p}} V^N(p_s^S) d p_s^B. \quad (2)$$

The first item of Equation (2) corresponds to the instrumental purposes of information. It is minimized at $\lambda = 1$ regardless of the value of γ^* . Thus, if the second item of Equation (2) does not exist, the optimal belief updating rule is to update according to Bayes' rule, which highlights the importance of belief-based utility (Loewenstein and Molnar, 2018). In contrast, in the presence of the non-instrumental purposes of information, the individual trades off between minimizing the two items in Equation (2). In particular, as shown in Equation (2), the instrumental value of information increases when $\text{Var}(p_s^B)$ increases, i.e., when the information is precise such that the Bayesian belief is more dispersed, or when γ^* increases, i.e., when the individual's action is more sensitive to his belief. The latter gives rise to our main theoretical result.

Proposition 2. A stronger risk aversion implies that individuals tailor their beliefs more to the non-instrumental than the instrumental purpose of information. For example,

1. if $V^N(p_s^S) = -w(p_s^S - 0.5)^2$, $\frac{\partial \lambda^*}{\partial \theta} \geq 0$, i.e., individuals with stronger risk aversion updates more;
2. if $V^N(p_s^S) = w(p_s^S - 0.5)^2$, $\frac{\partial \lambda^*}{\partial \theta} < 0$, i.e., individuals with stronger risk aversion updates less.

Proposition 2 is driven by the result in Proposition 1. As individuals with stronger risk-aversion choose more conservative actions, i.e., as γ^* decreases, there is a lower cost of belief distortion, i.e., the first item of Equation (2) decreases. As a result, they have more incentive to update their belief catering to the non-instrumental purpose, i.e., the second item of Equation (2). In the first bullet point, the belief-based element of the loss function, i.e., $-w(p_s^S - 0.5)^2$ decreases as p_s^S is more extreme, thus representing the presence of utility for reduced uncertainty: utility loss decreases when the individual is more confident about the state. In such case, learning rate increases in risk aversion. In the second bullet point, the belief-based element of the loss function, i.e., $w(p_s^S - 0.5)^2$ decreases as p_s^S is closer to 0.5, thus representing the presence of an updating cost: utility loss increases when the individual's belief is more away from her prior belief. In such case, learning rate decreases in risk aversion.

Proposition 2 thus shows that the relationship between risk aversion and belief updating is context dependent. In the next section, we present our experimental result that tests our theory. We hypothesize our experimental results based on the two cases in Proposition 2.

3 Experimental design

We run the following experiment with two experimental conditions corresponding to the first and second bullet points of Proposition 2, which we call “SELF” and “NON-SELF” settings, respectively. The instruction could be found in the [Online Appendix](#).

In both settings, subjects first fill out a demographic survey on their age, gender, and have to report their confidence about their own performance in a 20-question Raven Progressive Matrices test with a 5-point scale. Afterwards, we elicit the subjects' degree of risk aversion using a multiple-price list shown in [Figure 2](#) (Holt and Laury, 2002). Option As are “safer” than option Bs. The subjects essentially decide on which row they switch from choosing option A to option B, the lower down they switch, the more risk averse they are.¹⁰ After that, subjects have to complete a 20-question Raven Progressive Matrices test within 20 min. Lastly, subjects have to guess and report their beliefs about a random variable that differs in the “SELF” and “NON-SELF” settings.¹¹

In the “SELF” condition, subjects have to form a belief about their performance in the Raven Progressive Matrices test. Therefore, the uncertainty is self-related.¹² Without knowing their test results, they have to report their probabilistic belief that their result is among the top half of the session.¹³ We elicit their belief once right after the Raven test. We then provide them six pieces of information consecutively, and after each pieces of information, we elicit again their beliefs to track how they change. Thus, we elicit their belief seven times, which we denote as p_0, p_1, \dots, p_6 , using the table form of the binarized scoring rule as shown in [Figure 3](#) (Hossain and Okui, 2013). Subjects have to indicate their beliefs using the slider, and the choices between option 1s and 2s are automatically selected which help to illustrate consequences of the binarized scoring rule. It is important to point out that risk preference does not affect belief elicitation using the binarized scoring rule.¹⁴ Between each elicitation, we provide them with a piece of information, which could be either a thumbs-up or a thumbs-down. If their result is among the top half, we show them a thumbs-up with a probability of 60%; if their result is among the bottom half, we show them a thumbs-down with a probability of 60%. The information structure is shown in [Tables 1, 2](#) and is explained to the subjects.

¹⁰ Note that a rational individual should never choose option B in the first row and option A in the last row. Two of our 148 subjects chose option B in the first row, while six subjects chose option A on the last row. Our results do not change after excluding those eight subjects. The robustness test is presented in the [Online Appendix](#).

¹¹ Although subjects do not have to pick an action as in the theoretical model, as the belief elicitation is incentivized, there is still an instrumental value of belief.

¹² See Eil and Rao (2011), Castagnetti and Schmacker (2022), and Oprea and Yuksel (2022) for similar setup.

¹³ We break all ties randomly, and it is conveyed to the subjects.

¹⁴ See Hossain and Okui (2013) for the mathematical proof and experiment. By contrast, with quadratic scoring rule, subjects with stronger risk aversion report beliefs closer to 0.5, as shown in both theoretical and experimental analysis (Erkal et al., 2020).

Please choose option A or B in every row. Remember as you go down the rows, you can only switch from A to B once.

Option A		Option B	Expected Value of Option A minus that of Option B
0% winning 12, 100% winning 8	<input type="radio"/> <input type="radio"/>	0% winning 20, 100% winning 2	6
10% winning 12, 90% winning 8	<input type="radio"/> <input type="radio"/>	10% winning 20, 90% winning 2	4.6
20% winning 12, 80% winning 8	<input type="radio"/> <input type="radio"/>	20% winning 20, 80% winning 2	3.2
30% winning 12, 70% winning 8	<input type="radio"/> <input type="radio"/>	30% winning 20, 70% winning 2	1.8
40% winning 12, 60% winning 8	<input type="radio"/> <input type="radio"/>	40% winning 20, 60% winning 2	0.4
50% winning 12, 50% winning 8	<input type="radio"/> <input type="radio"/>	50% winning 20, 50% winning 2	-1
60% winning 12, 40% winning 8	<input type="radio"/> <input type="radio"/>	60% winning 20, 40% winning 2	-2.4
70% winning 12, 30% winning 8	<input type="radio"/> <input type="radio"/>	70% winning 20, 30% winning 2	-3.8
80% winning 12, 20% winning 8	<input type="radio"/> <input type="radio"/>	80% winning 20, 20% winning 2	-5.2
90% winning 12, 10% winning 8	<input type="radio"/> <input type="radio"/>	90% winning 20, 10% winning 2	-6.6
100% winning 12, 0% winning 8	<input type="radio"/> <input type="radio"/>	100% winning 20, 0% winning 2	-8

FIGURE 2

A multiple price list to elicit risk aversion. Subjects do not see the expected values. Note that a risk neutral individual should switch in the 6th row.

Next, we outline the “NON-SELF” condition. Rather than asking subjects to guess whether their performance in the cognitive ability test is among the top half or not, we ask subjects to guess whether a self-irrelevant, randomly drawn number is among the top half within the session. Formally, each subject is assigned a randomly drawn number from 1 to 100 with a uniform distribution, and the subjects are aware of this prior distribution. Similar to the “SELF” treatment, without telling the subjects their random number, we elicit subjects’ probabilistic belief that their number is among the top half within the session seven times (once without information, and six times with information). The information, i.e., thumbs-up and thumbs-down, is generated by the same information structure in the “SELF” condition and is explained to the subjects. Note that the “NON-SELF” condition is “essentially equivalent” to the “SELF” treatment, except for the fact that the nature of the uncertainty is self-relevant in “SELF” and self-irrelevant in “NON-SELF.”

Given the extensive evidence, on the behavioral and neural level, that self-relevant information receives preferential attention (Bargh, 1982; Shapiro et al., 1997; Symons and Johnson, 1997; Gray et al., 2004; Sui et al., 2006; Turk et al., 2011), we hypothesize that the “SELF” setting resembles the utility for reduced uncertainty, i.e., the first bullet point of Proposition 2. Thus, subjects with stronger risk aversion update more in the “SELF” setting. Conversely, in the “NON-SELF” setting, as the utility for reduced uncertainty is absent (or at least reduced), updating cost becomes more (relatively)

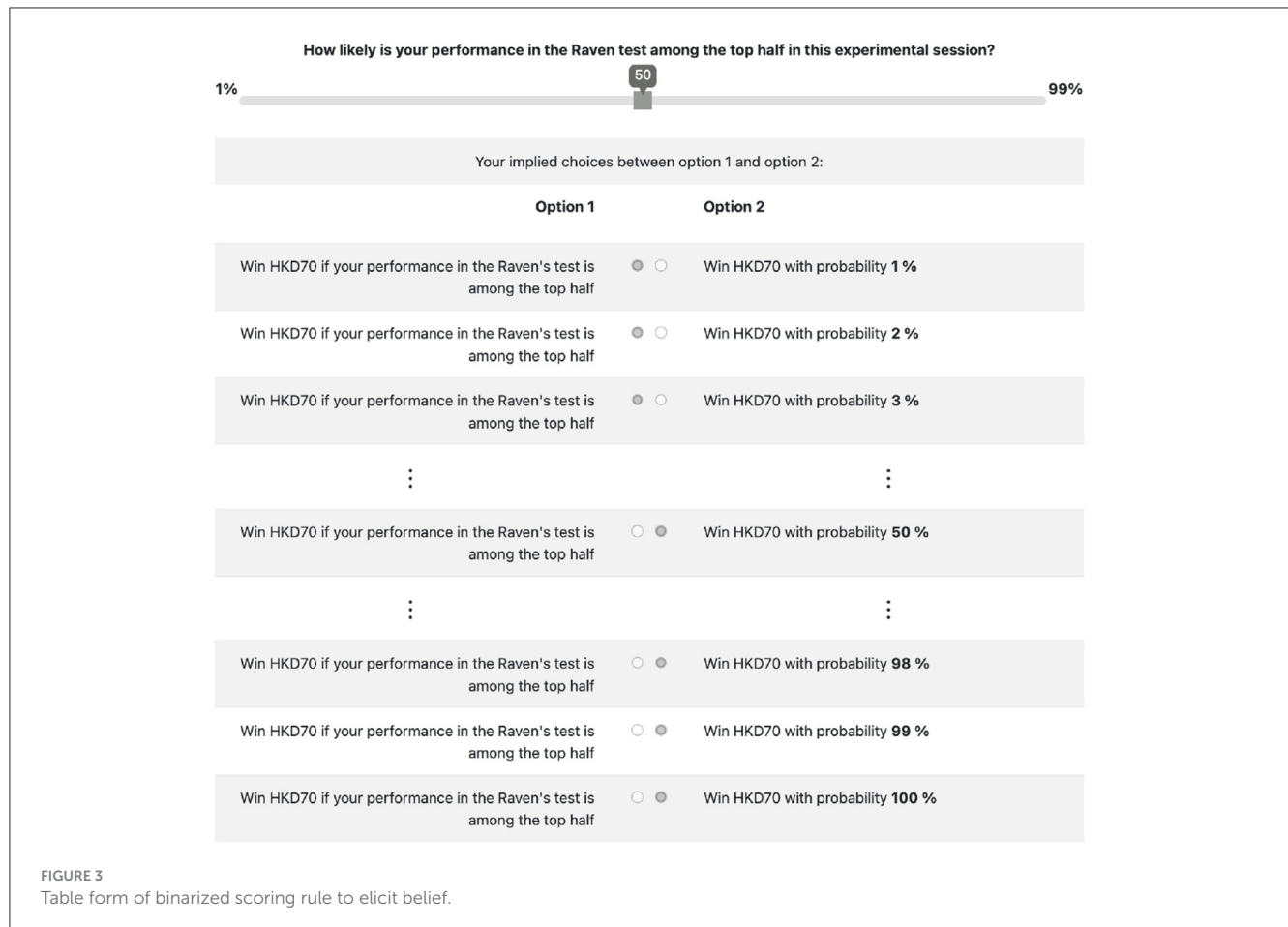
important.¹⁵ Thus, by the second bullet point of Proposition 2, subjects with stronger risk aversion update less in the “NON-SELF” setting.

4 Results

We have recruited 148 subjects via the university subject pool sign-UP system (Sona Systems; <https://www.sona-systems.com>). We run the “SELF” and “NON-SELF” sessions consecutively. In total, 74 subjects are in the “NON-SELF” and another 74 are in the “SELF” condition, giving us $148 \times 6 = 888$ data points of belief updating. The average age is 22.89, and 96 subjects are female. The summary of the demographics, along with other omitted statistical tests, can be found on the [Online Appendix](#). We conducted the experiment in the behavioral laboratory at the university. Each session lasts around 1 h, and each subject earns 75 HKD. The experiment is approved by the Research Ethics Committee at Hong Kong Baptist University (REC/22-23/0023).

Our key variable of interest is the extent of belief updating of individuals, i.e., the “distance” between their prior and posterior

¹⁵ We do not posit that updating cost is higher in the “NON-SELF” setting than in the “SELF” setting, but that the decrease in or absence of utility for reduced uncertainty implies that the updating cost affects individuals more in the “NON-SELF” setting than in the “SELF” setting.



beliefs, in which we quantify using the log-odds form of the Bayesian formula. With a prior belief p_0 and upon receiving a thumbs-up, a Bayesian individual should update his belief to p_1^B which follows:

$$\log \frac{p_1^B}{1 - p_1^B} = \log \frac{p_0}{1 - p_0} + \log \frac{0.6}{0.4}$$

where $\log \frac{0.6}{0.4}$ is the log-likelihood ratio of seeing a thumbs-up when the individual's performance/random number is among the top half versus when it is among the bottom half. Similarly, upon receiving a thumbs-down, a Bayesian individual should update his belief to p_1^B which follows:

$$\log \frac{p_1^B}{1 - p_1^B} = \log \frac{p_0}{1 - p_0} - \log \frac{0.6}{0.4}.$$

Therefore, a Bayesian individual should update her belief by the magnitude of $\log \frac{0.6}{0.4}$ (upwards with good news and downwards with bad news). We denote this ratio ($\log \frac{0.6}{0.4}$) as $y_{\text{Objective}}$ or log objective ratio. We denote the subjective analog of this log objective ratio by $y_{\text{Subjective}}$ or log subjective ratio. With p_i denoted as the elicited belief after the i -th signal, and p_0 denoted as the first elicited belief without any information, $y_{\text{Subjective}}$ is defined as

$$y_{\text{Subjective}} = \begin{cases} \log \frac{p_i}{1 - p_i} - \log \frac{p_{i-1}}{1 - p_{i-1}} & \text{upon receiving a thumbs-up} \\ \log \frac{p_{i-1}}{1 - p_{i-1}} - \log \frac{p_i}{1 - p_i} & \text{upon receiving a thumbs-down} \end{cases}$$

for $i = 1, 2, 3, 4, 5, 6$. $y_{\text{Subjective}}$ thus measures how much the individual updates her belief upwards upon receiving a thumbs-up, and how much the individual updates her belief downwards upon receiving a thumbs-down. For a Bayesian individual, $y_{\text{Subjective}} = y_{\text{Objective}}$.

4.1 Sanity check

We first check, using the data, whether subjects understand the information structure. More specifically, we regress the log subjective ratio with a regressor of log objective ratio¹⁶:

$$y_{\text{Subjective}} = \beta_1 \times y_{\text{Objective}} + \epsilon. \quad (3)$$

If the subjects do not understand the experiment and their belief updating process is totally random, β_1 should be 0; if the subjects update their belief in the same direction as a Bayesian individual, β_1 should be positive; if the subjects are perfectly Bayesian, both β_1 and R^2 should be equal to 1. The result is presented in Table 3.¹⁷ In both the "SELF" and "NON-SELF" condition, β_1

¹⁶ Note that there is no intercept term in the regression as $y_{\text{Objective}} = \log \frac{0.6}{0.4}$ for all data points.

¹⁷ The regression tables in this study are generated using the Stargazer package in R (Hlavac and Hlavac, 2022).

TABLE 1 If subject's performance/random number is among the top half of the session.

Generated signal		
Probability of the signal	60%	40%

TABLE 2 If subject's performance/random number is among the bottom half of the session.

Generated signal		
Probability of the signal	40%	60%

is positive and significant. On average, subjects update upwards their belief upon receiving good news and downwards their belief upon receiving bad news. The subjects update their belief in the same direction suggested by Baye's formula, meaning that they understand the experiment setting and the information content of signals. Moreover, although both β_1 in "SELF" and "NON-SELF" conditions are close to 1, the low R^2 implies that there is significant heterogeneity across subjects on their belief-updating behavior. The significant heterogeneity is also shown in the box plot of log subjective ratio divided by log objective ratio in the [Online Appendix](#).

4.2 Risk attitude and belief updating

Next, for our main experimental result, we estimate the following regression¹⁸:

$$y_{\text{Subjective}} = \beta_1 \times y_{\text{Objective}} + \beta_2 \times \text{high risk aversion} \times y_{\text{Objective}} + \epsilon \quad (4)$$

where "high risk aversion" is a dummy variable and is equal to 1 if the subjects' level of risk aversion is higher than the median.¹⁹ β_2 thus measures the average difference between an subject with higher-than-median level risk aversion and an subject with lower-than-median risk aversion. We focus on the estimation of β_2 , where $\beta_2 > 0$ implies that stronger risk aversion leads to more belief change and the subject's belief is more reactive to the received information, and $\beta_2 < 0$ implies that stronger risk aversion leads to less belief change. The result is presented in [Table 4](#). Our estimation shows that $\beta_2 = 0.623$ ($p < 0.01$) in the "SELF" condition, and $\beta_2 = -0.597$ ($p < 0.05$) in the "NON-SELF"

¹⁸ Note that we do not add "high risk aversion" as a separate predictor because of collinearity, as $y_{\text{Objective}} = \log \frac{0.6}{0.4}$ for all data points and "high risk aversion $\times y_{\text{Objective}}$ " is perfectly correlated with high risk aversion. Note that $y_{\text{Objective}}$ is constant because of our simple information structure illustrate in [Tables 1, 2](#). We add the constant $y_{\text{Objective}}$ in the regression for the ease of interpretation: β_1 measures the extent to which subjects update their belief vis-a-vis a Bayesian, and β_2 is the impact of risk attitude on β_1 .

¹⁹ The median subject switches from option A to option B in the eighth row of the multiple price list shown in [Figure 2](#). As a risk neutral individual should switch in the sixth row, our median subject is risk averse.

TABLE 3 Regression analysis of log subjective ratio on log objective ratio (with standard errors in parentheses).

	Dependent variable	
	Log subjective ratio	
	"SELF"	"NON-SELF"
Log objective ratio	0.951*** (0.119)	1.009*** (0.145)
Observations	444	444
R^2	0.125	0.099
Adjusted R^2	0.123	0.097
Residual Std. Error (df = 443)	1.020	1.237
F Statistic (df = 1; 443)	63.442***	48.537***

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 4 Regression analysis on how risk aversion affect belief updating.

	Dependent variable:	
	Log subjective ratio	
	"SELF"	"NON-SELF"
Log objective ratio	0.648*** (0.165)	1.356*** (0.223)
Log objective ratio \times high risk aversion	0.623*** (0.237)	-0.597** (0.292)
Observations	444	444
R^2	0.139	0.107
Adjusted R^2	0.135	0.103
Residual Std. Error (df = 442)	1.013	1.233
F Statistic (df = 2; 442)	35.588***	26.528***

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

condition. The result thus provides evidence for our theoretical prediction, in which individuals with stronger risk aversion update more when the information is self-relevant, corresponding to a setting with utility for reduced uncertainty, and update less when the information is not self-relevant, where updating cost is more influential. The magnitude of the effect is also substantial: in the "SELF" condition, subjects who has high risk aversion updates almost twice ($\frac{0.623+0.648}{0.648} = 1.96$ times) as much as the subjects who has low risk aversion; in the "NON-SELF" condition, subjects who has high risk aversion updates about half ($\frac{1.356-0.597}{1.356} = 0.56$ times) as much as the subjects who has low risk aversion.

4.3 Demographics

We also conduct regression analysis with demographic variables, including age, gender, and subjects' self-reported confidence in their Raven test. The result is shown in [Table 5](#), and the interactive plot in [Figure 4](#). First note that our main result remains significant: in the "SELF" condition, subjects with higher risk aversion update more ($\beta_2 = 0.634, p < 0.01$); while in the

TABLE 5 Regression analysis on how risk aversion affects belief updating, with demographic variables.

	Dependent variable	
	Log subjective ratio	
	"SELF"	"NON-SELF"
Log objective ratio	1.252 (1.018)	4.665*** (1.114)
Gender	0.165 (0.110)	−0.592*** (0.127)
Age	−0.006 (0.013)	−0.007 (0.013)
Confidence	−0.122** (0.054)	−0.080 (0.062)
Log objective ratio × high risk aversion	0.634*** (0.238)	−0.543* (0.289)
Observations	444	444
R^2	0.157	0.150
Adjusted R^2	0.147	0.141
Residual Std. Error (df = 439)	1.006	1.207
F Statistic (df = 5; 439)	16.318***	15.528***

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

"NON-SELF" condition, subjects with higher risk aversion update less ($\beta_2 = -0.543, p < 0.1$).

In the "SELF" condition, only confidence significantly affects belief updating. In other words, subjects who are more confident about their Raven score update less (coefficient = $-0.122, p < 0.05$). The result is verified with an ANOVA test [$F_{(1,441)} = 8.47, p = 0.0038$]. By contrast, confidence does not play a role in the "NON-SELF" condition, but gender does affect belief updating. More specifically, males significantly update more than females in the "NON-SELF" setting (coefficient = $-0.592, p < 0.01$), where the result is supported by an ANOVA test [$F_{(1,441)} = 21.16, p < 0.01$].

Note that both result on confidence and gender support our theory and the trade-off between the instrumental and non-instrumental value of information. Subjects with higher confidence have less demand of self-information, and less non-instrumental value, and thus update less with information. On the other hand, as males are more competitive than females (Croson and Gneezy, 2009; Buser et al., 2014; Saccardo et al., 2018), it suggests that males have a higher non-instrumental need of being precise in belief formation even when the information is self-irrelevant and therefore update more.

5 Potential concerns

In this section, we discuss potential concerns on our experimental setup and alternative explanations. The omitted tables of statistical tests can be found in the [Online Appendix](#).

5.1 Risk aversion as a binary variable

Note that we use a binary variable to avoid making extra parametric assumptions, in particular on the linear relationship between risk aversion and belief updating. While we show in our theoretical model a monotonic relationship between risk aversion and belief updating, the model is silent on the precise parametric relationship, e.g., it depends on the functional form of $V^N, \frac{\partial y^*}{\partial \theta}$, etc. Assuming, for example, a linear relationship essentially makes our prediction extra sensitive to subjects extreme level of risk-seeking/risk-aversion comparing to subjects with moderate level of risk attitude, which is particularly problematic given that the majority ($\approx 70\%$) of our subjects switch in the 6th, 7th, or 8th row in the risk-elicitation task.²⁰ Given that most subjects' level of risk aversion is 6, 7, or 8, in an extension, we model the level of risk aversion as a 3-levels variable: 0 when subject switches in or before the sixth row, 2 when subject switches in or after the eighth row, and 1 otherwise. In the "SELF" condition, $\beta_2 = 0.23$ ($p = 0.013$). In the "NON-SELF" condition, the result is less significant, i.e., $\beta_2 = -0.1854$ ($p = 0.104$) but the direction remains consistent with our main result. In another extension, we exclude all subjects whose level of risk aversion is strictly lower than 6 or strictly higher than 8. Similarly, in the "SELF" condition, $\beta_2 = 0.4723$ ($p = 0.001$). In the "NON-SELF" condition, $\beta_2 = -0.3591$ ($p = 0.147$). Lastly, we use the level of risk aversion as a 12-levels variables as elicited. The result is less significant but the direction remains consistent with our main result. In the "SELF" condition, $\beta_2 = 0.107$ ($p = 0.23$). In the "NON-SELF" condition, i.e., $\beta_2 = -0.05303$ ($p = 0.52$). All extensions exclude subjects who always choose option A or option B, and the results are shown in the [Online Appendix](#).

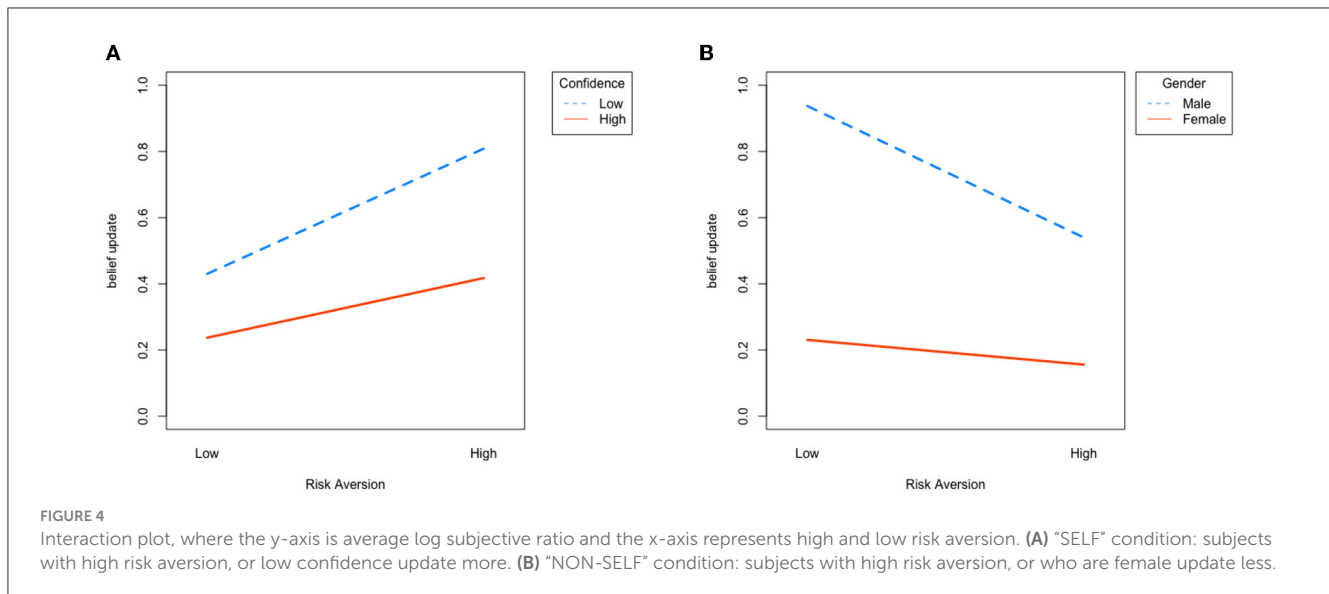
5.2 Overconfidence and motivated belief

In the "SELF" condition, we do find a better-than-average effect: subjects' average prior belief that they are in the top half of their experimental session is 61% (larger than 50%). However, we do not find evidence of motivated belief or asymmetric belief updating toward good and bad news (Coutts, 2019). First, the average last elicited belief (after six signals) is also roughly 61%, which is not larger than their prior belief (i.e., 61%). Our ANOVA analysis additionally shows that subjects do not significantly update more when they received good signals compared to bad signals [$F_{(1,441)} = 0.05, p = 0.8255$]. The equal updating between good and bad signals rules out motivated belief and supports a utility for reduced uncertainty as mentioned in previous sections.

5.3 Risk preference elicitation

We are aware that the multiple price list in [Figure 2](#) is an imbalance between risk-seeking and risk-averse preferences. This, however, does not affect our results. More specifically, we only

²⁰ Almost always fewer than 5 subjects switch in the other rows, in both the "SELF" and "NON-SELF" settings. The distributions are presented in the [Online Appendix](#).



require an ordinal elicitation of risk aversion: subjects who are more risk averse switch from option A to option B in the lower rows of the multiple price list but not a cardinal elicitation of risk aversion.

5.4 Decision errors

One potential confounding variable of our result is the correlation of decision errors between the risk elicitation and the belief formation task. However, we believe that it is highly unlikely, given the differences in results in the “SELF” and “NON-SELF” conditions. For example, if subjects who mistakenly report a higher risk aversion also mistakenly report a higher belief, it will induce a positive correlation between risk aversion and belief updating in *both* “SELF” and “NON-SELF” conditions. To explain the opposite results in the “SELF” and “NON-SELF” conditions, subjects who mistakenly report a higher risk aversion have to mistakenly report a higher belief in the “SELF” condition, but a lower belief in the “NON-SELF” condition, which we find highly unlikely.

6 General discussion and conclusion

In this study, we theoretically and experimentally show that higher risk aversion leads to a low instrumental need and a higher sensitivity to the non-instrumental need for information. With a psychological need for self-knowledge, i.e., in the “SELF” condition, where subjects receive self-relevant information about their IQ, stronger risk aversion leads to more belief updating. In contrast, when subjects receive self-irrelevant information such that updating cost is more influential, stronger risk aversion leads to less belief updating. Our experiment thus shows a context-dependent relationship between risk attitude and belief updating and also provides supportive evidence for the theory of belief-based utility (Loewenstein and Molnar, 2018).

Contributing to the research program of decision-making under certainty, our results suggest that risk preference and

belief formation are inherently related, and thus, information intervention could have a heterogeneous impact on different individuals. The results speak to the practice and designs of future research on information provision experiments (Haaland et al., 2023). In particular, future research could benefit from collecting data on (elicited or self-reported) risk attitudes as it allows researchers to identify the heterogeneous treatment effects on individuals with different risk attitudes. In contrast, the absence of data on risk attitudes will likely mute the estimated treatment effect, as individuals with stronger (resp. weaker) risk aversion update their beliefs with self-irrelevant (resp. self-relevant) information to a lesser extent. Estimating such heterogeneous treatment effects is particularly important in health economics (e.g., Nyhan et al., 2014; Nyhan and Reifler, 2015) as the target audience includes vulnerable, elderly, citizens who are typically more risk averse.

Conceptually, this study complements previous research about risk aversion and information acquisition/avoidance (Mehrez, 1985; Willinger, 1989; Ho et al., 2021). For example, Ho et al. (2021) finds that more risk-averse participants choose to avoid information to avoid risks of acquiring unfavorable or inaccurate information. Our study supplements their findings as we analyze how individuals update their belief upon receiving information. Our findings therefore apply in many situations where information is involuntarily received, for example, via social media, advertisements, or political campaigns. Our results additionally offer a potential alternative explanation to the result in Ho et al. (2021): as individuals with stronger risk averse anticipate their over-reaction to self-relevant information, when information quality is unknown, they have more incentive to avoid information in advance to protect themselves from inaccurate or unfavorable information.

Lastly, our results have important implications on advertisement, communication, and persuasion, and on how to better persuade or convey information to risk-averse individuals. We expect our results to provide firms guidance for advertisement strategies as well as inspire future marketing research. For example, as the relationship between risk attitude and belief updating

is context-dependent, our results suggest different framing of advertisement is needed to target more- or less-risk-averse consumers. In particular, relating information to oneself (more personally) compels more risk-averse individuals to learn more, while “context-neutral” information compels more risk-averse individuals to learn less. Thus, for firms that target risk-averse individuals, for example, insurance companies, a plain “facts and statistics” type of advertisement might not be as effective as advertisements that connect the product to the consumers on a personal level. More research needs to be done on how effective different advertisement works on different groups of consumers.

Similarly, our results have important implications for political campaigns. For example, to target older constituencies who are more risk averse (Albert and Duffy, 2012), politicians should have their messages framed with a higher self-relevance so as to emphasize the utility for reduced uncertainty, such as using metaphors in political campaigns (Musolf, 2017). Similarly, as older constituencies with stronger risk aversion update more in a self-relevant context, it potentially explains the increased polarization among elderly citizens (Boxell et al., 2017).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Research Ethics Committee at Hong Kong Baptist University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

BL: Methodology, Project administration, Writing—original draft, Writing—review & editing. EH: Conceptualization,

Methodology, Software, Formal analysis, Writing—original draft, Writing—review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We thank David Ahlstrom, Dino Levy, and participants in various seminars and conferences for their valuable comments and suggestions. The experiment was approved by the Research Ethics Committee at Hong Kong Baptist University (ref. no: REC/22-23/0023).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1281296/full#supplementary-material>

References

- Albert, S. M., and Duffy, J. (2012). Differences in risk aversion between young and older adults. *Neurosci. Neuroecon.* 2012, 3–9. doi: 10.2147/NAN.S27184
- Alós-Ferrer, C., and Garagnani, M. (2023). Part-time Bayesians: incentives and behavioral heterogeneity in belief updating. *Manage. Sci.* 69, 5523–5542. doi: 10.1287/mnsc.2022.4584
- Argaw, B., Maier, M. F., and Skriabikova, O. J. (2017). “Risk attitudes, job mobility and subsequent wage growth during the early career,” in *ZEW-Centre for European Economic Research Discussion Paper*, 17–23. doi: 10.2139/ssrn.2977217
- Arrieta, A., García-Prado, A., González, P., and Pinto-Prades, J. L. (2017). Risk attitudes in medical decisions for others: an experimental approach. *Health Econ.* 26, 97–113. doi: 10.1002/hec.3628
- Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *J. Pers. Soc. Psychol.* 43:425. doi: 10.1037/0022-3514.43.3.425
- Bénabou, R., and Tirole, J. (2004). Willpower and personal rules. *J. Polit. Econ.* 112, 848–886. doi: 10.1086/421167
- Bergemann, D., and Morris, S. (2019). Information design: a unified perspective. *J. Econ. Literat.* 57, 44–95. doi: 10.1257/jel.20181489
- Berlin, N., and Dargnies, M.-P. (2016). Gender differences in reactions to feedback and willingness to compete. *J. Econ. Behav. Organ.* 130, 320–336. doi: 10.1016/j.jebo.2016.08.002
- Bonin, H., Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2007). Cross-sectional earnings risk and occupational sorting: the role of risk attitudes. *Lab. Econ.* 14, 926–937. doi: 10.1016/j.labeco.2007.06.007
- Boxell, L., Gentzkow, M., and Shapiro, J. M. (2017). Greater internet use is not associated with faster growth in political polarization among us demographic groups. *Proc. Natl. Acad. Sci. U.S.A.* 114, 10612–10617. doi: 10.1073/pnas.1706588114
- Brunnermeier, M. K., and Parker, J. A. (2005). Optimal expectations. *Am. Econ. Rev.* 95, 1092–1118. doi: 10.1257/0002828054825493
- Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *Q. J. Econ.* 129, 1409–1447. doi: 10.1093/qje/qju009

- Castagnetti, A., and Schmacker, R. (2022). Protecting the ego: motivated information selection and updating. *Eur. Econ. Rev.* 142:104007. doi: 10.1016/j.euroecorev.2021.104007
- Coffman, K., Collis, M. R., and Kulkarni, L. (2023). Stereotypes and belief updating. *J. Eur. Econ. Assoc.* jvad063. doi: 10.1093/jea/jvad063
- Compte, O., and Postlewaite, A. (2019). *Ignorance and Uncertainty*. Cambridge University Press. doi: 10.1017/9781108379991
- Coutts, A. (2019). Good news and bad news are still news: experimental evidence on belief updating. *Exp. Econ.* 22, 369–395. doi: 10.1007/s10683-018-9572-5
- Croson, R., and Gneezy, U. (2009). Gender differences in preferences. *J. Econ. Literat.* 47, 448–474. doi: 10.1257/jel.47.2.448
- Eil, D., and Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *Am. Econ. J.* 3, 114–38. doi: 10.1257/mic.3.2.114
- Erkal, N., Gangadharan, L., and Koh, B. H. (2020). Replication: belief elicitation with quadratic and binarized scoring rules. *J. Econ. Psychol.* 81:102315. doi: 10.1016/j.joep.2020.102315
- Gaba, A., and Kalra, A. (1999). Risk behavior in response to quotas and contests. *Market. Sci.* 18, 417–434. doi: 10.1287/mksc.18.3.417
- Gray, H. M., Ambady, N., Lowenthal, W. T., and Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *J. Exp. Soc. Psychol.* 40, 216–224. doi: 10.1016/S0022-1031(03)00092-1
- Haaland, I., Roth, C., and Wohlfart, J. (2023). Designing information provision experiments. *J. Econ. Literat.* 61, 3–40. doi: 10.1257/jel.20211658
- Hlavac, M., and Hlavac, M. M. (2022). *Package “stargazer.”*
- Ho, E. H., Hagmann, D., and Loewenstein, G. (2021). Measuring information preferences. *Manage. Sci.* 67, 126–145. doi: 10.1287/mnsc.2019.3543
- Holt, C. A., and Laury, S. K. (2002). Risk aversion and incentive effects. *Am. Econ. Rev.* 92, 1644–1655. doi: 10.1257/000282802762024700
- Hossain, T., and Okui, R. (2013). The binarized scoring rule. *Rev. Econ. Stud.* 80, 984–1001. doi: 10.1093/restud/rdt006
- Jaeger, D. A., Dohmen, T., Falk, A., Huffman, D., Sunde, U., and Bonin, H. (2010). Direct evidence on risk attitudes and migration. *Rev. Econ. Stat.* 92, 684–689. doi: 10.1162/REST_a_00020
- Jeong, H. J., and Kwon, K.-N. (2012). The effectiveness of two online persuasion claims: limited product availability and product popularity. *J. Promot. Manage.* 18, 83–99. doi: 10.1080/10496491.2012.646221
- Jindal, P. (2015). Risk preferences and demand drivers of extended warranties. *Market. Sci.* 34, 39–58. doi: 10.1287/mksc.2014.0879
- Krpan, D., Makki, F., Saleh, N., Brink, S. I., and Klauznicer, H. V. (2021). When behavioural science can make a difference in times of COVID-19. *Behav. Public Policy* 5, 153–179. doi: 10.1017/bpp.2020.48
- Loewenstein, G., and Molnar, A. (2018). The renaissance of belief-based utility in economics. *Nat. Hum. Behav.* 2, 166–167. doi: 10.1038/s41562-018-0301-z
- Massin, S., Nebout, A., and Ventelou, B. (2018). Predicting medical practices using various risk attitude measures. *Eur. J. Health Econ.* 19, 843–860. doi: 10.1007/s10198-017-0925-3
- Matzler, K., Grabner-Kräuter, S., and Bidmon, S. (2008). Risk aversion and brand loyalty: the mediating role of brand trust and brand affect. *J. Prod. Brand Manage.* 17, 154–162. doi: 10.1108/10610420810875070
- Mehrez, A. (1985). The effect of risk aversion on the expected value of perfect information. *Oper. Res.* 33, 455–458. doi: 10.1287/opre.33.2.455
- Musolf, A. (2017). Truths, lies and figurative scenarios: metaphors at the heart of brexit. *J. Lang. Polit.* 16, 641–657. doi: 10.1075/jlp.16033.mus
- Noussair, C. N., Trautmann, S. T., and Van de Kuilen, G. (2014). Higher order risk attitudes, demographics, and financial decisions. *Rev. Econ. Stud.* 81, 325–355. doi: 10.1093/restud/rdt032
- Nyhan, B., and Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine* 33, 459–464. doi: 10.1016/j.vaccine.2014.11.017
- Nyhan, B., Reifler, J., Richey, S., and Freed, G. L. (2014). Effective messages in vaccine promotion: a randomized trial. *Pediatrics* 133, e835–e842. doi: 10.1542/peds.2013-2365
- Oehler, A., Horn, M., and Wedlich, F. (2018). Young adults’ subjective and objective risk attitude in financial decision making: evidence from the lab and the field. *Rev. Behav. Fin.* 10, 274–294. doi: 10.1108/RBF-07-2017-0069
- Okada, E. M. (2010). Uncertainty, risk aversion, and WTA vs. WTP. *Market. Sci.* 29, 75–84. doi: 10.1287/mksc.1080.0480
- Oprea, R., and Yuksel, S. (2022). Social exchange of motivated beliefs. *J. Eur. Econ. Assoc.* 20, 667–699. doi: 10.1093/jea/jvab035
- Rosen, A. B., Tsai, J. S., and Downs, S. M. (2003). Variations in risk attitude across race, gender, and education. *Med. Decis. Mak.* 23, 511–517. doi: 10.1177/0272989X03258431
- Saccardo, S., Pietrasz, A., and Gneezy, U. (2018). On the size of the gender difference in competitiveness. *Manage. Sci.* 64, 1541–1554. doi: 10.1287/mnsc.2016.2673
- Shapiro, D. (2011). Profitability of the name-your-own-price channel in the case of risk-averse buyers. *Market. Sci.* 30, 290–304. doi: 10.1287/mksc.1100.0622
- Shapiro, K. L., Caldwell, J., and Sorensen, R. E. (1997). Personal names and the attentional blink: a visual “cocktail party” effect. *J. Exp. Psychol.* 23:504. doi: 10.1037/0096-1523.23.2.504
- Sinclair, A. H., Stanley, M. L., and Seli, P. (2020). Closed-minded cognition: right-wing authoritarianism is negatively related to belief updating following prediction error. *Psychon. Bull. Rev.* 27, 1348–1361. doi: 10.3758/s13423-020-01767-y
- Sui, J., Zhu, Y., and Han, S. (2006). Self-face recognition in attended and unattended conditions: an event-related brain potential study. *Neuroreport* 17, 423–427. doi: 10.1097/01.wnr.0000203357.65190.61
- Symons, C. S., and Johnson, B. T. (1997). The self-reference effect in memory: a meta-analysis. *Psychol. Bull.* 121:371. doi: 10.1037/0033-2909.121.3.371
- Turk, D. J., Van Bussel, K., Brebner, J. L., Toma, A. S., Krigolson, O., and Handy, T. C. (2011). When “it” becomes “mine”: attentional biases triggered by object ownership. *J. Cogn. Neurosci.* 23, 3725–3733. doi: 10.1162/jocn_a_00101
- Wakebe, T., Sato, T., Watamura, E., and Takano, Y. (2012). Risk aversion in information seeking. *J. Cogn. Psychol.* 24, 125–133. doi: 10.1080/20445911.2011.596825
- Weber, E. U., Blais, A.-R., and Betz, N. E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *J. Behav. Decis. Mak.* 15, 263–290. doi: 10.1002/bdm.414
- Willinger, M. (1989). Risk aversion and the value of information. *J. Risk Insurance* 56, 104–112. doi: 10.2307/253017
- Wilson, A. (2014). Bounded memory and biases in information processing. *Econometrica* 82, 2257–2294. doi: 10.3982/ECTA12188

A Proofs

A.1 Proof of proposition 1

Proof. The first derivative of U_2 w.r.t. γ is

$$\begin{aligned}\frac{\partial U_2(a)}{\partial \gamma} &= \int [p_s^S u'(\pi^1(a)) \pi^{1'}(a) \\ &\quad + (1 - p_s^S) u'(\pi^0(a)) \pi^{0'}(a)] \frac{p_s^S - 0.5}{\lambda(\bar{p} - p)} d p_s^S \\ &= \int [p_s^S (\pi^1)^{-\theta} (2 - 2a) \\ &\quad - (1 - p_s^S) (\pi^0)^{-\theta} (2a)] \frac{p_s^S - 0.5}{\lambda(\bar{p} - p)} d p_s^S.\end{aligned}$$

And the second derivative of U_2 w.r.t. a is

$$\begin{aligned}\frac{\partial^2 U_2(a)}{\partial \gamma^2} &= \int \frac{p_s^S - 0.5}{\lambda(\bar{p} - p)} \left[p_s^S (\pi^1)^{-\theta} (-2) - \right. \\ &\quad \left. (1 - p_s^S) (\pi^0)^{-\theta} (2) - \theta p_s^S (\pi^1)^{-\theta-1} (2 - 2a)^2 - \right. \\ &\quad \left. \theta (1 - p_s^S) (\pi^0)^{-\theta-1} (2a)^2 \right] d p_s^S < 0.\end{aligned}$$

Thus, γ^* is uniquely pinned down by the first order condition. Next, using implicit differentiation, we have

$$\begin{aligned}\frac{\partial \gamma^*}{\partial \theta} &= - \frac{\partial^2 U_2(p_s^S, a)}{\partial \gamma \partial \theta} / \frac{\partial^2 U_2(p_s^S, a)}{\partial \gamma^2} \\ &= \int \frac{p_s^S - 0.5}{\lambda(\bar{p} - p)} \left[p_s^S \log(\pi^1) (\pi^1)^{-\theta} (2 - 2a) \right. \\ &\quad \left. - (1 - p_s^S) \log(\pi^0) (\pi^0)^{-\theta} (2a) \right] d p_s^S / \frac{\partial^2 U_2(a)}{\partial \gamma^2} \\ &= \int \frac{p_s^S - 0.5}{\lambda(\bar{p} - p)} p_s^S (\pi^1)^{-\theta} (2 - 2a) [\log(\pi^1) \\ &\quad - \log(\pi^0)] d p_s^S / \frac{\partial^2 U_2(a)}{\partial \gamma^2}\end{aligned}$$

where the last equality is implied by the first order condition. When $p_s^S > 0.5$, the first derivative $\frac{\partial U_2(a)}{\partial \gamma} |_{\gamma=0} > 0$ and, thus, $a > 0.5$ and $\pi^1 > \pi^0$. Thus, $\frac{\partial \gamma^*}{\partial \theta} < 0$ and the results follow.

A.2 Proof of proposition 2

Proof. We first prove the first bullet point of the proposition. Equation (2) in the main text can be rewritten as

$$\min_{\lambda} (\gamma^* (1 - \lambda))^2 \text{Var}(p_s^B) - w \lambda^2 \text{Var}(p_s^B)$$

The first order condition is

$$\begin{aligned}-2\gamma^* (1 - \lambda^*) \text{Var}(p_s^B) - 2w \lambda^* \text{Var}(p_s^B) &= 0 \\ \Leftrightarrow \lambda^* &= \frac{\gamma^*}{\gamma^* - w}\end{aligned}$$

and the second-order derivative is $(2\gamma^* - 2w) \text{Var}(p_s^B)$ which is positive if and only if w is small enough. When the second-order derivative is positive, as $\frac{\partial \gamma^*}{\partial \theta} < 0$, the result follows. On the other hand, when the second-order derivative is negative, $\lambda^* = \infty$, and the result trivially follows.

Similarly, for the second bullet point of the proposition, Equation (2) in the main text can be rewritten as

$$\min_{\lambda} (\gamma^* (1 - \lambda))^2 \text{Var}(p_s^B) + w \lambda^2 \text{Var}(p_s^B)$$

The first order condition is

$$\begin{aligned}-2\gamma^* (1 - \lambda^*) \text{Var}(p_s^B) + 2w \lambda^* \text{Var}(p_s^B) &= 0 \\ \Leftrightarrow \lambda^* &= \frac{\gamma^*}{\gamma^* + w}\end{aligned}$$

and the second-order derivative is $(2\gamma^* + 2w) \text{Var}(p_s^B)$ which is positive. Again, as $\frac{\partial \gamma^*}{\partial \theta} < 0$, the result follows.



OPEN ACCESS

EDITED BY

Riccardo Viale,
University of Milano-Bicocca, Italy

REVIEWED BY

Elisabet Tubau,
University of Barcelona, Spain
Ulrich Hoffrage,
Université de Lausanne, Switzerland

*CORRESPONDENCE

Nathalie Stegmüller
✉ nathalie.stegmueller@ur.de

RECEIVED 18 September 2023

ACCEPTED 15 January 2024

PUBLISHED 10 April 2024

CITATION

Stegmüller N, Binder K and Krauss S (2024)
How general is the natural frequency effect?
The case of joint probabilities.
Front. Psychol. 15:1296359.
doi: 10.3389/fpsyg.2024.1296359

COPYRIGHT

© 2024 Stegmüller, Binder and Krauss. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

How general is the natural frequency effect? The case of joint probabilities

Nathalie Stegmüller^{1*}, Karin Binder² and Stefan Krauss¹

¹Mathematics Education, Faculty of Mathematics, University of Regensburg, Regensburg, Germany,

²Mathematics Education, Institute of Mathematics, Ludwig Maximilian University Munich, Munich, Germany

Natural frequencies are known to improve performance in Bayesian reasoning. However, their impact in situations with two binary events has not yet been completely examined, as most researchers in the last 30 years focused only on conditional probabilities. Nevertheless, situations with two binary events consist of 16 elementary probabilities and so we widen the scope and focus on joint probabilities. In this article, we theoretically elaborate on the importance of joint probabilities, for example, in situations like the Linda problem. Furthermore, we implemented a study in a 2×5×2 design with the factors information format (probabilities vs. natural frequencies), visualization type ("Bayesian text" vs. tree diagram vs. double tree diagram vs. net diagram vs. 2×2 table), and context (mammography vs. economics problem). Additionally, all four "joint questions" (i.e., $P(A \cap B)$, $P(\bar{A} \cap B)$, $P(\bar{A} \cap \bar{B})$, $P(A \cap \bar{B})$) were asked for. The main factor of interest was whether there is a format effect in the five visualization types named above. Surprisingly, the advantage of natural frequencies was not found for joint probabilities and, most strikingly, the format interacted with the visualization type. Specifically, while people's understanding of joint probabilities in a double tree seems to be worse than the understanding of the corresponding natural frequencies (and, thus, the frequency effect holds true), the opposite seems to be true in the 2×2 table. Hence, the advantage of natural frequencies compared to probabilities in typical Bayesian tasks cannot be found in the same way when joint probability or frequency tasks are asked.

KEYWORDS

joint probabilities, Bayesian reasoning, natural frequencies, visualization, net diagram

1 Introduction

There is an interesting tension in empirical research on the understanding of *joint probabilities* (formal: e.g., $P(A \cap B)$). On one hand, researchers have stressed the importance of comprehending joint probabilities, e.g., in the legal context (O'Grady, 2023) and conducted empirical studies (e.g., Tversky and Kahneman, 1974; Donati et al., 2019). On the other hand, psychological studies mostly just ask for a *qualitative comparison* of $P(A)$ and $P(A \cap B)$ without the need for participants to assess a concrete joint probability. Let us, for example, consider the most famous instance of the so-called conjunction fallacy, namely the Linda problem (introduced by Tversky and Kahneman, 1983).

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and she also participated in anti-nuclear demonstrations. Which is more probable?

1. Linda is a bank teller.
2. Linda is a bank teller and is active in the feminist movement.

Let “A” be the event “being active in the feminist movement” and “B” “being a bank teller.” Since $B \cap A$ (being a bank teller *and* being active in the feminist movement) is a subset of B (being a bank teller), the single event B is more probable than both events at the same time. Formally, the multiplication rule concerning joint probabilities is $P(B \cap A) = P(B) \cdot P(A|B)$ and because $P(B)$ must be multiplied with a probability, i.e., a number between 0 and 1, $P(B \cap A)$ cannot be larger than $P(B)$.

Yet, the fact that no concrete probability has to be estimated or calculated stands in strong contrast to the way *conditional probabilities* are examined in cognitive psychology, for example, in the framework of *Bayesian reasoning* in which specific estimates have to be given by participants (see Theoretical Framework).

For requesting a concrete joint probability in the Linda task, participants, for instance, might be asked:

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and she also participated in anti-nuclear demonstrations. Assume that the probability that Linda is a bank teller is 5%. Assume that the probability that she is active in the feminist movement, if she is a bank teller, is 20%. What is the probability that she is a bank teller and active in the feminist movement?

Now, the *multiplication rule* based on the given information yields $P(B \cap A) = P(B) \cdot P(A|B) = 5\% \cdot 20\% = 1\%$. Considering this rule, it becomes clear that joint probabilities, i.e., $P(A \cap B)$, are deeply interwoven with conditional probabilities, i.e., $P(A|B)$. Joint probabilities are even used for defining conditional probabilities in mathematics ($P(A|B) = P(A \cap B)/P(B)$). The tension in psychological research is that joint probabilities are stressed as very relevant, but at the same time concrete joint probabilities usually do not have to be calculated by participants. In the present study, we investigate people's assessment of concrete numerical values of joint probabilities. The main aim is to explore, whether the so-called “natural frequency effect” (that helps participants assess conditional probabilities) can also be found for joint probability judgments.

2 Theoretical framework

In the following, we first embed the structure of the Linda problem in the larger framework of Bayesian reasoning situations consisting of two binary events. In general, in the statistical world of two binary events A and B (with the counter events \bar{A} and \bar{B}), one can consider 16 different elementary probabilities:¹

- Four *marginal probabilities*: $P(A)$, $P(\bar{A})$, $P(B)$, $P(\bar{B})$
- Four *joint probabilities*: $P(A \cap B)$, $P(\bar{A} \cap B)$, $P(\bar{A} \cap \bar{B})$, $P(A \cap \bar{B})$
- Eight *conditional probabilities*:
 $P(A|B)$, $P(A|\bar{B})$, $P(\bar{A}|B)$, $P(\bar{A}|\bar{B})$, $P(B|A)$, $P(B|\bar{A})$, $P(\bar{B}|A)$, $P(\bar{B}|\bar{A})$

Note that in the case of stochastic independence of both events, $P(A|B)$ equals $P(A)$ and, thus, the multiplication rule can be simplified:

- A and B are stochastic dependent: $P(B \cap A) = P(B) \cdot P(A|B)$
- A and B are stochastic independent: $P(B \cap A) = P(B) \cdot P(A)$

Ignoring the dependency of two events was, by the way, one of several problems in the famous miscarriage of justice concerning Sally Clark (Colmez and Schneps, 2013) or the one of Kathleen Folbigg (O'Grady, 2023), which again stresses the importance of understanding joint probabilities (including concrete values). After two infants of Sally Clark died shortly after birth, she was convicted of murdering her children. The court knew that the sudden infant death syndrome (SIDS) occurs with a chance of about 1 in 8500 cases. After not only one infant but two of her children died, it was considered to be very unlikely that this happened by chance, particularly under the wrong assumption that these two deaths were *independent* of each other.

Consequently, the chance for two children suffering from SIDS was

calculated as $\frac{1}{8500} \cdot \frac{1}{8500}$ ($\approx 0.0000014\%$), whereupon she was

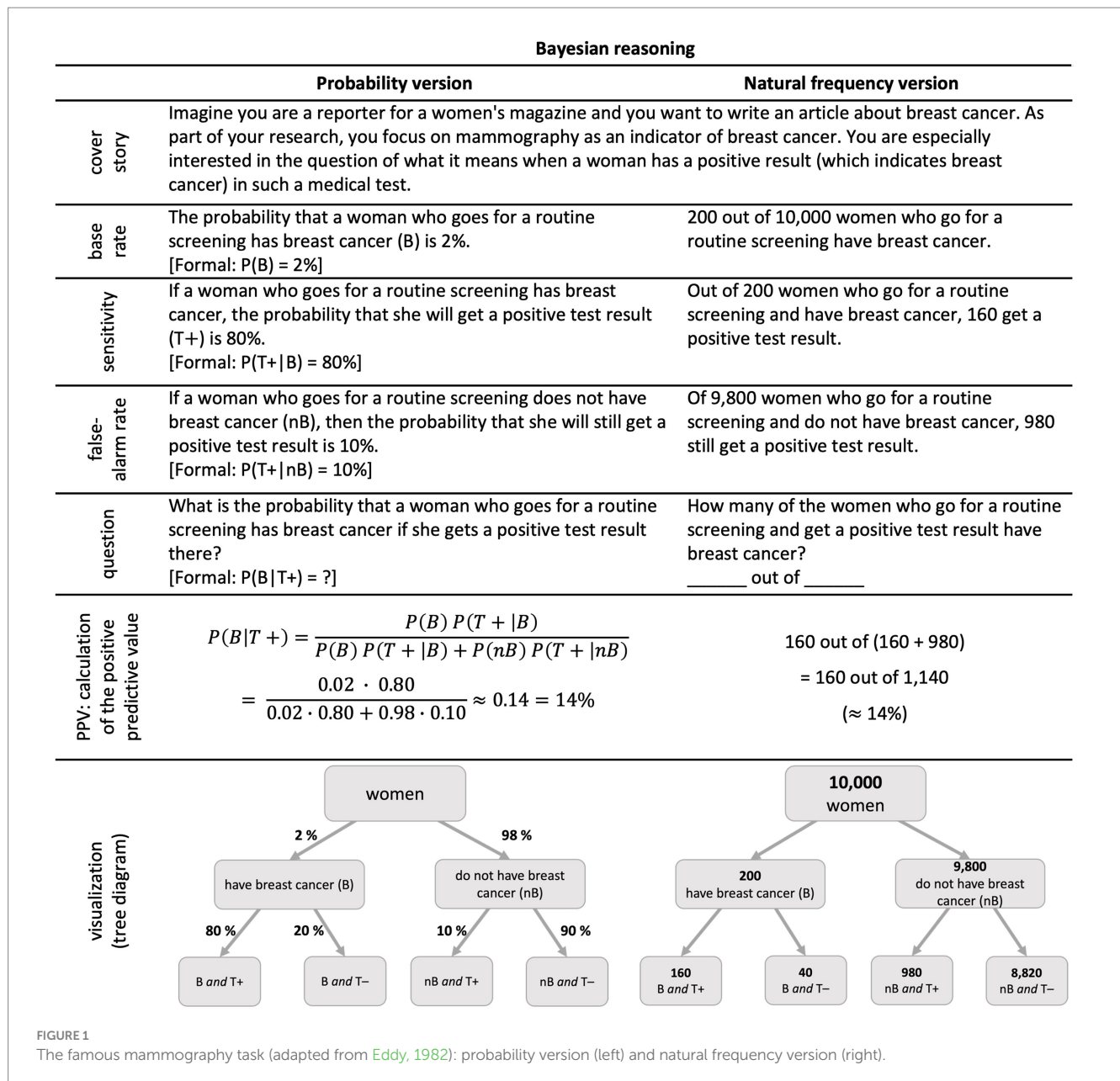
convicted of being a murderer. However, a second SIDS is more probable given a first one already happened (Glinge et al., 2023). As soon as this was stated clearly, Clark was released from prison (after three years of her sentence); nevertheless, her life had been destroyed (Colmez and Schneps, 2013). In a similar, more recent criminal case, Kathleen Folbigg was convicted of murdering three of her infant children and of manslaughter of her fourth child (Phillips, 2022). This verdict was based on the same misunderstanding as Clark's—the court assumed that four children could not independently die by accident but only by being murdered. After scientists, though, had analyzed the case for about 20 years and had proven a gene mutation in the family, Folbigg was finally released from prison in 2023 (Wells et al., 2023).

2.1 Bayesian reasoning and natural frequencies

In psychological research on situations with two binary events, typically *Bayesian reasoning* is investigated empirically. For this, a specific set of probabilities is given, and a concrete probability is required (Figure 1). In more detail, the “positive predictive value” $P(B|T+)$ has to be inferred from (1) the base rate $P(B)$, (2) the sensitivity $P(T+|B)$, and (3) the false-alarm rate $P(T+|nB)$, which reflects the typical setting of diagnostic situations. Figure 1 displays the famous mammography task (adapted from Eddy, 1982). Since the issue of joint probabilities is strongly related to such diagnostic reasoning, we first take a short look at the research area of Bayesian reasoning. Many studies documented the difficulties people—laymen and experts like physicians—have with such problems, especially when they are formulated in terms of probabilities (Figure 1, left; Gigerenzer and Hoffrage, 1995; Garcia-Retamero and Hoffrage, 2013; Binder et al., 2015; Bruckmaier et al., 2019).

In research on Bayesian reasoning, it turned out that a reformulation with so-called “natural frequencies” (Figure 1, right side) helps people to understand such situations (Gigerenzer and

¹ Of course, there are also the trivial probabilities $P(\emptyset)$ and $P(\Omega)$ as well as all probabilities regarding set unions, e.g., $P(A \cup B)$. An extensive overview and discussion of all possible cases can be found in Neth et al. (2021).



Hoffrage, 1995; Siegrist and Keller, 2011). Natural frequencies are a pair of natural numbers a and b ($a \leq b$), which are equivalent to percentages and used as “ a out of b ” (Krauss et al., 2020). Sometimes, people distinguish between “percentages” and “natural frequencies” instead of “probabilities” and “natural frequencies” (e.g., Knapp et al., 2009). In this article, we use the latter distinction. A meta-analysis revealed that on average in probability versions (without visualization) usually only 4% of people can solve such tasks correctly, while, in natural frequency versions (also without visualizations), 24% of people find the correct solution (McDowell and Jacobs, 2017).

Natural frequencies are helpful because the calculations are simpler compared to the probability version (Figure 1) and, thus, the solution can be accessed more easily (Gigerenzer and Hoffrage, 1995). The higher solution rates can, therefore, also be explained by the number of mental steps that are needed to solve the problem. In the

probability format, the correct solution has to be calculated using a sophisticated formula, while people only have to identify two correct numbers and do a simple addition in the frequency format. Studies show that Bayesian tasks are solved more correctly the less mental steps are needed (Ayal and Beyth-Marom, 2014).

Note that, in the tree diagram (Figure 1, left), conditional probabilities are depicted at the lower arrows, for instance the sensitivity $P(T+|B)$ of 80%, represented at the very left branch. Joint probabilities are *not* depicted. However, $P(B \cap T+)$, for example, might be calculated according to the multiplication rule above by $P(B \cap T+) = P(B) \cdot P(T+|B) = 2\% \cdot 80\% = 1.6\%$.

In typical Bayesian reasoning tasks, joint probabilities are neither given nor asked for. For an exception for *giving* joint probabilities see the “short menu” in Gigerenzer and Hoffrage (1995); for exceptions for *asking* for joint probabilities see Böcherer-Linder and Eichler (2017), Bruckmaier et al. (2019), or Binder et al. (2020).

From the perspective of the widespread research on Bayesian reasoning and the largely documented effect of natural frequencies, however, it is an interesting question, whether natural frequency formulations would also help understanding notorious joint probabilities. This is especially intriguing since Bayes formula (Figure 1, left) could alternatively be written as

$$P(B|T+) = \frac{P(B \cap T+)}{P(B \cap T+) + P(nB \cap T+)}$$

While 16 probabilities are available in statistical situations with two binary events, empirical research has, to a very large extent, primarily focused on Bayesian reasoning tasks. The enormous effect of natural frequencies in such basic diagnostic tasks motivates the question what happens in related or extended problem-solving situations.

2.2 Extensions of Bayesian reasoning—and the respective help of natural frequencies

Before we address a possible generalization of the natural frequency effect from Bayesian reasoning to joint probabilities in detail (see section 2.3), we first shed light on the potential of natural frequencies in alternative extensions of Bayesian reasoning. The following paragraphs summarize various possible extensions of Bayesian reasoning and whether studies document that natural frequencies also help in these cases. Interestingly, there seems to be a clear format effect as long as conditional probabilities are considered. When it comes to joint probabilities, though, there does not seem to be an overall format effect in favor of natural frequencies because the evidence is mixed.

To explain extensions 1–3, medical contexts are used in the following.

2.2.1 Increasing the number of tests (extension 1a)

One possible extension of Bayesian reasoning would be to vary the number of medical tests applied. In the context of breast cancer, for instance, after a mammography screening, an ultrasound test might be applied to verify the test results (which would yield another level in the tree diagram in Figure 1, e.g., Binder et al., 2018). Krauss et al. (1999), for example, found that natural frequency versions were more than four times as likely to be solved correctly than probability versions. Similar results can be found in Woike et al. (2017).

2.2.2 Increasing the number of test (or criterion) values (extension 1b)

Another way of altering Bayesian reasoning is to increase the number of test and/or criterion values. For instance, a medical test might have three different outcomes (positive, negative, unclear). In the same manner, a medical test can be sensitive to two different diseases, which would result in three possible criterion values (e.g., diabetes type 1, diabetes type 2, or healthy). Modeling three (or even more) possible test outcomes as well as three (or even more) possible health statuses would lead to three (or more) nodes in a tree diagram in the second or in the third level, respectively. Formulating tasks in such complex situations in natural frequencies leads to about 50% of

correct performances of participants (Hoffrage et al., 2015). Binder and Krauss (under review) confirm these results and give an extensive overview of studies on such types of generalization (i.e., 1a and 1b).

2.2.3 Covariational reasoning (extension 2)

Another interesting way of extending the classical Bayesian reasoning task would be to consider whether people are aware of the consequences of *changing* one of the three input variables (i.e., base rate, sensitivity, false-alarm rate) on the positive predictive value. Even though such kind of reasoning is very complex, some people, nevertheless, can correctly judge the direction of change of the positive predictive value after a respective training, when it is based on the natural frequency concept (Steib et al., 2023; Büchter et al., 2024).

2.2.4 Communication skills (extension 3)

The *communication quality* in Bayesian situations is a further aspect worth to consider. Since Bayesian situations often occur in medical contexts in which a physician is supposed to advise patients, the way of (verbally) communicating the meaning of a positive test result is very important (Gigerenzer et al., 1998; Brose et al., 2023). Unfortunately, counselors are not always communicating the results in a correct and comprehensible way (Gigerenzer et al., 1998; Ellis and Brase, 2015; Prinz et al., 2015) and medical students cannot even identify a high-quality communication with the correct value when it is presented as one out of several short video clips (Böcherer-Linder et al., 2022). To improve (pictorial) communication, the Harding Center for Risk Literacy developed fact boxes and icon boxes (Schwartz et al., 2007; McDowell et al., 2019), which are also based on the concept of natural frequencies. Clearly, verbal and pictorial communication can benefit from the frequency effect.

2.3 The issue of joint probabilities: Do natural frequencies help?

The extensions discussed so far (1–3) deal with *conditional probabilities*. However, there are 16 elementary probabilities available in Bayesian situations (see above). Thus, it is an interesting question whether natural frequencies help in a similar way when questions on *joint probabilities* are posed. In the following paragraphs, we analyze empirical evidence collected so far. First (in 2.3.1), we summarize experimental results concerning the *qualitative* comparison of $P(A \cap B)$ and $P(A)$. Afterwards (in 2.3.2), we turn to *quantitative* tasks in which a *concrete* probability is asked for. Finally, we conclude that the evidence regarding the help of natural frequencies concerning joint probabilities is mixed and explain the limitations of the studies conducted so far.

2.3.1 Qualitative comparison of $P(A \cap B)$ and $P(A)$

Besides the original study of the Linda problem by Tversky and Kahneman (1983), many studies document that people consider the second option with two events at the same time as more likely as the first option with only one event (e.g., Charness et al., 2009; Donati et al., 2019). However, as demonstrated above, one single event is *always* more probable than the simultaneous occurrence of this event and an additional event.

Since the background information on Linda, which is irrelevant for the multiplication rule, seems to make option 2 more plausible,

Tversky and Kahneman (1983) explain people's difficulties by the *representativeness heuristic*, which can sometimes lead to misjudgments. Yet, there are alternative explanations, for instance, that the word “and” in everyday communication has many different meanings (Mellers et al., 2001; Hertwig et al., 2008). Another explanation of the fallacy is that people interpret the first event “Linda is a bank teller” in reminiscence to the second option as “Linda is a bank teller and is NOT active in the feminist movement” (Hertwig et al., 2008).

Nonetheless, similar difficulties occur in related tasks like for example in “rolling the dice” (Tversky and Kahneman, 1983) in which the events are not formulated literally, and, therefore, such linguistic problems cannot explain participants' misconceptions.

Consider a regular six-sided dice with four green faces and two red faces. The dice will be rolled 20 times and the sequence of greens (G) and reds (R) will be recorded. You are asked to select one sequence from a set of three and you will win \$25 if the sequence you chose appears on successive rolls of the dice. Please check the sequence of greens and reds on which you prefer to bet.

1. RGRRR
2. GRGRRR
3. GRRRRR

In this task, three options (instead of two) are given, but, again, one (1.) is a subset of another (2.). Most participants orientated themselves on the probabilities of rolling a green face (4/6) and of rolling a red face (2/6) and, therefore, chose sequence 2, which includes more green faces compared to sequence 1, both absolutely and relatively, and is, therefore, more representative regarding the provided information (Tversky and Kahneman, 1983). The first sequence, though, again is more probable than the second one since the latter includes the first one.

To what extent can natural frequencies help in both problems? Note that neither in the “Linda problem” nor in “rolling the dice” concrete probabilities are asked for.² However, at least a “frequentist formulation” of both problems is possible, for instance: “Which option occurs most often?” In the Linda task, such a formulation does not seem possible at first sight, since the task is about a single event probability (Linda is only one person). Even in this case, though, one can imagine, for example, 200 people, who fit Linda's description (Fiedler, 1988). Picturing these 200 people while asking oneself, how many are (1) bank tellers or (2) bank tellers and simultaneously active in the feminist movement, makes it easier to understand the task regardless of whether such 200 people exist or not (Fiedler, 1988).

Wedell and Moro (2007) investigated the effect of such frequentist questions in multiple similar scenarios (including rolling the dice), but found no systematic differences between probability and frequentist questions. Interestingly, already Inhelder and Piaget (1964) implemented a frequentist question for investigating their so-called *class-inclusion problem*. They concluded that children who are asked

whether there are more red roses or roses in a bouquet often choose the answer “red roses,” although the latter ones clearly are included in the answer “roses.”

Note that in all examples so far only a *qualitative* comparison of $P(A \cap B)$ and $P(A)$ was asked for. While Fiedler (1988) found increased performances based on a frequency question, Wedell and Moro (2007) did not. Also, Inhelder and Piaget (1964) did not identify a frequentist formulation as beneficial, which overall results in mixed evidence.

2.3.2 Calculating $P(A \cap B)$ based on concrete given probabilities

Basically, there are two options for displaying *concrete* probabilities that allow assessing a joint probability. One of them is presenting several concrete pieces of information in a *text* and the other one is to provide statistical information in *visualizations* (also see Figure 2).

Concerning a textual representation, the question arises, which pieces of information should or must be given to determine a correct joint probability answer. In the Bayesian reasoning paradigm both the given pieces of information *and* the specific question are predefined. Interestingly, based on the typical three given pieces of information in a “Bayesian text,” namely $P(B)$, $P(A|B)$, and $P(A|\bar{B})$, not only the positive predictive value, but also all four joint probabilities can be calculated in principle. This set of information is, in so far, “complete” because it allows for the calculation of all 16 elementary probabilities.

It is important to note that for calculating *one specific* joint probability, i.e., $P(A \cap B)$, only two probabilities are needed (e.g., $P(A)$ and $P(B|A)$ or $P(B)$ and $P(A|B)$, respectively). Yet, if *all four* joint probabilities were asked for, more information would be necessary (for a case-by-case analysis see Stegmüller, 2020). For this reason, it is evident that providing a “Bayesian text” allows some generalization potential regarding the judgment of joint probabilities.

When asking for all four joint probabilities based on the mentioned set of given information $P(B)$, $P(A|B)$, $P(A|\bar{B})$, the four types ($P(A \cap B)$, $P(\bar{A} \cap B)$, $P(\bar{A} \cap \bar{B})$, $P(A \cap \bar{B})$) see Table 1) require a different number of mental steps. Looking at Table 1, it becomes clear that, for the first type, all needed factors for answering this “joint question” are directly given in the “Bayesian text” (Figure 1), whereas, for the third type, even two counter probabilities have to be assessed first. In the frequency version, the first and the last type can be inferred by “skipping one level” and reading off the correct numbers only (for the example in Figure 1, e.g., 160 out of 10,000 and 980 out of 10,000, respectively), while the counter events need to be assessed first for the other two joint frequencies (e.g., 40 out of 10,000 and 8,820 out of 10,000, respectively).

A first attempt to ask for a joint probability based on such a “Bayesian text” was made by Binder et al. (2020), however, in this study, only one joint probability was asked for (type 2 in Table 1). Although there was no substantial frequency effect (see Table 2: “Bayesian text”), this finding cannot simply be transferred to the other three joint probabilities, since the different questions require a different number of mental steps (Table 1; Ayal and Beyth-Marom, 2014) and are, thus, not directly comparable.

Another way to provide concrete probabilities that allow to assess a joint probability is to present them in a visualization. Figure 2 displays four visualizations that were already used for joint probability judgements in prior studies (yet, not

² Even though neither probabilities are given nor asked for explicitly, in “rolling the dice,” the probability of all three sequences can be calculated concretely: $P(RGRRR) \approx 0.82\%$, $P(GRGRRR) \approx 0.54\%$ and $P(GRRRRR) \approx 0.27\%$.

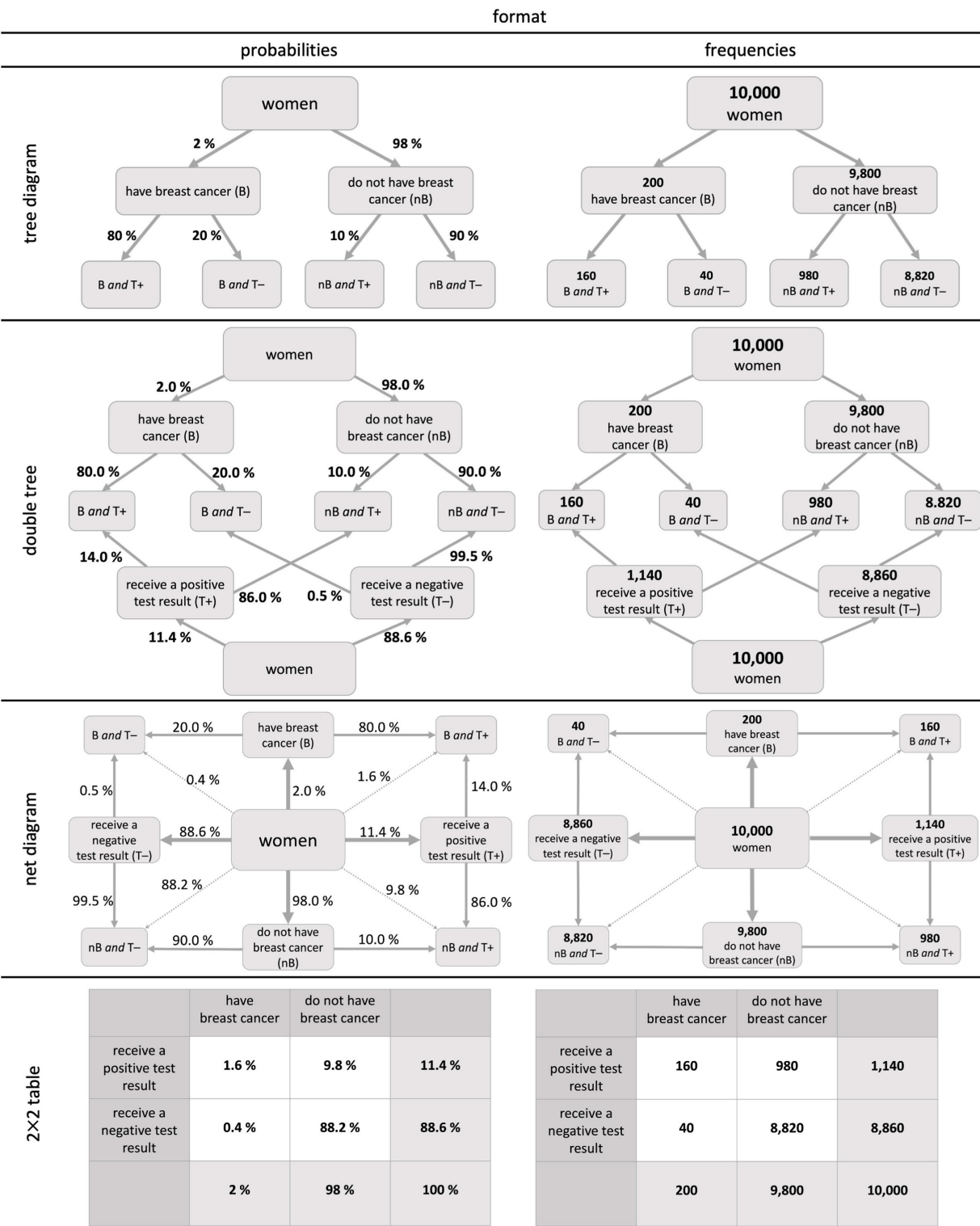


FIGURE 2 Visualizations of two binary events in the context of the mammography problem: Probability versions (left) and frequency versions (right).³

³ Note that because of the plural “women” in our probability trees (e.g., in Figures 1, 2) these trees are basically percentage trees. However, since research in Bayesian reasoning mostly distinguishes between probability and frequency format, we call them probability trees.

TABLE 1 Information given or not given in the “Bayesian text” in both formats; each “X” requires an additional mental step.

	Joint question	Probability format			Frequency format
		Needed calculation	First factor	Second factor	Both needed absolute frequencies
Type 1	$P(A \cap B)$	$P(B) \cdot P(A B)$	✓	✓	✓
Type 2	$P(\bar{A} \cap B)$	$P(B) \cdot P(\bar{A} B)$	✓	X	X
Type 3	$P(\bar{A} \cap \bar{B})$	$P(\bar{B}) \cdot P(\bar{A} \bar{B})$	X	X	X
Type 4	$P(A \cap \bar{B})$	$P(\bar{B}) \cdot P(A \bar{B})$	X	✓	✓

Note that also the questions with the switched event order (e.g., $P(B \cap A)$) have the same calculation steps as the listed ones (e.g., $P(A \cap B)$).
✓: directly given; X: not directly given; probability/frequency of counter event needs to be inferred first.

TABLE 2 Results in previous studies with questions on joint probabilities.

Information format	Bruckmaier et al. (2019)		Binder et al. (2020)			
	Tree diagram	2 × 2 table	“Bayesian text”	Double tree	Net diagram	2 × 2 table
Probabilities	46%	96%	16%	16%	59%	78%
Natural frequencies	50%	79%	22%	48%	45%	52%

systematically) based on the context and numbers given in Figure 1 (Bruckmaier et al., 2019; Binder et al., 2020). Note that the first two (tree diagram and double tree) display *conditional* but no *joint* probabilities. The opposite is true for the 2 × 2 table. Only the net diagram has the advantage of displaying *both* conditional and joint probabilities.

The tree diagram, the double tree, and the net diagram have a node-branch structure in which probabilities can be entered at the branches (Figure 2, left) and frequencies in the nodes (Figure 2, right). Nevertheless, frequencies and probabilities can, in principle, also be included simultaneously (imagine putting the left and the right visualization on top of each other), which makes it possible to depict both formats into the visualization at once (Binder et al., 2023). Thereby, the net diagram is the only visualization that can display all 16 probabilities. This versatility of the net diagram (i.e., all 16 probabilities and all 9 frequencies can be inserted), however, raised the concern that it would lead to a cognitive overload for students or study participants (Henze and Vehling, 2021). In 2 × 2 tables, cells normally *either* include probabilities *or* frequencies.

In both probability trees (simple and double), the answer to all joint questions cannot be read off directly but must be calculated first (e.g., in Figure 2: $P(B \cap T+) = P(B) \cdot P(T+|B) = 2\% \cdot 80\% = 1.6\%$). In the net diagram and the 2 × 2 table in probability format, only the correct numbers have to be read off, which results in fewer mental steps than in the tree diagrams. In the frequency format, however, all visualizations directly deliver the same information, since in each visualization, only the correct two numbers have to be combined without a calculation.

Table 2 presents previous results, when a joint probability question was asked explicitly based on these visualizations (Bruckmaier et al., 2019; Binder et al., 2020). While there seems to be no format effect in the “normal” tree diagram (Bruckmaier et al., 2019), natural frequencies appear to have a positive effect when placed in a double tree (Binder et al., 2020). Interestingly, in 2 × 2

tables, natural frequencies even seem to deteriorate the performance (see both studies in Table 2).

However, Bruckmaier et al. (2019) conducted an eye-tracking study with only 24 participants and Binder et al. (2020) focused predominantly on conditional probabilities (i.e., Bayesian reasoning).

In both studies, previously posed conditional probability questions might have framed participants toward thinking of conditional probabilities and, thereby, might have had an influence on the answer to the following joint probability question. Furthermore, in both studies, only one of the four possible joint probabilities was asked for, namely the one without the need to infer counter events first. Taken together, the findings in Table 2 must be interpreted very carefully.

In the present article, the understanding and assessing of joint probabilities and frequencies in situations with two binary events is examined for the first time systematically. Note that we are not primarily interested in which visualization is better than the other to foster understanding of joint probabilities. Rather, different visualization types have the potential to display statistical information in various ways and, thus, allow exploring possible format effects on a more differentiated level. In principle, we are, therefore, interested in potential interactions of a possible frequency effect with (a) the underlying representation of statistical information and (b) the type of probability question asked $P(A \cap B)$, $P(\bar{A} \cap B)$, $P(B \cap \bar{A})$, $P(\bar{B} \cap A)$. Both perspectives aim at generalizing possible frequency effects regarding the assessment of “joint information.”

3 Present approach

In the present study, we investigate people’s ability to assess concrete joint probabilities or frequencies based on various ways to represent statistical information. To study format effects, we considered five different “visualization types,” namely the “Bayesian text” (no visualization) and the four completely filled visualizations from Figure 2.

Next to each visualization, no additional text with statistical information was given. Since each of the five visualization types (“Bayesian text,” 2×2 table, tree diagram, double tree, net diagram) can be equipped with both information formats (probability or natural frequency), we implemented 10 different stimuli. Based on all visualization types, we, furthermore, ask for all four possible joint probabilities or frequencies.

Our research question is:

RQ: What is the effect of information format (i.e., probabilities vs. natural frequencies) for assessing all four concrete joint probabilities/frequencies when statistical information is presented as

- a. “Bayesian text,” i.e., the three pieces of information (base rate, sensitivity, and false-alarm-rate) typically presented in Bayesian reasoning tasks are provided in textual form

or in a completely filled visualization (Figure 2), namely as

- b. tree diagram
- c. double tree
- d. net diagram
- e. 2×2 table?

Furthermore, we want to know whether the type of joint probability ($P(A \cap B)$, $P(\bar{A} \cap B)$, $P(\bar{B} \cap \bar{A})$, $P(\bar{B} \cap A)$) that was asked substantially changed participants’ performance.

3.1 Hypotheses regarding research question (a) Bayesian text

In the probability version, answers need to be calculated, for example, by applying the multiplication rule (e.g., “2% · 80%”). In the natural frequency format, most absolute frequencies that must be combined for the correct answer are already available (depending on the type of question; see Table 1). For instance, in the “Bayesian text” in Figure 1, the first two provided natural frequencies (“200 out of 10,000” and “160 out of 200”) have to be combined correctly to receive the answer “160 out of 10,000.” Note that in both formats some of the given information has to be ignored. Since a calculation with probabilities seems to be more difficult than choosing and combining the right frequencies, we assume—in contrast to the results of Binder et al. (2020)—a substantial format effect here. Consequently, a natural frequency formulation should enhance the performance for questions on joint probabilities. Moreover and regarding the four types (Table 1), it is expected that the more counter events from the “Bayesian text” have to be inferred first, the less correct solutions will be given.

3.2 Hypotheses regarding research question (b) – (e) visualizations

Neither in the tree diagram nor in the double tree, joint probabilities are displayed, meaning that they must be calculated (e.g., by the multiplication rule). In the frequency versions of both tree diagrams, the two relevant absolute frequencies can be read off directly and only have to be combined, which is why we expect a

positive format effect here. All four joint probabilities can be directly read off from the net diagram and the 2×2 table, so high solution rates can be expected even in probability versions (these performances might be probably higher in the 2×2 table because less other possibly interfering probabilities are displayed as compared to the net diagram). According to Bruckmaier et al. (2019) and Binder et al. (2020), even a reverse format effect might be expected for the net diagram and the 2×2 table, since two relevant frequencies have to be identified first and then combined correctly.

In sum, concerning (b) and (c), we expect a format effect in favor of natural frequencies, while concerning (d) and (e), we expect no or even an opposite format effect.

Since in each implemented visualization, all statistical information is presented in a “symmetrical way” and no counter events have to be inferred, no differences are expected regarding the different type of probability question. Yet, the various types of joint probabilities still differ in a linguistic way since the number of negations in the question varies.

4 Method

4.1 Design

Participants had to work on two different contexts (i.e., mammography problem and economics problem; the first adapted from Eddy, 1982, and the second from Ajzen, 1977). In each context, they had to assess all four possible joint probabilities or frequencies. So, every participant had to work on eight tasks.

The study design (see Table 3) includes three factors (information format, visualization type, and context). This leads to a $2 \times 5 \times 2$ design:

- Factor 1: information format: probabilities vs. natural frequencies
- Factor 2: visualization type: “Bayesian text” (no visualization) vs. 2×2 table vs. tree diagram vs. double tree vs. net diagram
- Factor 3 (not a factor of interest): context: mammography vs. economics problem

Factor 1 is the main factor of interest by considering possible interactions with factor 2, while factor 3 was not a factor of interest but only implemented for mutual validation. Furthermore, each participant answered all four possible joint questions ($P(A \cap B)$, $P(\bar{A} \cap B)$, $P(\bar{B} \cap \bar{A})$, $P(\bar{B} \cap A)$) in both contexts. To control for effects of the event order (i.e., asking for $P(A \cap B)$ vs. asking for $P(B \cap A)$), two questions always first included the event A (e.g., getting a positive test result or not) and the other two the event B (e.g., having breast cancer or not).

4.2 Instruments and administration

For each context, 10 stimuli were constructed according to Table 3. In the testlets, one context (for both contexts see Table 4) per participant was always presented in natural frequencies and the other one in probabilities. If the first context processed was based on one out of five visualization types (“Bayesian text,” tree diagram, double tree, net diagram, 2×2 table), the second context was presented in one out of the remaining four visualization types. Thus, the

TABLE 3 Study design.

First context processed*			Second context processed*		
Probabilities	×	“Bayesian text”	Probabilities	×	“Bayesian text”
		Tree diagram			Tree diagram
		Double tree			Double tree
Natural frequencies		Net diagram	Natural frequencies		Net diagram
		2 × 2 table			2 × 2 table
All four possible joint questions (order of the events within a question was varied)			All four possible joint questions (order of the events within a question was varied)		

Each participant worked on both contexts. If the first context was presented, e.g., in natural frequencies and a net diagram, these both conditions were excluded for the second context.
*= order of contexts, formats, and the two visualization conditions were counterbalanced.

instruments were systematically constructed from the modules in Table 3. The rule was: If a participant worked on context X, information format Y, and visualization type Z, exactly these three conditions were forbidden for the second context processed. Every context comprised all four possible joint questions.

Besides the eight joint probability or frequency judgements, several covariates were collected from all participants (see 4.3): level of education ("Fachsemester"), grade point average from high school (German "Abiturnote"), the highest school degree, the field of study, gender, and age.

We varied the first three factors between participants (yielding 160 different testlets) and gave two participants that were sitting next to each other always different contexts for the first task. The two different scenarios (Table 4) were handed out one after the other to track the order of processing. The participants did not have a time limit, but they could use as much time as they wanted to. It took them between 5 and 25 minutes to complete all eight tasks. Further, they were given calculators since the study was on their understanding of the tasks and not on their ability to calculate.

4.3 Participants

Data analysis was based on $N=335$ students who were examined during university classes in Bavaria (Germany) in the year 2022. Students of social work ($N=251$), biomedical engineering ($N=53$), and business classes ($N=31$) participated. $N=271$ students were female, $N=62$ male, and $N=2$ nonbinary. The average age was $M=22.5$ ($SD=4.0$).

The study was carried out in accordance with the Research Ethics Standards of the university. Students were informed that their participation was voluntary, and anonymity was guaranteed. Initially, we had $N=339$ students attending, but only $N=335$ were considered for the analysis because two withdrew their consent and two more mentioned that they did not really think about the tasks and did not put any effort in trying to solve them.

Note that in German schools, only 2×2 tables (either filled with probabilities or frequencies) and tree diagrams (only with probabilities) were taught, so students probably were familiar with these types of visualizations.

4.4 Coding

An overview of the correct answers for each of the eight questions (for both contexts and both formats) is given in Table 5. For the probability versions, the correctness of a response is classified according to whether the participant gave the correct answer within a certain

interval of rounding ($\pm 0.1\%$). For the natural frequency version, both absolute frequencies had to be correct (no rounding occurs). Interrater reliability between two raters was calculated based on 15% of the data and yielded a Cohens Kappa of $\kappa=1$ (Cohen, 1960), therefore answers could be coded with a maximum of objectivity.

5 Results

5.1 Descriptive results regarding the four types of questions

Unexpectedly, there were almost no substantial differences regarding the special type of joint probability that was asked ($P(A \cap B)$, $P(\bar{A} \cap B)$, $P(\bar{B} \cap A)$, $P(B \cap A)$; always in this order). In Supplement S1, all descriptive results are displayed for each single stimulus. Across all versions, the type of question asked and, thus, the number of counter events that first had to be assessed as well as the number of negations in the question do not seem to make a substantial difference.

Another perspective on this fact is given by Figure 3, which illustrates the number of correct joint inferences (0–4). According to the bar diagrams, participants rather predominantly answered none or all of the four questions correctly. Thus, they either understood how to calculate or read off the answer or they did not at all, regardless of which information format was given. In the following, we will, therefore, report results aggregated across the four joint questions.

5.2 Results regarding research questions (a)–(e)

There seems to be a highly differential format effect regarding each visualization type (Figure 4). Because the response patterns in both contexts were very similar, Figure 4 displays the results across contexts. By considering the visualizations separate from each other, two opposite results can be observed already at a descriptive level: the expected frequency effect for the double tree and a reverse effect for the 2×2 table in which the probabilities lead to better performances.

To analyze the effects of information format, visualization type, and their interaction effects by means of inferential statistics, we estimated a generalized linear mixed model (GLMM) with a logit link function to predict the probability that participants solve a question for joint probabilities or frequencies correctly (as a

TABLE 4 Stimuli that emerged by systematically varying factors 1–3 (see Figure 2 for the visualizations).

		Mammography problem		Economics problem	
		Probabilities	Natural frequencies	Probabilities	Natural frequencies
Cover story		Imagine you are a reporter for a women’s magazine and you want to write an article about breast cancer. As a part of your research, you focus on mammography as an indicator of breast cancer. You are especially interested in the question of what it means if a woman has a positive result (which indicates breast cancer) in such a medical test. Please answer the following questions using the statistical information provided below:		Imagine that you are interested in the question of whether students at a boys’ school are more likely to choose economics courses or other courses at their school. For this purpose, you refer to a study conducted by the school psychology service on the connection between personality traits in students and the choice of subjects. Please answer the following questions using the statistical information provided below:	
Statistical information (visualization type)	“Bayesian text”	<ul style="list-style-type: none">• The probability that a woman who goes for a routine screening has breast cancer is 2%. If a woman who goes for a routine screening has breast cancer, the probability that she will get a positive test result is 80%. If a woman who goes for a routine screening does not have breast cancer, then the probability that she will still get a positive test result is 10%.	<ul style="list-style-type: none">• 200 out of 10,000 women who go for a routine screening have breast cancer. Out of 200 women who go for a routine screening and have breast cancer, 160 get a positive test result. Out of 9,800 women who go for a routine screening and do not have breast cancer, 980 still get a positive test result.	<ul style="list-style-type: none">• The probability that a student attends the economics course is 32%. If a student attends the economics course, the probability that he is career-oriented is 64%. If a student does not attend the economics course, the probability that he is still career-oriented is 60%.	<ul style="list-style-type: none">• 320 out of 1,000 students attend the economics course. Out of 320 students who attend the economics course, 205 are career-oriented. Out of 680 students who do not attend the economics course, 408 are still career-oriented.
	Visualization	<ul style="list-style-type: none">• 2 × 2 table with probabilities, or	<ul style="list-style-type: none">• 2 × 2 table with natural frequencies, or	<ul style="list-style-type: none">• 2 × 2 table with probabilities, or	<ul style="list-style-type: none">• 2 × 2 table with natural frequencies, or
		<ul style="list-style-type: none">• Tree diagram with probabilities, or	<ul style="list-style-type: none">• Tree diagram with natural frequencies, or	<ul style="list-style-type: none">• Tree diagram with probabilities, or	<ul style="list-style-type: none">• Tree diagram with natural frequencies, or
		<ul style="list-style-type: none">• Double tree with probabilities, or	<ul style="list-style-type: none">• Double tree with natural frequencies, or	<ul style="list-style-type: none">• Double tree with probabilities, or	<ul style="list-style-type: none">• Double tree with natural frequencies, or
		<ul style="list-style-type: none">• Net diagram with probabilities	<ul style="list-style-type: none">• Net diagram with natural frequencies	<ul style="list-style-type: none">• Net diagram with probabilities	<ul style="list-style-type: none">• Net diagram with natural frequencies
1 st question $P(A \cap B)$		What is the probability that a woman who goes for a routine screening will get a positive test result <i>and</i> has breast cancer?	How many of the women who go for a routine screening get a positive test result <i>and</i> have breast cancer?	What is the probability that a student is career oriented <i>and</i> chooses the economics course?	How many of the students are career oriented <i>and</i> choose the economics course?
2 nd question $P(\bar{A} \cap B)$		What is the probability that a woman who goes for a routine screening will get a negative test result <i>and</i> has breast cancer?	How many of the women who go for a routine screening get a negative test result <i>and</i> have breast cancer?	What is the probability that a student is not career oriented <i>and</i> chooses the economics course?	How many of the students are not career oriented <i>and</i> choose the economics course?
3 rd question $P(\bar{B} \cap \bar{A})$		What is the probability that a woman who goes for a routine screening does not have breast cancer <i>and</i> will get a negative test result?	How many of the women who go for a routine screening do not have breast cancer <i>and</i> get a negative test result?	What is the probability that a student does not choose the economics course <i>and</i> is not career oriented?	How many of the students do not choose the economics course <i>and</i> are not career oriented?
4 th question $P(\bar{B} \cap A)$		What is the probability that a woman who goes for a routine screening does not have breast cancer <i>and</i> will get a positive test result?	How many of the women who go for a routine screening do not have breast cancer <i>and</i> get a positive test result?	What is the probability that a student does not choose the economics course <i>and</i> is career oriented?	How many of the students do not choose the economics course <i>and</i> are career oriented?
Answer format		_____ (please specify to one decimal place)	____out of _____	_____ (please specify to one decimal place)	____out of _____

TABLE 5 Coding of the correct answers regarding all questions.

		Probabilities		Natural frequencies
		Correct answer	Interval in which answers were coded correct	Both absolute numbers must be exact
Mammography	Having breast cancer joint with a positive test result	1.6%	[1.5%; 1.7%] or a decimal fraction in [0.00; 0.02]	The correct answer is 160 out of 10,000.
	Having breast cancer joint with a negative test result	0.4%	[0.3%; 0.49%] or a decimal fraction in [0.00; 0.0049]	The correct answer is 40 out of 10,000.
	Not having breast cancer joint with a negative test result	88.2%	[88.1%; 88.3%] or a decimal fraction in [0.88; 0.89]	The correct answer is 8,820 out of 10,000.
	Not having breast cancer joint with a positive test result	9.8%	[9.7%; 9.9%] or a decimal fraction in [0.09; 0.10]	The correct answer is 980 out of 10,000.
Economics problem	Choosing the economics course joint with interest in a career	20.5%	[20.4%; 20.6%] or a decimal fraction in [0.20; 0.21]	The correct answer is 205 out of 1,000.
	Choosing the economics course joint with no interest in a career	11.5%	[11.4%; 11.6%] or a decimal fraction in [0.10; 0.12]	The correct answer is 115 out of 1,000.
	Not choosing the economics course joint with no interest in a career	27.2%	[27.1%; 27.3%] or a decimal fraction in [0.27; 0.30]	The correct answer is 272 out of 1,000.
	Not choosing the economics course joint with interest in a career	40.8%	[40.7%; 40.9%] or a decimal fraction in [0.40; 0.41]	The correct answer is 408 out of 1,000.

The problem of different rounding only occurs in the probability version, which is why only in these versions (and not in the frequency versions) answers within a certain interval were accepted. If we allowed the same interval for natural frequencies, though, nothing in the coding would change.

binary dependent variable with 0 = wrong, 1 = correct). We decided for a *mixed* analysis and against a, for instance, generalized linear model (i.e., a logistic regression) due to our between-within-subject design since each participant solved several tasks. To take this aspect into account, we modeled a generalized linear *mixed* model with the participants' ID as a random factor, so that the participant-specific error is also modeled (Figure 3 shows dependencies between the responses). In the generalized linear mixed model, we specified the probability version of the “Bayesian text” as the reference category and included the possible explanatory factors “frequencies,” on the one side and, on the other side, “tree diagram,” “double tree,” “net diagram,” and “2 × 2 table” via dummy coding. Furthermore, since the performance in the different formats was expected to vary depending on the visualization type, four interaction terms *visualization* × *format* were modeled as fixed effects.

Because the answers of the participants were dependent on each other (Figure 3) and to exclude sequence effects, we also controlled for the fact that one participant worked on more than one task. Specifically, we implemented participants' ID (w_1) and the order of the questions: 1st, 2nd, 3rd, 4th, 5th, 6th, 7th, and 8th (w_2) as random factors in the generalized linear mixed model:

$$\hat{y} = \beta_0 + \beta_1 \cdot \text{frequencies} + \beta_2 \cdot \text{tree diagram} + \beta_3 \cdot \text{double tree} + \beta_4 \cdot \text{net diagram} + \beta_5 \cdot 2 \times 2 \text{ table} + \beta_6 \cdot \text{tree diagram} \times \text{frequencies} + \beta_7 \cdot \text{double tree} \times \text{frequencies} + \beta_8 \cdot \text{net diagram} \times \text{frequencies} + \beta_9 \cdot 2 \times 2 \text{ table} \times \text{frequencies} + w_1 + w_2$$

The regression coefficient for the frequencies was significantly negative (Table 6), which means that, in the “Bayesian text,” tasks in probabilities are better solved than the ones in natural frequencies. This “probability effect” also holds true for the 2 × 2 table and the net diagram but does not become substantially bigger as can be seen from the regarding interactions that are not significant. In contrast, for the tree diagram and the double tree, this interaction was significantly positive, meaning that the negative format effect observed in the “Bayesian text” is outweighed in these two versions. As a side effect of the findings, we can observe that each visualization compared to the text version—except the double tree—has significant regression coefficients, which means that all of these visualizations in the probability version improved participants' performance. All fixed effects of the model explain 16.3% of the variance, whereas fixed and random effects together explain 75.9% of the variance.

If the question type ($P(A \cap B)$, $P(\bar{A} \cap B)$, $P(\bar{B} \cap \bar{A})$, $P(\bar{B} \cap A)$) is additionally implemented in the model (not displayed in Table 6), it can be observed that none of the other question types is solved correctly significantly rarer than the (easiest) question for $P(A \cap B)$. Moreover, the implementation of this variable, as well as other covariates such as age, gender, level of education, mathematics grade, and school degree, does not lead to substantial changes in the results presented.

Note that some of the results displayed in Table 6 at first seem to contradict the results in Figure 4. Concerning the “Bayesian text,” for example, there was a descriptive advantage of frequencies in Figure 4, while, with inferential statistics, the outcome is the opposite. The results differ because we controlled for order and

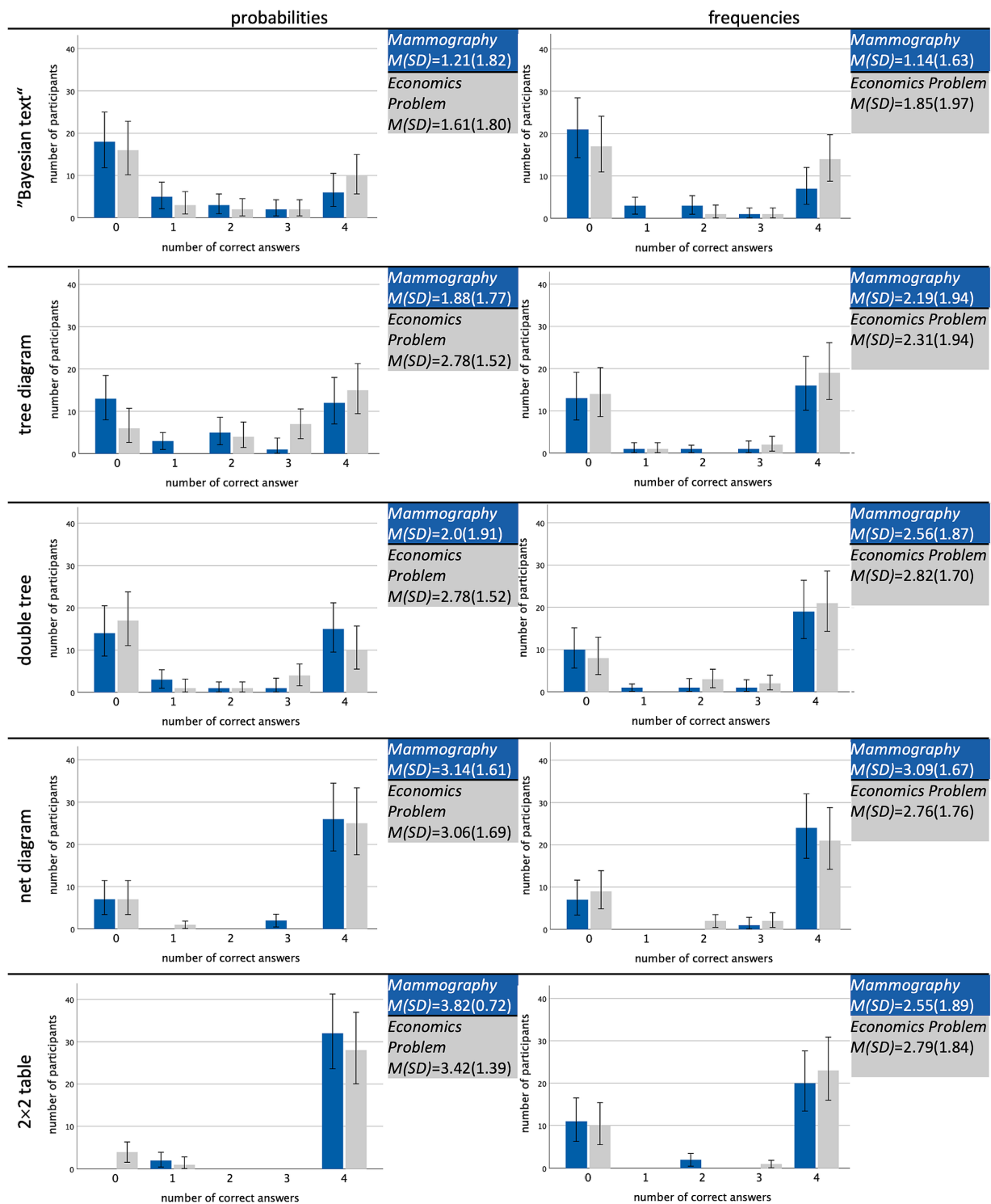


FIGURE 3

Overview of the absolute numbers of participants achieving no, one, two, three, or all four correct answers regarding all four types (Table 1), separated for all 20 stimuli.

ID in the GLMM, which we did not in the descriptive results. Of course, we varied all versions systematically when collecting the data, but, obviously, there are still “group” effects. This demonstrates the need for multi-level modeling since these more precise results cannot be obtained from the descriptive results alone.

6 Discussion

6.1 Summary

In the present study, we systematically investigated participants’ assessment of concrete joint probabilities in Bayesian reasoning

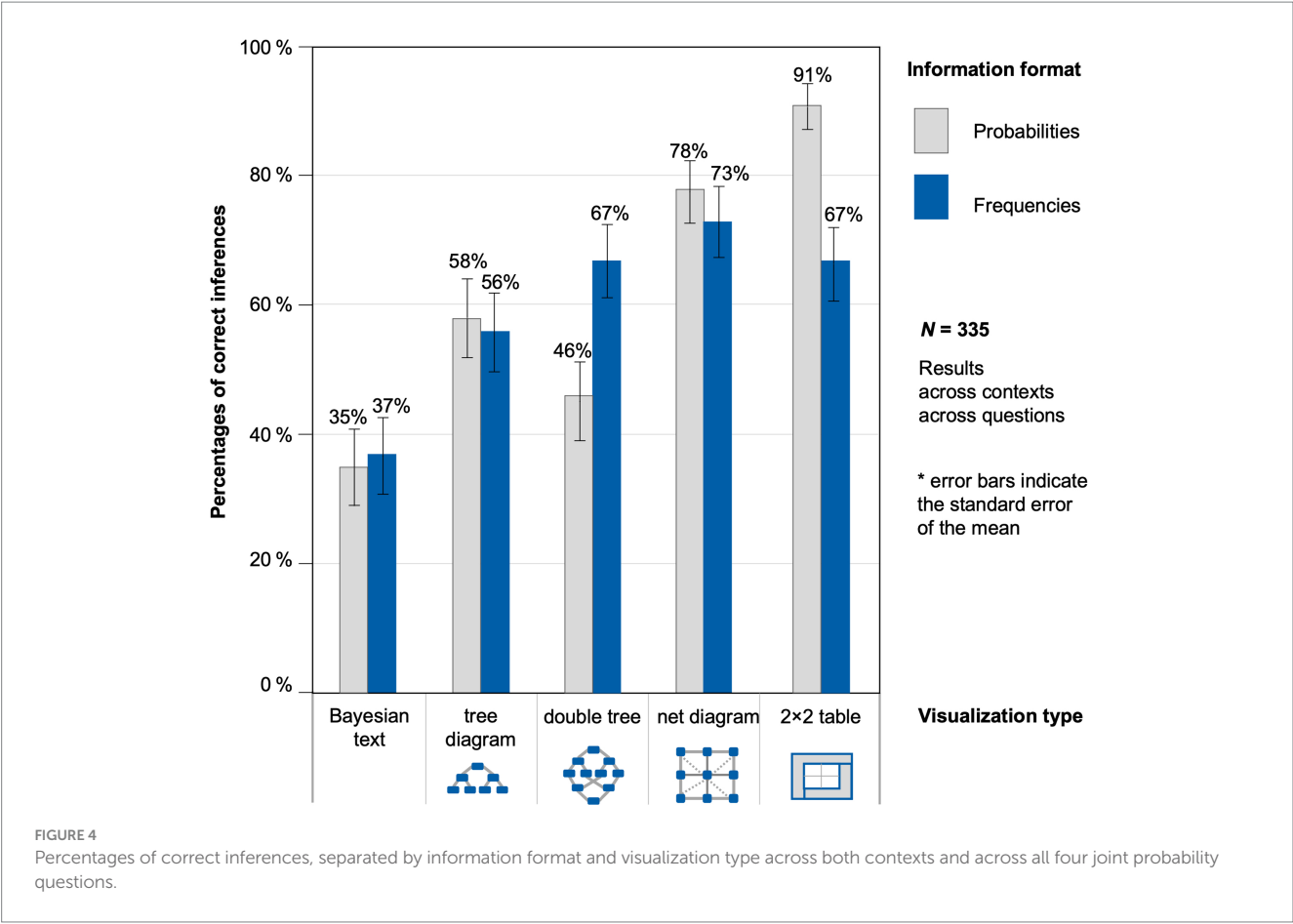


TABLE 6 Regression coefficients for information format, visualization type, and their interactions.

		Estimate	SE	z	p
β_0	Intercept	-0.27	0.30	-0.90	0.37
β_1	Frequencies	-1.18	0.40	-2.97	0.003
β_2	Tree diagram	0.71	0.34	2.08	0.04
β_3	Double tree	-0.05	0.37	-0.14	0.89
β_4	Net diagram	3.46	0.38	9.00	< 0.001
β_5	2 × 2 table	3.75	0.45	8.30	< 0.001
β_6	Tree diagram × frequencies	2.26	0.58	3.89	< 0.001
β_7	Double tree × frequencies	2.79	0.61	4.54	< 0.001
β_8	Net diagram × frequencies	-0.86	0.58	-1.50	0.13
β_9	2 × 2 table × frequencies	-0.17	0.63	-0.27	0.79

Note that bold regression coefficients are significant at $\alpha = 0.05$.
 $R^2_{\text{marginal}} = 16.3\%$, $R^2_{\text{conditional}} = 75.9\%$.

situations. In the theoretical part, we distinguished between paradigms that ask for a qualitative comparison of $P(A)$ and $P(A \cap B)$ and paradigms in which, principally, the whole “Bayesian situation” consisting of 16 probabilities is considered and, therefore, (all) joint probabilities can be assessed. After summarizing pertinent literature, we concluded that the evidence on a possible format effect with respect to joint probabilities is mixed.

In the empirical part of the paper, we reported a study with a $2 \times 5 \times 2$ design with the factors information format (probabilities vs. natural frequencies), visualization type (“Bayesian text” vs.

tree diagram vs. double tree diagram vs. net diagram vs. 2×2 table), and context (mammography vs. economics problem). Furthermore, each participant answered all four joint questions ($P(A \cap B)$, $P(A \cap \bar{B})$, $P(\bar{B} \cap A)$, $P(\bar{B} \cap \bar{A})$). Information format was the main factor of interest, and it was investigated which representation of a Bayesian situation shows which format effect.

First of all, looking at interactions between visualizations and information format, there were some opposite format effects. While tasks with probabilities improved participants’ performance in three visualization conditions (“Bayesian text,” net diagram, and 2×2 table),

this effect cannot be observed with tree diagrams and double trees. Second and compared to the "Bayesian text", participants' performance improved with the probability versions of the tree diagram, the net diagram, and the 2×2 table. However, it was not of our interest *per se* to examine which visualization improves the performance the most. Nevertheless, we found tendencies that suggest which visualizations should be used when explaining situations with joint probabilities, which will be shown in the following section.

6.2 Open questions: Linda and Sally Clark

Although we did not explicitly contribute to these two situations by our experimental setting, let us, nevertheless, recapitulate these situations shortly. With respect to the visualizations in Figure 2, Linda as well as Sally Clark "happen" in only one branch (or in one column of a 2×2 table) because only $P(A)$ and $P(A \cap B)$ are considered, which are depicted in one "line of branches." The difference between both situations is that Sally Clark has a stronger sequential structure because the second child always succeeds the first one.

6.2.1 Linda

Our results would suggest explaining the Linda problem with a 2×2 table in probability format (left in Figure 5). So, it might become obvious that it is more probable to be a bank teller than to be a bank teller *and* to be active in the feminist movement, since 1% is smaller than 5% (which, of course, stays true for any other chosen imaginary numbers).

Comparing the 2×2 tables in probability and frequency format, in the latter one (center of Figure 5), whole persons and no percentages appear (see also, for example, Brase et al., 1998). This is why the 2×2 table with frequencies also seems to be rather intuitive. Indeed, to answer the Linda problem, in both tables, the same two cells have to be compared. Fiedler (1988) could foster his participants' insight by letting them imagine 200 women fitting Linda's description but without providing the other numbers. In any case, it must be noted that for answering the Linda question, marginal probabilities or frequencies (i.e., $P(A)$) have to be considered in addition, but the understanding of them was not subject of our study.

Perhaps the visualization of the general situation (right in Figure 5) in which no imaginary concrete numbers are given, would also enhance the performance in the Linda problem. The general 2×2 table would be more analog to the initial problem (no numbers are given) and it can be easily transferred into a filled-out version by, e.g., requesting the participants to complete the table with imaginary numbers. Thereby, it could either result in a probability or in a natural

frequency version, so, alternatively, the abstract 2×2 table might be a good starting point for teaching in school.

6.2.2 Sally Clark

In the case of Sally Clark, information may be visualized in a tree diagram because of the sequential character of this situation. However, because our results would suggest an advantage of the net diagram and because this sequential character is served by the node-branch-structure, we display the net diagram here (Figure 6). In this visualization, joint probabilities can additionally be included. The red numbers show the situation that was wrongly assumed in court first, while the green numbers show the actual situation. The probability that the 2nd child dies of SIDS (S_2), if the 1st child already died of SIDS (S_1), is 4.3 times as likely as the probability that the 1st child dies of SIDS (Glinge et al., 2023). In the case of Sally Clark, this would result

in a probability of $\frac{1}{8500} \cdot \frac{4.3}{8500} = \frac{4.3}{72000000} \approx 0.000006\%$. Although

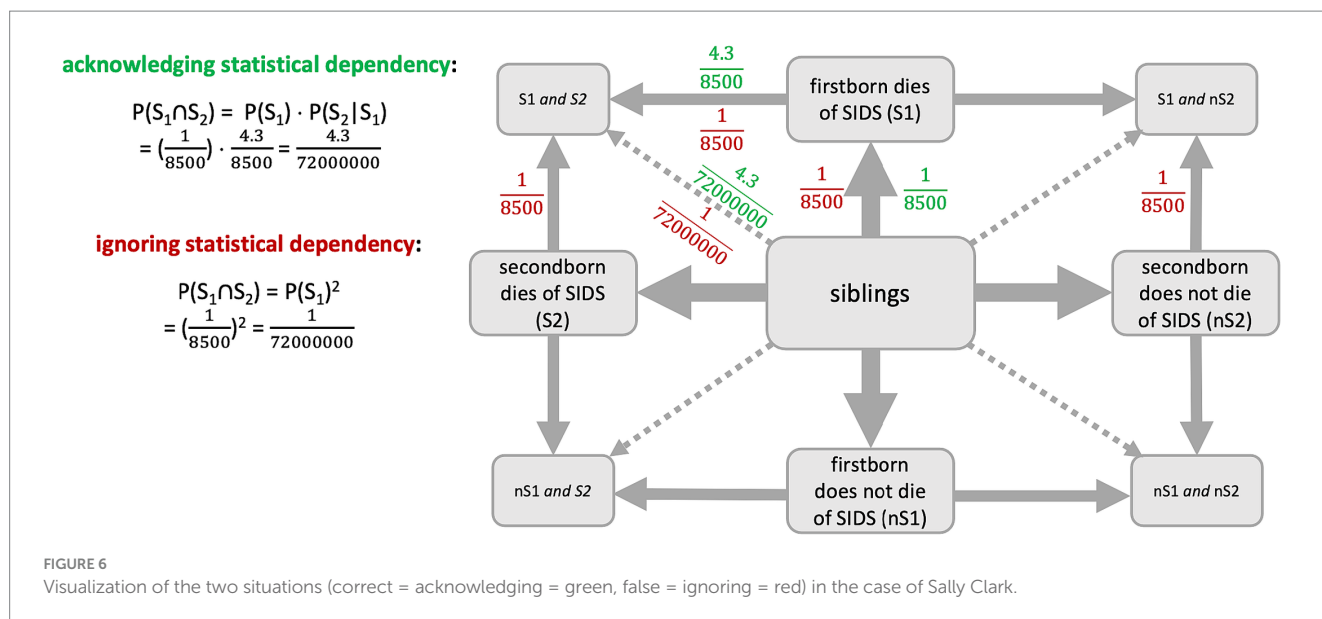
the disregard of the stochastic dependency is often named as the reason for the misjudgment, the calculation shows that this cannot be the only reason since the probability is still very small. The mistrial, in fact, also ignored, for example, that even a very small probability never is equal to 0% and, thus, *does happen* sometimes (Colmez and Schneps, 2013). The medical expert, Roy Meadow, furthermore, assumed that mothers kill their children more often than one might think and, therefore, made this very clear as an expert during trial, which made people—and the jury—think that Clark killed her children (Colmez and Schneps, 2013). This shows that people thought to understand the situation, but apparently not all of them did.

6.3 Limitations and future research

Since we chose typical Bayesian contexts, we might have caused priming toward conditional probabilities among participants, although we did not ask for a conditional probability at any time. By taking the mammography context, for example, most people want to know what a positive or negative test result actually means and not how many people receive a positive test result *and* have breast cancer. Furthermore, in the text version, only the base rate, the sensitivity, and the false-positive-rate—the pieces of information that are typically given in Bayesian inference tasks—were given. This information might prime questions for conditional probabilities and not for joint probabilities. However, the economics context does not lead to a certain kind of question, which mitigates this claim. Still, we might

	bank teller	not a bank teller			bank teller	not a bank teller			bank teller	not a bank teller	
active in the feminist movement	1%	9%	10%	active in the feminist movement	1	9	10	active in the feminist movement	F and B	F and nB	F
not active in the feminist movement	4%	86%	90%	not active in the feminist movement	4	86	90	not active in the feminist movement	nF and B	nF and nB	nF
	5%	95%	100%		5	95	100		B	nB	Ω

FIGURE 5
Visualization of the Linda version with 2×2 tables (probabilities, frequencies, abstract).



have triggered different assumptions of the participants (e.g., the need for a conditional probability), which might have led to specific errors like answering with a conditional probability.

Furthermore, some participants might have also wondered why we “just” asked for *all four* joint probabilities and have not included conditional probability questions. Moreover, the fact that the visualizations included much more information might have made some participants evaluate their answers as “too easy,” which could have made them change their initial answer. By including only joint probabilities, we also cannot judge the format effect regarding marginal probabilities.

Future research could look more deeply into variations. At first, it would be interesting to vary the given pieces of information (especially in the textual version). Then, it would also be interesting to implement further contexts—especially ones that make perfectly sense concerning joint probabilities (e.g., gambling).

In addition, note that the efficacy of natural frequencies always also depends on more factors than the ones mentioned above: Ayal and Beyth-Marom (2014) showed that if the presented and requested format is not compatible (e.g., the information is in probabilities and the question in natural frequencies), the performance is lower than, for example, if both are in probabilities. However, highest performance levels can be observed, if information is presented in natural frequencies and participants *also work* with natural frequencies instead of translating them “back” into probabilities (Weber et al., 2018; Feufel et al., 2023). It also has an impact on the performance, whether the given information and the question are “aligned,” which means that the presented and requested information should be attached to the same subset (Tubau et al., 2019; Tubau, 2022; Brose et al., 2023). Furthermore, the performance also improves if the task format is formulated “explicitly” (the intersecting set is explicitly named, i.e., “How many of the positive tested women are ill and test positive?”) instead of “implicitly” (i.e., “How many of the positive tested women are ill?”; Böcherer-Linder et al., 2018). Future research should also consider these factors to be able to derive conclusions about their effect on joint probabilities.

Finally, we want to propose a fifth extension of Bayesian reasoning, namely, to explicitly address *all* possible 16 probabilities in future research. There are *eight* conditional probabilities; two of them are just complemented probabilities of the given sensitivity and

false-alarm-rate. All four inverse conditional probabilities, nevertheless, belong to the full situation. From a mathematical viewpoint, all 16 probabilities are equally relevant and, furthermore, at school, of course, all of them are taught.

6.4 Conclusion

Our answer to the question “How general is the natural frequency effect?” is: There is no general statement possible concerning questions for joint probabilities. Whether natural frequencies improve participants’ performance in joint probability tasks highly depends on the way the statistical information is presented.

Data availability statement

The data of the study can be found here: https://epub.uni-regensburg.de/54717/1/Datensatz_open.xlsx.

Ethics statement

Ethical approval was not required for the studies involving humans in accordance with the local legislation and institutional requirements. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

NS: Conceptualization, Data curation, Formal analysis, Funding, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing. KB: Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing – review & editing. SK: Conceptualization, Methodology, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The publication of this work was supported by the German Research Foundation (DFG) within the funding program Open Access Publishing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *J. Pers. Soc. Psychol.* 35, 303–314. doi: 10.1037/0022-3514.35.5.303
- Ayal, S., and Beyth-Marom, R. (2014). The effects of mental steps and compatibility on Bayesian reasoning. *Judgm. Decis. Mak.* 9, 226–242. doi: 10.1017/S1930297500005775
- Binder, K., and Krauss, S. (under review). Generalizations of the Bayesian reasoning paradigm.
- Binder, K., Krauss, S., and Bruckmaier, G. (2015). Effects of visualizing statistical information – an empirical study on tree diagrams and 2×2 tables. *Front. Psychol.* 6:1186. doi: 10.3389/fpsyg.2015.01186
- Binder, K., Krauss, S., Bruckmaier, G., and Marienhagen, J. (2018). Visualizing the Bayesian 2-test case: the effect of tree diagrams on medical decision making. *PLoS One* 13:e0195029. doi: 10.1371/journal.pone.0195029
- Binder, K., Krauss, S., and Wiesner, P. (2020). A new visualization for probabilistic situations containing two binary events – the frequency net. *Front. Psychol.* 11, 1–21. doi: 10.3389/fpsyg.2020.00750
- Binder, K., Steib, N., and Krauss, S. (2023). Von Baumdiagrammen über Doppelbäume zu Häufigkeitsnetzen – kognitive Überlastung oder didaktische Unterstützung? [Moving from tree diagrams to double trees to net diagrams – cognitively overwhelming or educationally supportive?] *J. Math. Didakt.* 44, 471–503. doi: 10.1007/s13138-022-00215-9
- Böcherer-Linder, K., Binder, K., Büchter, T., Eichler, A., Krauss, S., Steib, N., et al. (2022). “Communicating conditional probabilities in medical practice” in *Bridging the gap: Empowering and educating Today's learners in statistics. Proceedings of the Eleventh International Conference on Teaching Statistics*. ed. S. Peters (Rosario (Argentina): International Association for Statistical Education)
- Böcherer-Linder, K., and Eichler, A. (2017). The impact of visualizing nested sets. An empirical study on tree diagrams and unit squares. *Front. Psychol.* 7:2026. doi: 10.3389/fpsyg.2016.02026
- Böcherer-Linder, K., Eichler, A., and Vogel, M. (2018). Die Formel von Bayes: Kognitionspsychologische Grundlagen und empirische Untersuchungen zur Bestimmung von Teilmenge-Grundmenge-Beziehungen [Bayes' formula: cognitive psychological basics and empirical investigation of determining subsets.]. *J. Math. Didakt.* 39, 127–146. doi: 10.1007/s13138-018-0128-1
- Brase, G. L., Cosmides, L., and Tooby, J. (1998). Individuation, counting, and statistical inference: the role of frequency and whole-object representations in judgment under uncertainty. *J. Exp. Psychol. Gen.* 127, 3–21. doi: 10.1037/0096-3445.127.1.3
- Brose, S. F., Binder, K., Fischer, M. R., Reincke, M., Braun, L. T., and Schmidmaier, R. (2023). Bayesian versus diagnostic information in physician-patient communication: effects of direction of statistical information and presentation of visualization. *PLoS One* 18:e0283947. doi: 10.1371/journal.pone.0283947
- Bruckmaier, G., Binder, K., Krauss, S., and Kufner, H.-M. (2019). An eye-tracking study of statistical reasoning with tree diagrams and 2×2 tables. *Front. Psychol.* 10:632. doi: 10.3389/fpsyg.2019.00632
- Büchter, T., Eichler, A., Böcherer-Linder, K., Vogel, M., Binder, K., Krauss, S., et al. (2024). Covariational reasoning in Bayesian situations. *Educ. Stud. Math.*
- Charness, G., Karni, E., and Levin, D. (2009). On the conjunction fallacy in probability judgment: new experimental evidence regarding Linda, working paper, no. 552, the Johns Hopkins University, Department of Economics, Baltimore, MD.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104
- Colmez, C., and Schneps, L. (2013). *Math on trial: How numbers get used and abused in the courtroom*. New York: Basic Books.
- Donati, C., Guazzini, A., Gronchi, G., and Smorti, A. (2019). About Linda again: how narratives and group reasoning can influence conjunction fallacy. *Future Internet* 11:10. doi: 10.3390/fi11100210
- Eddy, D. M. (1982). “Probabilistic reasoning in clinical medicine: problems and opportunities” in *Judgment under uncertainty*. eds. D. Kahneman, P. Slovic and A. Tversky (Cambridge: Cambridge University Press)
- Ellis, K. M., and Brase, G. L. (2015). Communicating HIV results to low-risk individuals: still hazy after all these years. *Curr. HIV Res.* 13, 381–390. doi: 10.2174/1570162X13666150511125629
- Feufel, M. A., Keller, N., Kendel, F., and Spies, C. D. (2023). Boosting for insight and/or boosting for agency? How to maximize accurate test interpretation with natural frequencies. *BMC Med. Educ.* 23, 1–10. doi: 10.1186/s12909-023-04025-6
- Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychol. Res.* 50, 123–129. doi: 10.1007/BF00309212
- Garcia-Retamero, R., and Hoffrage, U. (2013). Visual representation of statistical information improves diagnostic inferences in doctors and their patients. *Soc. Sci. Med.* 83, 27–33. doi: 10.1016/j.socscimed.2013.01.034
- Gigerenzer, G., and Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychol. Rev.* 102, 684–704. doi: 10.1037/0033-295X.102.4.684
- Gigerenzer, G., Hoffrage, U., and Ebert, A. (1998). AIDS counselling for low-risk clients. *AIDS Care* 10, 197–211. doi: 10.1080/09540129850124451
- Glinge, C., Rossetti, S., Bruun Ostergaard, L., Kjær Stampe, N., Hadberg Lynge, T., Skals, R., et al. (2023). Risk of sudden infant death syndrome among siblings of children who dies of sudden infant death syndrome in Denmark. *JAMA Netw. Open* 6. doi: 10.1001/jamanetworkopen.2022.52724
- Henze, N., and Vehling, R. (2021). Im Vordergrund steht das Problem - oder: Warum ein Häufigkeitsnetz. [The problem comes first - or: Why the net diagram?] *Stochastik in der Schule* 41.
- Hertwig, R., Benz, B., and Krauss, S. (2008). The conjunction fallacy and the many meanings of and. *Cognition* 108, 740–753. doi: 10.1016/j.cognition.2008.06.008
- Hoffrage, U., Krauss, S., Martignon, L., and Gigerenzer, G. (2015). Natural frequencies improve Bayesian reasoning in simple and complex inference tasks. *Front. Psychol.* 6:1473. doi: 10.3389/fpsyg.2015.01473
- Inhelder, B., and Piaget, J. (1964). *The early growth of logic in the child: Classification and seriation*. New York: Harper and Row.
- Knapp, P., Gardner, P. H., Carrigan, N., Raynor, D. K., and Woolf, E. (2009). Perceived risk of medicine side effects in users of a patient information website: a study of the use of verbal descriptors, percentages and natural frequencies. *Br. J. Health Psychol.* 14, 579–594. doi: 10.1348/135910708x375344
- Krauss, S., Martignon, L., and Hoffrage, U. (1999). *Simplifying Bayesian inference: The general case*. United States: Springer, 165–179.
- Krauss, S., Weber, P., Binder, K., and Bruckmaier, G. (2020). Natürliche Häufigkeiten als numerische Darstellungsart von Anteilen und Unsicherheit – Forschungsdesiderate und einige Antworten. [Natural frequencies as numerical representation of proportions and uncertainty - research desiderata and some answers.] *J. Math.-Didakt.* 41, 485–521. doi: 10.1007/s13138-019-00156-w
- McDowell, M., Gigerenzer, G., Wegwarth, O., and Rebitschek, F. G. (2019). Effect of tabular and icon fact box formats on comprehension of benefits and harms of prostate cancer screening: a randomized trial. *Med. Decis. Mak.* 39, 41–56.
- McDowell, M., and Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychol. Bull.* 143, 1273–1312. doi: 10.1037/bul0000126

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1296359/full#supplementary-material>

- Mellers, B., Hertwig, R., and Kahnemann, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychol. Sci.* 12, 269–275. doi: 10.1111/1467-9280.00350
- Neth, H., Gradwohl, N., Streeb, D., Keim, D. A., and Gaissmaier, W. (2021). Perspectives on the 2x2 matrix: solving semantically distinct problems based on a shared structure of binary contingencies. *Front. Psychol.* 11:567817. doi: 10.3389/fpsyg.2020.567817
- O'Grady, C. (2023). Unlucky numbers. *Science* 379, 228–233. doi: 10.1126/science.adg6746
- Phillips, N. (2022). She was convicted of killing her four children. Could a gene mutation set her free? *Nature* 611, 218–223. doi: 10.1038/d41586-022-03577-9
- Prinz, R., Feufel, M., Gigerenzer, G., and Wegwarth, O. (2015). What counselors tell low-risk clients about HIV test performance. *Curr. HIV Res.* 13, 369–380.
- Schwartz, L. M., Woloshin, S., and Welch, H. G. (2007). The drug facts box: providing consumers with simple tabular data on drug benefit and harm. *Med. Decis. Mak.* 27, 655–662. doi: 10.1177/0272989X07306786
- Siegrist, M., and Keller, C. (2011). Natural frequencies and Bayesian reasoning: the impact of formal education and problem context. *J. Risk Res.* 14, 1039–1055. doi: 10.1080/13669877.2011.571786
- Stegmüller, N. (2020). *Bayes ins NETZ gegangen: Alle Häufigkeiten und Wahrscheinlichkeiten auf einen Blick im Häufigkeitsnetz*. [All frequencies and probabilities at one sight in the net diagram.] Unpublished admission work. Regensburg: Universität Regensburg.
- Steib, N., Krauss, S., Binder, K., Büchter, T., Böcherer-Linder, K., Eichler, A., et al. (2023). Measuring people's covariational reasoning in Bayesian situations. *Front. Psychol.* 14:1184370. doi: 10.3389/fpsyg.2023.1184370
- Tubau, E. (2022). Why can it be so hard to solve Bayesian problems? Moving from number comprehension to relational reasoning demands. *Think. Reason.* 28, 605–624. doi: 10.1080/13546783.2021.2015439
- Tubau, E., Rodríguez-Ferreiro, J., Barberia, I., and Colomé, À. (2019). From reading numbers to seeing ratios: a benefit of icons for risk comprehension. *Psychol. Res.* 83, 1808–1816. doi: 10.1007/s00426-018-1041-4
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases: biases in judgments reveal some heuristics of thinking under uncertainty. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124
- Tversky, A., and Kahneman, D. (1983). Extensional vs. intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* 90, 293–315. doi: 10.1037/0033-295X.90.4.293
- Weber, P., Binder, K., and Krauss, S. (2018). Why can only 24% solve bayesian reasoning problems in natural frequencies: frequency phobia in spite of probability blindness. *Front. Psychol.* 9:1833. doi: 10.3389/fpsyg.2018.01833
- Wedell, D. H., and Moro, R. (2007). Testing boundary conditions for the conjunction fallacy: effects of response mode, conceptual focus, and problem type. *Cognition* 107, 105–136.
- Wells, J., Malone, U., Parkes-Hupton, H., Stonehouse, G., Wakatama, G., Coote, G., et al. (2023). Kathleen Folbigg pardoned after 20 years in jail over killing her four children. ABC News. Available at: <https://www.abc.net.au/news/2023-06-05/kathleen-folbigg-attorney-general-provides-update/102440136>
- Wolke, J. K., Hoffrage, U., and Martignon, L. (2017). Integrating and testing natural frequencies, Naïve Bayes, and fast-and-frugal trees. *Decision* 4, 234–260. doi: 10.1037/dec0000086



OPEN ACCESS

EDITED BY

Samuel Shye,
Hebrew University of Jerusalem, Israel

REVIEWED BY

Hidehito Honda,
Otemon Gakuin University, Japan
Eldad Yechiam,
Technion Israel Institute of Technology, Israel

*CORRESPONDENCE

Leonidas Spiliopoulos
✉ spiliopoulos@mpib-berlin.mpg.de

RECEIVED 26 May 2024

ACCEPTED 15 July 2024

PUBLISHED 06 August 2024

CITATION

Spiliopoulos L and Hertwig R (2024)
Stochastic heuristics for decisions under risk
and uncertainty. *Front. Psychol.* 15:1438581.
doi: 10.3389/fpsyg.2024.1438581

COPYRIGHT

© 2024 Spiliopoulos and Hertwig. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Stochastic heuristics for decisions under risk and uncertainty

Leonidas Spiliopoulos* and Ralph Hertwig

Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

Models of heuristics are often predicated on the desideratum that they should possess no free parameters. As a result, heuristic implementations are usually deterministic and do not allow for any choice errors, as the latter would require a parameter to regulate the magnitude of errors. We discuss the implications of this in light of research that highlights the evidence supporting stochastic choice and its dependence on preferential strength. We argue that, in principle, the existing models of deterministic heuristics should, and can, be quite easily modified to stochastic counterparts through the addition of an error mechanism. This requires a single free parameter in the error mechanism, whilst otherwise retaining the parameter-free cognitive processes in the deterministic component of existing heuristics. We present various types of error mechanisms applicable to heuristics and discuss their comparative virtues and drawbacks, paying particular attention to their impact on model comparisons between heuristics and parameter-rich models.

KEYWORDS

errors, decision making under risk and uncertainty, model comparison, stochastic heuristics, bounded rationality

1 Introduction

Heuristics, though there are many varying definitions of them, viewpoints (cf. [Gigerenzer and Goldstein, 1996](#); [Kahneman and Tversky, 1996](#)) and different classes ([Mousavi and Gigerenzer, 2017](#)), are typically defined as models with clearly spelled-out cognitive processes. Their aim is to describe and approximate the actual processes as opposed to as-if models of behavioral outcomes, such as optimization theories (e.g., Bayesian decision theory, expected utility maximization). Another aspect of models of heuristics is that they eschew complex calculations that overtax human abilities and they ignore some of the available information ([Gigerenzer and Gaissmaier, 2011](#)), yet often still may manage to outperform significantly more complex models (e.g., [Gigerenzer and Brighton, 2009](#); [Katsikopoulos et al., 2010](#)). Beyond the evidence from the lab, heuristics often perform very well in the field ([Şimşek, 2013](#); [Katsikopoulos et al., 2021](#)) including the business world as even CEOs rely on heuristics to navigate exceptional uncertainty (e.g., see the overview in [Mousavi and Gigerenzer, 2014](#)). Another feature of models of heuristics is that they are usually constructed without free parameters to be estimated from data; in essence, they are deterministic models. This is particularly true of fast and frugal or ecologically rational heuristics (see [Gigerenzer et al., 1999, 2011](#); [Todd et al., 2012](#), and a comparative discussion of how ecological rationality is considered in economics and psychology; [Mousavi and Kheirandish, 2014](#)). This contrasts the majority of choice and inference models that include free parameters (e.g., expected utility model, cumulative prospect theory, and drift-diffusion models). There are several reasons for eschewing free parameters, perhaps the most important one is that they risk to unduly increase

the flexibility of a model, thereby accounting for many different data patterns including noise. If the data is noisy or the training data limited, flexible free-parameter models are vulnerable to over-fitting to the noise in the data, thereby resulting in worse out of sample predictive performance than models without free parameters (e.g., see Gigerenzer and Brighton, 2009).

Fully deterministic models of heuristics, however, exact a cost. They make it difficult to model both between-participants and within-participant heterogeneity. People are known to invoke different cognitive process across one and the same task, and the same person may switch to different processes even within the same class of task, depending on contextual factors such as time pressure (e.g., Svensson and Maule, 1993; Spiliopoulos and Ortmann, 2018), incentives (e.g., Payne et al., 1997) or task characteristics (e.g., choice difficult; see Brandstätter et al., 2008). Flexibility in the use of heuristics, both across individuals and within an individual and across environments, is inherent to the notion of the adaptive decision maker (Payne et al., 1997), and the adaptive tool box of heuristics as the basis of adaptive behavior (Gigerenzer et al., 2011). Given the unavoidable between-participants heterogeneity and the theoretically postulated within-person flexibility across properties of the choice environment, how can deterministic models of heuristics be allowed some flexibility without falling into the trap of too much flexibility? In the interest of full disclosure, we are sympathetic to a parameter-free approach (e.g., Spiliopoulos and Hertwig, 2020). Even if only to explore how much predictive power deterministic models have. Nevertheless, in this manuscript we are predominantly interested in a type of stochasticity arising within a person and within a task from errors in cognitive processes, to which no cognitive model, not even heuristics, are immune.

The desideratum of models of heuristics to avoid free parameters has consequently led to the majority of heuristics being implemented as deterministic, for instance, choosing a single option with certainty. This is because allowing for stochastic choice through errors inadvertently requires a free parameter to modulate the magnitude of errors. It is not clear how to avoid this without the arbitrary choice of such an error value that would not be fitted to data and likely not representative of its true value. We argue that the science of heuristics needs to seriously consider the pros and cons of the existing strict adherence to no free parameters (see also, Ortmann and Spiliopoulos, 2023), and allowing flexibility in perhaps the most important place, namely, with respect to stochastic choice arising from errors.

We will argue that transitioning toward some flexibility offers several opportunities, including methodological improvements particularly in model comparisons. Furthermore, modifying heuristics so that they are able to predict a strength of preference over options, rather than a deterministic choice will increase the empirical content of heuristics and make them more falsifiable. We will discuss how this will level the playing field when comparing flexible models with free parameters against models of heuristics, as current practices involving deterministic models of the latter may be problematic. Given the strong procedural foundations of heuristics, one can consider errors in a more principled and structured way than is possible with as-if behavioral models. This is because the clearly defined and transparent processes in heuristics suggest how errors come about and constrain the error

distributions, whereas with as-if models it is harder to arrive at a priori reasonable constraints.

Let us briefly define some terms that are used throughout. We will think of models as consisting of two components: the first one is indispensable and is the *core deterministic* component; the second one represents an *error-mechanism* (or stochastic) component and is often referred to as a choice rule. A *deterministic* model always chooses one of the available options with certainty (i.e., probability 1.0), wholly rejecting all other options (i.e., probability zero)—this choice distribution is *discrete*. *Continuous* choice distributions, in contrast, imply that at least one choice is made with a probability greater than zero and less than one. As mentioned before, most implementations of models of heuristics are deterministic; in this manuscript we will refer to them as deterministic models of heuristics, as opposed to stochastic models of heuristics that permit continuous choice distributions. Finally, a *flexible* behavioral model is one that has free parameters in the core component that are typically estimated from data. That is, according to our terminology, a model without free parameters in the core but with an error mechanism that includes one or more free parameters, is not a flexible model. For our purposes, such models are *stochastic* models of heuristics.

Models with free parameters are inferred from data using an *estimation technique*, which requires the specification of a *loss function*, e.g., mean-squared-deviation or a likelihood function. To avoid issues of flexible models over-fitting empirical data, we only consider model performance out-of-sample as derived from a performance metric. Such metrics may also be *discrete* (as the prediction is an extreme choice of 0 or 1) or *continuous*—note, discreteness or continuity of the metric refers to each individual choice, not to the average metric applied over many choices. For example, consider a metric such as the percentage correct predictions that is the average of the values of 0 (if a choice was not correctly predicted) or 1 (if it was). At the individual choice level, this is discrete, but the final metric constructed from the average of these values is continuous. We refer to this metric as discrete, to differentiate it from another metric that may make probabilistic (continuous) individual predictions, say that one option is chosen with probability 0.8 and the other 0.2, but which again when averaging over choice predictions would also return a final continuous measure.

Models consist of various processes ranging from information search to information integration and each process may be prone to error. We refer to the final process that leads to a choice, as the *valuation stage* and it typically involves the comparison of a set of values, one for each of the options available—errors that happen during this stage are referred to as *valuation errors*. Earlier processes will also often involve numerical comparisons, but typically these numbers would not represent a final valuation—these are coined *procedural errors*.

The manuscript is structured as follows. First, we briefly overview the overwhelming evidence that points to significant stochasticity in choices and how it relates to a decision maker's strength of preference over the available options. We then proceed with methodological arguments for stochastic models of heuristics, highlighting some of the problems that may arise if analyses are based solely on deterministic variants. In a subsection, we will

deal specifically with issues that may arise in model comparisons between flexible models and heuristics if the latter are not modeled as stochastic. Having laid the foundations for why we consider stochastic models of heuristics to be important, we lay out a classification of applicable error mechanisms. The ensuing comparative discussion about the advantages and shortcomings of each will allow us to make practical recommendations about their implementation. We will illustrate these by presenting possible stochastic variants of the popular maximin heuristic. Our emphasis throughout is on models of choice heuristics under risk and uncertainty. Yet, our arguments are easily translated to models of heuristics in general. Last but not least, let also emphasize that the question of whether, and how, to incorporate flexibility in deterministic models of heuristic models has been discussed before (see [Rieskamp, 2008](#); [Schulze et al., 2021](#)); we will discuss this work below.

2 Arguments for stochastic heuristics

2.1 Decision making is stochastic

There is strong evidence in favor of the proposition that choice is inherently stochastic—see [Rieskamp \(2008\)](#) for a detailed discussion. In choice under risk, participants presented with identical lotteries under risk often make different choices when repeatedly responding to them ([Hey, 2001](#); [Mata et al., 2018](#)). Some choice theories such as cumulative prospect theory are often amended by adding a choice rule that accommodates such errors. The underlying cognitive processes that may underlie choice behavior are also error prone or noisy, for example, memory retrieval and attention. Consequently, some theories are constructed to be inherently stochastic by nature, such as evidence accumulation models of behavior (e.g., [Ratcliff, 1988](#); [Busemeyer and Townsend, 1993](#); [Usher and McClelland, 2001](#)), where the accumulation process itself is stochastic, but the final step of hitting a decision threshold is error free (i.e., the choice corresponding to the threshold is chosen with certainty).

Choice stochasticity from the viewpoint of an observer (such as researchers) may also be attributed to other causes. Even if a deterministic decision-maker were to exist, to an observer that does not have access to the exact states of all variables entering the decision processes, choices will appear stochastic due to the (unobservable) latent variables. This is analogous to the example of a die roll being essentially deterministic, yet appearing as stochastic to observers that do not have access to the exact initial conditions and physical values. The argument for extending choice models to be stochastic is therefore not just one of modeling realism (due to internal noisy cognitive processes), but is also related to the methodology of model estimation: How unobservable latent variables are accounted for, even indirectly, is important.

Ultimately, the research goal may dictate whether adding an error mechanism is desirable or not. If it is solely for prediction, then a deterministic heuristic with no free parameters may be preferable and adequate. Of course, the allure of deterministic models of heuristics is that they are powerful exactly because no data is needed for them to make behavioral predictions. A caveat is that if there is heterogeneity among decision makers or

a decision maker flexibly uses different heuristics, then one would need empirical data to obtain an estimate of the proportional use of different heuristics. On the other hand, for robust model comparisons we believe it is advisable to accept the addition of free parameters in the error mechanism, whilst retaining the hallmark of models of heuristics—a deterministic and parameter-free model core. In model comparisons, ignoring errors risks being problematic as the models are essentially misspecified.

2.2 Preferential strength affects choice consistency

Choice consistency in a wide variety of tasks is a monotonically increasing function of the (absolute) relative strength of preference of an option over the remaining options. That is, errors are increasingly more likely and more substantial when options are relatively similar in their valuations. At a cognitive level, this can be understood in terms of just-noticeable differences or signal detection theory. Error-mechanisms and choice rules have a long history in cognitive psychology ([Thurstone, 1927](#); [Mosteller and Nogee, 1951](#); [Luce, 1959](#)) and economics ([McFadden, 2001](#)). The choice rules most often used in the literature are based on exactly this monotonicity assumption, e.g., logit and probit choice rules. In drift-diffusion models the magnitude of the drift is derived from the evidence in favor of each option, and the higher the magnitude of the drift rate, the more extreme the choice predictions and, correspondingly, the higher the choice consistency. Independent of specific parametric forms, empirical evidence for a strong monotonic relationship between consistency and preferential strength in choice under risk and intertemporal choice is presented in [Alós-Ferrer and Garagnani \(2021, 2022\)](#). Further indirect evidence of the important role of preferential strength is evident from the finding that response times are typically longer the closer the valuation of the options is ([Moffatt, 2005](#); [Chabris et al., 2009](#); [Spiliopoulos, 2018](#); [Spiliopoulos and Ortmann, 2018](#); [Alós-Ferrer and Garagnani, 2022](#)).

Flexible models have been implemented with error-mechanisms more often than models of heuristics for several reasons. In flexible models, options typically receive some absolute value in the final valuation stage. From here it was but a small step to consider preferential strength and how this may map to continuous choice probabilities. An example of this is the expected utility of each prospect in a pair of lotteries, which is typically translated into a probability distribution over options using an error mechanism that is a function of option valuations. Perhaps the hesitation in considering stochastic models of heuristics is the concern that it requires one of two things: (a) complex parametric forms to calculate continuously-valued option valuations in combination with a choice rule and/or (b) multiplicative integration of probabilities and outcome values in contrast to the simpler comparative and logical operations found in models of heuristics (e.g., comparison of magnitudes). We will show later that this concern may be unwarranted in many cases, as option valuations and preferential strength can be trivially inferred from existing heuristics for choices under risk and uncertainty,

without changing their deterministic parameter-free core and the assumption of simple processes.

2.3 Methodological arguments

Heuristics have been presented as procedural models of behavior that are more realistic than their parameter-rich as-if adversaries. Scholars advocating for models of heuristics have correctly, in our opinion, asserted that comparisons of heuristics and flexible models should be done on the basis of out-of-sample or cross-validation performance. The argument is that good performance by flexible models with many free parameters is illusive if their performance is estimated in sample. Ultimately, comparisons between the two types of models comes down to their out of sample performance on the same sets of tasks; however, the difference in their need for estimation may be problematic when it comes to such a model comparison. Can flexible stochastic models and deterministic models of heuristics be directly compared without unduly handicapping one or the other? We wish to draw attention to some issues with existing methods of comparing these models and suggest a viable alternative that may alleviate them—see related arguments about model comparisons in [Spiliopoulos and Hertwig \(2020\)](#).

The first issue concerns the fact that flexible models are usually implemented with an error mechanism, and are therefore stochastic, admitting continuous-valued predictions (on the probability scale) derived from valuations, whereas heuristics are deterministic, admitting discrete-valued predictions only. How is this difference typically reconciled in the literature? For a direct comparison, both models must be scored according to the same performance metric leading to three possible solutions:

1. Use a discrete performance metric and convert the continuous-valued predictions of a stochastic flexible model to discrete predictions to be compared against a discrete heuristic.
2. Use a discrete performance metric and deterministic flexible models and heuristics, so that the above conversion need not to be made.
3. Use a continuous performance metric with both flexible models and heuristics implemented with a stochastic error mechanism.

We believe that the first two, which are predominantly used in the literature, may be problematic in various respects, and recommend the third option—let us explore the reasoning behind our assertion.

2.3.1 Option 1

This option is problematic because of the mismatch between the *continuous* loss function necessitated by the estimation of the stochastic flexible model and the subsequent application of a *discrete* performance metric. Converting a stochastic prediction to a deterministic one is usually achieved by assuming that the option with the highest predicted likelihood is chosen with probability 1 and the other options with probability 0. Having done this conversion, both flexible models and heuristics can be compared using the percentage of correct choice metric, ignoring any probabilistic information that existed in the flexible models

(and by extension in the choice data). This is clearly inefficient and may have put flexible models at a relative disadvantage to heuristics, as they are estimated using a procedure with a different goal or metric than the one that their comparison to heuristics is based on, possibly leading to poorer performance than would otherwise be the case. This occurs because the nature of the loss function determines the parameter estimates, which in turn affect the choice predictions. Consider how the continuous L2 and log-likelihood loss functions penalize errors during estimation. Since the penalty for the error is a continuous function of the error magnitude, the errors between a continuous-valued prediction of 0.49 and 0.51 are very similar in value (assuming two options). Now consider the discrete performance metric, which requires that those two predictions are discretized to values of 0 and 1, respectively. The errors are now diametrically opposed, one prediction has an error of 1 and the other 0. In general, the true (continuous) magnitude of the error is irrelevant under discrete loss functions (and performance metrics) as long as it is on the same side of 0.5. Under continuous loss functions, larger errors are always penalized more, whereas under discrete loss functions errors are penalized more only when the threshold prediction of 0.5 is crossed, jumping discontinuously at this point. These significant differences substantially influence the estimation procedure and the resulting parameter estimates, possibly leading to worse predictive performance than if the parameters were fitted with a loss function identical to the performance metric.

There are important drawbacks to using a discrete performance metric. As the link between preferential strength and choice probabilities is severed by this metric, deterministic heuristics may be placed at a relative advantage to flexible models, as the advantage of the latter in accounting for preferential strength is ignored. Also, as discrete model predictions are less precise than continuous predictions, this makes models less identifiable or distinguishable, less falsifiable and more prone to model mimicry, thereby hampering efficient model comparison.

The topic of model mimicry has received increasing attention in the methodological literature, particularly with respect to its impact on model comparisons. Sets of flexible models can often exhibit significant model mimicry exactly because if endowed with numerous free parameters they can fit almost any data. Deceivingly, significant model mimicry can be found even across models that appear to have very different foundations and non-linear parametric forms, if there are enough parameters to interact with each other. Recall von Neumann's quip to Fermi, that "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk." Concerning model comparisons between flexible models such as cumulative prospect theory and choice heuristics, significant model mimicry has been found even for such different models ([Brandstätter et al., 2006](#), Table 5; [Pachur et al., 2013](#)). One perspective is that this is a feature of heuristics, since it implies that they can have approximately the same predictive performance as flexible models with much simpler functional form and a lack of free parameters. We agree, but wish to point out that these model mimicry comparisons have typically been performed on discrete prediction metrics, which necessarily preclude any informativeness that may be derived from strength-of-preference (and by extension, from stochastic variants of said models). An exception is the model recovery analysis by [Pachur et al. \(2013\)](#), who showed how

Cumulative Prospect Theory can mimic a wide variety of heuristics with very different processes, through the flexibility afforded by the probability weighting function parameters. In this study, CPT and heuristics were rendered stochastic through a fixed error mechanism, and the success of model recovery was determined for varying levels of errors. We suspect that comparing models not on a discrete metric but on a continuous metric (and stochastic variants of the models) such as choice probability after the introduction of a choice rule, will reveal less model mimicry than previously observed. This is a corollary to the argument that stochastic models are more falsifiable than their deterministic counterparts due to their more precise predictions covering the full probability range.

How large can this difference be theoretically? Consider the following example presented in Table 1. Let us take a simple case where decision-makers are presented with two different pairs of lotteries, each with two prospects A and B. The lotteries are repeated 100 times each, so that consistency and stochasticity can be revealed. If a discrete performance metric is used, then perfect model mimicry would be observed under the following conditions. Consider the empirical choice data first (the last column in the table), and let us assume that A was chosen in Lottery pair 1 99% of the time and 51% of the time in Lottery pair 2. Note that this would likely be the case if the two prospects in Pair 1 had valuations that were very different, leading to few errors. In Pair 2, in contrast, the two valuations of both prospect result in very similar values, leading to many errors and a near 50–50 choice proportion.

Now consider deterministic versions of two models, both of which predict prospect A as being chosen with certainty. In this case, both models would have the same predictive accuracy of 99 and 51% for Pairs 1 and 2, respectively, implying perfect model mimicry—see the first two columns in the table. That is, the two models' predictions are perfectly positively correlated across the lotteries. Suppose that the stochastic version of Model 1 predicts $p(A) = 0.51$ for Lottery 1 and $p(A) = 0.99$ for Lottery 2, whereas these values are flipped for Model 2. This is entirely consistent with the numbers used for the deterministic models above, as long as the stochastic versions both predict a choice probability for Prospect A greater than 0.5 for both lotteries—this would imply choosing A with certainty under a discrete metric. The predictive accuracy of Model 1 is now 99 and 51% for lotteries 1 and 2, whereas that of Model 2 is 51 and 91%, respectively. Examining the correlation between the two models and across the lottery predictions reveals that they are perfectly *negatively* correlated, in contrast to the perfect *positive* correlation between the deterministic models. Consequently, model mimicry was significantly over-estimated in the latter case. It is clear from the empirical (true) choice data, that Model 1 is preferable, however this can only be concluded by comparing the stochastic model variants with a continuous metric, not by their deterministic models.

2.3.2 Option 2

The second option is also problematic for numerous reasons—note, the arguments made above regarding discrete performance metrics continue to hold in this case. The strong empirical evidence that choice behavior is generally stochastic implies that deterministic models are strongly misspecified during estimation. Consequently, inferring that one deterministic model or the other

TABLE 1 A comparison of model mimicry and performance between deterministic and stochastic models.

Task	Model predictions				Choice data
	Deterministic		Stochastic		
	Model 1	Model 2	Model 1	Model 2	
#1	0.99	0.99	0.99	0.51	0.99
#2	0.51	0.51	0.51	0.99	0.51

has been invalidated by a model comparison is wrought with difficulty, as deviations of model predictions from the empirical data cannot necessarily be attributed to the core deterministic model being wrong, but may be due to the lack of an error mechanism. This is particularly problematic for studies that employ the axiomatic approach to invalidating a model (or that include specifically designed tasks to stress-test axioms). For example, deterministic EUT assumes transitivity of choices, which is not supported by the empirical data as we often observe violations. However, this does not preclude the deterministic component of EUT being correct, and that any violations of the transitivity axiom arise solely due to errors. Similarly, violations of stochastic dominance may arise either in the core deterministic model or from an error component (or both).

How problematic can this become? We perform a simple recovery simulation where the true choices or data are generated using a stochastic model and perform a model comparison analysis using deterministic models. If the deterministic model recovered matches the core deterministic component of the true stochastic choice model, then we deem this as a correct recovery. For example, if we generate the choice data using a stochastic Expected Value (EV) model, do we conclude often enough that the core component was the EV model even when our model comparison assumes a deterministic EV model and other competing deterministic heuristics?

We implement the simulation using heuristics that are often used in the choice under risk literature (Thorngate, 1980; Payne et al., 1988; Hertwig et al., 2019). The set of models consists of the Expected Value model, which uses all information (probabilities and outcomes), and the following heuristics that either ignore or process probability information in a non-multiplicative way: Maximax, Maximin, Least likely (LL), Most likely (ML), Equiprobable (EQ), Probable (Prob), and the Priority heuristic.¹ Assume that a decision-maker uses the same stochastic heuristic to make choices in N choice tasks or lotteries. Given the practical limitations of experiments, setting $N = 50$ is a reasonable

¹ Maximax chooses the prospect with the highest maximum value, Maximin the one with the highest minimum value, Least likely the prospect with the lowest probability of the worst outcome, Most likely the prospect with the highest most-likely outcome. Equiprobable assumes that each outcome has the same probability of occurring and chooses the prospect with the highest expectation. Equiprobable eliminates outcomes whose probability are less than the inverse of the number of outcomes, and then assumes equal-weighting of the surviving outcomes to calculate a prospect's expectation. See Section 3.3 for the definition of the Priority heuristic.

assumption. Two prospects for each lottery are randomly drawn in the following fashion. Both prospects consist of two outcomes each and probabilities are drawn from a uniform distribution drawn over $[0, 1]$ and the outcomes are drawn from a uniform distribution over $[0, 100]$.

We examine the case of two different stochastic models as the true choice models, EV and Maximin. Stochasticity is modeled as noise in the outcome values. This is a simplification, of course, as noise could also affect probabilities. However, since many heuristics ignore probabilities, but not outcome information, we settled on the latter. We vary the degree of noise or stochasticity in the data generation process by adding errors to each outcome value that are normally distributed with mean 0 and standard deviation equal to 5, 10, 20, 50, or 100.

Summarizing, for each true choice model and associated noise level, we calculate the recovery rate for each of the eight decision models under investigation. Model comparison and recovery is based on the following performance metric: The best performing model that we infer is used by the decision-maker is the one with the highest percentage of correct predictions across the N tasks. The choice data generation and model prediction is simulated 10,000 times, and the recovery rate is defined as the percentage of those simulations for which the correct model was inferred. If the outcome noise is zero, then the recovery rate will necessarily be 100% as the stochastic and deterministic versions of a model are identical.

Tables 2, 3 present the recovery rates for each of the two true models (stochastic EV and Maximin, respectively) for each noise level. If the true model is stochastic EV, for low noise ($\sigma = 5$) the recovery rate is very high (97%); however, as the level of noise increases recovery falls significantly to 70% for a noise level of 20. At high levels of noise (50 and 100), the recovery rate falls to 41 and 26%, indicating a significant failure in recovering the true model by deterministic models of choice behavior. When recovery fails, the most commonly inferred (incorrect) models are ML and Probable. This constitutes a significant failure as they are in principle quite different models from EV, ignoring some of a prospect's events and not fully utilizing probabilistic information. For high levels of noise (50 and 100), even more parsimonious heuristics may be incorrectly inferred as the true model, in particular ones that completely ignore probabilistic information (i.e., Maximax and Maximin).

Let us now turn to the case where a stochastic Maximin model generates choices. At the lowest level of noise (5), the recovery rate is 86%, falling to 71% for the next noise level (10), and only 50% for a noise level of 20. Compared to the case where the stochastic EV was the true model, recovery rates for stochastic Maximin are generally worse at the corresponding noise levels, and drop more quickly even for intermediate noise levels. The most common wrongly inferred model is the Priority heuristic, which is understandable as the latter shares a very similar first step in the lexicographic decision tree. The recovery rate of 50% at a noise level of 20 is quite poor, and even more problematic is the fact that some of the wrongly inferred models are significantly different to Maximin. The second most commonly inferred wrong model is Equiprobable, followed by EV. Equiprobable is a significantly different heuristic to Maximin in principle, as it examines all outcomes rather than just the minimum outcomes in each prospect. Even more concerning is that in the presence of noise the EV model

may be inferred as the true model, even though it is the antithesis of the Maximin heuristic. Our recovery simulations—while relatively simple abstractions of more complex model comparisons—have shown that there is cause for concern regarding the accuracy of model inference when heuristics are incorrectly assumed to be deterministic instead of stochastic. Further simulations seem warranted to investigate the accuracy of model recovery: (a) in a broader set of tasks where lotteries are sampled differently, (b) for a broader range of models, including flexible models such as CPT, (c) and for various categories of error mechanisms as defined in the next section.

Another issue with this option is that deterministic models necessitate discrete loss functions and performance metrics. This leads to a loss in the informational content of the empirical data, which by its nature is stochastic. Finally, using a discrete error function to estimate a flexible model (whether deterministic or stochastic) is extremely problematic due to key properties relating to the behavior of the loss function with respect to the estimation technique. For example, estimation based on minimizing the percentage of correct predictions is generally avoided as there is no guarantee of a unique solution in the parameter values due to the discreteness and lack of continuity of this loss function, i.e., different parameter values can lead to the same percentage of correct predictions, and it is not guaranteed that the estimation algorithm will converge to a global rather than local optimum. Since it is clearly desirable to estimate flexible models using continuous loss functions and to use an identical loss function and performance metric, this leaves only the next option as a viable candidate.

2.3.3 Option 3

This option in our opinion dominates the two previously discussed ones, yet to the best of our knowledge has not been extensively used in the literature for a wide range of flexible models and heuristics, only for a limited number of models in rare cases (e.g., Rieskamp, 2008). First, it deals with the misspecification issue as both types of models are implemented as stochastic and prone to errors. Secondly, the mismatch between the loss function and the performance metric can be eliminated for both types of models, by using a continuous error function with an identical performance metric. Unifying the loss function, estimation technique and performance metric for both models minimizes the auxiliary assumptions involved in any model comparison (in the spirit of the Duhem-Quine problem), lending further credibility to the comparison conclusions. The stochastic specifications will make the models more identifiable and falsifiable, by making more precise predictions on the continuous probability interval compared to discrete predictions with certainty, as argued above.

2.4 Discussion

To conclude, there are important reasons to consider stochastic models of heuristics and to compare flexible models and heuristics using continuous performance metrics, in contrast to the majority of studies that have used discrete metrics (e.g., Pachur et al., 2013). First, they are cognitively more realistic as choice is stochastic

TABLE 2 Recovery rates (%) if the true model is stochastic EV.

Noise σ	EV	Maximax	Maximin	LL	ML	Eq	Prob	Priority
5	97	0	0	0	6	2	6	0
10	89	1	1	0	15	7	15	0
20	70	5	4	0	26	15	26	2
50	41	15	13	7	29	21	29	9
100	26	19	16	16	24	19	24	14

TABLE 3 Recovery rates (%) if the true model is stochastic maximin.

Noise σ	EV	Maximax	Maximin	LL	ML	Equip	Prob	Priority
5	0	0	86	0	0	3	0	26
10	3	0	71	0	0	11	0	38
20	10	1	50	0	3	26	3	38
50	19	15	28	2	11	31	11	28
100	18	22	22	10	16	24	16	24

(or noisy) and dependent on preferential strength. Second, such models would allow for a more equitable comparison of heuristics versus flexible models by doing away with differences in auxiliary assumptions, using the full informational content of data and the more precise predictions of continuous choice probabilities.

Furthermore, stochastic models of heuristics will be more falsifiable than their deterministic counterparts, as they will be forced to also account for preferential strength to perform well. This is a crucial test for heuristics that has not been empirically conducted yet. It may lead to further innovation in the field if the existing models of heuristics are not found to predict preferential strength well.

It is conceivable that some deterministic models of heuristics that have been rejected as not predicting behavior well in past studies, may in fact have fallen prey to their lack of an error mechanism that could “explain” some deviant choices. Simply put, the misspecification of heuristics as deterministic may invalidate conclusions drawn from deterministic heuristic modeling comparisons, as we showed in our recovery simulation. To be fair, all models are misspecified, but given how elemental stochastic choice seems to be in every facet of human behavior, the omission of an error mechanism may be more important than other sources of misspecification, such as a parametric form that is not exactly faithful to its true form.

3 A classification of error mechanisms

We now classify various types of error mechanisms that are applicable to models of heuristics, and discuss their advantages and shortcomings. We will use the maximin heuristic as a case study of how to define a stochastic variant. It is well-known both in individual choice under risk and uncertainty, and also in strategic decision making, such as games where the choices of other influence one’s own payoffs (see [Spiliopoulos and Hertwig, 2020](#)). The maximin heuristic recommends that the chosen prospect is the one that has the most attractive worst-possible outcome. This heuristic is non-probabilistic and as it only compares outcome

values across prospects and is thus an instance of the class of fast-and-frugal heuristics.

3.1 Fixed (or independent) errors

Stochasticity arising from fixed errors is not conditional on any of the processes involved at arriving at a choice. Alternatively, they are sometimes referred to as naive errors, as they simply stipulate that the deterministically derived choice is mistakenly not chosen in $\epsilon\%$ of choices. Thus, if two options are available, the stochastic model of a heuristic will predict the choice of the deterministic model of the heuristic $100 - \epsilon\%$ of the time and the other choice $\epsilon\%$. If more than two options exist, then one must stipulate how the errors are spread to the other option. The most obvious choice that retains independence is to apportion the $\epsilon\%$ of errors uniformly over the other options. A stochastic version of maximin would therefore choose the option with the best worst-case scenario $100 - \epsilon\%$ of the time.

The advantage of fixed errors is that they are quite simple to implement and can be useful in cases where there may be multiple errors occurring prior to the final choice, but which would be too difficult to estimate and effectively identify during estimation. Thus, the cumulative effect of the errors during the decision processes will be estimated, with the cost that this distribution may not effectively capture the true error distribution. Note that fixed errors can be used even with models without a final valuation stage from which a strength of preference could be inferred.

A useful extension of fixed errors are *conditionally* fixed errors, where a fixed error parameter may be valued differently conditional on characteristics of the task. For example, errors may be more likely for more difficult tasks than for easier tasks. Returning to the example above, suppose a decision-maker must choose between two options in one case and four options in another—the probability of making an error is likely higher in the latter case than in the former. This could be modeled by allowing the value of the error rate ϵ to be conditional on the number of available options.

Of course, this comes at the cost of additional free parameters to be estimated.

A disadvantage of this fixed error mechanism is that it is still not ideal when used in conjunction with a continuous performance metric. For all tasks the choice predictions take on only two possible values, ϵ and $1 - \epsilon$, whereas a continuous metric can take on the full range of values between 0 and 1. This happens exactly because fixed errors are not conditional on preferential strength, which will typically vary across tasks allowing model predictions to take on a broader range of values instead of two discrete values. If the core model includes a valuation stage, then the next type of errors would be more desirable, as they allow the size of the error to be conditional on the measure of preferential strength derived from the valuations. This, in turn, would enable probabilistic predictions that are not constrained to just two values, ϵ and $1 - \epsilon$.

3.2 Valuation errors

Valuation errors are perhaps the most commonly employed error-mechanisms for flexible models. Valuation based error mechanisms are conditional on the relative magnitude of the valuations, which can be interpreted as a strength of preference. This error mechanism is more sophisticated and realistic than a fixed error mechanism—recall the evidence we presented earlier about the link between preferential strength and errors (or consistency).

How could such a mechanism be implemented in a deterministic heuristic, which usually do not have an explicit valuation stage? Let us turn again to the maximin heuristic. A prospect i is defined by the n possible outcomes and associated probabilities p_n . The maximin heuristic can be procedurally calculated in three steps:

1. Determine the minimum value in each option.
2. Compare the minimum values and find the option with the larger minimum value.
3. Choose this option.

The choice rule is based on the comparison between these two minimum values. Regardless of the magnitude of the differences between the minimum values, the heuristic uses an all-or-nothing rule in the final choice. What if a rule is used that depends on the difference between the two minimum values? That is, let us define the valuation of a prospect as its minimum outcome value, and interpret the difference in the two minimum values as defining the continuous strength-of-preference for one prospect over another. The higher the preferential strength, the more likely the prospect is to be chosen, meaning that choice probability is an increasing function of strength of preference. Consequently, a stochastic maximin heuristic could be defined as follows, where $\lambda = \epsilon^{-1}$ is the consistency parameter:

$$p_A = f(\min X_A, \min X_B, \lambda)$$

$$\frac{\partial p_A}{\partial (\min X_A) > 0}$$

$$\frac{\partial p_A}{\partial (\min X_B) < 0}$$

It is convenient to choose a parametric function f such that if the error parameter ϵ is zero the function will return the same prediction as the deterministic maximin heuristic. The advantage of this is that by estimating ϵ it is possible to actually ascertain *how* stochastic choice is, and it also includes the special case of the deterministic heuristic, if warranted by the data. An obvious candidate is the logit (or probit) function alluded to earlier, see Equation 1. As λ approaches infinity, the probability of choosing one of the prospects tends to 1 and the other to 0, i.e., identical to that made by deterministic maximin.

$$p_A = \frac{e^{\lambda(\min X_A)}}{e^{\lambda(\min X_A)} + e^{\lambda(\min X_B)}} \quad (1)$$

The parametric form of the error mechanism f has been shown to be very important, affecting not only the estimated parameters as we have already discussed above, but also the predictive performance of decision models and the informativeness of model comparisons (Zilker, 2022). Using cumulative prospect theory as the core deterministic component, Zilker rigorously examined various forms of error mechanisms and concluded that independent or fixed errors are eclipsed by the informativeness of the valuation error mechanism that we propose. Schulze et al. (2021) also concluded in their probabilistic model of the social circle heuristic that valuation error mechanisms, logit and probit, significantly outperformed a fixed error mechanism. Stott (2006) performed an extensive comparison of all possible combinations of different parameterizations for cumulative prospect theory's probability weighting functions, value functions and error functions, also concluding that the logit outperformed fixed errors and was the best performing parameterization. Consequently, wherever possible, we recommend using a valuation error mechanism instead of a fixed error mechanism. Further research should be directed at considering the appropriate functional form of the error mechanism for models of heuristics because, as we discuss below, other more sophisticated alternatives exist.

3.3 Procedural errors

Procedural errors can occur at any processing level or step (with the exception of the valuation stage, which was covered above as a special case). A prerequisite for such an error mechanism is that a procedural model be clearly defined in terms of the requisite cognitive operations. An obvious approach for interpreting such a model is to define it in terms of elementary information processing units (EIPs) and to allow for an error in multiple, but ultimately, all of the EIPs. That is, errors occur at every level of information integration (and possibly search) instead of after integration is complete and a valuation returned. The resultant choice errors are caused by the propagation of the procedural errors throughout the model. For example, an error at an early EIP can interact with an error at a later EIP, thereby leading to a very rich distribution of final choice errors that may even be multimodal. This contrasts the unimodal error distributions associated with valuation error mechanisms as a result of the assumption that preferential strength is monotonically related to errors. While not the focus here, finding multi-modal (and a more discretized) rather than uni-modal (and

continuous) error distributions may be a strong indication that the true core behavioral model is a heuristic rather than a flexible model. This conjecture may warrant further investigation as it may be a powerful way of identifying when heuristics are used by decision-makers.

Let us turn again to our Maximin example. The valuation error mechanism we implemented above assumed that the first step in Maximin—determining the minimum value in each option—was error free. A procedural error mechanism would introduce an error at this step. The procedural error that occurs in comparing outcomes within each option could be implemented as an independent error with fixed probability of occurring or as an error conditional on the difference between the compared values (e.g., the minimum and maximum outcomes in a two-outcome option).

For simplicity, let us present the procedurally stochastic maximin heuristic under the assumption of fixed procedural errors at the first step (occurring with probability ζ in both options) and valuation dependent errors as recommended above. This can be considered as a hybrid valuation and procedural error model. We assume that each of the two options consists of two outcomes each, therefore there are four possible combinations of errors in correctly ascertaining minimum and maximum values:

$$p_A = \begin{cases} \frac{e^{\lambda(\min X_A)}}{e^{\lambda(\min X_A)} + e^{\lambda(\min X_B)}} & \text{with prob. } (1 - \zeta)^2 \\ \frac{e^{\lambda(\max X_A)}}{e^{\lambda(\max X_A)} + e^{\lambda(\min X_B)}} & \text{with prob. } \zeta(1 - \zeta) \\ \frac{e^{\lambda(\min X_A)}}{e^{\lambda(\min X_A)} + e^{\lambda(\max X_B)}} & \text{with prob. } \zeta(1 - \zeta) \\ \frac{e^{\lambda(\max X_A)}}{e^{\lambda(\max X_A)} + e^{\lambda(\max X_B)}} & \text{with prob. } \zeta^2 \end{cases}$$

A more sophisticated implementation of procedural errors for the priority heuristic can be found in Rieskamp (2008). The priority heuristic is lexicographic and considers attributes of the prospects in the following order (first to last): minimum gain, probability of minimum gain, maximum gain, and probability of maximum gain. Stopping rules, diverting to a final choice, at each step are defined by setting minimum thresholds.

1. If the minimum gains of the two prospects differ by 1/10 (or more) of the (global) maximum gain, choose the prospect with the highest minimum gain; otherwise continue to step 2.
2. If the probabilities of the minimum gains differ by 1/10 (or more) of the probability scale, choose the prospect with the highest probability of the minimum gain; otherwise continue to step 3.
3. If the maximum gains differ by 1/10 (or more) of the (global) maximum gain, choose the prospect with the highest probability of the maximum gain; otherwise continue to step 4.
4. Choose the prospect with the highest probability of the maximum gain.

Each of the steps involves a comparison between two values, which in the deterministic version occur without error. By contrast, Rieskamp (2008) assumes that the subjective difference in the two values compared at each step is a random variable with a mean equal to the real difference and non-zero variance capturing errors in the comparison. Consequently, comparing the subjective difference to the threshold of each step ultimately leads to stochastic

or noisy choices. Thus, this stochastic model of the priority heuristic implements procedural errors according to our definition that are dependent on the magnitude of differences (in contrast to our maximin example above). Note that Rieskamp (2008) also estimates different threshold values, which are fixed in the deterministic version, and allows for the order of the steps to vary leading to between-participant stochasticity. However, we are here concerned with error mechanisms and stochasticity that arises within-participants.

3.4 Discussion

We consider procedural errors to be the most cognitively realistic error mechanism. Yet, there are disadvantages to implementing this type of mechanism relative to fixed or valuation error mechanism. The primary disadvantage is probably already apparent. It is the increase in model complexity introduced by the addition of more parameters at every processing step. The more parameters that need to be estimated, the more data are needed to identify those parameters well in the estimation and to avoid the curse of in-sample over-fitting. At some point, if too many arbitrary error parameters are introduced, this will blur the line between models of simple heuristics and flexible models. Two methodological tools may be useful in taming the problem of model complexity and identification if procedural errors are used. Instead of increasing the number of tasks in an experiment to collect more data, it may be useful to collect additional non-choice data, such as response times and process-tracing data. This data will also reduce model mimicry, as some models making similar choices may have very different implications for response times and/or information search and integration. A better understanding of the decision processes will be conducive to the addition of more appropriate procedural errors.

The advantage of the fixed and valuation mechanisms is that they can be implemented with only a single error parameter to be estimated. Let us return to our stochastic maximin example: the independent error version requires the estimation only of λ whereas the procedural version requires both λ and ζ . There is thus a tradeoff between cognitive plausibility, which we believe dictates an error mechanism at every information search or integration step (EIP) and estimation practicality. Given the constraints in the length of experiments and the number of tasks that can be reasonably presented to participants, in many cases procedural error mechanisms may not be a viable solution.

For the majority of studies, we anticipate that the most practical solution will be valuation mechanisms that implicitly aggregate the procedural errors into a single error at the valuation stage, albeit with some loss of information and misspecification of the true error distribution. Compared to fixed errors, the valuation mechanism has the advantage of being conditional on valuations, which given the existing empirical evidence cited earlier is highly likely to be relevant, and has been shown to be a significant improvement over fixed errors. At the very least, we would recommend empirical researchers to compare their deterministic heuristic to at least one stochastic version of it, following the example of Schulze et al.

(2021), who actually went further by considering two stochastic versions based on fixed and valuation error mechanisms.

While researchers should decide upon which type of mechanism to employ based on the merits of each particular study and tasks, we anticipate that valuation mechanisms will often represent the best tradeoff. However, wherever possible we would encourage consideration of a simple procedural mechanism, such as the one we presented for the Maximin heuristic that only adds one more parameter. Unfortunately, for procedural models with many EIPs, the complexity and number of free parameters may quickly increase, unless all types of EIPs are assumed to have an identical error mechanism and error parameter. Even though such an assumption is not realistic, it may be a reasonable approximation and a practical solution as it avoids additional error parameters.

In general, the heuristics commonly used in the literature on decision making under risk and uncertainty are all amenable to the valuation-based error mechanism adopted in our maximin example. More specifically, any heuristic at some point must make a comparison across options. It is simple to assume that a valuation-based error mechanism operates on those values that are compared when leading to the final decision (of the deterministic heuristics). For example, for the stochastic model of the maximax heuristic, the comparison would be across the maximum values of the two options. For the stochastic model of the equiprobable heuristic, it would be the sums of all outcomes of each option. Note, that while we refer to this as a valuation stage, our suggestion remains true to the simplicity of heuristics as this “valuation” is not derived from multiplicative and probabilistic calculations (as in expected utility theory), but is simply a comparison of two values that are not transformed in any way. The approach of treating the final values that are compared by a heuristic at a decision node as a form of valuation is virtually universally applicable, and is a practical way of generating preferential strength predictions from heuristics.

This approach can also be trivially extended to lexicographic heuristics with more than one final decision node. Here the valuation error mechanism is added to whichever node makes the final decision for a specific decision problem. However, it is not clear that the same error parameter would be appropriate at each decision node, especially if the compared values are scaled very differently or even refer to very different entities. For instance, the first and third decision nodes in the priority heuristic compare outcome values, whereas the second and fourth nodes compare the outcomes’ likelihoods. This is not prohibitive, but would mean that it may be necessary to estimate a different error parameter for different nodes.

4 Conclusion

The majority of models of choice heuristics in the literature make deterministic predictions. That is, they predict a specific choice with certainty. However, empirical evidence regarding choice stochasticity in general challenges this practice and raises important questions about whether stochastic variants of heuristics may be desirable. The few instances of stochastic heuristics, the stochastic priority model (Rieskamp, 2008) and the social circle

model (Schulze et al., 2021) have confirmed the superiority of stochastic variants over their deterministic counterparts.

We have presented a simple method for converting most heuristics for choice into stochastic variants. Crucially, this technique allows heuristics to determine a strength of preference for the options under consideration, thereby allowing for errors to be conditioned on the magnitude of preferential strength. This places heuristics on a more level playing field with the flexible models using free parameters that are often used in the literature and which are typically implemented with an error mechanism that induces choice stochasticity. Stochastic variants of heuristics address the problem of misspecification when error distributions are not included and also various methodological issues that arise particularly in model comparisons. Other advantages include making heuristics more falsifiable and allowing for more informative predictive metrics that encompass probabilistic predictions instead of an all-or-nothing metric (such as % correct responses).

There is of course a tradeoff to the above, and this comes in the form of the addition of at least one free parameter to capture the magnitude of errors. While anathema to parts of the heuristic literature that reject free parameters, our proposed technique allows researchers to retain the deterministic and parameter-free core component of existing heuristics, so that the free parameters enter only through the additional error mechanism. We believe that this is an acceptable tradeoff and that the advantages will outweigh the disadvantages. Importantly, we highlighted the possibility that some existing heuristics in the literature may have been erroneously discarded as not predictive of behavior due to the fact that errors were not accounted for. Ultimately, however, the pros and cons of stochastic versions of models of heuristics should be assessed empirically and in model competitions involving flexible models, and deterministic and stochastic models of heuristics.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

LS: Writing – review & editing, Writing – original draft. RH: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Alós-Ferrer, C., and Garagnani, M. (2021). Choice consistency and strength of preference. *Econ. Lett.* 198:109672. doi: 10.1016/j.econlet.2020.109672
- Alós-Ferrer, C., and Garagnani, M. (2022). Strength of preference and decisions under risk. *J. Risk Uncertain.* 64, 309–329. doi: 10.1007/s11166-022-09381-0
- Brandstätter, E., Gigerenzer, G., and Hertwig, R. (2006). The priority heuristic: making choices without trade-offs. *Psychol. Rev.* 113, 409–432. doi: 10.1037/0033-295x.113.2.409
- Brandstätter, E., Gigerenzer, G., and Hertwig, R. (2008). Risky choice with heuristics: reply to Birnbaum (2008), Johnson, Schulte-Mecklenbeck, and Willemsen (2008), and Rieger and Wang (2008). *Psychol. Rev.* 115, 281–289. doi: 10.1037/0033-295x.115.1.281
- Busemeyer, J. R., and Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychol. Rev.* 100, 432–459.
- Chabris, C. F., Laibson, D., Morris, C. L., Schuldt, J. P., and Taubinsky, D. (2009). The allocation of time in decision-making. *J. Eur. Econ. Assoc.* 7, 628–637. doi: 10.1162/jeea.2009.7.2-3.628
- Gigerenzer, G., and Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Top. Cogn. Sci.* 1, 107–143. doi: 10.1111/j.1756-8765.2008.01006.x
- Gigerenzer, G., and Gaissmaier, W. (2011). Heuristic decision making. *Ann. Rev. Psychol.* 62, 451–482. doi: 10.1146/annurev-psych-120709-145346
- Gigerenzer, G., and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychol. Rev.* 103, 650–669.
- Gigerenzer, G., Hertwig, R., and Pachur, T., editors (2011). *Heuristics: The Foundations of Adaptive Behavior*. Oxford: Oxford University Press.
- Gigerenzer, G., Todd, P. M., and the ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press.
- Hertwig, R., Woike, J. K., Pachur, T., Brandstätter, E., and Center for Adaptive Rationality. (2019). “The robust beauty of heuristics in choice under uncertainty,” in *Taming Uncertainty*, eds. R. Hertwig, T. J. Pleskac, and T. Pachur (Cambridge, MA: MIT Press), 29–50.
- Hey, J. D. (2001). Does repetition improve consistency? *Exp. Econ.* 4, 5–54. doi: 10.1007/bf01669272
- Kahneman, D., and Tversky, A. (1996). On the reality of cognitive illusions. *Psychol. Rev.* 103, 582–591.
- Katsikopoulos, K. V., Schooler, L. J., and Hertwig, R. (2010). The robust beauty of ordinary information. *Psychol. Rev.* 117, 1259–1266. doi: 10.1037/a0020418
- Katsikopoulos, K. V., Şimşek, Ö., Buckmann, M., and Gigerenzer, G. (2021). *Classification in the Wild: the Science and Art of Transparent Decision Making*. Boston, MA: The MIT Press.
- Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York, NY: John Wiley & Sons.
- Mata, R., Frey, R., Richter, D., Schupp, J., and Hertwig, R. (2018). Risk preference: a view from psychology. *J. Econ. Perspect.* 32, 155–172. doi: 10.1257/jep.32.2.155
- McFadden, D. (2001). Economic choices. *Am. Econ. Rev.* 91, 351–378. doi: 10.1257/aer.91.3.351
- Moffatt, P. G. (2005). Stochastic choice and the allocation of cognitive effort. *Exp. Econ.* 8, 369–388. doi: 10.1007/s10683-005-5375-6
- Mosteller, F., and Nogue, P. (1951). An experimental measurement of utility. *J. Polit. Econ.* 59, 371–404.
- Mousavi, S., and Gigerenzer, G. (2014). Risk, uncertainty, and heuristics. *J. Bus. Res.* 67, 1671–1678. doi: 10.1016/j.jbusres.2014.02.013
- Mousavi, S., and Gigerenzer, G. (2017). Heuristics are tools for uncertainty. *Homo Oeconomicus* 34, 361–379. doi: 10.1007/s41412-017-0058-z
- Mousavi, S., and Kheirandish, R. (2014). Behind and beyond a shared definition of ecological rationality: a functional view of heuristics. *J. Bus. Res.* 67, 1780–1785. doi: 10.1016/j.jbusres.2014.03.004
- Ortmann, A., and Spiliopoulos, L. (2023). Ecological rationality and economics: where the Twain shall meet. *Synthese* 201:135. doi: 10.1007/s11229-023-04136-z
- Pachur, T., Hertwig, R., and Gigerenzer, G. (2013). Testing process predictions of models of risky choice: a quantitative model comparison approach. *Front. Psychol.* 4, 1–22. doi: 10.3389/fpsyg.2013.00646/abstract
- Payne, J. W., Bettman, J. R., and Johnson, E. J. (1988). Adaptive strategy selection in decision making. *J. Exp. Psychol.* 14, 534–552.
- Payne, J. W., Bettman, J. R., and Johnson, E. J. (1997). “The adaptive decision maker: effort and accuracy in choice,” in *Research on Judgment and Decision Making: Currents, Connections, and Controversies*, eds. R. M. Hogarth and W. M. Goldstein (Cambridge: Cambridge University Press), 181–204.
- Ratcliff, R. (1988). Continuous versus discrete information processing: modeling accumulation of partial information. *Psychol. Rev.* 95, 238–255.
- Rieskamp, J. (2008). The probabilistic nature of preferential choice. *J. Exp. Psychol.* 34, 1446–1465. doi: 10.1037/a0013646
- Schulze, C., Hertwig, R., and Pachur, T. (2021). Who you know is what you know: modeling boundedly rational social sampling. *J. Exp. Psychol.* 150, 221–241. doi: 10.1037/xge0000799
- Şimşek, Ö. (2013). Linear decision rule as aspiration for simple decision heuristics. *Adv. Neural. Inf. Process. Syst.* 26, 2904–2912.
- Spiliopoulos, L. (2018). The determinants of response time in a repeated constant-sum game: a robust Bayesian hierarchical dual-process model. *Cognition* 172, 107–123. doi: 10.1016/j.cognition.2017.11.006
- Spiliopoulos, L., and Hertwig, R. (2020). A map of ecologically rational heuristics for uncertain strategic worlds. *Psychol. Rev.* 127, 245–280. doi: 10.1037/rev0000171
- Spiliopoulos, L., and Ortmann, A. (2018). The BCD of response time analysis in experimental economics. *Exp. Econ.* 21, 383–433. doi: 10.1007/s10683-017-9528-1
- Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *J. Risk Uncertain.* 32, 101–130. doi: 10.1007/s11166-006-8289-6
- Svenson, O., and Maule, A. J. (1993). *Time Pressure and Stress in Human Judgment and Decision Making*. Berlin: Springer.
- Thorngate, W. (1980). Efficient decision heuristics. *Behav. Sci.* 25, 219–225.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychol. Rev.* 34, 273–286.
- Todd, P. M., Gigerenzer, G., and Group, A. R. (2012). *Ecological Rationality: Intelligence in the World*. Oxford: Oxford University Press.
- Usher, M., and McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev.* 108, 550–592. doi: 10.1037/0033-295x.108.3.550
- Zilker, V. (2022). Choice rules can affect the informativeness of model comparisons. *Comput. Brain Behav.* 5, 397–421. doi: 10.1007/s42113-022-01042-5



OPEN ACCESS

EDITED BY

Samuel Shye,
Hebrew University of Jerusalem, Israel

REVIEWED BY

Elias L. Khalil,
Doha Institute for Graduate Studies, Qatar
Geoff Kushnick,
Australian National University, Australia

*CORRESPONDENCE

Gerd Gigerenzer
✉ gigerenzer@mpib-berlin.mpg.de

RECEIVED 12 March 2024

ACCEPTED 15 July 2024

PUBLISHED 05 September 2024

CITATION

Gigerenzer G and Garcia-Retamero R (2024)
Uncertainty about paternity: a study on
deliberate ignorance.
Front. Psychol. 15:1399995.
doi: 10.3389/fpsyg.2024.1399995

COPYRIGHT

© 2024 Gigerenzer and Garcia-Retamero.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Uncertainty about paternity: a study on deliberate ignorance

Gerd Gigerenzer^{1*} and Rocio Garcia-Retamero²

¹Max Planck Institute for Human Development, Berlin, Germany, ²Department of Experimental Psychology, University of Granada, Granada, Spain

Deliberate ignorance is the willful choice not to know the answer to a question of personal relevance. The question of whether a man is the biological father of his child is a sensitive issue in many cultures and can lead to litigation, divorce, and disinheritance. Thanks to DNA tests, men are easily able to resolve the uncertainty. Psychological theories that picture humans as *informavores* who are averse to ambiguity suggest men would do a DNA test, as does evolutionary theory, which considers investing in raising a rival's offspring a mistake. We conducted two representative studies using computer-based face-to-face interviews in Germany ($n = 969$) and Spain ($n = 1,002$) to investigate whether men actually want to know and how women would react to this desire. As a base line, Germans (Spanish) estimated that 10% (20%) of fathers mistakenly believe that they are the biological father of their child. Nevertheless, in both countries, only 4% of fathers reported that they had performed a DNA paternity test, while 96% said they had not. In contrast, among men without children, 38% (33%) of Germans (Spanish) stated they would do a DNA test if they had children, mostly without telling their partners. Spanish women with children would more often disapprove of a paternity test or threaten their husbands with divorce (25%) than would German women (13%). We find that a simple test of risk aversion, measured also by the purchase of non-mandatory insurances, is correlated with not wanting to know.

KEYWORDS

anticipated regret, deliberate ignorance, DNA paternity tests, Germany, insurance, paternity, risk aversion, Spain

Introduction

In August Strindberg's (2014) *The Father*, a cavalry captain learns that he is not the father of the daughter he adores. Without a biological link, he laments, paternal love is without foundation. He finds consolation in his childhood nursemaid and, his head nestled in her lap, speaks of the comfort of his "mother," the role the nursemaid assumed for him. Strindberg's play seizes on the conflicting forces of biological and social paternity.

The question of whether a man is the biological father of his child is a sensitive issue in many cultures and can lead to litigation, divorce, disinheritance, and disputes about child support (Anderson et al., 2007). Because of internal fertilization and live birth, a human female can be practically certain about her biological parenthood, whereas a male has to live with the uncertainty that someone else might be the biological father of his child. Until recently, men had to rely on uncertain cues such as physical resemblance or ABO blood tests that could exclude but not prove paternity. Modern DNA technology ("genetic fingerprinting") can resolve this uncertainty with practical certainty. Paternity is typically concluded if the probability that two individuals are biologically parent and child is estimated at 99.99% or

higher. The necessary material is easy to obtain (mouth swab, hair with roots, or used Kleenex), and the test is relatively cheap, approved by courts, and available for purchase on the internet. Now that paternal certainty is only “one click away,” do men want to find out?

In this article, we begin with theoretical perspectives that suggest different answers to the question. Then we report the first nation-wide representative studies in two large European countries, Germany and Spain, where we asked men with (without) children whether they had performed (would perform) a DNA paternity test, and women with (without) children about how they would react if their husband or partner asked for a DNA test. To see whether not wanting to know is associated with risk aversion, as it has been reported in previous studies on deliberate ignorance, we conducted a standard risk aversion test and also obtained data about real-life risk aversion as expressed by purchasing non-mandatory insurances.

The case for wanting to know

Much of philosophy and psychology has assigned a positive value to the power of knowing, and sometimes deemed it a moral obligation. Aristotle began his *Metaphysics* (Aristoteles, 1953) with the dictum “All men by nature desire to know.” Locke (1690/1953) listed ignorance as the first cause of wrong judgment. Logical positivists such as Rudolf Carnap (1969) argued that valid information should not be left on the table, and Bayesian statisticians such as I. J. Good (1967) reasoned that one's prior probabilities should be updated by new information. Similarly, modern psychological theories on information search assume that people want to know. Psychological theories generally picture humans as *informavores* (Miller, 1983) who are averse to ambiguity (Hogarth, 1987) and in need of closure (Kruglanski and Webster, 1996). Likewise, most theories in neo-classical economics assume that rational choice requires all relevant information to be known, and if not, actively searched for, until the costs of search exceed its expected benefits (Rizzo and Whitman, 2020). The desire for information appears to be the natural condition of humankind, whereas not wanting to know seems irrational and has often been linked to self-deception and shirking responsibility, as when women refuse to participate in breast cancer screening and people at risk for HIV do not pick up their test results (Thornton, 2008; Hertwig and Engel, 2020).

Given that the ability to invest in children is a limited resource, evolutionary theories focusing on inclusive fitness arrive at a similar conclusion. Altruistic behavior such as parental investment is assumed to be proportional to the genetic relatedness between donor and recipient (Hamilton, 1964; Alexander, 1974; Trivers, 1974; Anderson, 2006). These various *parental investment theories* predict that men's investment in children is a function of their confidence in paternity. In the words of Daly and Wilson (2006), “From the gene's eye view, laboring to raise a rival's offspring is a disastrous mistake” (p. 195), and “we might therefore expect men to be sensitive to available information about paternity” (p. 196).

In this view, a man can make two kinds of error: invest in a rival's offspring because he mistakenly believes himself to be the biological father or invest in his own child insufficiently because he mistakenly suspects that he is not the biological father. In terms of signal detection theory, the first error is a false positive, the second a miss. Today, a DNA paternity test can reduce both errors to practically zero. Thus,

various philosophical, psychological, and biological theories converge to the conclusion that it is rational for men to do a DNA test in order to eliminate paternal uncertainty.

The case for not wanting to know

Research on *deliberate ignorance* has documented cases where the expected desire for information does not hold and a substantial proportion of people willfully remain uninformed. For instance, after East Germany's Stasi records were opened in 1991, many citizens declined the opportunity to read their personal files. In their seminal analysis, Hertwig and Ellerbrock (2022) estimated that although about 40% of adult citizens believed that a Stasi file on them existed, more than half of these did not access it. Interviews uncovered a variety of reasons for this choice, including the anticipation of negative emotions and personal conflict if personal files were to reveal colleagues, friends, or family members who had spied on them (Hertwig and Ellerbrock, 2022).

The concept of deliberate ignorance refers to the willful decision not to know, as opposed to the inability to access information or mere disinterest. Deliberate ignorance requires two conditions (Gigerenzer and Garcia-Retamero, 2017, p. 180):

- 1 Choice of ignorance even when information is free or search costs are negligible.
- 2 Choice of ignorance notwithstanding personal interest.

Thus, deliberate ignorance is neither a result of another party withholding information nor the result of indifference or forgetting. Nor does it resemble a search for confirmatory information, as studied in the selective exposure literature (see Sweeny et al., 2010). The study of deliberate ignorance is also to be distinguished from the study of agnotology (Proctor and Schiebinger, 2008) and the sociology of ignorance (McGoey, 2014), which investigate the systematic production of ignorance by obscuring knowledge or disseminating fake news, as in generating and supporting public ignorance about global climate change.

Four key motives for deliberate ignorance have been identified (Hertwig and Engel, 2020; Gigerenzer and Garcia-Retamero, 2017). Three of these do not apply to the present study: achieving fairness and impartiality (as embodied by blindfolded Lady Justice), gaining strategic advantage (as in bankers' willful blindness to risks that led to the financial crisis of 2008; see Admati and Hellwig, 2013), and suspense and surprise (e.g., 40% of Germans do not want to know the sex of their child before birth, and instead wish to maintain the suspense and surprise; Gigerenzer and Garcia-Retamero, 2017).

The fourth motive is relevant for the present study: to avoid potentially bad news and subsequently regret having to live with it, particularly in situations that one cannot change. For instance, when agreeing to have his genome sequenced, James Watson, the co-discoverer of DNA, requested that information about his ApoE4 genotype, which indicates risk of Alzheimer's disease, be deleted from his published genome and not revealed to himself (Wheeler et al., 2008). Watson had perhaps concluded that because the disease is incurable, the anticipated regret of living with bad news would be larger than the meager benefits of knowing (Hertwig and Engel, 2020). The decision of many citizens not to read their personal Stasi records is another case in point. This motive is known as *anticipatory regret*.

Regret is a negative emotion that people may experience *after* choosing option A (e.g., not buying fire insurance) and later learning that option B (buying insurance) would have resulted in a more favorable outcome. *Anticipated regret* is an emotion that occurs *before* the choice is made (Luce and Raiffa, 1957). Anticipating possible regret may itself influence the choice. One imagines what would happen if an outcome were known and then decides not to know.

For the present topic, men might prefer deliberate ignorance because they anticipate regret about having performed a DNA test. If the test shows non-paternity, they might regret facing this new situation, in particular, their relation with spouse and child; if the test confirms paternity, they might regret having done the test and offended their partner by mistrusting her. Thus, to the degree that men have anticipatory regret, they should prefer deliberate ignorance about paternity.

The regret theory of deliberate ignorance (Gigerenzer and Garcia-Retamero, 2017) is based on Luce and Raiffa's (1957) classical regret theory and makes several general predictions, which are formally derived in Gigerenzer and Garcia-Retamero (2017). The first is that anticipatory regret increases the nearer the event is, that is, the nearer regret can occur.

Are men with or without children more likely to consider a paternity test?

According to parental investment theories, men with children should be most interested in doing a paternity test, while theories that picture humans in general as informavores do not make a specific prediction, so men without children might be equally interested in doing a test if they were a parent. In contrast, the regret theory of deliberate ignorance specifically predicts that the closer in time to the critical event that could generate regret, the higher the anticipated regret and the lower the number of individuals who want to know (Gigerenzer and Garcia-Retamero, 2017). For instance, the older people are, the less likely they want to know when they and their partner will die (Gigerenzer and Garcia-Retamero, 2017). This dependence of the rate of deliberate ignorance in a population on the time to the possible regret is the *time-to-event hypothesis*. In the case of paternity, it leads to this prediction:

Prediction 1: Men without children are more likely to say that they would want to know, whereas those who have children are less inclined to actually find out.

The rationale is that men without children can less likely imagine the anticipated regret of knowing than can men with children.

Risk aversion

The regret theory of deliberate ignorance is a direct extension of Luce and Raiffa's (1957) regret theory, which was formulated for risky choices. It facilitates deducing predictions about the relation between risk aversion and deliberative ignorance (Gigerenzer and Garcia-Retamero, 2017). Here we apply this theory for the first time to uncertainty about paternity.

Risk aversion test

People are said to be risk averse for gains if they choose a certain gain $v = \$X$ over a gamble with a higher expected gain. To measure

risk aversion, we used a standard paradigm, where participants can choose between a sure gain and a gamble. The rationale for Prediction 2 lies in the asymmetry of the possibility of the experience of regret in the standard risk aversion paradigm. If the risky gamble is chosen, it is played out and regret can occur if the result is less than the certain gain. If the certain gain is chosen, the risky gamble is not played out, meaning that it is not possible to know whether choosing the risky option would have led to a better or worse outcome. Regret is possible only when people choose the risky option and the result is unfavorable. By selecting the certain gain, an individual can thus avoid regret.

In other words, the same motivation—avoiding anticipatory regret—underlies both risk aversion and deliberate ignorance. Hence, we predict that if deliberate ignorance is due to regret avoidance, it should be more frequent among men who are risk averse.

Prediction 2: Men who are risk averse for gains are more likely to exhibit deliberate ignorance.

Consider now losses. People are said to be risk averse for losses if they choose a certain loss $v = \$X$ over a gamble with a smaller expected loss. As explained above, regret is only possible if a person chooses the risky option. Thus, by choosing the certain loss, one can avoid the possibility of regret.

Prediction 3: Men who are risk averse for losses are more likely to exhibit deliberate ignorance.

Note that Predictions 2 and 3 assume that risk aversion applies to both gains and losses, unlike the hypothesis that people are risk averse for gains and risk seeking for losses (Kahneman and Tversky, 1979).

Purchasing non-mandatory insurance

Buying non-mandatory insurance such as life and property insurance is equivalent to choosing a sure loss $v = \$X$ (the insurance premium) over a probable loss with a lower expected loss. Thus, buying non-mandatory insurance is equivalent to risk aversion for losses, which leads to the following prediction:

Prediction 4: Men who buy non-mandatory insurance are more likely to exhibit deliberate ignorance.

Note that the predictions state correlations between deliberate ignorance and measures of risk aversion, including purchasing non-mandatory insurance, not causations. We also do not postulate that the two feelings—anticipatory regret in the case of paternity and risk aversion in the case of gambles and insurance—are of the same subjective quality or currency, but only that they are correlated. To the best of our knowledge, Predictions 1 to 4 are new and have never been tested in the context of paternity uncertainty. Confirming them would provide support for the regret theory of deliberate ignorance. Moreover, it would demonstrate that the classical measure of risk aversion is a valid diagnostic test for men's attitudes toward wanting to know about paternity.

Women's willingness to agree

Women's reaction to their partners' request for a paternity test likely depends on the cultural context. One might thus expect

differences between the two countries, but it was not clear to us in which direction. Risk aversion, in contrast, allows for a prediction. If the classical risk test has diagnostic power, risk-averse women should more likely agree to their partners' request for a paternity test than risk-seeking women. For instance, women with small children may be financially dependent on their partners and might anticipate that openly disagreeing with the request would only heighten their partner's suspicion and endanger emotional and financial support. In this way, they might anticipate regret for having disagreed openly.

We measured risk aversion for women in the same two ways as for men, by a classical risk aversion test and by the possession of non-mandatory insurance.

Method

Population and sample

We hired the international survey company GfK Group, based in Nuremberg, Germany, with an office in Valencia, Spain. GfK obtained nationwide quota samples of 1,016 adults in Germany and 1,002 adults in Spain. The samples were representative of the population in each country in terms of four variables: age, gender, region, and size of settlement. In the German sample, 47 participants did not complete the questions, which reduced the sample size to 969. Table 1 shows the characteristics of the two samples. The paternity study was part of a larger survey on deliberate ignorance (Gigerenzer and Garcia-Retamero, 2017). We report 95% confidence intervals (CIs) for sample statistics. When 95% CIs are used, our sample size of approximately 1,000 participants per country provides a power of 0.99 to detect a small effect size (corresponding to Cohen's $h = 0.2$) and a power of over 0.995 to detect a medium effect size (corresponding to Cohen's $h = 0.5$; Cohen, 1988). The ethics committee of the Max Planck Institute for Human Development approved the methodology.

Procedure

To ensure the quality of data for this sensitive topic, we invested in computer-based face-to-face interviews and risk aversion tests rather than a less expensive telephone or internet survey. After a first telephone contact was established, all participants were interviewed individually in their homes. Participants could enter their responses directly into the computer. To begin with, they were asked to estimate the frequency of non-paternity in their countries. Males were asked whether they had performed DNA testing or, for those who did not have children, whether they intended to perform DNA testing when they had children. Females were asked about their reactions to their partner's wanting to know.

All participants took two tests of risk aversion, one for gains and one for losses.

Risk aversion for gains:

You won a contest and have to choose between two alternatives: a lottery and a sure gain. The lottery has 10 items, five of which win 100 euros, the others nothing. Would you prefer the sure gain to the lottery?

Win 20 euros for sure instead of the lottery. yes/no

TABLE 1 The German and the Spanish sample by gender, age, religious practice, education, marital status, risk aversion, and non-mandatory insurances bought.

	Germany		Spain	
	<i>n</i>	%	<i>n</i>	%
Total	969	100.0	1,002	100.0
Gender				
Male	471	48.6	491	49.0
Female	498	51.4	511	51.0
Age				
18–35	306	31.6	322	32.1
36–50	294	30.3	304	30.4
51+	369	38.1	376	37.6
Religious services per month				
0 times	683	70.5	699	69.8
1–2 times	187	19.3	175	17.5
3+ times	99	10.2	127	12.7
Education				
1	45	4.7	53	5.3
2	374	39.3	139	14.0
3	317	33.3	328	32.9
4	129	13.6	321	32.2
5	87	9.1	155	15.6
Marital status				
Married	391	40.4	409	40.8
Not married	578	59.6	593	59.2
Risk aversion				
1	216	31.8	271	35.9
2	283	41.6	132	17.5
3	154	22.6	287	38.0
4	27	4.0	65	8.6
Insurance				
Life	575	59.3	423	42.2
Household	757	78.1	712	71.1
Personal	751	77.5	227	22.7
Legal	440	45.4	51	5.1

Education: 1 = primary/lower secondary school without vocational training; 2 = primary/lower secondary school with vocational training; 3 = further education without secondary school leaving qualification (US: high school diploma); 4 = secondary school leaving qualification; 5 = university. Percentages do not add up to 100% because 18 Germans and 6 Spaniards were still in school. Risk aversion: 1 = risk averse for gains and risk seeking for losses; 2 = risk averse for gains and losses; 3 = risk seeking for gains and losses; 4 = risk seeking for gains and risk averse for losses (numbers do not add up to total sample size because risk neutrals are not included). Insurance: Life = life insurance; Household = household insurance ("Hausratsversicherung"); Personal = personal liability insurance ("Privathaftpflicht"); Legal = legal expenses insurance.

Win 30 euros for sure instead of the lottery. yes/no
Win 40 euros for sure instead of the lottery. yes/no
Win 50 euros for sure instead of the lottery. yes/no
Win 60 euros for sure instead of the lottery. yes/no
Win 70 euros for sure instead of the lottery. yes/no

The interviewer presented the options, in the order shown above, successively to the participant on a computer screen until the participant answered “yes.” If a person preferred a sure gain to a lottery with a higher expected value, they were classified as risk averse for gains. An example would be a participant who preferred a sure win of 40 euros to a lottery whose expected value is 50 euros. If a person preferred a lottery to a sure gain despite the lottery having a smaller expected gain, they were classified as risk seeking for gains. An example would be a participant who preferred the lottery to a sure win of 60 euros.

Risk aversion for losses:

You lost a contest and have to choose between two alternatives: a lottery and a sure loss. The lottery has 10 items, for five of which you have to pay 100 euros, for the others nothing. Would you prefer the sure loss to the lottery?

Pay 70 euros for sure instead of the lottery. yes/no

Pay 60 euros for sure instead of the lottery. yes/no

Pay 50 euros for sure instead of the lottery. yes/no

Pay 40 euros for sure instead of the lottery. yes/no

Pay 30 euros for sure instead of the lottery. yes/no

Pay 20 euros for sure instead of the lottery. yes/no

A person is said to be risk averse for losses if preferring a sure loss to a lottery with a smaller expected loss, and risk seeking for losses if preferring a lottery to a sure loss when the lottery has a higher expected loss.

Results

Perceived prevalence of non-paternity

We obtained base rate estimates from both the German and the Spanish sample regarding their perceived rate of non-paternity:

What is your estimate of how many fathers in Germany [Spain] mistakenly believe that they are the biological father of their child? _____ out of every 1,000.

Among Germans, the average estimate was 96 in 1,000; among Spaniards, it was twice as high, 199 in 1,000. Thus, the general public in both countries appears to understand the potential magnitude of this eventuality. The average estimates of 10 and 20% are at the high end of scientific estimates reported in the literature, with the caveat that objective figures are hard to obtain (see Discussion Section).

Do fathers want to know?

Next, we asked men with children:

A DNA test can determine paternity with high certainty. All one needs is a hair from the child’s head. Have you ordered a test of one or more of your children to be sure that you are the biological father?

Only 4% (4%) of the men with one child in Germany (Spain) said that they had done a paternity test. Among men with several children, the proportion was similar, 4% (5%). In contrast, 96% of fathers reported not having done a DNA test.

We found that those who said they had performed or would perform a DNA test had higher estimates of non-paternity. Among Germans, the average estimate in this group was 106 in 1,000 (SEM=5.3) compared with 91 in 1,000 (SEM=5.3) among those who would not perform or had not performed the test. The same pattern occurred among Spaniards, with estimates of 224 in 1,000 (SEM=10.1) and 179 in 1,000 (SEM=8.9), respectively. This association could indicate that beliefs about the frequency of non-paternity influence the decision about conducting a paternity test, but it could also mean that the decision influences the estimates.

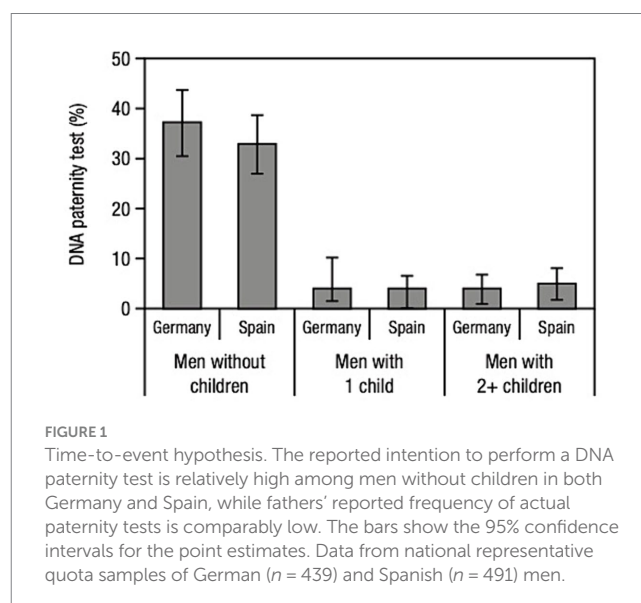
In sum, 96% of men in both countries with children reported that they had not performed a DNA paternity test to determine whether or not they were the biological father of their children.

Men with and without children

According to Prediction 1, men without children are more likely to say that they would want to know, whereas those who have children are less inclined to actually find out. To test this, we asked men without children the same question as for men with children, except that it was phrased “Assume you are married and have a 3-year-old child. A DNA test can determine paternity with high certainty. All one needs is a hair from the child’s head. Would you order...”

Thirty-eight percent of German men and 33% of Spaniards without children answered “yes” (Figure 1, left two panels). This substantial difference to the average 4% of fathers who reported that they had actually done a DNA test (Figure 1, right four panels) is consistent with the time-to-event hypothesis (Prediction 1), but not with the hypothesis of an increasing desire to know. We checked whether this effect could be due to the age difference between those with and without children, with the former being on average older. But, consistent with the hypothesis, a logistic regression analysis using all variables in Table 1 showed that age was not a valid predictor, whereas marital status and having children were.

With respect to honesty, the majority of men without children who said that they would conduct a DNA test indicated that they



would not tell their wives. Among Germans, the 38% figure splits into 23% who would not inform their wives and 15% who would; among Spaniards, the 33% figure splits into 21 and 12%, respectively. Note that “secret” paternity testing without both parents’ full consent is illegal in Germany under the Gene Diagnostics Act of 2009. The current Spanish law also requires consent, although it does not specify what happens if the mother does not consent (Barrot et al., 2014).

Risk aversion

Previous research reported that deliberate ignorance is more frequent among people who are risk averse than among those who are risk seeking. This phenomenon was observed in contexts other than paternity, both for negative events such as wanting to know the time one will die and for positive events such as wanting to know what presents one will get for Christmas (Gigerenzer and Garcia-Retamero, 2017). Does a similar association hold for paternity as well?

Table 2 reports the results aggregated across the two countries and across men with and without children because these were consistent. Among men who were risk averse for gains, 85.3% (422 of 495) stated they did not want to know, compared with 76.6% among those classified as risk seeking, resulting in a difference of 8.7 percentage points (95% CI = 2.9–14.4). This result is consistent with Prediction 2.

TABLE 2 Risk aversion and purchase of non-mandatory insurance are diagnostic for men’s choice of deliberate ignorance about biological paternity.

Proportion of men (n/N) who do not want to know			
Risk attitude	Risk averse	Risk seeking	Difference (risk averse – risk seeking) in percentage points [95% CI]
Gains	85.3% (422/495)	76.6% (229/299)	8.7 [2.94 to 14.39]
Losses	83.6% (225/269)	81.0% (439/542)	2.6 [–2.87 to 8.16]
Total (gains + losses)	84.7% (647/764)	79.4% (668/841)	5.3 [1.52 to 9.00]
Purchase of non-mandatory insurance	Insurance	No insurance	Difference (Yes – No) in percentage points [95% CI]
Life insurance	85.2% (403/473)	78.1% (382/489)	7.1 [2.22 to 11.95]
Property insurance	85.5% (572/669)	72.7% (213/293)	12.8 [7.05 to 18.56]
Personal insurance	86.4% (412/477)	76.9% (373/485)	9.5 [4.61 to 14.32]
Legal insurance	90.9% (219/241)	78.5% (566/721)	12.4 [7.66 to 17.08]
Total (all insurances)	86.3% (1,606/1,860)	77.2% (1,534/1,988)	9.2 [6.76 to 11.60]

Men with and without children are included. *n* = number of men in the subgroup who did not want to know, *N* = total number of men in subgroup. For instance, among the 495 men who were risk averse for gains, 422 did not want to know (85.3%), while among the 299 men who were risk seeking for gains, only 229 (76.6%) did not want to know. Men who were neither risk averse nor risk seeking but risk neutral are not included.

Among men who were risk averse for losses, the difference is also consistent with Prediction 3, but smaller in size, with a difference of 2.7 percentage points and the confidence interval including zero. Across both risk attitudes, gains and losses, deliberate ignorance is 5.3 percentage points higher among risk-averse men (95% CI = 1.5–9.0). Thus, overall, risk aversion among men is associated with not wanting to know about paternity.

Purchase of non-mandatory insurance

We asked participants whether they had bought life, property, personal, and legal insurances (the four most frequent non-mandatory insurances in Germany and Spain). According to Prediction 4, men who buy these insurances are more likely to exhibit deliberate ignorance. This prediction is correct for each of the four insurances (Table 2). For instance, 85.2% of men who had bought life insurance said they did not want to know whether they are the biological father of their child, compared with 78.1% of men who had not purchased life insurance, resulting in a difference of 7.1 percentage points (95% CI = 2.2–11.9). Across all four insurances, the percentage who reported not wanting to know was 9.2 percentage points higher among men who had purchased insurance (95% CI = 6.8–11.6).

In sum, the tests of Predictions 1 to 4 were consistent with the regret theory of deliberate ignorance. That is, risk aversion, as measured by a simple risk test or by the possession of non-mandatory insurances, can serve as a diagnostic test of men’s willingness not to know about paternity.

Women’s reaction to husband’s request for a paternity test

How would women react if their husband or partner wanted to find out whether he is the biological father? We asked the female participants:

Assume you are married and have a 3-year-old child. A DNA test can determine paternity with high certainty. All one needs is a hair from the child’s head. Your husband wants to conduct a test to be sure that he is the biological father of the child. How would you react?

- 1 I would agree because I have nothing to hide.
German women without children: 50%; with children: 57%.
Spanish women without children: 35%; with children: 45%.
- 2 I would agree but I would be offended.
German women without children: 33%; with children: 30%.
Spanish women without children: 33%; with children: 30%.
- 3 I would not agree.
German women without children: 8%; with children: 9%.
Spanish women without children: 19%; with children: 16%.
- 4 I would threaten with divorce or separation.
German women without children: 8%; with children: 4%.
Spanish women without children: 13%; with children: 9%.

The responses reveal two major results. The first is a pattern similar to the time-to-event hypothesis (Figure 1). Women

without children, in comparison to women with children, less often said they would agree and more often said they would threaten with divorce or separation. We checked whether this effect could be due to the age difference between those with and without children. As in the case of the men, a regression analysis using all variables in Table 1 showed that age was not a valid predictor but marital status and having children were. Thus, among couples with children, more women said they would tolerate men's wish to be certain about paternity, while few men actually have this wish.

The second result is a cultural difference between German and Spanish women. About twice as many Spanish women with children would not agree to a paternity test or would threaten their husbands with divorce or separation (25%) than German women with children (13%). Correspondingly, more German than Spanish women said they would agree to testing because they had nothing to hide (a difference of 15 percentage points for women without children, and 12 percentage points for those with children). A regression analysis showed that culture remained a valid predictor when controlled for the other variables in Table 1.

Thus, Spanish women were less likely to accept paternity tests than their German counterparts, which may have to do with traditional values of honor and marital integrity, or also reflect more recent developments such as that Spain has surpassed Germany in the number of women in full-time employment, leadership positions, and active military service.

Risk aversion and insurance

We compared women who would agree because they had nothing to hide (response alternative 1) with those who would be offended, not agree, or threaten with divorce or separation (other response alternatives), for short, “agree” versus “offended.” Women who were risk averse for gains were more likely to agree to a paternity DNA test than those who were risk seeking (Table 3). The difference was 12.1 percentage points (95% CI = 5.1–19.2). A similar difference replicates for women who were risk averse for losses. Across both gains and losses, the willingness to accept the husband's request for a DNA test was 10 percentage points higher among risk-averse women (95% CI = 5.3–14.7).

Buying non-mandatory insurance was also associated with women's willingness to accept a paternity test, but the absolute effect size was smaller (Table 3). Across all four insurances, the willingness to accept the husband's request for a paternity test was 5.9 percentage points higher among insured women (95% CI = 2.8–8.9).

In sum, risk aversion, as measured by a simple test or by the possession of non-mandatory insurance, is diagnostic of women's willingness to agree. Women who said they would be offended, not consent to testing, or threaten with divorce were more likely risk seeking and had not purchased non-mandatory insurances. The effect sizes are quite substantial, up to 12 percentage points, and similar for men and women.

It is surprising that a simple test of risk aversion can capture so well the attitudes of both men and women toward paternity testing.

TABLE 3 Risk aversion and purchase of non-mandatory insurance are diagnostic for women's willingness to consent to a DNA paternity test.

Proportion of women (n/N) who would agree with a paternity test			
Risk attitude	Risk averse	Risk seeking	Difference (risk averse – risk seeking) in percentage points [95% CI]
Gains	52.3% (301/575)	40.2% (111/276)	12.1 [5.05 to 19.21]
Losses	55.7% (156/280)	45.0 (260/578)	10.7 [3.64 to 17.82]
Total	53.4% (457/855)	43.4% (371/854)	10.0 [5.29 to 14.72]
Purchase of non-mandatory insurance	Insurance	No insurance	Difference (Yes – No) in percentage points [95% CI]
Life insurance	49.1% (258/525)	46.5% (225/484)	2.7 [–3.50 to 8.78]
Property insurance	49.4% (395/800)	42.1% (88/209)	7.3 [–0.34 to 14.63]
Personal insurance	53.3% (267/501)	42.5% (216/508)	10.8 [4.61 to 16.82]
Legal insurance	53.2% (133/250)	46.1% (350/759)	7.1 [–0.05 to 14.12]
Total	50.7% (1,053/2,076)	44.8% (879/1,960)	5.9 [2.80 to 8.95]

n = number of women in subgroup who would agree to DNA testing, N = total number of women in subgroup. For instance, among the 575 women who were risk averse for gains, 301 would agree to a DNA test without being offended (52.3%), while among the 276 women who were risk seeking for gains, only 111 (40.2%) would do so. Percentages are rounded.

Discussion

The present study addressed the phenomenon of deliberate ignorance—the decision not to know particular information of personal relevance despite low search costs. Contrary to cognitive theories that picture humans as informavores, 96% of fathers in Germany and Spain reported that they had not performed a DNA test, and thus did not want to know. This finding clashes with expectations from a spectrum of theories, from philosophy to evolutionary biology, that emphasize the value of knowledge and the dangers of not knowing. Why would so many men not want to know? We suggested anticipatory regret as one of the motivations, derived four predictions from regret theory (Luce and Raiffa, 1957; Gigerenzer and Garcia-Retamero, 2017), and found support for these. Deliberate ignorance is higher (1) for men with children than for men without children (time-to-event hypothesis), (2) for men who are risk averse for gains, (3) for men who are risk averse for losses, and (4) for men who buy non-mandatory insurances. The results indicate that risk aversion is diagnostic for deliberate ignorance regarding paternity.

We showed that risk aversion is also diagnostic for women's willingness to agree to a DNA paternity test. Women who are risk

seeking would more likely not consent and would threaten with divorce or separation. This consistent finding supports the interpretation that women, like men, try to avoid situations for which they anticipate regret.

Alternative explanations

Surveys, even with nationally representative quota samples, cannot provide a unique answer to what motivates deliberate ignorance. But we can use the evidence to exclude some alternative explanations. The first is that men might believe that non-paternity is so rare that it is not worth the effort of conducting a DNA test. We can exclude this explanation on the basis of participants' estimates that 10% (Germans) or 20% (Spaniards) of fathers mistakenly believe that they are the biological father of their children. In the literature, the frequency of actual non-paternity has also sometimes been estimated between 10 and 20% (e.g., Baker and Bellis, 1995; Gaulin et al., 1997; Alfred, 2002), but these figures appear to be inflated as estimates for the general population because of selection biases, such as mistrustful fathers visiting paternity-testing laboratories. A meta-analysis with more than 24,000 subjects from mostly Caucasian populations estimated a rate of 2–3% (Voracek et al., 2008). A study in Germany estimated the non-paternity rate as around 1% (Wolf et al., 2012). Whatever the true rates are, the participants in our study estimated the frequency of non-paternity at the high end. Thus, the explanation that men consider non-paternity a negligible phenomenon has little support.

A second explanation is that men might not believe that DNA testing is reliable. There might be too many misses and false alarms. To determine whether our participants were aware of the high accuracy of DNA profiling, we asked whether the result of a DNA test is “absolutely certain.” Seventy-eight percent of the participants in Germany and 89% in Spain thought so. This result is consistent with an earlier representative survey, where 78% of Germans also thought that the result of DNA test is absolutely certain, compared with 63% for fingerprints and HIV tests (Gigerenzer et al., 2007). The fact that the large majority believed that DNA tests are absolutely certain makes lack of trust in the reliability of the test an unlikely explanation for why so many men did not use the test.

Finally, we checked whether religion or education could explain men's wanting to know and women's willingness to accept. Religious belief was measured by the number of attendances of religious services per month (Table 1). A regression analysis using all variables in Table 1 showed that neither religion nor education was associated with whether men wanted to know, when controlled for other factors. The same result was obtained for women. That education made no difference may come as a surprise, yet it is consistent with studies of deliberate ignorance in other contexts (Gigerenzer and Garcia-Retamero, 2017).

The four hypotheses we tested were derived from the assumption that anticipatory regret is a key factor for deliberate ignorance. That is, a man imagines that after seeing the test result, he might wish he had not done the test. Anticipatory regret increases the closer one is to the point at which regret can

occur. Regret can be avoided by being risk averse, as measured by the classic test for risk aversion, both for gains and for losses. Similarly, purchasing non-mandatory insurance is motivated by anticipatory regret. All of these factors proved to be diagnostic, suggesting that a key motivation for deliberate ignorance is anticipated regret, consistent with previous results in other domains (Gigerenzer and Garcia-Retamero, 2017).

Strengths and limits

A unique strength of this study is that we obtained two representative quota samples from two countries and conducted the computer-assisted interviews in person by visiting the participants' homes. The downside was the substantially higher cost of this survey method compared with telephone or internet surveys. We decided upon this more expensive and labor-extensive procedure to secure the quality of the data, given the sensitivity of the topic.

One limit of the present study is that it relies on reported behavior rather than on measurements of actual behavior. We sought to reduce potential reporting bias by using computer-based face-to-face interviews with guaranteed anonymity. Nevertheless, the true number of paternity tests could be larger than the self-reported cases if some men did not admit to testing. To check this possibility, we obtained estimates of the actual sales of DNA paternity tests, which are difficult to verify given the multitude of companies that sell them. The best estimates for Germany seem to be in the order of 30,000 tests per year (Hipp, 2007). In relation to the approximately 680,000 newborns per year, this amounts to 4–5% of children being tested, which is consistent with the self-reports. Another limit is that we do not explicitly deal with how a man's decision depends on his trust in his wife and how other members of the family would be impacted by a positive DNA test. However, one can consider the negative impact on the family, especially on the child, as part of the anticipated regret. A final limitation is that these two representative studies allow generalization to the population of Germany and Spain, but not necessarily to different cultures.

Can deliberate ignorance be rational?

For those who believe that more information is always better, the majority of the men in both countries decide irrationally. As mentioned at the beginning of this article, philosophers such as Rudolf Carnap and Bayesian statisticians such as I. J. Good have proposed principles of rationality that imply one should not leave information on the table if it costs little or nothing. Anticipatory regret, in contrast, provides a reasonable explanation of this seemingly irrational behavior. Many do not want to know information that could become a disturbing problem. In the case of paternity, men's decision not to know provides protection for the wellbeing of the children and the family, preferring trust to the objective potential of technology.

According to Greek mythology, Cassandra, the daughter of the king of Troy, was cursed by Apollo to foresee the future. Cassandra foresaw the death of her father, the hour of her own death, and the name of her murderer. If she had had the choice to stay deliberately ignorant, she would have been spared a life of

incessant pain and suffering. Those of us who have that option can decide not to know. The logic of deliberate ignorance is to avoid the regret of knowing the worst possible outcome and to instead learn to live with uncertainty.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ethics Committee Max Planck Institute for Human Development. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

GG: Validation, Supervision, Project administration, Investigation, Formal analysis, Conceptualization, Writing – review & editing, Writing – original draft. RG-R: Writing – review & editing, Writing – original draft, Software, Methodology, Data curation.

References

- Admati, A., and Hellwig, M. (2013). *The bankers' new clothes: what's wrong with banking and what to do about it*. Princeton, NJ: Princeton University Press.
- Alexander, R. (1974). The evolution of social behavior. *Annu. Rev. Ecol. Syst.* 5, 325–383. doi: 10.1146/annurev.es.05.110174.001545
- Alfred, J. (2002). Flagging non-paternity. *Nat. Rev. Genet.* 3:161. doi: 10.1038/nrg757
- Anderson, K. G. (2006). How well does paternity confidence match actual paternity? *Curr. Anthropol.* 47, 513–520. doi: 10.1086/504167
- Anderson, K. G., Kaplan, H., and Lancaster, J. B. (2007). Confidence in paternity, divorce, and investment in children. *Evol. Hum. Behav.* 28, 1–10. doi: 10.1016/j.evolhumbehav.2006.06.004
- Aristoteles, (1953). *Metaphysics* (W. D. Ross, Trans.). Oxford, UK: Clarendon Press.
- Baker, R. R., and Bellis, M. A. (1995). *Human sperm competition: copulation, masturbation, and infidelity*. New York: Chapman & Hall.
- Barrot, C., Sánchez, C., Ortega, M., De Alcaraz-Fossoul, J., Carreras, C., Medallo, G., et al. (2014). DNA paternity testing in Spain without the mother's consent: the legal responsibility of the laboratories. *Forensic Sci. Int. Genet.* 8, 33–35. doi: 10.1016/j.fsigen.2013.06.016
- Carnap, R. (1969). *The logical structure of the world* (R. A. George, Trans.). Berkeley, CA: University of California Press. (Original work published 1928).
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd Edn. New York: Erlbaum Associates.
- Daly, M., and Wilson, M. (2006). "Selfish genes and family relations" in Richard Dawkins: *How a scientist changed the way we think*. eds. A. Grafen and M. Ridley (Oxford, UK: Oxford University Press), 191–202.
- Gaulin, S. J. C., McBurney, D. H., and Brakeman-Wartell, S. L. (1997). Matrilateral biases in the investment of aunts and uncles. *Hum. Nat.* 13, 391–402. doi: 10.1007/s12110-002-1022-5
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., and Woloshin, S. (2007). Helping doctors and patients to make sense of health statistics. *Psychol. Sci. Public Interest* 8, 53–96. doi: 10.1111/j.1539-6053.2008.00033.x
- Gigerenzer, G., and Garcia-Retamero, R. (2017). Cassandra's regret: the psychology of not wanting to know. *Psychol. Rev.* 124, 179–196. doi: 10.1037/rev0000055
- Good, I. J. (1967). On the principle of total evidence. *Br. J. Philos. Sci.* 17, 319–321. doi: 10.1093/bjps/17.4.319
- Hamilton, W. (1964). The genetical evolution of social behavior. *Science* 156, 477–488. doi: 10.1126/science.156.3774.477
- Hertwig, R., and Ellerbrock, D. (2022). Why people choose deliberate ignorance in times of societal transformation. *Cognition* 229:105247. doi: 10.1016/j.cognition.2022.105247
- Hertwig, R., and Engel, C. (2020). (eds.) *Deliberate ignorance: Choosing not to know*. Cambridge, MA: MIT Press.
- Hipp, D. (2007). Kuckucksei im Nest. *Der Spiegel*. Nr. 7. Available at: <http://www.spiegel.de/spiegel/print/d-50503694.html>
- Hogarth, R. (1987). *Judgment and choice*. 2nd Edn. New York: Wiley.
- Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291. doi: 10.2307/1914185
- Kruglanski, A. W., and Webster, D. M. (1996). Motivated closing of the mind: seizing and freezing. *Psychol. Rev.* 103, 263–283. doi: 10.1037/0033-295X.103.2.263
- Locke, J. (1690/1953) in *An essay concerning human understanding*. ed. A. C. Fraser (New York: Dover). (Original work published 1690)
- Luce, R. D., and Raiffa, H. (1957). *Games and decisions*. New York: Dover.
- McGoey, L. (2014). *An introduction to the sociology of ignorance: essays on the limits of knowing*. Abingdon, UK: Routledge.
- Miller, G. A. (1983). "Informavores" in *The study of information: interdisciplinary messages*. eds. F. Machlup and U. Mansfield (New York: Wiley-Interscience), 111–113.
- Proctor, R. N., and Schiebinger, L. (2008). *Agnotology: the making and unmaking of ignorance*. Stanford, CA: Stanford University Press.
- Rizzo, M. J., and Whitman, G. (2020). *Escaping paternalism*. New York: Cambridge University Press.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We would like to thank Ralph Hertwig and Rona Unrau for helpful comments.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Strindberg, J. A. (2014). *The father* (E. Oland and W. Oland, Trans.) (Original work published 1887)
- Sweeny, K., Melnyk, D., Miller, W., and Shepperd, J. A. (2010). Information avoidance: who, what, when, and why. *Rev. Gen. Psychol.* 14, 340–353. doi: 10.1037/a0021288
- Thornton, R. L. (2008). The demand for, and impact of, learning HIV status. *Am. Econ. Rev.* 98, 1829–1863. doi: 10.1257/aer.98.5.1829
- Trivers, R. L. (1974). Parent-offspring conflict. *Am. Zool.* 14, 249–264. doi: 10.1093/icb/14.1.249
- Voracek, M., Haubner, T., and Fisher, M. L. (2008). Recent decline in nonpaternity rates: a cross-temporal meta-analysis. *Psychol. Rep.* 103, 799–811. doi: 10.2466/pr0.103.3.799-811
- Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., Mcguire, A., et al. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452, 872–876. doi: 10.1038/nature06884
- Wolf, M., Musch, J., Enczmann, J., and Fischer, J. (2012). Estimating the prevalence of nonpaternity in Germany. *Hum. Nat.* 23, 208–217. doi: 10.1007/s12110-012-9143-y



OPEN ACCESS

EDITED BY

Riccardo Viale,
University of Milano-Bicocca, Italy

REVIEWED BY

Elisabetta Pisanu,
International School for Advanced Studies
(SISSA), Italy
Kaileigh Byrne,
Clemson University, United States

*CORRESPONDENCE

Leyla Loued-Khenissi
✉ leyla.loued-khenissi@chuv.ch

RECEIVED 15 January 2024

ACCEPTED 26 August 2024

PUBLISHED 19 September 2024

CITATION

Loued-Khenissi L and
Corradi-Dell'Acqua C (2024) Gambling on
others' health: risky pro-social
decision-making in the era of COVID-19.
Front. Psychol. 15:1370778.
doi: 10.3389/fpsyg.2024.1370778

COPYRIGHT

© 2024 Loued-Khenissi and
Corradi-Dell'Acqua. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Gambling on others' health: risky pro-social decision-making in the era of COVID-19

Leyla Loued-Khenissi^{1,2*} and Corrado Corradi-Dell'Acqua^{1,3}

¹Theory of Pain Laboratory, Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland, ²Department of Clinical Neuroscience, University Hospital of Lausanne, Lausanne, Switzerland, ³Center for Mind/Brain Sciences, University of Trento, Rovereto, Italy

Introduction: In the early days of the COVID-19 pandemic, individuals were asked to perform costly actions to reduce harm to strangers, even while the general population, including authorities and experts, grappled with the uncertainty surrounding the novel virus. Many studies have examined health decision-making by experts, but the study of lay, non-expert, individual decision-making on a stranger's health has been left to the wayside, as ordinary citizens are usually not tasked with such decisions.

Methods: We sought to capture a snapshot of this specific choice behavior by administering two surveys to the general population in the spring of 2020, when much of the global community was subject to COVID-19-related restrictions, as well as uncertainty surrounding the virus. We presented study participants with fictitious diseases varying in severity that threatened oneself, a loved one or a stranger. Participants were asked to choose between treatment options that could either provide a sure, but mild improvement (sure option) or cure the affected person at a given probability of success (risky option).

Results: Respondents preferred gambles overall, but risk-seeking decreased progressively with higher expected severity of disease. This pattern was observed regardless of the recipient's identity. Distinctions between targets emerged however when decisions were conditioned on a treatment's monetary cost, with participants preferring cheaper options for strangers.

Discussion: Overall, these findings provide a descriptive model of individual decision-making under risk for others; and inform on the limits of what can be asked of an individual in service to a stranger.

KEYWORDS

risk, decision-making, other-regarding behavior, COVID-19 pandemic, uncertainty

1 Introduction

In December 2019, an unprecedented outbreak of pneumonia caused by the SARS-CoV-2 virus, emerged in the city of Wuhan (China). This disease, known as COVID-19, rapidly spread throughout the globe, finding authorities, healthcare providers and lay individuals woefully unprepared, leading to a high degree of uncertainty (Loretto et al., 2021). By March 2020, many countries had put restrictive measures in place such as lockdowns, curfews, social distancing and quarantines to contain the virus' spread. These measures had economic and psychological consequences on the people involved (Kunzler et al., 2021; Lima et al., 2020; Nochaiwong et al., 2021; Santomauro et al., 2021; Wu et al., 2021), with some individuals more affected than others (Adams-Prassl et al., 2020; Blundell et al., 2020; Daly et al., 2020). At the same time, the effect of the virus was not uniformly distributed, changing as a function of age and pre-existing medical

conditions (Du et al., 2020; Yancy, 2020). As such, the early days of the pandemic presented a stark dilemma to lay-people: how to act for the sake of another in the face of uncertainty, and at what cost? This problem spilled over into socio-political discourse (Barbieri and Bonini, 2020; McKee and Stuckler, 2020; Wolff, 2022) and even prompted acts of violence (Choi and Lee, 2021; Elfrink, 2020; Taylor and Asmundson, 2021). But the pandemic also offered a real-time opportunity to assess costly, individual, pro-social, risky decision-making in instances where even authorities were subject to uncertainty.

1.1 Economic decision-making under risk

Uncertainty has been extensively studied in economic decision-making (Johnson and Busemeyer, 2010; Loued-Khenissi and Preuschoff, 2020; Preuschoff et al., 2013) commonly in the form of “risk,” an expected uncertainty based on event probability. Within this framework, an event is deemed riskier the more unpredictable it is. Hence, high risk differs from predictable danger, which refers to conditions where individuals can easily foresee what will occur. Theoretical and empirical accounts within the domain of economic decision-making (Allais, 1953; Bernoulli, 1738/1954; Cox and Sadiraj, 2008; Tversky and Kahneman, 1992) show that risk steers decision-making away from maximizing predicted gains: agents are risk-averse in the gain domain, showing a preference for sure options, while they are risk-seeking when facing loss. Importantly, risk-seeking decreases linearly with expected monetary loss, a robust phenomenon that has been replicated across different countries (Huck and Müller, 2012; Ruggeri et al., 2020). It remains unclear whether such risk preferences are conserved when deciding for others.

Several empirical studies (Andreoni and Miller, 2002; Cutler and Campbell-Meiklejohn, 2019; Rand and Nowak, 2013) as well as human social structure (Tomasello et al., 2012), provide evidence that other-oriented decisions often violate the *homo economicus* model (Samuelson, 1993). Individuals commonly engage in costly actions to cooperate with others (Diekmann, 2004; Fehr and Fischbacher, 2004) or to help people in need (Soetevent, 2005), albeit by investing less resources than those usually mobilized for one's own benefit (Lockwood et al., 2017). Furthermore, several studies have found that decisions made on behalf of others resemble those made for the self (Civai et al., 2010, 2012; Corradi-Dell'Acqua et al., 2013, 2016), thus suggesting an individual ability to put oneself in the shoes of unknown others. When looking specifically at risk preferences, a recent meta-analysis reveals a strong variability in the effects described in the literature, all converging toward an overall trend of slightly enhanced risk-seeking for others relative to one's self (Polman and Wu, 2020).

1.2 Health decision-making under risk

The role of uncertainty in decision-making also extends to health choices (Reis and Spencer, 2019). In this context, experts, such as authorities or physicians specifically trained in assessing risk, are usually those tasked with making decisions. For instance, authorities may rely on the Precautionary Principle, assuming a risk-averse stance toward public health (Goldstein, 2001; Gollier and Treich, 2003; Sunstein, 2005). Other health evaluations commonly use expected utility (Abellan-Perpiñan et al., 2009; Evans and Viscusi, 1991; Levy

and Nir, 2012; Meltzer, 2001; Russell and Schwartz, 2012). For example, the Quality-Adjusted Life-Years (QALY) is a form of expected utility integrating time that guides cost analysis (Bleichrodt, 1997) and health policy implementations (Mooney, 1989; Pinto-Prades et al., 2019). With both the Precautionary Principle and QALYs, the calculus employed in the service of making a decision is explicit. However, the novelty and virulence of Sars-Cov-2 imposed the burden of costly decision-making on people's health under uncertainty on lay people, leaving the question open as to what strategy is used in such a context. Tversky and Kahneman (1981) speculated that individuals would be risk-seeking for treatment when facing an infectious disease, mirroring choice behavior when facing monetary loss. However, whereas some studies investigating pain-management choices confirm this prediction (Loued-Khenissi et al., 2022), others show that individuals are risk-averse for their own treatment options (Hellinger, 1989) or for choices that could influence their life expectancy (Attema et al., 2013). Finally, meta-analyses suggest that health decision-making in medical contexts lead to a shift toward more cautious (risk-averse) decisions for others relative to the self, in contrast to what is found in monetary/managerial scenarios (Atanasov, 2015; Polman and Wu, 2020). However, the medical contexts included in these meta-analyses involved primarily physicians (or people acting as physicians) choosing for patients (Atanasov, 2015; Polman and Wu, 2020), thus still focusing on decision-making processes in professionals, rather than the lay individual. To the best of our knowledge no studies have tested how ordinary people make risky decisions on disease treatments for unknown others.

1.3 The present research

Here, we investigated uncertainty's role in costly decision-making for people's health by applying an expected value model of disease severity in a probabilistic task. We administered two anonymous surveys to the general population between May and July, 2020, when at least half the global population was under confinement and grappling with questions on how to personally respond to the COVID-19 pandemic and how much to sacrifice in service of the greater good. We presented respondents with fictitious diseases and their associated risks of contraction, along with different, costly treatment options. Participants were asked to choose between a treatment that avoids contracting the disease at a given probability (risky option) or one that mitigates symptoms with 100% effectiveness (sure option). Building on a well-established literature from economic decision-making (Johnson and Busemeyer, 2010; Loued-Khenissi and Preuschoff, 2020; Preuschoff et al., 2013), we define a treatment as riskiest when outcome (negatively or positively valenced) probability approached 0.5. Respondents made decisions for themselves (*Self*), a loved one (*Beloved*), and a *Stranger*. In this design, the *Self* condition represents the baseline against which we compare choices made for others (either the *Beloved* or a *Stranger*). In particular, choices made for an unknown person (relative to one's self) are of key interest, as they provide a snapshot on individual, costly, risky decisions for a stranger's wellbeing. We included a loved one as a target to further characterize self-other differences, and to investigate effects related to the social proximity of the deciding agent. Following the literature reviewed above, we sought to test the following three hypotheses. First, we expect that, in the *Self* condition, individuals would display

TABLE 1 Surveys 1 and 2 cohort details.

	Survey 1	Survey 2	Comparison
Gender	43% females	64% females	$\chi^2 = 20.45^{***}$
Participant's age	25 (iqr = 10)	26 (iqr = 10)	$t = -1.29$
Beloved's age	30 (iqr = 29)	32 (iqr = 28)	$t = -0.49$
# countries represented	42	44	$\chi^2 = 0.06$
Monthly income	\$2,499 (iqr = \$3250)	\$3,499 (iqr = \$6500)	$t = -4.25^{***}$
Pandemic-related monetary loss	\$200 (iqr = \$1000)	\$3.5 (iqr = \$2000)	$t = 0.99$
Perceived adequacy of confinement measures (<i>most frequent response</i>)	Adequate (69.60%)	Adequate (63.75%)	$\chi^2 = 1.67$
Job-loss due to COVID-19	27	16	$\chi^2 = 0.44$
Positive to COVID-19	4	2	$\chi^2 = 0$

Continuous variables are described in terms of median and inter-quartile range (iqr). Group differences are estimated in terms of χ^2 test (for proportions) or independent sample t -test (for continuous variables). Significant effects are highlighted in bold, with *** corresponding to $p < 0.001$.

an overall risk-seeking stance that progressively declines with increasing expected disease severity (*Hypothesis 1*). This prediction is directly derived from studies arguing that choices on disease contraction mirror those observed for monetary losses (Tversky and Kahneman, 1981). Second, we predict that decisions for strangers would be more risk-averse than those associated with the self (*Hypothesis 2*). This is motivated by previous studies meta-analyses testing risk decision-making in medical contexts (Atanasov, 2015), and pain management (Loued-Khenissi et al., 2022). Third, and consistent with our prior research (Loued-Khenissi et al., 2022), we expect social proximity between self and other to influence the results, with agents acting for their loved ones as they would for themselves (*Hypothesis 3*).

2 Survey 1

2.1 Methods

2.1.1 Population

The survey was made available online to individuals 18 years and older. Respondents were recruited through the Prolific.co platform.¹ Participation was voluntary and compensated between £1.5 and £2 (i.e., £5/h on an average completion time of 22 min). The Ethics Committee of the Faculty of Psychology and Educational Sciences of the University of Geneva approved the study.

Survey 1 was an exploratory investigation to obtain a first snapshot of decision-making for others' well-being. Within a week (in May 2020), 381 participants in the Prolific.co platform began the survey, and 366 completed the questionnaire. We excluded an additional 22 participants for making attentional errors (see data analysis below for more details) or providing implausible answers (e.g., listing a non-human as a loved one), leaving $n = 344$ for analysis [43% F; mean age 25 (IQR = 10)]. Cohort characteristics are described in Table 1.

2.1.2 Procedure

As a first step, participants accessed an informed consent page. By selecting the option "I accept," they were then directed to the main

survey. This was an adaptation of standard lottery tasks from economic decision making (e.g., Tversky and Kahneman, 1992) to the context of a pandemic, where respondents had to choose between different treatments that could either dampen or cure disease at given effectiveness rates. We specifically chose scenarios involving treatments for fictitious diseases, so as to freely manipulate probabilities in a plausible fashion. This manipulation would not have been possible had we employed real diseases (which participants might have prior knowledge of), or non-pharmacological protective behavior like wearing masks or abiding to self-confinement (for which precise probabilities might have appeared implausible). The survey contained 45 items, each describing a risk of contracting a given disease. For each item, respondents chose one of 3 treatment options. Below is an example:

"Your loved one has a 25% chance of contracting and falling severely ill with disease F. Symptoms include high fever, muscle pain and vomiting of blood. Standard treatment requires a two-week hospital stay in intensive care. The illness leaves minor but lasting cardiac deficits and a slight but permanent hearing impairment. You can:

- *Do nothing and let your loved one face the initial chance of falling severely ill with disease F*
- *Pay half your monthly salary for additional treatment that will certainly reduce the severity of the illness, such that it leaves only a slight but permanent hearing impairment*
- *Pay one tenth of your monthly salary to halve the risk of contracting the illness altogether with an experimental treatment."*

The scenarios described 5 possible diseases, with different risks of contraction (p_D ; in the example above, 0.25), and levels of severity (S_D ; either death or severe lasting deficits). Diseases were also described according to their symptoms, which were loosely based on real world infectious diseases (C for Chikungunya; D for Dengue; E for Ebola; F for Flu, etc.), with, however, fictitious morbidity and mortality rates. Each item threatened one of three possible targets (*Self*, *Beloved*, or *Stranger*), and were followed by 3 possible options:

- Do nothing and let the person face the initial risk of disease;
- Pay an amount of money for a known treatment that partially reduces disease severity (sure option);
- Pay an amount of money for an experimental treatment that reduces the risk of initial contraction (the gamble).

¹ <https://www.prolific.co/>

TABLE 2 Disease items and treatment options.

	Disease	Prognosis (SD)	Contraction risk (pD)	Sure option prognosis	Gamble success
Survey 1	C	Death	5%	Minor, lingering cardiac deficits and slight, but permanent hearing impairment	75%
	D	Death	25%	Minor, lingering respiratory deficits and slight, but permanent visual impairment	75%
	E	Death	5%	Minor, lingering respiratory deficits and slight, but permanent visual impairment	50%
	F	Minor but lasting cardiac deficits, and a slight but permanent hearing impairment	25%	Slight but permanent hearing impairment	50%
	M	Minor but lasting respiratory deficits and a slight but permanent visual impairment	10%	Slight but permanent visual impairment	75%
Survey 2	C	Death	5%	Minor, lingering cardiac deficits and slight, but permanent hearing impairment	75%
	D	Death	25%	Minor, lingering respiratory deficits and slight, but permanent visual impairment	75%
	P	Minor but lasting cardiac deficits and a slight but permanent hearing impairment	50%	Slight but permanent hearing impairment	50%
	M	Minor but lasting respiratory deficits and a slight but permanent visual impairment	10%	Slight but permanent visual impairment	75%

The gamble cost was always set to one tenth of the respondent's monthly income. Sure treatment options for each disease varied in price, between 0.1, 0.5, or 1 unit of the respondent's monthly income. Selecting the sure option treatment for diseases with a mortal risk (C, D, and E from Table 2) reduced prognosis to severe lasting deficits. Known treatment reduced prognosis to mild lasting deficits for diseases that carried severe lasting deficits (F and M) (see Table 2). Four attentional catch questions were interspersed across the survey to ensure respondent engagement (e.g., "Is 7 > 3?"). All items were randomized across diseases and targets.

We also collected participants' non-identifying demographic information (country of residence, age and gender, and household monthly income and education), that could impact costly, pro-social decision-making (Boschini et al., 2018; Freund and Blanchard-Fields, 2014; Wiepking and Breeze, 2012). We also asked participants to identify a loved one (*Beloved*) by their role. The survey was designed using LimeWire software, and was fully anonymous; and available in English, French and Italian (English version available under the Open Science Framework <https://osf.io/9fjdq/>).

2.1.3 Data analysis

Analyses were performed using R 4.1.2 freeware software;² de-identified data and processing scripts are available at: <https://osf.io/9fjdq/>.

2.1.3.1 Expected (dis)utility of disease and risk preferences

We modeled responses to disease vignettes using expected utility theory. Specifically, we computed the expected value of disease and associated treatment options to address the main question of decision-making under risk. Each illness presented an expected disutility of

disease severity (EDS), computed from the expected utility theorem (Bernoulli, 1738/1954) as follows:

$$EDS = p_D \cdot S_D.$$

where p_D is the probability of contraction and S_D is disease severity. Each disease has a specific p_D (ranging from 0.05 to 0.50, see Tables 1, 2). S_D is a value on an ordinal scale ranging from 1 to 3, where 3 = death; 2 = severe lasting deficits and 1 = minor lasting deficits (the latter case was never used as a starting value, but only as the outcome of the sure option treatment).

2.1.3.2 Linear models

We analyzed participants' choices using a generalized linear mixed model with binomial distribution and Laplace approximation, to examine factors influencing decisions. First, we assessed the likelihood of making a costly choice (either certain or gamble) as opposed to inaction (Model 1a). Subsequently, we assessed, among costly choices, the likelihood of selecting a gamble over a sure option (Model 1b). In both models, we specified *EDS*, the sure option's *Price* (0.1, 0.5 and 1 unit of monthly income), disease *Target* (*Self*, *Beloved* and *Stranger*), and the interaction thereof, as fixed factors. *EDS* and *Price* were specified as continuous predictors, whereas *Target* was treated as a categorical factor with three levels. Finally, we designed a linear mixed model to fit the treatment cost (i.e., the chosen treatment prices) as a function of the fixed factors *EDS*, *Target* and their interactions (Model 1c). In all three models, participant identity was specified as a random factor, with random intercept and slope for all within-subject predictors. In modeling the random components, we always chose the most complex random structure (slope of simple effects and high order interactions), except in cases of misconvergence, where a simpler structure was adopted (full details on the models implemented are provided in Supplementary Table A1). The analysis was performed using the *lmerTest* package of R (Kuznetsova et al., 2017).

² <https://cran.r-project.org/>

2.2 Results

2.2.1 Choice analysis

Overall, costly actions were significantly more frequent than inactions [76.53%, 26.67 interquartile range (IQR); test against 50%: $t_{(343)} = 25.47$, $p < 0.001$]. This effect was driven by *Self* and *Beloved* conditions, where costly actions were chosen most often (88.06%, IQR: 13.33 and 93.72%, IQR: 6.66, respectively). When choosing for a stranger (*Other* condition), costly options were selected less often, ~47.81% (IQR = 87.67; $t_{(343)} = -1.03$, $p = 0.300$; see Figure 1A). Among costly actions, participants preferred gambles (56.04%, IQR: 40.96; $t_{(341)} = 4.25$, $p < 0.001$). This effect was driven by *Self* and *Other* conditions, where gambles occurred 56.00% (IQR: 51.83) and 71.50% (IQR: 45.63) of the time, respectively. For the *Beloved*, gambles were chosen ~48.22% (IQR: 50.26; $t_{(340)} = -1.10$, $p = 0.273$) of the time.

We extended our analysis in a generalized linear mixed model with a binomial distribution to assess factors affecting choice (Model 1a; Model 1b; Table 3). First, we found a positive effect of *EDS* on costly (Model 1a) and sure choices (Model 1b) (see Figure 1C; Table 3). We also found an effect of *Price*, with preferences for the cheaper option (Inaction in Model 1a; Gambles in Model 1b) as the sure option's price increased; and an effect of *Target*, with fewer costly (Model 1a) and sure choices (Model 1b) made for the *Other* (see Figures 1A–C). In Model 1b, when participants chose a costly

option, they gambled less for the *Beloved*. Finally, the three factors of interest interacted with one another, suggesting *EDS* and *Price* effects were conditioned on the *Target*. To further explore these interactions we repeated the previous models in each *Target* separately (Figure 2). Results confirm both *EDS* and *Price* influence self-regarding decisions in opposite directions. Whereas *EDS* promotes costly (Model 1a) and sure (Model 1b) choice selection, *Price* promotes inaction and gambles (Figure 2, left column). *Price* influences decisions for the *Beloved* less (Figure 2, middle column) while decisions for the *Other* are less influenced by *EDS* and more by *Price* (Figure 2, right column).

2.2.2 Chosen treatment price

In a follow-up model, the cost of the selected option was modeled as a dependent variable. Results (see Table 3 and Figure 3A) confirmed that participants spent more on the *Beloved*, but less on the *Stranger*. Furthermore, participants spent more for high *EDS* across targets, but this effect was less pronounced for the *Stranger*.

2.2.3 Follow-up analyses

2.2.3.1 Nuisance variables

We repeated all above analyses by accounting for *Sex*, *Age*, *Monthly Income* and COVID-19 information (e.g., log-transformed USD financial loss) as nuisance variables. Results confirm effects

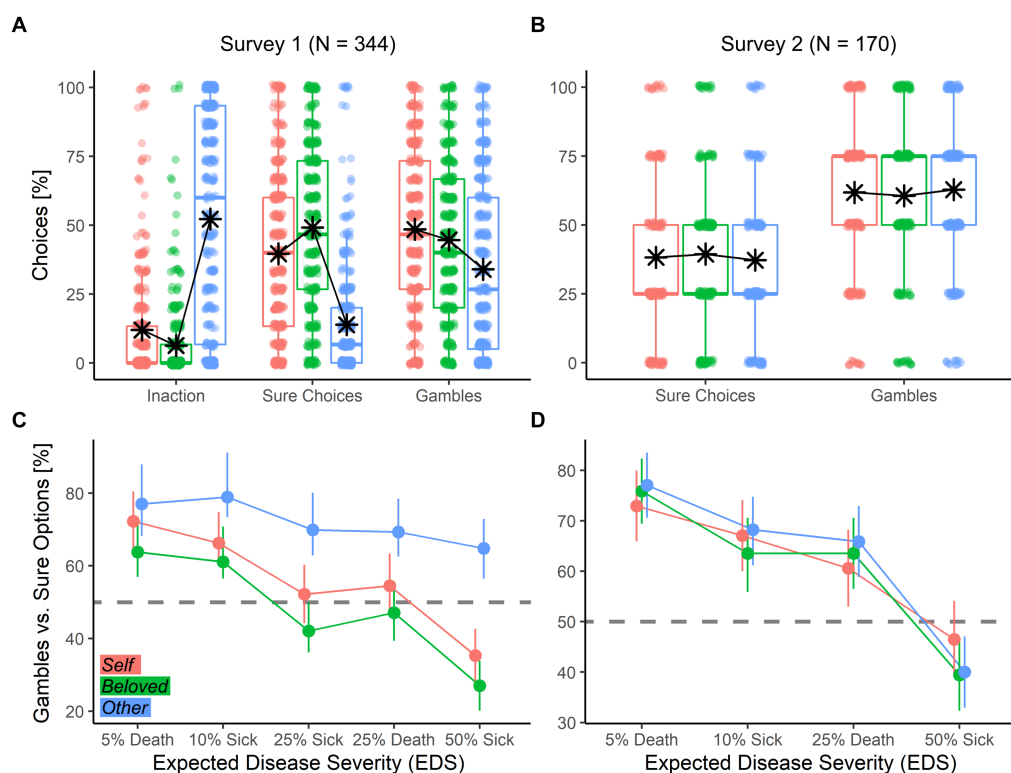


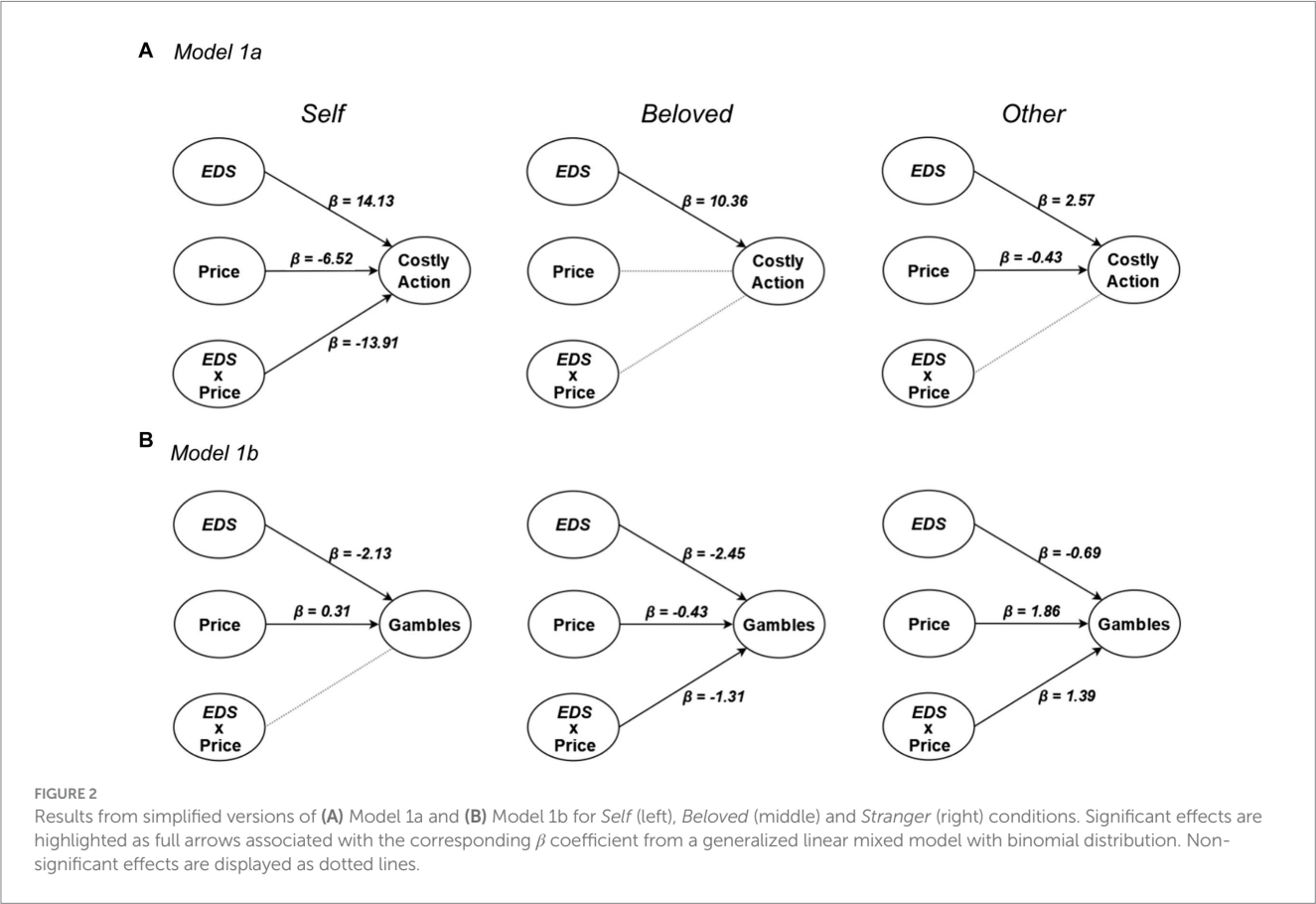
FIGURE 1

(A,B) Boxplots describing the percentage of each kind of choice across decision targets. For each boxplot, the horizontal line represents the median value of the distribution, the star represents the average, the box edges refer to the inter-quartile range, and the whiskers to the data range within 1.5 of the inter-quartile range. Individual data-points are also displayed as dots. (C,D) Line-graphs describing the relative percentage of Gambles vs. Sure Options across EDS and Target. Each condition is represented by the overall mean with bootstrap-based 95% confidence intervals. The horizontal dashed gray line shows the indifference point.

TABLE 3 Results of Survey 1.

Predictor	Model 1a		Model 1b		Model 1c	
	(Act. vs. inact.)		(Gamble vs. sure)		(Treatment cost)	
	β	Z	β	Z	β	t
Intercept	9.07	10.30***	0.38	2.59**	0.26	-33.61***
EDS	12.97	6.94***	-2.29	-12.44***	0.13	14.83***
Price	-0.73	-2.35*	0.32	2.31*	-	-
Target <i>Beloved</i>	-0.89	-0.80	-0.71	-8.23***	0.06	10.52***
Target <i>Stranger</i>	-8.94	-9.99***	1.64	9.16***	-0.15	-19.98***
EDS*Price	-2.06	-3.13**	0.25	0.75	-	-
EDS*Target <i>Beloved</i>	-3.90	-1.61	-0.22	-1.16	~0	-0.07
EDS*Target <i>Stranger</i>	-10.50	-5.63***	1.43	5.53***	-0.09	-7.60***
Price*Target <i>Beloved</i>	1.07	2.34*	-0.84	-4.06***	-	-
Price*Target <i>Stranger</i>	0.39	1.16	1.69	4.54***	-	-
EDS*Price*Target <i>Bel.</i>	3.21	3.15**	-1.47	-2.84***	-	-
EDS*Price*Target <i>Str.</i>	1.58	2.19*	1.04	1.63	-	-

For each model, each fixed factor is described in terms of β coefficient and a statistical test (Z for binomial models, t for linear models) testing potential deviations from 0. Significant effects are highlighted in bold, with *** corresponding to $p < 0.001$, ** to $p < 0.01$, and * to $p < 0.05$.



observed in the main analysis (Supplementary Table A2). We also found a significant effect of Sex, where males chose gambles less frequently (Model 1b) and accepted higher treatment prices (Model 1c).

2.2.3.2 Alternative approach to EDS

All analyses reported were performed by modeling the predictor EDS, an adaptation of expected utility theory to health-based decision-making. In particular, EDS is defined as the product of the probability

of contraction (p_D) with disease severity (S_D), where the latter is treated as a ratio-value although resulting from an ordinal predictor (e.g., 3 = death; 2 = severe lasting deficits and 1 = minor lasting deficits; Methods). As it could be argued that imposing linearity on S_D biases the analysis, we repeated the main analyses above, this time modeling the severity of disease as $p_D + S_D$, as two independent predictors. Here, p_D was specified as a continuous predictor, and S_D as a categorical factor (in all diseases, initial S_D is either 2 or 3). Full results are displayed in [Supplementary Table A3](#), and reveal that the effects originally attributed to EDS are now associated with p_D . In some instances (albeit non-systematically) participants' choices were also influenced by S_D .

2.3 Discussion

This survey tested individual risky decision-making on other's health using an expected disutility of disease framework. We found evidence supporting *Hypothesis 1* as individuals prefer gambles in the disease domain overall, with this preference decreasing linearly with expected severity of disease. This prediction was confirmed by our data ([Figures 1A–C](#), rose data-points), similar to what found in the economic loss domain (Tversky and Kahneman, 1981) and pain management (Loued-Khenissi et al., 2022). More specifically, EDS and treatment price heavily influenced participants' choices: whereas a disease of higher expected severity increased sure option selection, higher treatment cost increased gamble selection. Our prediction that individuals would be more risk averse for unknown others (*Hypothesis 2*) was not observed in our data, as participants selected sure options less frequently when acting for the *Stranger* ([Figures 1A–C](#), blue data-points). Finally, *Hypothesis 3* predicted that choices for the *Beloved* would differ from those made for the *Stranger*, and be more similar to those observed for the *Self*. This was confirmed in our data, with increased risk aversion in the *Beloved* relative to the *Stranger* condition ([Figures 1A–C](#), green data-points). In addition, participants exhibited the most risk-aversion when choosing for a loved one. The *Target* also affected the role played by EDS and Price on the decision, with choices for the *Beloved* more strongly influenced by EDS, while those for the *Other* were primarily price-based ([Figure 2](#)).

Finally, although our main analysis was framed on the estimation of a (dis)utility score for the disease (EDS), the results were not conditional to this choice. Similar effects were also observed when modeling the raw probability of disease contraction (p_D).

3 Survey 2

In Survey 2, we explicitly differentiated risk preferences from cost concerns. For this purpose, we devised a modified version of Survey 1, where participants chose between sure and risky treatment options, and were subsequently asked to bid (in their own currency) on their chosen option. This measure of willingness-to-pay (WTP) has previously been used to value health interventions (Olsen and Smith, 2001). Importantly, removing the factor "price" made for a shorter survey of only 15 scenarios (5 diseases * 3 Targets). We further shortened the questionnaire to 12 scenarios (4 diseases * 3 Targets) to a survey that could be filled in ~15 min.

3.1 Methods

Unless otherwise stated, the set-up and analysis of Survey 2 were identical to those of Survey 1.

3.1.1 Population

In Survey 2, respondents were not remunerated and were recruited through social media (Facebook, Twitter and LinkedIn) and survey swapping platforms.³ Within a time-window of 2 months (June–July 2020), 273 participants began the survey, and 175 completed the questionnaire. Five participants were excluded from analysis for providing implausible answers, leaving a sample of $n = 170$ for analysis. Cohort 2 was comparable to Cohort 1 for age (of both respondents and chosen *Beloved*), education and number of countries represented ([Table 1](#)). However, the cohorts differed on sex ratio and income (Survey 2 included more females and respondents reported a higher income).

3.1.2 Procedure

Survey 2 included 12 scenarios with 4 diseases ([Supplementary Table A1](#)) affecting one of 3 possible targets (*Self*, *Beloved* and *Stranger*). Participants were asked to select between a sure option and a gamble, as in Survey 1. Respondents were not allowed to forgo action. Furthermore, no cost was associated with the options. Following choice, participants were asked to name a price [Willingness to Pay (WTP)], in their own currency for the chosen option. Respondents were explicitly told they could enter a value of 0 if they wished. Given its brevity, Survey 2 did not include any catch trials to assess attention.

3.1.3 Data analysis

As in Survey 1, we first performed a choice analysis (Model 2a) testing whether EDS, *Target* or their interaction affected choice. Then we examined (WTP) as a dependent variable (Model 2b), with EDS, *Target*, previous *Choice* (Gamble vs. Sure Option) and their interaction as predictors of interest ([Supplementary Table A1](#)). The WTPs were converted from participants' local currency to USD (based on the official exchange rate on the day of their response), and log-transformed to account for the large range in responses.

Each model was associated with a power analysis to test whether the current design at a given sample size would be sufficiently powered to replicate findings from Survey 1. Estimates of fixed factors coefficients and random-effect terms for Model 2a were obtained by re-analyzing Models 1b from Survey 1 without the factor "price." As for Model 2b, we took the coefficients/terms obtained from the analysis of treatment price (Model 1c), although this model provides only partial information as no factor "choice" was specified in Survey 1. For each model, and each main/interaction effect, we ran 1,000 Monte-Carlo simulations aimed at replicating the same fixed factors coefficients and random-effect terms observed in Survey 1 on the design and sample size from Survey 2. Power was then estimated from the frequency of significant effects from the simulated data, as implemented in the *simr* package of R (Green and MacLeod, 2016). This analysis showed that the design and sample size were sufficiently

³ <https://www.surveycircle.com/>

sensitive to replicate the effects of *EDS* and *Target* observed in Model 1b from Survey 1 with a probability of at least 0.88.

3.2 Results

3.2.1 Choice analysis

As in Survey 1, participants preferred gambles over sure options (62.72%, IQR: 25.00; test against 50%: $t_{(169)} = 6.18$, $p < 0.001$), an effect observed in all three targets with comparable percentages. We further inspected choice preferences through a generalized linear mixed model, under binomial distribution (Model 2a). Results confirmed the same effect of *EDS* observed in Survey 1 (Table 4), where gambles decreased with increasing *EDS* (Figure 1D). We found no effect of *Target*. Overall, the analysis of Survey 2 revealed that when risk preferences are dissociated from cost, *Target* effects disappear.

3.2.2 Willingness-to-pay

In contrast to risk preferences, *Target* influences *WTP*, with participants bidding less on the *Stranger* (Table 4; Figure 3B). Additionally, we found an effect of previous *Choice*, with participants bidding more on sure options than gambles (Figure 3C). However, only 24.71% of trials listed a 0 bid for the *Stranger*, indicating a persistence of altruism and prosocial motivation.

3.2.3 Follow-up analyses

3.2.3.1 Nuisance variables

We repeated analyses including Sex, Age, Monthly Income and COVID-19 information as nuisance variables. Results confirmed all effects observed in the main analysis (Supplementary Table A4), with the exception of the Choice effect from Model 2c ($\beta = -0.31$, $t_{(627.62)} = -1.70$, $p = 0.090$). When analyzing the effect of the nuisance variables, we found a significant positive effect of COVID-19 financial loss on *WTP* (Supplementary Figure A1), suggesting that participants

who sustained a higher financial loss due to the pandemic were willing to pay more for others. This was observed by specifying both monthly income and financial loss in the same model, indicating this effect was not confounded with personal wealth.

3.2.3.2 Alternative approach to EDS

As in Survey 1, we repeated the analyses of the main models by replacing *EDS* with $p_D + S_D$, as two independent predictors. Full results are displayed in Supplementary Table A5, and reveal that all effects originally attributed to *EDS* are now associated with p_D . No effect was associated with S_D .

3.3 Discussion

Survey 2 confirms both *Hypothesis 1* and the first result from Survey 1 in that, in the *Self* condition, individuals display risk-seeking behavior in the disease domain, as highlighted by the preference toward gambles vs. sure options. Furthermore, this preference for gambles decreases linearly with *EDS*. However, when differentiating risk preferences from cost concerns, Survey 2 results show no *Target* difference, going against the predictions of *Hypotheses 2 and 3* (Figures 1B–D). Instead, target differences were observed only in the analysis of *WTP*, with participants bidding less to treat the *Stranger* (Figure 3B). This result disambiguates an open issue from Survey 1, suggesting target differences in other-regarding decision-making under risk are conditioned on cost considerations, and not risk preferences. Finally, while respondents preferred gambles overall, *WTP* analyses reveal a higher monetary value placed on sure options (Figure 3C).

4 General discussion

The goal of this study was to probe decision-making under risk in health interventions for self and others in the context of the COVID-19

TABLE 4 Results of Survey 2.

Predictor	Model 2a		Model 2b	
	(Gamble vs. sure)		(Willingness to pay)	
	β	Z	β	t
Intercept	0.71	4.57***	7.99	25.67***
EDS	−1.57	−4.75***	0.40	1.51
Choice	–	–	−0.36	−2.63**
Target <i>Beloved</i>	0.09	0.51	0.11	0.81
Target <i>Stranger</i>	0.08	0.50	−2.00	−7.40***
EDS*Choice	–	–	0.33	0.91
EDS*Target <i>Beloved</i>	−0.69	−1.48	0.18	0.50
EDS*Target <i>Stranger</i>	−0.48	−1.11	−0.67	−1.67
Choice*Target <i>Beloved</i>	–	–	0.15	0.85
Choice*Target <i>Stranger</i>	–	–	0.31	1.52
EDS*Choice*Target <i>Bel.</i>	–	–	−0.27	−0.57
EDS*Choice*Target <i>Str.</i>	–	–	0.04	0.08

For each model, each fixed effect is described in terms of β coefficient and a statistical test testing potential deviations from 0. Significant effects are highlighted in bold, with *** corresponding to $p < 0.001$, ** to $p < 0.01$, * to $p < 0.05$.

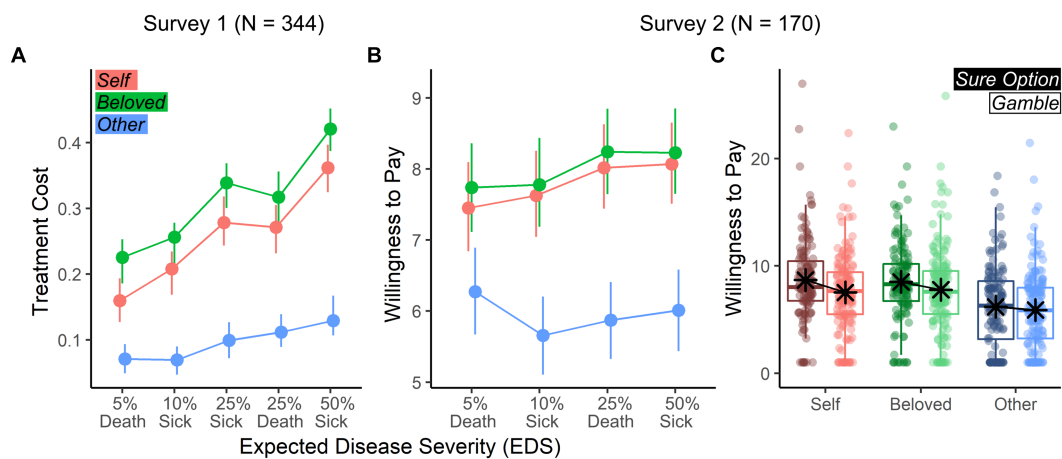


FIGURE 3

Treatment cost and willingness to pay. (A) Survey 1: Line-graphs describing the average cost (and bootstrap-based 95% confidence intervals) of the chosen treatment across EDS (horizontal axis) and Target (different color-coded lines). Costs are described as proportions of participants' monthly salary. (B) Survey 2: Line-graphs describing the average willingness to pay of the chosen treatment described as log-transformed USD units. (C) Survey 2: Boxplots describing the willingness to pay across Target and previous choice. Values (in log-transformed USD) are displayed in different colors (to discriminate Targets) and luminance (to discriminate Choice).

pandemic's early days. The study was specifically aimed at gaining a cross-sectional description of how a lay individual selects costly actions on behalf of another person's health, under risk. As authorities called on the public to act for others' sake during the pandemic, the burden of uncertain decision-making was thrust onto individual shoulders.

4.1 Individuals are risk-seeking for health treatments

Across two surveys, we confirm the first hypothesis of this study, according to which individuals prefer to gamble to prevent disease for themselves. This effect is in line with studies on different kinds of negative rewards, ranging from monetary loss (Tversky and Kahneman, 1981), to pain prevention (Loued-Khenissi et al., 2022). It is possible that the framing of decision outcomes in the present study as negative (getting sick) could have influenced the results toward a pattern similar to that of monetary losses (Tversky and Kahneman, 1981). In this perspective, an alternative framing with a positive outcome (being healed) could in principle lead to diverging results. However, surveys targeted the general population that, while in good health, was confronted with the risk of contracting COVID-19 (as the survey was conducted in the early days of pandemic, almost no-one of our participants contracted SARS-Cov-2 virus, Table 1). Casting the retention of improved health as a positive outcome in this population may have stretched responders' credulity. This may differ in patient populations, where treatments for different disease scenarios could be realistically framed as a positive shift from their present condition.

Critically however, though participants displayed an overall preference for gambles, they became progressively more risk-averse with higher expected disease severity. This finding is consistent with previous research in the domain of economic decision-making

(Tversky and Kahneman, 1981) and pain management (Loued-Khenissi et al., 2022). In addition to confirming our predictions, these results put forward the effectiveness of expected utility models in explaining health decision-making (Cohen, 1996). Although several studies criticize such an approach for health decisions (Abellan-Perpiñan et al., 2009; Dolan and Kahneman, 2008), we argue that expected utility is a useful tool for modeling individual behavior, in line with what is known on life quality (Attema et al., 2016), pain management (Loued-Khenissi et al., 2022), as well as brain activity (Knutson et al., 2005; Loued-Khenissi et al., 2020; Schultz, 2016). Importantly, our results are not idiosyncratic to the theoretical framework adopted in our study, as similar effects were obtained when replacing *expected disease severity* with the raw probability of disease contraction (p_D). Hence, absent any explicit requests to compute probabilistic outcomes, individuals in our study appear to choose according to those quantities nonetheless. In this perspective, concerns over individuals' difficulty in understanding probabilities, particularly in the context of the pandemic (Aguilar and Castaneda, 2021; Muñoz-Rodríguez et al., 2020) are not supported, and therefore authorities should consider informing the public in accurate, probabilistic terms (Kahlenberg et al., 2023).

Although individuals selected gambles more often, they simultaneously assigned a higher monetary value to sure treatment options (Figure 3C). This effect is known as *preference reversal* (Safra et al., 1990; Seidl, 2002), prevalent in economic frameworks but also observed in the domain of pain management (Loued-Khenissi et al., 2022). Although the cause of preference reversals is still debated, scholars attribute it to the sequential nature of many paradigms that distort pricing estimates, or to a general tendency to overprice options with high probability and low benefit at the expense of options with low probability and high benefit (see Seidl, 2002, for a review). Both these explanations fit the case of Survey 2, further stressing how choices in the context of disease prevention dovetail with predictions based on theories of economic decision-making.

4.2 Target differences are explained by cost considerations

In both surveys, participants' behavior differed as a function of disease target, especially for strangers. In Survey 1, respondents selected gambles more often for strangers than for themselves or their loved ones (Figure 1C). *Prima facie*, these results suggest a stronger risk-seeking stance for other-regarding decisions. However, Survey 1 results may also reflect the fact that the price associated with gambles in the survey was (1) stable across trials, and (2) cheaper or equal to that of sure options. Participants' behavior toward strangers was also characterized by a high amount of inactions (Figure 1A) where participants refrained from choosing to avoid incurring personal cost. In Survey 2, where choices were embedded in a cost-free context, individuals risk preference for others was the same as that observed for self-regarding behavior. Individuals diverged in action for themselves and unknown others only with respect to willingness-to-pay. We therefore propose that, when cost is not a factor in decision-making, risk biases do not have a differential impact on self and others. However, when cost is a factor, decisions differ between targets by pushing agents toward a value-based heuristic, where cheaper options are preferred for strangers.

It is unclear why, in the present study, risk preferences remain the same across the self-other boundary, something that contrasts with prior research that have found a dissociation (Atanasov, 2015; Loued-Khenissi et al., 2022; Polman and Wu, 2020). Two considerations emerge from this finding. First, results provide evidence that individuals deploy a simple self-referential strategy when computing uncertainty for others (at least in contexts that are cost-free and anonymous), possibly to minimize cognitive demand (Tomova et al., 2020). Second, although previous meta-analyses report overall self-other differences in risk preferences these effects are extremely variable between studies, pointing to a wide range of moderators that influence participants' choices (Atanasov, 2015; Polman and Wu, 2020). Among these are the framing of the context (e.g., involving positive vs. negative outcomes, financial vs. medical decisions) and, most critically, the identity of the target (adult, child, patient) and his/her personal relationship with the deciding agent (family member, colleague, stranger) (Atanasov, 2015; Polman and Wu, 2020). It is possible that any of these moderating factors (or the combination thereof) might have influenced our results with respect to prior literature.

Instead, the fact that target differences (across the self-stranger boundary) emerge only when monetary cost becomes a relevant parameter to an agent is consistent with current theoretical accounts. For instance, Lockwood et al. (2017) found that effortful (costly) pro-social choices triggered apathy, thereby suggesting that one main discriminant between self vs. others decision-making lies in resource mobilization. Most importantly, these results are also in line with predictions from evolutionary theory on kinship and indirect fitness (Kay et al., 2019), according to which costly behavior might be evolutionarily advantageous only when benefiting close ones, at the expense of strangers. It should be stressed, however, that strangers still received treatments that were more expensive than the cheapest option: in Survey 1, costly choices were chosen ~48% of the time, while in Survey 2 participants consistently made bids higher than 0. Based on these results, individual behavior for strangers does not mirror self-regarding behavior—but neither does it reflect purely or

even mostly selfish motivations. On the contrary, participants were willing to incur costly prosocial behavior even when kinship motives are absent, as for the *stranger* condition in our study. These altruistic tendencies have previously been found in several researches from behavioral economics (Fehr and Fischbacher, 2004; Fehr and Schmidt, 2006; Frey and Meier, 2004) and more generally in the field (Sisco and Weber, 2019). They offer a valuable insight into the boundaries one can expect from individuals for the sake of others' wellbeing, especially in scenarios with high uncertainty, rather than relying on heuristics that may backfire (Bonaccorsi et al., 2020; Fink et al., 2022; Wood et al., 2022) and compromise political trust (Jørgensen et al., 2022).

It could be argued that these donations are the result of a so-called *experimenter effect*, where participants are motivated by reputation concerns (Hoffman et al., 1996). The experimenter effect has been observed in tasks such as the dictator game, where anonymity between the participant and the experimenters decreased the amount of free donations (Hoffman et al., 1996). However, as our study guaranteed full anonymity, we believe that the risk of such confounds are negligible. Furthermore, participants' WTP in Survey 2 was positively influenced by real confinement-related financial loss (while controlling for personal wealth). This hints toward a genuine pro-social disposition held by respondents to provide for others' wellbeing, including strangers. These results are in line with previous environmental research measuring willingness-to-pay for options that benefit members of a future generation. Even in those scenarios where delay-discounting can dampen pro-social motivations, people exhibit a positive attitude toward others' wellbeing (Graham et al., 2019).

4.3 Choices for loved ones resemble those made for the self

Our third hypothesis predicted that individuals' behavior toward a stranger would differ from that made for a loved one, in that the latter would trigger decisions more similar to those made for the self. When considering risk preferences alone, we found no target differences, thus providing no support for our prediction. However, when taking into account cost-considerations, we find evidence supporting this hypothesis. Whereas participants chose cheaper options for a stranger, the chosen cost for treating a loved one was either higher than (Survey 1) or comparable to (Survey 2) that chosen for the self. This effect is reminiscent of what is found in the literature on pain decisions, where individuals' behavior and susceptibility to risk differs strongly between an unknown other and a loved one, with the latter resembling those made for the self (Loued-Khenissi et al., 2022). This result also conforms to the empathy model from social neuroscience literature, where individuals treat others' suffering as their own by triggering the same neural processes that mediate direct pain experience (Bernhardt and Singer, 2012; Corradi-Dell'Acqua et al., 2011, 2016, 2023). This model acknowledges a strong role played by social proximity, with less pronounced empathic responses for those deemed distant from the self (Cheng et al., 2010; Hein et al., 2010; Xu et al., 2009).

Although similar to one another, responses associated with the self and a loved one were not identical. In particular, in Survey 1, choices made for a loved one differed in several ways from the self-condition: they were more risk averse, less influenced by price and

more by *EDS* (or p_D , depending on the analysis), and assigned a higher monetary value. These results show that social proximity shifts one's behavior from wellbeing-oriented (for close others) to price-oriented (for strangers; see Figure 2B). These effects were not observed in Survey 2, though a power analysis established that the sample collected was adequate to reproduce *Target* differences from Survey 1. It is possible that these effects manifest themselves only in complex settings where price and *EDS* are integrated together. However, it should also be mentioned that the power analysis tested effects of *Target* as a whole: i.e., across all three levels. It is therefore possible that the results were influenced by the strong modulations of the *Stranger* condition, and that a more sensitive cohort would have been necessary to replicate subtle *Self* vs. *Beloved* differences.

4.4 Limitations of the study and future implications

This study has three limitations that need to be acknowledged. First, Survey 2 differed slightly from Survey 1 in that: it was shorter, it was targeted to unpaid volunteers recruited outside Prolific.co platform, and no catch trials were implemented to monitor participants' attentional level. It is in principle possible that some participant in Survey 2 lost focus during the task despite its brevity. More critically, the two surveys might have probed slightly different populations, by attracting individuals with different financial status (Table 1). Second, recruitment for unpaid volunteers for Survey 2 was more time consuming. Given the rapid pandemic progression, it is possible that perception of risk for people's health changed across time. Hence, a much delayed recruitment time would have exposed us to the risk of probing a different situational cohort with respect to Survey 1. We minimized such possibility by interrupting data collection following 2 months so that participants from Survey 1 (tested on May 2020) and those from Survey 2 (June–July) were tested in close proximity. The drawback of this choice was that the sample size was imbalanced between the surveys, with Survey 2 being limited to 170 participants. However, rigorous simulation-based power analysis insured that such sample was sufficiently sensitive to replicate the effects of interest from Survey 1. Third, studies on economic decision making and prosocial behavior often report big inter-individual differences, explainable in terms of personality or empathic traits (Thielmann et al., 2020), prosocial beliefs (Carlson and Zaki, 2022) as well as COVID-19 information (disease contraction, regional death rate, etc.; Fang et al., 2022). Unfortunately, personality/social traits were not collected in this study, preventing us to assess these effects also in our dataset. We did collect information about individual COVID-19 experience (see methods for the measures collected), but these measures were either unsuitable for statistical analyses (positive cases of SARS-Cov-2 virus were negligible; Table 1) or did not reveal reliable influence on choice behavior.

Keeping these considerations aside, our study provides novel and replicable evidence on how people make decisions about one's own other people's health under risk. In particular, we found that individuals act for their own health as is observed in the monetary loss domain, by displaying an overall risk-seeking stance that progressively declines as the expected value of a negative event increases. This effect did not differ statistically when choices under risk were made for others, at least when cost considerations were put

aside. However, distinctions between decision targets emerged when choices were conditioned on monetary cost, with participants preferring cheaper treatment options for unknown others, but not for loved ones.

The COVID-19 pandemic imposed the burden of costly decision-making under risk for others' health on ordinary people. Most of the restrictive measures implemented across the globe had negative economical, but also psychological consequences on the people involved (Kunzler et al., 2021; Lima et al., 2020; Nochaiwong et al., 2021; Santomauro et al., 2021; Wu et al., 2021). These restrictions negatively impacted sensitivity to others' suffering, and empathy traits (Antico and Corradi-Dell'Acqua, 2023; Cao et al., 2022), thus raising the question on what cost lay individuals were willing to incur for the sake of a stranger, and this under uncertainty. In this perspective, our study provides a descriptive model of individual risky decision-making in health-contexts; and inform on the limits of what can be asked of an individual in service to a stranger. Furthermore, as global society faces looming events such as climate change (Hornsey et al., 2022) or migration (Denniston, 2021) that demand individual participation for their mitigation, it is crucial to gain an understanding of factors influencing other-regarding decision-making under uncertainty. As respondents showed a readiness to incur cost for the sake of strangers, authorities can assume a general goodwill and willingness to help others (albeit at a lower rate of cost to that observed in self-regarding decisions or decisions made for loved ones), underscoring our tendency toward pro-sociality.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: Open Science Framework: <https://osf.io/9fjdq/>.

Ethics statement

The studies involving humans were approved by The Ethics Committee of the Faculty of Psychology and Educational Sciences of the University of Geneva, Geneva, Switzerland. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

LL-K: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. CC-D'A: Funding acquisition, Supervision, Writing – review & editing, Conceptualization, Validation.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was

funded by the Swiss National Science Foundation (SNSF) grant PP00P1_183715. CCD is further supported by SNSF grant n. 320030_182589.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

References

- Abellan-Perpiñan, J. M., Bleichrodt, H., and Pinto-Prades, J. L. (2009). The predictive validity of prospect theory versus expected utility in health utility measurement. *J. Health Econ.* 28, 1039–1047. doi: 10.1016/j.jhealeco.2009.09.002
- Adams-Prassl, A., Boneva, T., Golin, M., and Rauh, C. (2020). Inequality in the impact of the coronavirus shock: evidence from real time surveys. *J. Public Econ.* 189:104245. doi: 10.1016/j.jpubeco.2020.104245
- Aguilar, M. S., and Castaneda, A. (2021). What mathematical competencies does a citizen need to interpret Mexico's official information about the COVID-19 pandemic? *Educ. Stud. Math.* 108, 227–248. doi: 10.1007/s10649-021-10082-9
- Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica* 21, 503–546. doi: 10.2307/1907921
- Andreoni, J., and Miller, J. (2002). Giving according to GARP: an experimental test of the consistency of preferences for altruism. *Econometrica* 70, 737–753. doi: 10.1111/1468-0262.00302
- Antico, L., and Corradi-Dell'Acqua, C. (2023). Far from the eyes, far from the heart: COVID-19 confinement dampened sensitivity to painful facial features. *Q. J. Exp. Psychol.* 76, 554–567. doi: 10.1177/17470218221094772
- Atanasov, P. D. (2015). Risk preferences in choices for self and others: Meta analysis and research directions. SSRN scholarly paper 1682569. doi: 10.2139/ssrn.1682569
- Attema, A. E., Brouwer, W. B. F., and l'Haridon, O. (2013). Prospect theory in the health domain: a quantitative assessment. *J. Health Econ.* 32, 1057–1065. doi: 10.1016/j.jhealeco.2013.08.006
- Attema, A. E., Brouwer, W. B. F., Lharidon, O., and Pinto, J. L. (2016). An elicitation of utility for quality of life under prospect theory. *J. Health Econ.* 48, 121–134. doi: 10.1016/j.jhealeco.2016.04.002
- Barbieri, P., and Bonini, B. (2020). Political orientation and adherence to social distancing during the COVID-19 pandemic in Italy. SSRN scholarly paper 3640324. doi: 10.2139/ssrn.3640324
- Bernhardt, B. C., and Singer, T. (2012). The neural basis of empathy. *Annu. Rev. Neurosci.* 35, 1–23. doi: 10.1146/annurev-neuro-062111-150536
- Bernoulli, D. (1738/1954). Exposition of a new theory on the measurement of risk. *Econometrica* 22, 23–36. doi: 10.2307/1909829
- Bleichrodt, H. (1997). Health utility indices and equity considerations. *J. Health Econ.* 16, 65–91. doi: 10.1016/S0167-6296(96)00508-5
- Blundell, R., Costa Dias, M., Joyce, R., and Xu, X. (2020). COVID-19 and inequalities*. *Fisc. Stud.* 41, 291–319. doi: 10.1111/1475-5890.12232
- Bonaccorsi, G., Pierri, F., Cinelli, M., Flori, A., Galeazzi, A., Porcelli, F., et al. (2020). Economic and social consequences of human mobility restrictions under COVID-19. *Proc. Natl. Acad. Sci.* 117, 15530–15535. doi: 10.1073/pnas.2007658117
- Boschini, A., Dreber, A., von Essen, E., Muren, A., and Ranehill, E. (2018). Gender and altruism in a random sample. *J. Behav. Exp. Econ.* 77, 72–77. doi: 10.1016/j.socec.2018.09.005
- Cao, S., Qi, Y., Huang, Q., Wang, Y., Han, X., Liu, X., et al. (2022). Emerging infectious outbreak inhibits pain empathy mediated prosocial behaviour. *Res. Square*. [Preprint] (Version 2) doi: 10.21203/rs.3.rs-530170/
- Carlson, R. W., and Zaki, J. (2022). Belief in altruistic motives predicts prosocial actions and inferences. *Psychol. Rep.* 125, 2191–2212. doi: 10.1177/00332941211013529
- Cheng, Y., Chen, C., Lin, C.-P., Chou, K.-H., and Decety, J. (2010). Love hurts: an fMRI study. *NeuroImage* 51, 923–929. doi: 10.1016/j.neuroimage.2010.02.047

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1370778/full#supplementary-material>

- Choi, H. A., and Lee, O. E. (2021). To mask or to unmask, that is the question: facemasks and anti-Asian violence during COVID-19. *J. Hum. Rights Soc. Work* 6, 237–245. doi: 10.1007/s41134-021-00172-2
- Civai, C., Corradi-Dell'Acqua, C., Gamer, M., and Rumiati, R. I. (2010). Are irrational reactions to unfairness truly emotionally-driven? Dissociated behavioural and emotional responses in the ultimatum game task. *Cognition* 114, 89–95. doi: 10.1016/j.cognition.2009.09.001
- Civai, C., Crescentini, C., Rustichini, A., and Rumiati, R. I. (2012). Equality versus self-interest in the brain: differential roles of anterior insula and medial prefrontal cortex. *NeuroImage* 62, 102–112. doi: 10.1016/j.neuroimage.2012.04.037
- Cohen, B. J. (1996). Is expected utility theory normative for medical decision making? *Med. Decis. Mak.* 16, 1–6. doi: 10.1177/0272989X9601600101
- Corradi-Dell'Acqua, C., Civai, C., Rumiati, R. I., and Fink, G. R. (2013). Disentangling self- and fairness-related neural mechanisms involved in the ultimatum game: an fMRI study. *Soc. Cogn. Affect. Neurosci.* 8, 424–431. doi: 10.1093/scan/nss014
- Corradi-Dell'Acqua, C., Hofstetter, C., Sharvit, G., Hugli, O., and Vuilleumier, P. (2023). Healthcare experience affects pain-specific responses to others' suffering in the anterior insula. *Hum. Brain Mapp.* 44, 5655–5671. doi: 10.1002/HBM.26468
- Corradi-Dell'Acqua, C., Hofstetter, C., and Vuilleumier, P. (2011). Felt and seen pain evoke the same local patterns of cortical activity in insular and cingulate cortex. *J. Neurosci.* 31, 17996–18006. doi: 10.1523/JNEUROSCI.2686-11.2011
- Corradi-Dell'Acqua, C., Tusche, A., Vuilleumier, P., and Singer, T. (2016). Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nat. Commun.* 7:10904. doi: 10.1038/ncomms10904
- Cox, J. C., and Sadiraj, V. (2008). "Risky decisions in the large and in the small: theory and experiment" in Risk aversion in experiments. eds. J. C. Cox and G. W. Harrison, vol. 12 (Bingley, UK: Emerald Group Publishing Limited), 9–40.
- Cutler, J., and Campbell-Meiklejohn, D. (2019). A comparative fMRI meta-analysis of altruistic and strategic decisions to give. *NeuroImage* 184, 227–241. doi: 10.1016/j.neuroimage.2018.09.009
- Daly, M. C., Buckman, S. R., and Seitelman, L. M. (2020). The unequal impact of COVID-19: why education matters. *FRBSF Econ. Lett.* 17, 1–5.
- Denniston, L. (2021). "They just come and try to help": exploring the prioritization of downstream accountability in citizen-led humanitarianism in Calais" in Citizen humanitarianism at European Borders (New York, NY: Routledge), 66–82.
- Diekmann, A. (2004). The power of reciprocity: fairness, reciprocity, and stakes in variants of the dictator game. *J. Confl. Resolut.* 48, 487–505. doi: 10.1177/0022002704265948
- Dolan, P., and Kahneman, D. (2008). Interpretations of utility and their implications for the valuation of health. *Econ. J.* 118, 215–234. doi: 10.1111/j.1468-0297.2007.02110.x
- Du, R.-H., Liang, L.-R., Yang, C.-Q., Wang, W., Cao, T.-Z., Li, M., et al. (2020). Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: a prospective cohort study. *Eur. Respir. J.* 55:2000524. doi: 10.1183/13993003.00524-2020
- Elfrink, T. (2020). "Not handling the pandemic well: man fires at officers with AK-47 after refusing to wear a mask, police say. August 6. *Washington Post*. Available at: <https://www.washingtonpost.com/nation/2020/08/03/mask-mandatory-pennsylvania-shooting-police/>
- Evans, W. N., and Viscusi, W. K. (1991). Utility-based measures of health. *Am. J. Agric. Econ.* 73, 1422–1427. doi: 10.2307/1242395
- Fang, X., Freyer, T., Ho, C.-Y., Chen, Z., and Goette, L. (2022). Prosociality predicts individual behavior and collective outcomes in the COVID-19 pandemic. *Soc. Sci. Med.* 308:115192. doi: 10.1016/j.socscimed.2022.115192

- Fehr, E., and Fischbacher, U. (2004). Social norms and human cooperation. *Trends Cogn. Sci.* 8, 185–190. doi: 10.1016/j.tics.2004.02.007
- Fehr, E., and Schmidt, K. M. (2006). “The economics of fairness, reciprocity and altruism—experimental evidence and new theories” in *Handbook on the economics of giving, reciprocity and altruism*. eds. S.-C. Kolm and J. M. Ythier, vol. 1 (Elsevier), 615–691. Available at: <http://www.sciencedirect.com/science/article/pii/S1574071406010086>
- Fink, G., Tediosi, F., and Felder, S. (2022). Burden of COVID-19 restrictions: national, regional and global estimates. *eClinicalMedicine* 45:101305. doi: 10.1016/j.eclinm.2022.101305
- Freund, A. M., and Blanchard-Fields, F. (2014). Age-related differences in altruism across adulthood: making personal financial gain versus contributing to the public good. *Dev. Psychol.* 50, 1125–1136. doi: 10.1037/a0034491
- Frey, B. S., and Meier, S. (2004). Pro-social behavior in a natural setting. *J. Econ. Behav. Organ.* 54, 65–88. doi: 10.1016/j.jebo.2003.10.001
- Goldstein, B. D. (2001). The precautionary principle also applies to public health actions. *Am. J. Public Health* 91, 1358–1361. doi: 10.2105/AJPH.91.9.1358
- Gollier, C., and Treich, N. (2003). Decision-making under scientific uncertainty: the economics of the precautionary principle. *J. Risk Uncertain.* 27, 77–103. doi: 10.1023/A:1025576823096
- Graham, H., de Bell, S., Hanley, N., Jarvis, S., and White, P. C. L. (2019). Willingness to pay for policies to reduce future deaths from climate change: evidence from a British survey. *Public Health* 174, 110–117. doi: 10.1016/j.puhe.2019.06.001
- Green, P., and MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol. Evol.* 7, 493–498. doi: 10.1111/2041-210X.12504
- Hein, G., Silani, G., Preuschoff, K., Batson, C. D., and Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron* 68, 149–160. doi: 10.1016/j.neuron.2010.09.003
- Hellinger, F. J. (1989). Expected utility theory and risky choices with health outcomes. *Med. Care* 27, 273–279. doi: 10.1097/00005650-198903000-00005
- Hoffman, E., McCabe, K., and Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *Am. Econ. Rev.* 86, 653–660.
- Hornsey, M. J., Chapman, C. M., and Oelrichs, D. M. (2022). Why it is so hard to teach people they can make a difference: climate change efficacy as a non-analytic form of reasoning. *Think. Reason.* 28, 327–345. doi: 10.1080/13546783.2021.1893222
- Huck, S., and Müller, W. (2012). Allais for all: revisiting the paradox in a large representative sample. *J. Risk Uncertain.* 44, 261–293. doi: 10.1007/s11166-012-9142-8
- Johnson, J. G., and Busemeyer, J. R. (2010). Decision making under risk and uncertainty. *WIREs Cogn. Sci.* 1, 736–749. doi: 10.1002/wcs.76
- Jorgensen, F., Bor, A., Rasmussen, M. S., Lindholt, M. F., and Petersen, M. B. (2022). Pandemic fatigue fueled political discontent during the COVID-19 pandemic. *Proc. Natl. Acad. Sci.* 119:e2201266119. doi: 10.1073/pnas.2201266119
- Kahlenberg, H., Williams, D., van Tilburg, M. A. L., and Jiroutek, M. R. (2023). Vaccine hesitancy for COVID-19: what is the role of statistical literacy? *Front. Public Health* 11:1230030. doi: 10.3389/fpubh.2023.1230030
- Kay, T., Lehmann, L., and Keller, L. (2019). Kin selection and altruism. *Curr. Biol.* 29, R438–R442. doi: 10.1016/j.cub.2019.01.067
- Knutson, B., Taylor, J., Kaufman, M., Peterson, R., and Glover, G. (2005). Distributed neural representation of expected value. *J. Neurosci. Off. J. Soc. Neurosci.* 25, 4806–4812. doi: 10.1523/JNEUROSCI.0642-05.2005
- Kunzler, A. M., Röthke, N., Günthner, L., Stoffers-Winterling, J., Tüscher, O., Coenen, M., et al. (2021). Mental burden and its risk and protective factors during the early phase of the SARS-CoV-2 pandemic: systematic review and meta-analyses. *Glob. Health* 17:34. doi: 10.1186/s12992-021-00670-y
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13
- Levy, M., and Nir, A. R. (2012). The utility of health and wealth. *J. Health Econ.* 31, 379–392. doi: 10.1016/j.jhealeco.2012.02.003
- Lima, C. K. T., Carvalho, P. M. D. M., Lima, I. D. A. A. S., Nunes, J. V. A. D. O., Saraiva, J. S., De Souza, R. I., et al. (2020). The emotional impact of coronavirus 2019-nCoV (new coronavirus disease). *Psychiatry Res.* 287:112915. doi: 10.1016/j.psychres.2020.112915
- Lockwood, P. L., Hamonet, M., Zhang, S. H., Ratnavel, A., Salmony, F. U., Husain, M., et al. (2017). Prosocial apathy for helping others when effort is required. *Nat. Hum. Behav.* 1:0131. doi: 10.1038/s41562-017-0131
- Loretto, L., Piu, D., and Bellizzi, S. (2021). “Uncertainty in pandemic times” in *Anxiety, uncertainty, and resilience during the pandemic period—anthropological and psychological perspectives* (London, UK: IntechOpen).
- Loued-Khenissi, L., Martin-Brevet, S., Schumacher, L., and Corradi-Dell'Acqua, C. (2022). The effect of uncertainty on pain decisions for self and others. *Eur. J. Pain* 26, 1163–1175. doi: 10.1002/ejp.1940
- Loued-Khenissi, L., Pfeuffer, A., Einhäuser, W., and Preuschoff, K. (2020). Anterior insula reflects surprise in value-based decision-making and perception. *NeuroImage* 210:116549. doi: 10.1016/j.neuroimage.2020.116549
- Loued-Khenissi, L., and Preuschoff, K. (2020). Information theoretic characterization of uncertainty distinguishes surprise from accuracy signals in the brain. *Front. Artif. Intell.* 3:5. doi: 10.3389/frai.2020.00005
- McKee, M., and Stuckler, D. (2020). If the world fails to protect the economy, COVID-19 will damage health not just now but also in the future. *Nat. Med.* 26, 640–642. doi: 10.1038/s41591-020-0863-y
- Meltzer, D. (2001). Addressing uncertainty in medical cost-effectiveness analysis: implications of expected utility maximization for methods to perform sensitivity analysis and the use of cost-effectiveness analysis to set priorities for medical research. *J. Health Econ.* 20, 109–129. doi: 10.1016/S0167-6296(00)00071-0
- Mooney, G. (1989). QALYs: are they enough? A health economist's perspective. *J. Med. Ethics* 15, 148–152. doi: 10.1136/jme.15.3.148
- Muñiz-Rodríguez, L., Rodríguez-Muñiz, L. J., and Alsina, Á. (2020). Deficits in the statistical and probabilistic literacy of citizens: effects in a world in crisis. *Mathematics* 8:1872. doi: 10.3390/math8111872
- Nochaiwong, S., Ruengorn, C., Thavorn, K., Hutton, B., Awiphan, R., Phosuya, C., et al. (2021). Global prevalence of mental health issues among the general population during the coronavirus disease-2019 pandemic: a systematic review and meta-analysis. *Sci. Rep.* 11:10173. doi: 10.1038/s41598-021-89700-8
- Olsen, J. A., and Smith, R. D. (2001). Theory versus practice: a review of 'willingness-to-pay' in health and health care. *Health Econ.* 10, 39–52. doi: 10.1002/1099-1050(200101)10:1<39::AID-HEC563>3.0.CO;2-E
- Pinto-Prades, J. L., Attema, A., and Sánchez-Martínez, F. I. (2019). “Measuring health utility in economics” in *Oxford Research Encyclopedia of Economics and Finance*.
- Polman, E., and Wu, K. (2020). Decision making for others involving risk: a review and meta-analysis. *J. Econ. Psychol.* 77:102184. doi: 10.1016/j.joep.2019.06.007
- Preuschoff, K., Mohr, P., and Hsu, M. (2013). Decision making under uncertainty. *Front. Neurosci.* 7:218. doi: 10.3389/fnins.2013.00218
- Rand, D. G., and Nowak, M. A. (2013). Human cooperation. *Trends Cogn. Sci.* 17, 413–425. doi: 10.1016/j.tics.2013.06.003
- Reis, J., and Spencer, P. S. (2019). Decision-making under uncertainty in environmental health policy: new approaches. *Environ. Health Prev. Med.* 24:57. doi: 10.1186/s12199-019-0813-9
- Ruggeri, K., Ali, S., Berge, M. L., Bertoldo, G., Bjørndal, L. D., Cortijos-Bernabeu, A., et al. (2020). Replicating patterns of prospect theory for decision under risk. *Nature human. Behaviour* 4, 622–633. doi: 10.1038/s41562-020-0886-x
- Russell, L. B., and Schwartz, A. (2012). Looking at patients' choices through the lens of expected utility: a critique and research agenda. *Med. Decis. Mak.* 32, 527–531. doi: 10.1177/0272989X12451339
- Safra, Z., Segal, U., and Spivak, A. (1990). Preference reversal and nonexpected utility behavior. *Am. Econ. Rev.* 80, 922–930.
- Samuelson, P. A. (1993). Altruism as a problem involving group versus individual selection in economics and biology. *Am. Econ. Rev.* 83, 143–148.
- Santomauro, D. F., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., et al. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet* 398, 1700–1712. doi: 10.1016/S0140-6736(21)02143-7
- Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. *Nat. Rev. Neurosci.* 17, 183–195. doi: 10.1038/nrn.2015.26
- Seidl, C. (2002). Preference reversal. *J. Econ. Surv.* 16, 621–655. doi: 10.1111/1467-6419.00184
- Sisco, M. R., and Weber, E. U. (2019). Examining charitable giving in real-world online donations. *Nat. Commun.* 10:3968. doi: 10.1038/s41467-019-11852-z
- Soeteven, A. R. (2005). Anonymity in giving in a natural context—a field experiment in 30 churches. *J. Public Econ.* 89, 2301–2323. doi: 10.1016/j.jpubeco.2004.11.002
- Sunstein, C. R. (2005). The precautionary principle as a basis for decision making. *Econ. Voice* 2, 1–9. doi: 10.2202/1553-3832.1079
- Taylor, S., and Asmundson, G. J. G. (2021). Negative attitudes about facemasks during the COVID-19 pandemic: the dual importance of perceived ineffectiveness and psychological reactance. *PLoS One* 16:e0246317. doi: 10.1371/journal.pone.0246317
- Thielmann, I., Spadaro, G., and Balliet, D. (2020). Personality and prosocial behavior: a theoretical framework and meta-analysis. *Psychol. Bull.* 146, 30–90. doi: 10.1037/bul0000217
- Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., and Herrmann, E. (2012). Two key steps in the evolution of human cooperation: the interdependence hypothesis. *Curr. Anthropol.* 53, 673–692. doi: 10.1086/668207
- Tomova, L., Saxe, R., Klöbl, M., Lanzenberger, R., and Lamm, C. (2020). Acute stress alters neural patterns of value representation for others. *NeuroImage* 209:116497. doi: 10.1016/j.neuroimage.2019.116497
- Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* 211, 453–458.
- Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5, 297–323. doi: 10.1007/BF00122574

- Wiepking, P., and Breeze, B. (2012). Feeling poor, acting stingy: the effect of money perceptions on charitable giving. *Int. J. Nonprofit Volunt. Sect. Mark.* 17, 13–24. doi: 10.1002/nvsm.415
- Wolff, J. (2022). The COVID-risk social contract is under negotiation. *The Atlantic*. Available at: <https://www.theatlantic.com/ideas/archive/2022/01/new-risk-social-contract-covid-ethics/621246/>
- Wood, R., Reinhardt, G. Y., Rezaeedyakenari, B., and Windsor, L. C. (2022). Resisting lockdown: the influence of COVID-19 restrictions on social unrest. *Int. Stud. Q.* 66:sqac015. doi: 10.1093/isq/sqac015
- Wu, T., Jia, X., Shi, H., Niu, J., Yin, X., Xie, J., et al. (2021). Prevalence of mental health problems during the COVID-19 pandemic: a systematic review and meta-analysis. *J. Affect. Disord.* 281, 91–98. doi: 10.1016/j.jad.2020.11.117
- Xu, X., Zuo, X., Wang, X., and Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *J. Neurosci.* 29, 8525–8529. doi: 10.1523/JNEUROSCI.2418-09.2009
- Yancy, C. W. (2020). COVID-19 and African Americans. *JAMA* 323, 1891–1892. doi: 10.1001/jama.2020.6548

Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

