



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/239725/>

Version: Published Version

Article:

Chmura, T., Nguyen, K. and Biermann, F. (2026) Evolutionarily optimal risk aversion. Risk Analysis, 46 (5). e70250. ISSN: 0272-4332

<https://doi.org/10.1111/risa.70250>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

ORIGINAL ARTICLE OPEN ACCESS

Evolutionarily Optimal Risk Aversion

Chmura¹  | Nguyen² | Biermann¹¹Nottingham Business School, Nottingham Trent University, Nottingham, UK | ²School of Economics, University of Sheffield, Sheffield, UK**Correspondence:** Nguyen (k.nguyen@sheffield.ac.uk)**Received:** 1 February 2025 | **Revised:** 10 February 2026 | **Accepted:** 27 March 2026**Keywords:** risky choice | risk attitudes | evolution of preferences | HS&E

ABSTRACT

In an experimental choice situation, we identify risk-acceptability thresholds and show how such thresholds are updated in response to benchmark information, a recurrent feature of health, safety, and environmental (HS&E) risk governance. We present a theoretical framework linking the observed behavior to an underlying evolutionary parameter, which in this case is (an abstract notion of) risk aversion. The theoretical model allows to predict how the experimental subjects adjust their risk aversion when informed about the risky choices of others. Applications of the framework arise naturally in HS&E settings, where individuals and organizations revise risk thresholds by observing peers, experienced coworkers, or acknowledged experts. By distinguishing confidence-driven inertia from trust-driven overreaction, the paper provides actionable guidance for HS&E risk communication and adaptive risk management.

1 | Introduction

In evolutionary settings, the maximization of the number of surviving offspring is the ultimate objective which even preferences have to serve (see, e.g., Lande and Arnold 1985; Krebs 2009; Breslin 2013; Portera and Bartalesi 2016).¹ This article formalizes and empirically tests the idea that preferences of people are associated with evolutionarily relevant parameters. Preferences are therefore not exogenous but result from a “hard-wired” drive of a decision maker to get as close as possible to an optimal value of the evolutionary parameter. For example, the degrees of *compassion*, *patience*, or *aggression* may matter evolutionarily in different situations, and the observed preferences of a decision maker may follow downstream of a more fundamental desire to reach the optimal value of the evolutionary parameter, that is, to be compassionate, patient, or aggressive at an evolutionarily optimal level.

This paper has theoretical and practical-empirical contributions. On a theoretical level, we model a decision maker for whom the adopted preferences depend on an underlying continuous

parameter. We introduce the concept of a *thrive function*, which is a mapping M of an evolutionary parameter into the set of preferences. If r^* is the evolutionarily optimal value of that parameter, then $M(r^*)$ are those preferences of the decision maker that would maximize her evolutionary outcome. We will identify properties a thrive function should fulfill in order for the preferences displayed by the decision maker to be linked in a logical way with the underlying parameter.

As a practical and empirical contribution, we analyze *risky choice* using the theoretical framework of our model. In this context, the underlying evolutionary parameter is *risk aversion*.² Throughout evolutionary history, the outcomes of risky situations were gains or losses in resources that could be used for mate search and kin feeding, and many such situations involved a possibility of injury or death (bringing further reproduction to an abrupt end). Therefore, the degree of risk aversion adopted in such critical situations had an impact on a human's survival and reproduction potential. As an illustration, consider distant ancestors in hunting or conflict situations. If they were reckless, they could perish in the pursuit of killing animals or in a fight with other

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Risk Analysis* published by Wiley Periodicals LLC on behalf of Society for Risk Analysis.

group members, in which case their expected offspring was less numerous than if they had adopted a more cautious approach. On the other hand, reluctance to take risks would entail feeble hunting success and giving away too much to others, leading to less resources and a potential loss of status. This illustrates that by the expected offspring criterion, it is both possible to be too risk-averse and too risk-loving in a given situation, implying that there should be a context-dependent *evolutionarily optimal* degree of risk aversion between the extremes.

When applied to risk preferences, our model yields empirically testable predictions about how the choice behavior of humans will be affected through learning about the *choices of others*. The model predicts that knowledge about the choices of others may lead to preference changes, resulting from agents updating their original opinions about what is the evolutionarily optimal risk attitude in the situation at hand. Specifically, it follows that the extent to which decision makers adjust their preferences depends on the confidence they have in their own judgment versus their confidence in the judgments of others. Based on these variables, we can, among other testable hypotheses, predict who will modify their preferences upon observing others and who will not. We test these predictions in a sequential lottery choice experiment devised for this purpose, and we find affirmative evidence.

Modern health, safety, and environmental (HS&E) contexts yield ample applications of our theory as individuals frequently adjust their willingness to accept health risks, workplace hazards, or environmental exposures in response to perceived threats, social information, and protective norms (see the discussions, e.g., in Kasperson et al. 1988; Dosman et al. 2001; Lindell and Hwang 2008; Ho et al. 2022). We will argue that some of the risk attitude changes observed in HS&E contexts reflect—on a fundamental level—the adaptive pressure to balance caution and risk-taking in the pursuit of maximizing the expected number of offspring, as hypothesized in this paper.

As illustrations, consider how vaccine uptake (for Covid and other viruses) is influenced by the behaviors of colleagues, neighbors, and friends (e.g., LaJoie et al. 2018; Patelarou et al. 2022; McCready et al. 2023), or how workers in hazardous occupations modify their safety practices when they observe colleagues either taking precautions or ignoring them (Wakefield et al. 2010). Individuals may initially underestimate or overestimate the health or safety risks they face, but their risk-taking behaviors changes rapidly when they observe the behavior of others. A worker who normally ignores protective equipment may become more cautious after seeing peers consistently adopt safety measures, whereas someone observing widespread rule-breaking may become less risk-averse (Dean 2014; Maglio et al. 2016; Wong et al. 2020). This mirrors the evolutionary logic underlying our model: natural selection would favor individuals who adjust their risk attitudes flexibly, drawing on both personal judgment and social information about what constitutes an advantageous level of risk-taking under prevailing conditions.

We offer this paper for publication in the special issue of *Risk Analysis* in commemoration of Daniel Kahneman (1934–2024). While not at the center of *Prospect Theory*, the question of what causes preferences to change was a topic that he was always interested in. The phenomenon of *preference reversal* had

been described by Lichtenstein and Slovic (1971), triggering a protracted debate about the reasons why this would happen. Kahneman joined his friends Amos Tversky and Paul Slovic relatively late in this debate in order to contribute an important (some would say: authoritative) explanation for this phenomenon (Tversky et al. 1990). Likewise, a prominent brainchild of Kahneman were *framing effects*. From the beginning, these were shown to have an impact on choice problems involving *lotteries*, where they caused people to *change their preferences*, as in the original paper of Tversky and Kahneman (1981). As our experiment demonstrates how (risk) preferences change, we believe that Daniel Kahneman would have been interested in our empirical findings—but would he have liked the evolutionary story?

While he was not an evolutionary psychologist in a strict sense, Kahneman was no stranger to evolutionary arguments in support of his behavioral observations. For example, he attributed *loss aversion* in Prospect Theory to evolutionary pressures: “When directly compared or weighted against each other, losses loom larger than gains. This asymmetry between the power of positive and negative expectations or experiences has an evolutionary history. Organisms that treat threats as more urgent than opportunities have a better chance to survive and reproduce” (Kahneman 2011). Against this backdrop, we believe that he would have been open-minded to consider the theoretical explanation we offer.

This paper contributes to the analysis of risk-related decision-making with a particular focus on HS&E contexts. We develop a formal decision-theoretic model in which observed choices reflect the adoption and updating of *risk-acceptability thresholds* in response to social and benchmark information, a central feature of HS&E risk governance. The model is tested experimentally using a sequential choice design that elicits how individuals adjust their risk thresholds when informed about the risky choices of others. The results reveal systematic heterogeneity in updating behavior that is related to proxies for confidence in one’s own judgment and trust in others, helping to explain why identical risk signals can lead to underadjustment in some settings and overadjustment in others. We further illustrate the relevance of the framework through applications to real-world HS&E decision problems, including the escalation of safety thresholds following near-miss events, responses to public health guidance, and the formation of environmental risk acceptability, thereby demonstrating how the model informs risk communication, organizational learning, and adaptive risk management.

The remainder of the paper is structured as follows. Section 2 presents a rigorous mathematical model on how the adoption of preferences can be linked to a continuous evolutionary parameter. Section 3 uses this framework to predict what will happen when agents observe the choices of *others*. Section 4 features a description of the experiment we have conducted and states hypotheses on the experimental results we expect to find. In Section 5, we present and interpret the experimental outcomes in view of our hypotheses. Section 6 offers a general discussion of how our theory, for example, how it squares with the well-known finding of *risky shift* (Wallach et al. 1962). We also consider what it means that one of our hypotheses was rejected in the experiment, and we respond to some potential objections against the assumptions of our model. Section 7 zooms in on the relevance of our findings in HS&E settings, and Section 8 concludes the

paper. The proofs of the mathematical results can be found in Appendix A and a description of the experiment in Appendix B.

2 | The Thrive Function

This section presents a formal decision-theoretic model in which the preferences adopted by the decision maker depend on a continuous parameter. This is motivated by our hypothesis that in choice situations with evolutionary significance, preferences are not exogenous but follow from deciding on an evolutionarily relevant value. The decision maker's displayed preferences are then endogenously determined by that prior choice.

When describing our model, we will call the elements of the choice set *lotteries* and the parameter the *degree of risk aversion*, but this is not the only possible interpretation. In other potential applications of this framework, the choice set could have other objects, and the evolutionary parameter could have different interpretations, for example, it could be the level of *aggression*, the level of *patience*, or the level of *compassion* inherent in the agent's choices.

Technically, our agent has to adopt preferences over a finite set L . By *preferences*, we mean a strict total relation $>$ over L .³ The set of all strict total relations over L is denoted by $\mathcal{P}$.

We assume that in a given decision problem, any preferences that can be adopted are associated with an evolutionary parameter $r \in (\underline{r}, \bar{r}) \subset \mathbb{R}$.⁴

The evolutionary significance of adopted preferences is formalized through a function $M : (\underline{r}, \bar{r}) \rightarrow \mathcal{P}$ that assigns an element $>_r \in \mathcal{P}$ to each $r \in (\underline{r}, \bar{r})$, with the interpretation that $>_r$ are the preferences that are evolutionarily most beneficial if r was the optimal evolutionary value in the decision situation at hand. We call M a *thrive function*.⁵

To simplify the analysis, we remove some elements from L that are fundamentally irrelevant but have the potential to complicate matters:

Definition 1 (Domination). If there are two lotteries $l', l'' \in L$ such that $l' >_r l''$ for all $r \in (\underline{r}, \bar{r})$, then l'' is called a *dominated* lottery.

When a lottery l'' is dominated, it would *never* be chosen by a decision maker who optimizes her evolutionary outcome, regardless of what she considers to be the correct value of r in the situation at hand. By an argument resembling *Independence of Irrelevant Alternatives* (Arrow 1951), removing l'' from L would not change the agent's actual choice whenever the other elements in L remain on offer. This simplifies some of our arguments, and we therefore assume for this analysis that there are no dominated lotteries in L .

A thrive function M can have the *single-crossing property* (Apesteguia et al. 2017; Chiappori et al. 2019; Barseghyan et al. 2021), which in the context of this model can be motivated as follows. Consider a decision maker who believes that the optimal level of risk aversion in the situation is r' , and assume that

according to the preferences $>_{r'}$, she prefers lottery l' to lottery l'' , that is, $l' >_{r'} l''$. Later, she updates her belief, maybe because she obtained new information about some characteristic of the situation, and now she feels that the optimal risk aversion is $r'' > r'$. Accordingly, her preferences change to $>_{r''}$. If under these new preferences, she would prefer l'' to l' , that is, $l'' >_{r''} l'$, it would suggest that l' is, in some sense, *riskier* than l'' because as her risk aversion increases, her choice switches from l' to l'' . Consequently, it would be reasonable to expect that for an even *higher* risk aversion $r''' > r''$, she would *keep* to preferring l'' to l' , that is, $l'' >_{r'''} l'$. By the same logic, if she initially prefers l' to l'' but reverses her preferences due to a *reduction* of risk aversion, her preferences should not switch again if her risk aversion decreases further. This idea is captured in:

Property 1 (Single-crossing property). Let $l', l'' \in L$ and $r' < r''$. If M is single-crossing, then from $l' >_{r'} l''$ and $l'' >_{r''} l'$ follows $l' >_r l''$ for all $r < r'$ and $l'' >_r l'$ for all $r > r''$.

Lemma 1. Assume that M is single-crossing and there are no dominated lotteries in L . Then, for $l', l'' \in L$ with $l' \neq l''$, exactly one of the following statements is true:

1. There exists a value $r(l', l'') \in (\underline{r}, \bar{r})$ such that $l' >_r l''$ if $r < r(l', l'')$ and $l'' >_r l'$ if $r > r(l', l'')$.
2. There exists a value $r(l'', l') \in (\underline{r}, \bar{r})$ such that $l'' >_r l'$ if $r < r(l'', l')$ and $l' >_r l''$ if $r > r(l'', l')$.

Note that we do not specify whether $l' >_{r(l', l'')} l''$ or $l'' >_{r(l', l'')} l'$. For our analysis, it does not matter which of the two is the case. Likewise, it does not matter whether $l' >_{r(l'', l')} l''$ or $l'' >_{r(l'', l')} l'$.⁶

For our next steps, we need to define a metric on L . In line with straightforward intuition, we will say that the more comparisons one agent makes differently than another agent, the more remote are their preferences. Formally, we define:

$$d(>', >'') := \frac{1}{2} \left| \left\{ (l', l'') \in L \times L \mid l' >' l'' \wedge l'' >'' l' \right\} \right|. \quad (1)$$

There are $d(>', >'')$ binary comparisons between lotteries in L that are made differently by $>'$ and $>''$. We multiply the number of different pairs by 1/2 as otherwise, (l', l'') and (l'', l') would both be counted for the same lotteries l', l'' even though it is the same comparison.⁷

Lemma 2. The function $d : L \times L \rightarrow \mathbb{N} \cup \{0\}$ defined by (1) is a metric on L .

Technically, the above metric is a two-dimensional version of a *Hamming metric* (Hamming 1950).^{8,9}

The metric $d(\cdot, \cdot)$ allows us to define another intuitive property a thrive function can have:

Property 2 (Linearity). Assume $r' < r''$ with $d(>_{r'}, >_{r'') > 1$. If M is linear, there are values r^1, r^2 with $r' \leq r^1 < r^2 \leq r''$ such that for all $r \in (r^1, r^2)$, it holds $d(r, r'') = 1$.¹⁰

Linearity means that if there is a distance greater than one between two relations $>_{r'}$ and $>_{r''}$ in the image of M , then we

can find a range of r between r' and r'' where the distance to $\succ_{r''}$ is exactly one. We consider this to be an intuitive property for many applications of this model, including the one where r is interpreted as risk aversion: linearity implies that there exists always a change in the value of r that is so small that the resulting change of the adopted preferences, if any, will also be minimal, meaning that it triggers only two lotteries to be swapped in the ranking. When r ranges over $(\underline{r}, \bar{r})$, the function values $\succ_r$ of a linear thrive function will not “jump” in terms of their distance from a given element in $\mathcal{P}$. Assume that r initially stands at r' and is mapped into $\succ_{r'}$. If r increases or decreases, the next relation into which r is mapped will have a distance of *one* from $\succ_{r'}$, the subsequent one a distance of *two*, and so on. We assume from here onward that the decision maker’s thrive function M is linear.

Linearity ensures that the values defined in Lemma 1 are distinct:

Lemma 3. *Let $\{l^1, l^2, l^3, l^4\} \in L$ with $\{l^1, l^2\} \neq \{l^3, l^4\}$. If the thrive function M is single-crossing and linear, then $r(l^1, l^2) \neq r(l^3, l^4)$.*

Proposition 1. *If the function M is single-crossing and linear, then $(\underline{r}, \bar{r})$ can be partitioned into n intervals $I_1, \dots, I_n$ (each of which may be open, closed, or semiclosed) such that:*

1. $\succ_{r'} = \succ_{r''}$ if and only if $r', r'' \in I_j$ for one $j \in \{1, \dots, n\}$.
2. The n intervals can be labeled in a way that fulfils:

$$r' \in I_i, r'' \in I_j \Rightarrow d(\succ_{r'}, \succ_{r''}) = |i - j|.$$

The above result provides conditions on when it is possible to rank preference relations associated with an underlying parameter such that: (a) higher ranked preferences correspond to higher parameter values and (b) neighboring preferences differ by exactly one comparison. An example where this is relevant is the menu of lotteries of Holt and Laury (2002), which has been used in countless experiments to establish the risk attitudes of experimental subjects¹¹ and is also deployed in our own experiment (see Section 4).

The lotteries in Holt and Laury (2002) are arranged in two groups A and B . There are 10 lotteries in A and 10 lotteries in B , and the payoffs of all lotteries in each group are identical, being \$2 and \$1.60 in Group A , and \$3.85 and \$0.10 in Group B . Within each group, the lotteries just differ by their probabilities. Denoting by p the probability of obtaining the first payoff, there is exactly one lottery in each group with $p = 0.1, p = 0.2, \dots, p = 0.9$, and $p = 1$. In setbuilder notation, we can write the two sets of lotteries as

$$A := \left\{ [p, 1-p \mid \$2, \$1.60] \mid p \in \{0.1, 0.2, \dots, 0.9, 1\} \right\}$$

and

$$B := \left\{ [p, 1-p \mid \$3.85, \$0.1] \mid p \in \{0.1, 0.2, \dots, 0.9, 1\} \right\}.$$

In experiments making use of this menu, subject are typically asked (explicitly or implicitly) for which probabilities they prefer the lotteries in Group A , and for which probabilities they prefer

the lotteries in Group B . If the decision makers have a Constant Relative Risk Aversion (CRRA) utility function (Arrow 1974; Wakker 2008),¹² there will be a unique *switching point* where they move from the lotteries in Group A to the lotteries in Group B . For example, for $p \leq 0.4$, the decision maker might prefer the lotteries in set A , but if $p \geq 0.5$, she switches to set B . In this case, the switching point, defined as the highest scenario in Table 3 where *Option A* is preferred, would be 4. Table 1 shows the approximate switching points as a function of the underlying risk aversion r , using the exact utility function $U(x)$ stated in footnote (not a transformation thereof).

Every switching point is associated with a unique preference relation over the set of 20 lotteries that are included in the table. To illustrate this point, label the A -lotteries $A_1, \dots, A_{10}$ with increasing probability p , and the B -lotteries $B_1, \dots, B_{10}$ with increasing probability p (see the “scenarios” in Table 3). If, for example, the switching point is 5, then we have $0.1 < r \leq 0.3$. From here, we infer that $A_4 > B_4, A_3 > B_3, A_2 > B_2$, and $A_1 > B_1$, as well as all other comparisons, by computing the expected utility of the lotteries with the utility function given in footnote, setting r to be a value in the specified interval.^{13,14}

As can be seen, implicit in the Holt and Laury (2002) menu of lotteries is a continuous variable r that is mapped into a set of preference relations over those lotteries. Moreover, there is an interval structure of r , having the same properties as the interval structure in Proposition 1. To illustrate the last point, consider Table 3, where the “scenarios” in the upper half of the table are the lotteries of Holt and Laury (2002).¹⁵ If someone chooses Asset A in scenarios 1–3, and Asset B in all other scenarios, her switching point is 3. If another decision maker has switching Point 4, then the preferences of the two persons are different by exactly one comparison:¹⁶ they only rank the lotteries in scenario 4 differently, while all other scenarios comparisons are made in the same way.

When we considered to run an experiment based on the lottery menu of Holt and Laury (2002), we were not sure whether their connection between risk aversion and preferences can be generalized. Is it required to design the lotteries in the choice menu a special way so as to obtain an interval mapping from the degree of risk aversion into the set of preferences? Obviously, we did not want to develop a theory that only explains the very choice situation constructed by Holt and Laury (2002) but does not carry over to other lottery choice problems. Proposition 1 addresses that problem by stating sufficient conditions for a choice problem to have the same structure as the lottery menu of Holt and Laury (2002). If the connection between the underlying parameter and the preferences can be represented by a linear and single-crossing thrive function, *any* choice set leads to a structure as in Holt and Laury (2002). This may be of relevance for other researchers concerned with the generalizability of results derived from the Holt and Laury (2002) lottery menu, and it may also be relevant in applications where r is not risk aversion but some other parameter.¹⁷

3 | Observing Others’ Choices

Up to this point, we have described how the decision maker forms preferences in response to a perceived optimal value of

TABLE 1 | How relative risk aversion r determines the switching points in Table 3.

Value of r	Switching point (= highest scenario in Table 3 where <i>Option A</i> is preferred)
$0 \leq r \leq 0.1$	4
$0.1 < r \leq 0.3$	5
$0.3 < r \leq 0.5$	6
$0.5 < r \leq 0.95$	7
$0.95 < r \leq 1.2$	8
$1.2 < r$	9 and 10

the evolutionary parameter r . However, we still need to formalize what it means to hold an opinion about the evolutionarily optimal level $r^* \in (r, \bar{r})$, and, for making empirical predictions, we need to model how this opinion is affected by new information. This section is devoted to this aspect.

Importantly, we assume that when multiple decision makers are confronted with the same decision situation, they all have the same thrive function. Heterogeneity between the agents stems from different views on what is the optimal value of r , not from different views on the preferences that should be adopted for a given value of r . Specifically, in view of Lemma 1, all agents will agree that if the evolutionarily optimal value of r was lower than $r(l', l'')$, one should rank $l' > l''$, and if it was higher than $r(l', l'')$, one should rank $l'' > l'$.

Our assumption makes sense if the individuals are sufficiently similar. For that case, an evolutionary argument can be made why in an evolutionary steady state, the thrive functions of all decision makers should be the same: if Agent A feels that for a given value r , the preferences $\succ_r^A$ should be adopted, whereas Agent B feels that preferences $\succ_r^B$ should be displayed, then if $\succ_r^A \neq \succ_r^B$, one of the two preferences would yield higher reproductive success in situations where the evolutionary optimal value was r , and agents who have the other preferences would, over the passage of many generations, eventually go extinct.¹⁸

We assume that the agent does not know with certainty what is the evolutionarily optimal level r^* of the parameter but holds a belief $p := \{p(i)\}_{i \in \mathbb{Z}}$ where $p(i)$ for $i \in \{1, \dots, n\}$ denotes the (subjective) probability of the event $r^* \in I_i$ (see Proposition 1). For technical reasons, it is convenient to define $p(i)$ for every integer value $i \in \mathbb{Z}$ and set $p(i) = 0$ if $i \notin \{1, \dots, n\}$. We may construe the probability mass function p to be derived from an underlying continuous density function θ of r^* over $(r, \bar{r})$ by setting $p(i) = \int_a^b \theta(r) dr$, where a and b are the left and right boundaries of I_i .

Let $M(r, \bar{r}) \subseteq \mathcal{P}$ be the image of the thrive function M , and denote by $\succ^i$ the element $\succ_r \in M(r, \bar{r})$ if $r \in I_i$. In view of the convention we made ($p(i) = 0$ if $i \notin \{1, \dots, n\}$) and the second statement of Proposition 1, we can express the expected distance of element $\succ^i$ from the optimal relation $\succ_{r^*}$ as

$$E_p[d(\succ^i, \succ_{r^*})] = \sum_{k=1, \dots, \infty} k(p(i+k) + p(i-k)). \quad (2)$$

For $k > 0$, the expression $p(i+k) + p(i-k)$ in (2) is the probability under the belief p that $\succ^i$ and $\succ_{r^*}$ have a distance of k .

$\sum_{k=1, \dots, \infty} k(p(i+k) + p(i-k))$ is thus the *expected distance* of $\succ^i$ from $\succ_{r^*}$.

We assume that a decision maker with belief p will adopt the preferences $\succ^{i^*} \in M(r, \bar{r})$ for which the expected distance to the evolutionarily optimal preferences is minimal, that is,

$$i^* := \arg \min_{i \in \{1, \dots, n\}} E_p[d(\succ^i, \succ_{r^*})]. \quad (3)$$

It is well known that the median minimizes the expected absolute deviation from the realization of a random variable (see Lemma 4 below). Therefore, the median of the probability mass function p solves the minimization problem associated with (3). As is common in the literature (e.g., Mittelhammer 2012, Definition 3.1.1, p. 139), we define the median of a probability mass function $p := \{p(i)\}_{i=1, \dots, n}$ to be $m \in \{1, \dots, n\}$ such that both $\sum_{i=1, \dots, m} p(i) \geq 0.5$ and $\sum_{i=m, \dots, n} p(i) \geq 0.5$. Unless n is even and the values $1, \dots, n/2$ have exactly the same joint probability mass as the values $n/2 + 1, \dots, n$, the median will be unique. While our model does not require a unique median, the possibility of multiple medians forces us to make awkward case distinctions throughout our remaining analyses without adding any additional qualitative insights. Pragmatically, we avoid these problems by making the simplifying assumption that all probability mass functions in our analyses have unique medians.¹⁹

Lemma 4 (Median minimizes expected absolute error). *If $m(p)$ is the unique median of the probability mass function p , then $i = m(p)$ minimizes the expression*

$$\sum_{k=1, \dots, \infty} k(p(i+k) + p(i-k)).$$

Now assume that the agent with belief p learns about another belief q , with $q := \{q(i)\}_{i \in \mathbb{Z}}$ where $q(i)$ for $i \in \{1, \dots, n\}$ denotes the subjective probability of the event $r^* \in I_i$. As before, we set $q(i) = 0$ for $i \notin \{1, \dots, n\}$. The origin of q could be the observation of another agent making choices compatible with preferences consistent with q , or it could simply be that our decision maker is told that other people have the belief q . Being confronted with q , the agent will update her own belief in accordance with a parameter $\lambda \in [0, 1]$, forming a new “synthesis belief” s_λ :

$$s_\lambda = \{\lambda q(i) + (1 - \lambda)p(i)\}_{i \in \mathbb{Z}}. \quad (4)$$

The parameter λ captures the agent’s uncertainty about the accuracy of her own judgment as well as the trust she has in the reliability and trustworthiness of the belief q . If $\lambda = 0$, finding

out about q does not affect the agent's existing belief about the value of r^* at all; her initial belief p is the same as her subsequent belief s_λ . This would be the extreme case where she has absolute confidence in her own perception, understanding, and expertise of the decision problem so that she dismisses any choice inconsistent with her initial belief p . On the other hand, if $\lambda = 1$, she would fully adopt the belief q . If $\lambda = 0.5$, she would put the same trust in her own judgment as well as in the judgment of whoever proposed q .

By Lemma 4, the median $m(p) \in \{1, \dots, n\}$ of the random variable p solves the minimization problem (3). Accordingly, $m(q)$ solves the analogous problem if the belief is q (i.e., $m(q) := \arg \min_{i \in \{1, \dots, n\}} E_q[d(>^i, >_{r^*})]$), and $m(s_\lambda)$ solves the problem if the belief is s_λ . With this notation, we state the following.

Lemma 5. *Let p and q be probability mass functions defined on $\mathbb{Z}$, supported on a finite set $\{1, \dots, n\} \subset \mathbb{N}$, and let $s_\lambda = \{\lambda q(i) + (1 - \lambda)p(i)\}_{i \in \mathbb{Z}}$ with $\lambda \in [0, 1]$ be a linear combination of p and q . If $m(q), m(p), m(s_\lambda)$ are the unique medians of the three functions and $m(p) \leq m(q)$, then $m(p) \leq m(s_\lambda) \leq m(q)$.*

Our next and final result shows that the distance between $>_{m(s_\lambda)}$ and $>_{m(q)}$ is weakly smaller than the distance between $>_{m(p)}$ and $>_{m(q)}$. This is, upon encountering a belief q , our agent, who previously adopted the preferences $>_{m(p)}$, tends to move closer toward $>_{m(q)}$.

Proposition 2. *Let $\lambda \in [0, 1]$. Then,*

$$d(>_{m(s_\lambda)}, >_{m(q)}) \leq d(>_{m(p)}, >_{m(q)}).$$

Table 2 illustrates this result with three examples for $n = 11$ (this number was chosen because it corresponds with the experiment we will describe in the next section). Each example features two distributions p and q and their synthesis distribution s_λ based on the values $\lambda = 0.3, \lambda = 0.5$, and $\lambda = 0.9$. Under each preference, we state the expected distance from the realization of the random variable r^* if those preferences are adopted. The last column shows the medians of the distributions, which are the minimizers of the expected distances to r^* . For illustration, consider the first case with $\lambda = 0.3$. Under belief p , the probability that $r^* \in I_1$, meaning that $>_1$ are the optimal preferences, is 0.07. For $r^* \in I_2$, the probability is close to 0, and for $r^* \in I_3$, the probability is 0.12, and so forth. The respective expected distances from $>_{r^*}$ are 5.69, 4.82, and 3.97. The median of belief p is 8, which is the choice where the expected distance from $>_{r^*}$ is minimal, standing at 2.8. The model predicts that Preferences $>_8$ will be those displayed by a decision maker with belief p .

Now the decision maker encounters the belief q , for example, she speaks with someone who reveals belief q , or she observes someone who displays belief q through the way she acts, or she gets other information which lets an alternative belief q emerge. Under q , the probabilities of $>_1, >_2$, and $>_3$ are 0.17, 0.15, and 0.08, respectively, and under q , it would be optimal to display preferences $>_4$. Our decision maker assigns reliability of $\lambda = 0.3$ to the new belief q , which means she has confidence that her current belief is correct at level $1 - \lambda = 0.7$. Forming the corresponding synthesis belief $s_{0.3}$, she obtains the probabilities 0.1, 0.05, and 0.1 for the preference relations $>_1, >_2$, and $>_3$,

respectively. The preference she will adopt under the synthesis belief is $>_6$.

In all three instances, the preferences adopted under the synthesis belief are, according to the metric $d(\cdot, \cdot)$, strictly “between” the preferences displayed under the beliefs p and q , meaning that the decision maker moves strictly closer toward the preferences she observes. However, this is not generally the case, and it is easy to find distributions p and q for the given values of λ where the choice based on the synthesis distribution is the same as the choice based on either p or q .

4 | Experiment Design and Empirical Hypotheses

To test the model, we designed an experiment based on the seminal work of Holt and Laury (2002) that was already described at the end of Section 2. However, in order to test the predictions of Section 3, we added three additional stages.

The experiment was conducted at the *Centre for Research in the Behavioural Sciences* (CRIBS) at *Nottingham University Business School*. We recruited 248 students from various disciplines using ORSEE (Greiner 2015). Participants had no prior experience with a Holt and Laury (2002) experiment and were restricted to participating only once. Each participant received a £3 show-up fee and an average of £3 for a 30-min experiment. Four participants did not complete the survey, so we excluded their results from the dataset, resulting in 244 observations.

The experiment had three stages. In each stage, we used a modified version of the Holt and Laury (2002) lottery menu to measure individuals' attitudes toward risk (see Table 3 and the description of the Holt and Laury (2002) setup at the end of Section 2). Participants had to choose one of the options in the lower part of Table 3, which revealed the switching point where they wanted to switch from “Asset A” to “Asset B.” As can be seen in Table 1, the higher the switching point, the more risk-averse is the participant.

Before the experiment started, participants received instructions about the experiment and their payoffs. We explained the three stages of the experiment, the decisions that participants would need to make in each stage, and how payoffs would be calculated. The payoff calculation proceeded as follows. One scenario would be selected randomly from any of the three rounds, and a fair die would determine the final payoffs of the selected decision (see Appendix B). This procedure was designed to prevent participants from attempting to play strategically. The experiment ran in z-tree (Fischbacher 2007) and was followed by a short questionnaire. We will now describe the three stages of the experiment in more detail.

Stage 1: Personal risk preference (we will refer to this stage as *Base*) This stage was designed to determine participants' initial risk preferences. Participants were presented with the lottery menu of Holt and Laury (2002) (in the version shown in Table 3) and asked to decide their switching point by choosing one of the switching points from 0 to 10.²⁰ The selected switching point indicates individual risk preferences: switching points from 0 to 3 imply risk-seeking behavior because the decision

TABLE 2 | Examples of belief updating for $\lambda = 0.3$, $\lambda = 0.5$, and $\lambda = 0.9$.

$\lambda = 0.3$	$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$	$\succ_6$	$\succ_7$	$\succ_8$	$\succ_9$	$\succ_{10}$	$\succ_{11}$	Minimizer
Probabilities p	0.07	0.00	0.12	0.19	0.04	0.05	0.01	0.21	0.03	0.14	0.14	8
Expected distance	5.69	4.82	3.97	3.34	3.10	2.92	2.85	2.80	3.17	3.60	4.31	
Probabilities q	0.17	0.15	0.08	0.10	0.08	0.02	0.01	0.03	0.11	0.08	0.16	4
Expected distance	4.52	3.87	3.53	3.35	3.36	3.54	3.74	3.98	4.28	4.80	5.48	
Probabilities s	0.10	0.05	0.10	0.16	0.05	0.04	0.01	0.16	0.06	0.12	0.15	6
Expected distance	5.34	4.54	3.84	3.34	3.18	3.11	3.12	3.15	3.50	3.96	4.66	
$\lambda = 0.5$	$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$	$\succ_6$	$\succ_7$	$\succ_8$	$\succ_9$	$\succ_{10}$	$\succ_{11}$	Minimizer
Probabilities p	0.03	0.20	0.27	0.02	0.15	0.03	0.13	0.03	0.01	0.11	0.01	3
Expected distance	3.76	2.82	2.28	2.29	2.34	2.70	3.11	3.77	4.50	5.26	6.24	
Probabilities q	0.01	0.03	0.13	0.13	0.13	0.13	0.07	0.06	0.08	0.11	0.11	6
Expected distance	5.54	4.56	3.64	2.97	2.56	2.41	2.52	2.77	3.14	3.69	4.46	
Probabilities s	0.02	0.12	0.20	0.08	0.14	0.08	0.10	0.05	0.05	0.11	0.06	5
Expected distance	4.65	3.69	2.96	2.63	2.45	2.55	2.81	3.27	3.82	4.47	5.35	
$\lambda = 0.9$	$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$	$\succ_6$	$\succ_7$	$\succ_8$	$\succ_9$	$\succ_{10}$	$\succ_{11}$	Minimizer
Probabilities p	0.01	0.03	0.08	0.03	0.00	0.11	0.18	0.06	0.20	0.19	0.11	9
Expected distance	6.74	5.77	4.85	4.08	3.38	2.68	2.20	2.07	2.07	2.48	3.26	
Probabilities q	0.08	0.18	0.07	0.00	0.17	0.09	0.18	0.10	0.00	0.03	0.09	5
Expected distance	4.44	3.61	3.12	2.79	2.46	2.47	2.67	3.22	3.98	4.74	5.56	
Probabilities s	0.07	0.16	0.07	0.01	0.15	0.09	0.18	0.10	0.02	0.04	0.09	6
Expected distance	4.67	3.82	3.30	2.92	2.55	2.49	2.62	3.11	3.79	4.51	5.33	

maker prefers lotteries with bigger spread and lower expected value to lotteries with lower spread and higher expected value.²¹ Conversely, switching points from 5 to 10 are consistent with risk aversion. A risk-neutral decision maker would therefore choose switching point 4, though subjects whose attitude is not exactly risk-neutral but sufficiently close to risk-neutral would choose switching point 4, too. Participants did not have a hard time constraint for making their decisions, and all participants made their decisions in a reasonable time frame so that there was no necessity to enforce a time limit.

Before stages 2 and 3, participants were given information about the switching points of a reference group as follows: “The average switching points of some participants in previous sessions was x .” The signal was predetermined to be fuzzy, that is, it was left open for participants to guess *how many* participants chose that switching point and which kind of reference group we were referring to. We set x to be 3 before stage 2 and 7 or 9 before stage 3. In this way, we avoided *deception*, as there had been a test run of the experiment where participants had chosen switching points 3, 7, and 9. When deciding on the signals that were communicated as signals to participants before stages 2 and 3, we centered around the most commonly found switching point of 5 (approximately, see Table 4 below) and then moved equidistantly into both directions. A share of 25% of students got signal nine. While a switching point of 7 lies just within the range of switching points found in the literature, the switching points of 3 and 9 lie clearly outside that range (see Table 4). A switching point of 3 clearly indicates that participants in a previous session were risk-

loving, while a switching point of nine suggests abnormally high risk aversion.

Table 4 provides an overview of switching points found in the literature. The lowest switching point is reported by He et al. (2012). They report switching points for males (4.48), females (5.25), and couples (4.84) separately.²² The highest value of 6.6 was found by Masclet et al. (2009) who study individual attitudes toward risk. Moreover, they report a risk attitude of 6.9 for decision-making in groups, suggesting that groups are slightly more risk-averse than individuals and contradicting the so-called *risky shift* phenomenon (Wallach et al. 1962; Aronson et al. 2017, p. 308). Most papers find a switching point of around 5, suggesting that the average decision maker is slightly risk-averse.

While switching points are commonly interpreted as measures of individual risk aversion, they can also be understood as decision thresholds analogous to those encountered in high-stakes HS&E contexts. In such settings, decision makers frequently face binary or discrete choices—such as whether to shut down a process, adopt a safety margin, or escalate a warning—under substantial uncertainty. The switching point thus captures the point at which a decision maker is willing to trade expected performance for increased safety, making it a natural proxy for precautionary behavior in HS&E risk management.

Stage 2: Risk-loving reference (we will refer to this stage as H3) Before stage 2, participants were told that the average switching point in a selected session was three and then asked

TABLE 3 | The choice task given to the experimental subjects.

Scenario #	Asset A	Asset B
1	10% chance of 20 ECU and 90% chance of 16 ECU	10% chance of 38.5 ECU and 90% chance of 1 ECU
2	20% chance of 20 ECU and 80% chance of 16 ECU	20% chance of 38.5 ECU and 80% chance of 1 ECU
3	30% chance of 20 ECU and 70% chance of 16 ECU	30% chance of 38.5 ECU and 70% chance of 1 ECU
4	40% chance of 20 ECU and 60% chance of 16 ECU	40% chance of 38.5 ECU and 60% chance of 1 ECU
5	50% chance of 20 ECU and 50% chance of 16 ECU	50% chance of 38.5 ECU and 50% chance of 1 ECU
6	60% chance of 20 ECU and 40% chance of 16 ECU	60% chance of 38.5 ECU and 40% chance of 1 ECU
7	70% chance of 20 ECU and 30% chance of 16 ECU	70% chance of 38.5 ECU and 30% chance of 1 ECU
8	80% chance of 20 ECU and 20% chance of 16 ECU	80% chance of 38.5 ECU and 20% chance of 1 ECU
9	90% chance of 20 ECU and 10% chance of 16 ECU	90% chance of 38.5 ECU and 10% chance of 1 ECU
10	100% chance of 20 ECU	100% chance of 38.5 ECU
Sw. point #	Subjects had to choose one of the following alternatives:	
0	Choose Asset B for ALL scenarios	
1	Choose Asset A for Scenario 1 and Asset B for Scenario 2 to 10	
2	Choose Asset A for Scenario 1 to 2 and Asset B for Scenario 3 to 10	
3	Choose Asset A for Scenario 1 to 3 and Asset B for Scenario 4 to 10	
4	Choose Asset A for Scenario 1 to 4 and Asset B for Scenario 5 to 10	
5	Choose Asset A for Scenario 1 to 5 and Asset B for Scenario 6 to 10	
6	Choose Asset A for Scenario 1 to 6 and Asset B for Scenario 7 to 10	
7	Choose Asset A for Scenario 1 to 7 and Asset B for Scenario 8 to 10	
8	Choose Asset A for Scenario 1 to 8 and Asset B for Scenario 9 to 10	
9	Choose Asset A for Scenario 1 to 9 and Asset B for Scenario 10	
10	Choose Asset A for ALL Scenarios	

TABLE 4 | Average switching points in the literature.

Source	Value
He et al. (2012)	4.48 (5.25, 4.84)
Bellemare and Shearer (2010)	4.62
Anderson and Mellor (2008)	4.98
Dave et al. (2010)	5.00
Holt and Laury (2002)	5.17
Harrison et al. (2005); Goeree et al. (2003)	5.3
Lusk and Coble (2005)	5.42
Baker et al. (2008)	5.67
Anderson and Mellor (2009)	5.68
Holt and Laury (2005)	6.10
Masclot et al. (2009)	6.6 (6.9)

to choose once more their switching point from the same design that they were given in stage 1.

From an HS&E perspective, the information provided in stage 2 mirrors situations in which decision makers receive benchmarks or summaries of peer behavior, such as industry safety averages,

best-practice reports, or expert committee recommendations. These signals often shape subsequent safety-related decisions, even when they are abstracted from the specific operational environment. Stage 2 therefore captures how exposure to comparatively risk-tolerant reference behavior may influence the revision of safety thresholds and precautionary standards.

Stage 3: Risk-averse reference (we will refer to this stage as H7/H9) In this stage, we split participants into two groups. The first group, 184 participants (75%), was told that the average switching point of a selected previous session was seven (H7), and the second group, 60 participants (25%), was informed that the switching point of a selected previous session was nine (H9). After receiving the information, participants were again asked to choose their switching point from the same setup that they were given in stage 1.

Stage 3 corresponds to a contrasting but equally prevalent HS&E scenario: the reception of highly precautionary reference information, such as reports following accidents, regulatory tightening, or evidence of near misses in comparable organizations. By varying the degree of risk aversion conveyed by the reference group, we are able to study whether decision makers respond proportionally to increasingly conservative signals, a key issue in understanding how safety standards escalate, or fail to escalate after adverse events.

Stage 4: Payoff preference. After the experiment, participants were queried about their preference regarding which of the three stages should determine their payoff. Given that participants understood the selection of the stage for calculating their payoff was random and their response would not influence their payout, their choice provides insight into their risk attitude after completing all three stages. This gives us an opportunity to separate the fleeting effects of information on their risk attitude from the more permanent ones.

In HS&E settings, organizations must often decide whether to adhere to initial safety assessments or adopt revised standards in light of new information. Stage 4 mirrors this trade-off by allowing participants to select which of their previously adopted preferences should determine their payoff. This choice provides insight into whether decision makers regard updated preferences as superior and more reliable, or whether they prefer to rely on their initial judgment. As such, this stage captures an essential dimension of risk governance: the balance between stability of rules and adaptability to new evidence.

After participants finished the experiment, they were asked to fill out a questionnaire asking for their demographic characteristics, such as age, nationality, major, gender, and level of trust (Glaeser et al. 2000). The questions concerning social trust were trust in traffic controllers, foreigners (both domestic, that is, “strangers,” and international), accountants, and pilots. Regarding institutional trust, we enquired trust levels toward the police, public authorities, banks, and churches. A four-level Likert scale was employed to measure the level of trust, ranging from “No trust at all” to “Little trust,” “Quite a bit of trust,” and “A lot of trust.” We utilize the factor analysis approach to construct the social trust index and institutional trust index.

Following Chmura et al. (2022), who found that finance students are less likely to follow leaders, we measured confidence by conducting an experiment with a separate group of finance students. In addition, we assessed social trust using a well-established survey introduced by (Brehm and Rahn 1997; Fehr et al. 2002; Zmerli and Newton 2008). For institutional trust, we adopted Gallup’s approach to measure trust levels in various institutions, such as the government, large corporations, the press, courts, and banks. To construct social trust and institutional trust indices, we employed the principal component analysis (PCA).

We will now state the theoretical predictions of the experimental results.

According to the model laid out in Section 2, if we confront somebody with a choice between lotteries, the decision maker aims at gauging the evolutionarily optimal level of risk aversion, and then she adopts the preferences associated with that level. In our experiment, this would be happening in the first stage. Accordingly, we interpret the preferences she displays in the first round as those associated with the level of r , she initially considers evolutionarily optimal for the choice problem at hand.

In the experiment, once the participant displays her preferences, we inform her about the average switching point of a reference group (the *signal*), and she is asked to choose again. Proposition 2

predicts that if her original choice corresponded to Interval $j \in \{1, \dots, n\}$, and she is informed that the average choice of the reference group corresponded to Interval $k \in \{1, \dots, n\}$, her new choice will correspond to an Interval weakly between j and k (i.e., j and k included). In terms of the associated switching points, this translates into:

Hypothesis 1. *The switching point of the choice after information was obtained will be (weakly) between her original switching point and the switching point of the reference group.*

In the model, the parameter λ captures the confidence in her a priori judgment relative to how much she trusts the choices of others. She may be very confident in her own assessment of the situation and mostly dismisses the information about others’ choices, in which case her λ is close to 0; or she is rather insecure and assigns a lot of significance to what others do, in which case her λ is close to 1. The simulation (Table 2) suggests that the strength of the movement in response to the signal positively correlates with the value of λ .

The parameter λ can be interpreted having two components and is, in a conceptual sense, their *ratio*: (1) trust in other agents’ judgments and (2) confidence in her own assessment.²³ These two components can be proxied by traits of the decision makers; for example, “trust in authorities” may be a proxy for (1), and experience with risky choice and expertise in statistics may be a proxy for (2). We therefore state:

Hypothesis 2.1. *Confidence in her own judgment is negatively correlated with how far the decision maker moves toward the signal.*

Hypothesis 2.2. *Trust in others is positively correlated with how far the decision maker moves toward the signal.*

In the context of HS&E risk analysis, the parameter λ can be interpreted as capturing a fundamental tension between expert confidence and institutional trust. High confidence may reflect expertise and experience, but it may also lead to insufficient responsiveness to external warnings. Conversely, high trust in others can facilitate organizational learning but may also result in excessive conformity. By linking λ to measurable traits such as trust in institutions and confidence in judgment, our experiment provides a framework for diagnosing when social learning improves safety decisions and when it may undermine them.

In the experiment, subjects were given information about the average choices of reference groups *twice*. In the end of the experiment, they were asked to choose from which of the three preferences they adopted they want to receive their payoffs.

Our model suggests that from an evolutionary perspective, the adopted preferences can only get “evolutionarily better,” if anything, when additional information comes in. Information can never lead to a worse estimate: if the decision maker believed that the information is worthless because, say, the reference group has less understanding of the situation than herself, this would correspond to $\lambda = 0$, and the information would have no impact on the preferences displayed. This leads to:

TABLE 5 | Individual risk preferences in the three stages, based on full sample.

	Base	H3	H7	H9
Switching point	6.475	6.209	6.793	6.466
<i>p</i> -value (Base vs. H3)	0.001***			
<i>p</i> -value (H3 vs. H7)	0.000***			
<i>p</i> -value (H3 vs. H9)	0.063*			
<i>N</i>	244	244	184	60

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ **TABLE 6** | Individual risk preferences in the three stages, based on subsample of participants with no finance background.

	Base	H3	H7	H9
Switching point	6.666	6.282	6.790	7.090
<i>p</i> -value (Base vs. H3)	0.000***			
<i>p</i> -value (H3 vs. H7)	0.000***			
<i>p</i> -value (H3 vs. H9)	0.046**			
<i>N</i>	195	195	162	33

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ **TABLE 7** | Individual risk preferences in the three stages, based on subsample of participants with finance background.

	Base	H3	H7	H9
Switching point	5.714	5.918	6.818	5.703
<i>p</i> -value (Base vs. H3)	0.945			
<i>p</i> -value (H3 vs. H7)	0.455			
<i>p</i> -value (H3 vs. H9)	0.543			
<i>N</i>	49	49	22	27

Hypothesis 3. *Subjects want to have the payoffs from the round where they displayed their most updated preferences.*

5 | Experimental Results

Tables 5–and 7 present the individual risk preferences observed across the three treatments: Base, H3, and H7/H9. Table 5 presents the results for all participants (244), Table 6 for participants without finance backgrounds (184), and Table 7 for participants with finance backgrounds (60). On average, participants exhibit a risk-averse mentality, with an average switching point of 6.475, which aligns with existing literature (e.g., Holt and Laury 2005; Masclet et al. 2009). The *p*-values in the tables correspond to the null hypothesis that risk preferences in the Base and H3; H3 and H7/H9 are equal, using the Mann–Whitney U test.

In order to measure the strength of changing risk attitudes, we employ two binary variables which we call weak herding (*h-weak*) and strong herding (*h-strong*). *h-weak* is 1 if a participant does not shift her risk preferences in response to the signal or moves

TABLE 8 | Preference changes among participants with and without financial training—primary variables.

	All	Nonfinance	Finance	<i>p</i> -values
<i>h-weak</i>	70.90%	73.85%	59.18%	0.043**
<i>h-strong</i>	22.54%	24.62%	14.29%	0.122
<i>N</i>	244	195	49	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ **TABLE 9** | Preference changes among participants with and without financial training—secondary variables.

	All	Non-finance	Finance	<i>p</i> -values
H Mag	0.803	0.953	0.204	0.037**
Herd (both)	26.23%	27.18%	22.45%	0.501
Contrarian (both)	4.92%	4.10%	8.16%	0.240
H012 weak	1.647	1.697	1.448	0.022**
- H0 weak	6.15%	4.10%	14.29%	
- H1 weak	22.95%	22.05%	26.53%	
- H2 weak	70.90%	73.85%	59.18%	
H012 strong	0.790	0.835	0.612	0.078*
- H0 strong	43.44%	41.03%	53.06%	
- H1 strong	34.02%	34.36%	32.65%	
- H2 strong	22.54%	24.62%	14.29%	
<i>N</i>	244	195	49	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

toward the signal, and 0 otherwise. *h-strong* assumes the value 1 only if a participant strictly changes her preference toward the signal (i.e., shifting toward 3 in stage 2 and toward 7 or 9 in stage 3), and 0 otherwise.

We define further binary variables to capture other dimensions of preference changes that allow for various interpretations and follow-up analyses. These are magnitude of risk shifting (herding magnitude, *h-mag*), direction of risk shifting (herding in both stages, *herd*), moving against the signals in both stages (*contrarian*), and the frequency of risk shifting (no strong herding once or twice, *H012-strong*; no weak herding, once or twice *H012-weak*). While we do not use these variables to test the hypotheses, they will serve to formulate some exploratory findings at the end of this section.

Table 8 presents the two main measures, and Table 9 exhibits the secondary measures.

Original switching point and reference's switching point

The findings reported in Tables 5–7 suggest that in treatments H3 and H7/H9, participants adjust their risk preferences in response to the provided signals, and these differences are statistically significant. This supports Hypotheses 1. More specifically, the average switching point in the full sample in treatment H3 was 6.209, while in H7/H9, it was 6.793/6.466 (Table 5). Figure 1

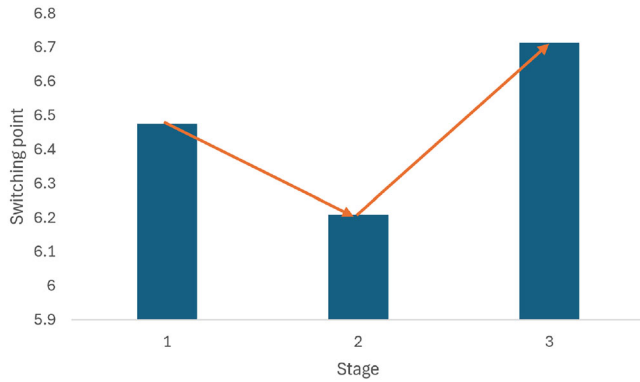


FIGURE 1 | Risk preferences in the three stages followed the signals.

shows how the average switching point chosen by the participants follows the signals.

Confidence, trust and risk preferences

Results in Table 8 that show that 70.90% of participants exhibit weak herding behavior, while 22.54% engaged in strong herding are also supportive of Hypothesis 1.

The findings reported Tables 6 and 7 suggest that the decision to shift risk preferences toward the signals was statistically significant among nonfinance participants but not among those with finance backgrounds. While the switching point of nonfinance individuals is 6.666 in the Base, 6.282 in H3, and 6.790/7.090 in H7/H9, these numbers are 5.714, 5.918, and 6.818/5.703, respectively, for individuals with finance training. This finding is in line with Hypothesis 2.1, that is, individual characteristics such as financial training increase a participant's confidence in her own judgment. This is also consistent with results in Table 8 that participants without finance backgrounds are more likely to conform to informed risk preferences. Table 9 reveals similar results across different measures. In terms of the magnitude of risk-shifting, h -mag is 0.803 with all participants but is significantly higher with nonfinance participants. Overall, our findings support Hypothesis 2.1.

Next, we explore individual trust and its relationship with preference-changing behavior. Specifically, we test if trust in others leads individuals to adjust their risk preferences stronger toward the signal. To explore this, we construct a social trust index and institutional trust index with PCA, based on various questions in the survey regarding both social trust and trust in institutions (Glaeser et al. 2000).

The results in Table 10 indicate that the level of trust has a negative effect on risk-shifting behavior. Participants are less inclined to alter their risk preferences if they exhibit a high level of social and institutional trust (Table 11 B). However, the picture is not clear if one looks at individual questions. The correlations between h -weak and "trust in courts," "trust in the EU," and "trust in foreigners" are 0.23, 0.15, and 0.22, all of which are significant with $p < 0.05$.

Taken all of this together, we *weakly reject* Hypothesis 2.2.

TABLE 10 | Adjustment of risk preferences, trust, and personality.

	Main risk-shifting measures	
	h -weak	h -strong
Age	−0.108*	−0.037
Maths	0.125*	0.017
Risk preferences	0.059	−0.019
Social trust	0.080	−0.059
Institutional trust	0.008	−0.084
Extraversion	−0.041	−0.009
Agreeableness	−0.048	−0.045
Conscientiousness	0.050	0.046
Neuroticism	0.002	0.068
Openness	−0.027	−0.107*
N	244	244

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Updated risk preferences

After the first three stages, participants were asked from which stage they would prefer to receive their payment. Overall, 29.10% of participants opted for payment based on their original decisions in the first stage, while 38.11% preferred payment based on the second stage results, and 32.79% preferred payment from the third stage. For testing Hypothesis 3, we excluded participants who made the same choice in all three stages.

For those who remain, the question is whether participants tended to choose the payoff from the last stage in which they shifted their risk preferences. The first two columns of Table 12 show that 50% of participants who shifted in stage 2 and remained the same in stage 3 selected the payoff from stage 2 (in fact, 83.33% chose the payoff from the last stage in which they shifted, since in this case, risk preferences in stages 2 and 3 were the same). Additionally, 43.10% of participants who shifted in both stage 2 and stage 3 selected the payoff from stage 3. This is consistent with Hypothesis 3. Results in the last column are not consistent with Hypothesis 3, only 26.92% of participants who remained the same in stage 2 and shifted in stage 3 selected the payoff from stage 3. However, it appears that the behavior of majority of participants is consistent with Hypothesis 3. It should also be noted that the behavior of those participants who did not change their preferences at all, whom we did not include in Table 12, is in line with Hypothesis 3. It is also worth pointing out that there was a relatively small sample size of participants who changed their preferences twice, namely, only 52 participants.

We also found some noteworthy results not directly related to the model. First of all, we examined whether gender influences the decisions to shift individual risk preferences. The results in Table 13 indicate that females are more inclined to shift their risk preferences. Specifically, 26.87% of females exhibit a strong shift, compared to only 17.27% of males, and this difference is statistically significant. Similarly, 30.60% of females follow the signals (herd) in both stages, while this percentage is 20.91% for males. Additionally, the frequency of strong risk shifting

TABLE 11 | Individual characteristics and risk preference shifting.

	Other risk-shifting measures				
	h-mag	Herd	Contrarian	H 012-weak	H 012 -strong
Age	−0.113*	−0.043	0.115*	−0.067	−0.116*
Maths	0.098	0.004	−0.061	0.133**	0.030
Risk preferences	0.085	−0.076	−0.166***	0.092	0.032
Social trust	−0.039	−0.106*	−0.054	0.083	−0.061
Institutional trust	−0.077	−0.127**	−0.060	0.020	−0.100
Extraversion	0.014	−0.009	−0.068	−0.028	0.039
Agreeableness	−0.021	−0.002	−0.009	−0.046	0.002
Conscientiousness	0.139**	0.015	0.051	0.048	0.124*
Neuroticism	−0.000	0.052	0.059	−0.005	−0.011
Openness	0.006	−0.105	0.029	−0.031	−0.020
<i>N</i>	244	244	244	244	244

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ **TABLE 12** | Payoff preferences among participants who shifted their risk preferences.

	Shift in S2 & stay same in S3	Shift in S2 & S3	Stay same in S2 & shift in S3
Payoff 1	16.67%	26.72%	23.08%
Payoff 2	50.00%	30.17%	50.00%
Payoff 3	33.33%	43.10%	26.92%
<i>N</i>	24	116	52

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ **TABLE 13** | Changing risk preferences and gender.

	All	Female	Male	<i>p</i> -values
<i>H</i> (weak)	70.90%	72.39%	69.09%	0.573
<i>H</i> (strong)	22.54%	26.87%	17.27%	0.075*
<i>N</i>	244	134	110	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

is significantly higher among females; 26.87% of females make strong shifts twice, compared to 17.27% of males. There is a body of literature on herd behavior and gender. For instance, studies such as Gupta and Goyal (2024) and Talpsepp and Tănav (2021) find no gender gap in following influencers, while Zheng et al. (2021) suggest that female investors are more likely to herd in the Chinese stock markets. Our result contributes to this literature by highlighting the gender gap in individual risk preference shifting.

Table 14 presents other correlations between individual characteristics and the change of risk preferences. The results show that age is negatively and significantly correlated with *h*(weak), *hmag*, and *h012weak*, while positively and significantly correlated with contrarian behavior. As participants grow older, they are less likely to shift their risk preferences toward the signals but against the signals. This finding supplements the literature regarding age and herd behavior. Lamont (2002) indicates that age is negatively

TABLE 14 | Individual characteristics and risk preference shifting.

	H (weak)	H (strong)
Age	−0.108*	−0.037
Maths	0.125*	0.017
Risk preferences	0.059	−0.019
Social trust	0.080	−0.059
Institutional trust	0.008	−0.084
Extraversion	−0.041	−0.009
Agreeableness	−0.048	−0.045
Conscientiousness	0.050	0.046
Neuroticism	0.002	0.068
Openness	−0.027	−0.107*
<i>N</i>	244	244

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

correlated with hearing decisions among American forecasters, while Ashiya and Doi (2001) claim that Japanese forecasters herd regardless of their age.

The Big Five personality traits also have significant impacts on risk preference shifting, wherein conscientious participants are more likely to conform to the informed risk preferences with

a greater magnitude, while open participants tend to make the opposite decisions.

6 | Discussion

We interpret the experimentally elicited switching points as *risk-acceptability thresholds*, analogous to the safety cutoffs that guide decisions in HS&E contexts. From this perspective, changes in switching points across stages reflect the updating of safety thresholds in response to social and benchmark information, a central mechanism in adaptive risk management. The experimental results therefore shed light on why decision makers respond heterogeneously to identical risk signals, with some exhibiting inertia and others substantial adjustment.

Analyzing the initial switching points in the Base treatment reveals that participants with financial training are more likely to take risks compared to nonfinance participants (5.714 vs. 6.666, p -value = 0.048). This suggests that before any information exchange, confidence in one's own judgment, that is, the inverse value of λ , is negatively correlated with risk aversion. This may reconcile our findings with the known phenomenon of *risky shift* (Wallach et al. 1962), which is the observation that group decisions tend to be less risk averse than individual decisions. Assuming that group decisions are based on information sharing among the group members, the "group λ " will usually be lower than the individual λ 's of the group members so that the group acts more risk-loving than the individual members.

Interpreted through an HS&E lens, the initial switching point can be read as a safety threshold—a cutoff at which a decision maker is willing to sacrifice expected performance (or convenience) for reduced downside risk. In many HS&E settings, analogous cutoffs determine whether a process continues, whether precautionary measures are adopted, or whether activities are halted under uncertainty. The observed difference between finance and nonfinance participants therefore suggests that domain training may shift the location of such thresholds even before any social information is introduced, with implications for heterogeneity in precautionary decision-making within teams and across roles (e.g., operators, engineers, and managers).

A question is why Hypothesis 2.2 was rejected. One possibility would be that the model parameter λ is determined only by confidence in one's own judgment, as suggested by the confirmation of Hypothesis 2.1, but not by trust in others. This would not challenge the model itself but rather our ex-ante interpretation of the λ parameter, not requiring a change to the architecture of our theory.

From the perspective of HS&E risk analysis, the rejection of Hypothesis 2.2 may reflect that the relevant construct is not "trust in others" in a general, survey-based sense, but rather credibility-weighting of specific information channels (e.g., trust in regulators, incident databases, safety leadership, or technical experts) (Cox et al. 2006). In operational environments, decision makers often exhibit high trust in some sources (e.g., internal engineers) and low trust in others (e.g., corporate messaging), and this source-specific credibility is precisely what governs whether safety thresholds are updated. Thus, our null finding may indicate

a measurement mismatch: the "trust" items capture a broad disposition, while the model's λ may be driven by contextual trust in the relevance and competence of the reference group.

There are also more general objections against the idea that people are optimizing an evolutionary parameter through the adoption of preferences. One is the apparent lack of introspective support. We, and most people with whom we discussed our theory, do not perceive decision situations, not least choices between *lotteries*, to be about maximizing offspring and survival. Introspectively, many people question whether they are investigating their preferences *at all* when choosing among lotteries. Rather, they think that they base their decisions on quantitative aspects of those lotteries, like their expected values. This cognitive understanding is built into our theory, where experience and cognitive understanding influence the parameter λ , but for many people, the cognitive component seems not to square with the idea that preferences result form an underlying evolutionary optimization.

Our response to these objections is that if the processes posited in this paper happen, then they must be subconscious, at least when it comes to their hypothesized evolutionary foundations. A person choosing among lotteries will not ask: "What risk should I take to optimise my evolutionary outcome?" Rather, she will ask: "What risk *feels right* for this situation?"

This mechanism is particularly plausible in HS&E contexts because many safety decisions are made under time pressure, ambiguity, and incomplete probabilistic knowledge (Pooladvand and Hasanzadeh 2022; Pawar and Velaga 2022). Even when formal risk matrices or quantitative risk assessments are available, day-to-day choices frequently rely on intuitive thresholding—whether a situation "looks unsafe," whether conditions "seem within limits," or whether a deviation "feels acceptable." Our framework provides a disciplined way to interpret such judgments as the output of preference formation and updating rather than as noise or purely cognitive calculation.

This does not mean, however, that her cognitive apparatus, her *understanding of the situation*, has no influence on her preferences. To the contrary, like many other preferences and emotions, risk preferences are almost surely *cognitively moderated* (see, e.g. Zajonc and Markus 1982; Schkade and Johnson 1989; Betto 2024). To illustrate what this means, consider another evolutionarily consequential decision the reader is certainly familiar with, namely, whether or not to eat another portion of powdered-sugar doughnuts.²⁴ If you have to make this decision, you will primarily interrogate your appetite and hunger and not consciously analyze the effects of the doughnut consumption on your reproductive success. Nonetheless, your desire for the doughnut may be *cognitively moderated*: awareness of the facts that eating the doughnuts may impact your body weight or blood sugar levels may make you resist the temptation, override the immediate craving for sugar, and possibly impact your preferences to an extent that you do not *want* to eat the doughnuts anymore (White et al. 2020; Calabro et al. 2023, survey research on the impact of cognitive processes on food preferences).²⁵ On the other hand, if you know that a 30-mile hike is ahead of you, you may have no qualms whatsoever to eat the doughnuts, and you may actually *enjoy* them more than if you were mostly concerned with your

weight. In both cases, conscious considerations feed back to the basic preference level.

In HS&E practice, this type of cognitive moderation corresponds to the effect of training, procedures, and decision aids: technical knowledge can shift which risks are attended to, how salient worst-case outcomes feel, and how strongly new information is weighted. In our model, this maps naturally into λ as the relative weight on one's own assessment versus socially transmitted information. The empirical differences by financial training are therefore consistent with the broader HS&E premise that expertise can stabilize thresholds, sometimes appropriately (robust judgment), sometimes inappropriately (excessive confidence).

It was first pointed out by Maslow (1943) that some preferences are highly instinctual, while others are more intellectual; along this spectrum, it seems likely that preferences for lotteries belong to the latter.²⁶ However, as the above examples show, even if sophisticated cognitive reasoning is involved in a decision-making process, the outcomes may still be genuine *preferences*, that is, *valence* attached to different options, in the same way as the desire for vengeance can be felt viscerally even though it may depend on a sophisticated cognitive interpretation of a complex social situation.²⁷

Cognitively moderated preferences can, like all other preferences, be grounded in evolutionary history. The optimal risk attitude when deciding whether or not to attack a dangerous but nutritious animal, like a mammoth, will depend on various parameters: How large and dangerous is the animal, and how likely are different outcomes that can result (e.g., the mammoth being killed, oneself being killed or injured, or oneself having to abandon the attack)? How are these probabilities affected by the weapon that is available (e.g., spear, stone, club) and by the current physical fitness? How likely is it that other members of the horde will join the attack? And, depending on their contribution to the project, what share of the spoil will they claim if the attack was successful? The answers to these questions depend on cognitive analysis, but they lead to the emergence of *genuine preferences*, namely, the appetite for taking the risk of attacking the mammoth.

Another aspect of our theory not intuitive to everyone is the *context dependence* of the evolutionarily optimal level of risk aversion. However, context dependence is an obvious fact for many other preferences:²⁸ the desire for food does not emerge if one has just eaten, a lust for aggression or wish to flee usually requires to be in a fighting or conflict situation, and social preferences, say, for harmony, may depend on who are the people one is interacting with.

The context dependence of risk preferences has been discussed for a long time (Wolf and Pohlman 1983), and there is considerable empirical affirmative evidence from different domains (e.g., Karni et al. 1983; Barseghyan et al. 2011; Einav et al. 2012; Schildberg-Hörisch 2018). There is, of course, also the related debate about the connection between risk aversion and a decision maker's *wealth* (which can be considered a part of the *context*),²⁹ although the empirical picture regarding the correlation between risk aversion and wealth is not unambiguous (see the findings and the discussion of the empirical literature in Rieger et al. 2015).

Summing up, we do not think that context dependence is an aspect of our theory that is difficult to accept.

Importantly, the contribution of our experimental results to HS&E is not that a lottery task directly reproduces industrial or public-health hazards, but that it isolates a general mechanism that those environments depend on how safety thresholds move when decision makers receive social information. Because HS&E risk governance frequently relies on benchmark comparisons, incident learning, and expert consensus signals, understanding when individuals update, how far, and which traits predict responsiveness, speaks directly to the design of effective risk communication and organizational learning processes.

7 | Relevance in Health, Safety, and Environment Contexts

In public health contexts, two largely separate determinants of behavior in situations involving risk have been identified. The first one is, as may have been expected, the *risk attitude*. Experimentally measured risk aversion predicts health-related behaviors such as smoking, heavy drinking, obesity, and seatbelt use (Anderson and Mellor 2008). Even subtle framing differences in risk-elicitation tasks can shift measured preferences and change behaviors (Yang et al. 2022; Van der Pol and Ruggeri 2008).

The second determinant are *peer effects*, sometimes leading to “contagiousness” of harmful activities like smoking, alcohol, and drug consumption (Alexander et al. 2001; Fergusson et al. 2002; Duncan et al. 2005; Powell et al. 2005; Lundborg 2006; Clark and Lohéac 2007). In psychological research, however, peer effects are typically not explained as the result of information transmission and learning from others, as it is the case in our theory, but as a result of, for example, social norms (Tajfel et al. 1971), reciprocity (Vaquera and Kao 2008; Lin and Weinberg 2014), and group identity (e.g., Goette et al. 2006; Charness et al. 2007; Chen and Li 2009; Gioia 2017, the last paper focusing on risk-taking behavior).³⁰ Our evolutionary hypothesis combines and extends both explanations: the existing risk preferences matter, but peer effects, which work through information sharing, contribute to the formation of those preferences.

Our model and experimental results can also explain the dynamics of whether and when risk preferences change. Arguably, this aspect has at times been neglected as much of the existing literature views or finds (risk) preferences to be basically stable (Stigler and Becker 1977; Andersen et al. 2008; Dasgupta et al. 2017).³¹

In contrast, and in line with the ideas we advance in this paper, a substantial body of research in HS&E shows that risk preferences and risk acceptability can shift frequently, even in routine, noncrisis settings. In line with our theory, risk-taking in these contexts is often shaped by local experiences, perceived exposure to hazards, and social cues and peer behavior. For example, views on pesticide policies in Canadian cities are not primarily driven by general environmental preferences but by residents' personal experiences with chemical use and conflict over the use of pesticides (Hirsch and Baxter 2011). In public health, shocks and framing effects can trigger short-run shifts in

risk attitudes (Decker and Schmitz 2016; Van der Pol and Ruggeri 2008), consistent with the idea that risk preferences are adjusted in response to new information.

Environmental valuation studies also demonstrate these dynamics: willingness to pay for mortality risk reductions varies with local pollution levels and the salience of specific causes of death (Lindhjem et al. 2011; Guignet and Alberini 2015). Likewise, workplace safety research demonstrates how information acquired through communicating with and observing others can alter perceived risks and protective behavior (Wakefield et al. 2010; Dean 2014; Maglio et al. 2016; Wong et al. 2020; Cakir et al. 2023).

Across these domains, HS&E research consistently shows that risk preferences and risk acceptability are dynamic and influenced by social learning, trust, and domain-specific knowledge, precisely the mechanisms for which our framework provides an evolutionary foundation.

7.1 | Escalation of Safety Thresholds Following Process Safety Deviations

A central challenge in occupational and industrial safety is deciding whether ongoing operations should be halted, modified, or allowed to continue following indications that safety margins have been eroded. In many high-risk industries, such as chemical processing or refining, catastrophic accidents are often preceded not by a single dramatic event but by a series of process deviations, warning signs, or near-miss incidents that signal elevated risk without immediately causing harm. Well-documented cases, such as the events preceding the Texas City refinery explosion, illustrate how such warning signals can be interpreted very differently across organizations and over time (Reason 1998; Vaughan 1996; Phimister et al. 2003).

From an HS&E risk analysis perspective, these decisions can be understood as threshold problems: managers and engineers implicitly compare their perceived probability of catastrophic failure against an acceptable risk level that determines whether operations are stopped, modified, or continued. If perceived risk exceeds this threshold, shutdowns or corrective actions are triggered; if not, operations proceed (Kang 2024; Grechuk and Zabaranin 2014). This structure closely parallels the experimental switching point in our design, which captures the cutoff at which decision makers are willing to trade expected performance for reduced downside risk.

Information about safety deviations rarely comes in isolation. Organizations typically interpret local warning signs in light of external reference information, such as incident statistics from peer facilities, industry benchmarks, regulatory guidance, or past operational experience. In terms of our model, these reference points constitute a *signal* about the operationally appropriate level of risk tolerance in the given context. Our framework predicts that the extent to which safety thresholds are revised in response to such signals depends on the parameter λ , which captures the relative weight placed on external information versus confidence in local judgment. In work-safety settings, λ can be interpreted as the balance between trust in industry-

wide safety knowledge and confidence in site-specific expertise, engineering assessments, and operational familiarity.

Two distinct failure modes follow naturally from the model. When confidence in local judgment is very high (i.e., λ is low), organizations may insufficiently update their safety thresholds despite accumulating warning signals, leading to the normalization of deviance and the gradual acceptance of elevated risk. Conversely, when reliance on external benchmarks is strong and confidence is low (i.e., λ is high), organizations may overadjust to salient but noisy signals, triggering overly conservative shutdowns or costly operational disruptions.

From a risk management perspective, the framework highlights that improving work-safety decisions is not only a matter of collecting more safety data, but also of understanding how different actors weight internal experience against external benchmarks. Effective HS&E governance therefore requires not only transparent reporting of safety signals, but also mechanisms that calibrate confidence appropriately, distinguishing between credible systemic warnings and transient noise, and tailoring communication to the expertise of decision makers.

7.2 | Extensions to Other HS&E Domains

Similar threshold-updating problems arise in public health decision-making. Individuals and policymakers must decide whether to adopt, maintain, or relax preventive measures, such as vaccination, mask usage, or dietary guidelines, based on perceived health risks. These decisions depend on implicit acceptability thresholds for adverse outcomes, such as infection or morbidity risks. Typical signals include public health guidelines, expert consensus statements, epidemiological benchmarks, and observed compliance or behavior within peer groups (Morgano et al. 2022; Neumann et al. 2022; Meng et al. 2023). In terms of our model, these thresholds correspond to switching points, the external information constitutes the signal, and the parameter λ captures the balance between confidence in one's own assessment and trust in public health authorities or peer behavior. A key implication is that uniform risk communication may induce heterogeneous responses: highly confident individuals may underreact to credible guidance, while less confident individuals may overadjust to salient but noisy signals, suggesting a role for targeted and differentiated communication strategies.

Environmental risk management likewise involves the formation and revision of risk-acceptability thresholds, for example, in setting permissible pollution levels, approving infrastructure projects, or determining acceptable failure probabilities for environmental hazards (Mourato 2006; Fei et al. 2023). Decision makers often rely on external reference points such as regulatory standards, international benchmarks, peer jurisdictions, or high-profile environmental incidents. Within our framework, these reference points act as signals that inform updates of acceptable risk thresholds, with λ reflecting confidence in local expertise relative to trust in external standards and institutions. The model predicts that differences in this balance help explain why some authorities rapidly align with external benchmarks, while others resist adjustment. From a design perspective, the framework highlights the importance of calibrating benchmark information

to institutional context in order to promote adaptive learning without triggering either inertia or excessive precaution.

8 | Conclusion

This paper offers a decision-theoretic model in which the adopted preferences depend on an underlying evolutionary parameter. Agents have different confidence in their ability to gauge that parameter and use the observation of other decision makers to improve their estimates. This allows for an experimental test of the model which we have carried out. By and large, the results are supportive of the model.

Interpreted through an HS&E lens, the model and experimental findings speak directly to how risk acceptability thresholds are formed and revised when individuals observe the behavior or judgments of others. In many HS&E domains, such as occupational safety, environmental regulation, and public health decision makers rely on benchmark information, incident statistics, expert panels, or peer practices to assess whether risks remain within acceptable bounds. Our results show that such information does not affect all decision makers uniformly: responsiveness depends systematically on confidence, experience, and the perceived value of others' judgments. This insight is central for understanding why safety standards sometimes update rapidly following new information and at other times remain surprisingly inert.

Our experiment also contributes empirical evidence on the questions of stability and influence factors of risk preferences. As future work, we may check whether the specific sequential setup of the experiment may have benefits for addressing some of the open question.

The framework developed in this paper directly contributes to HS&E research by offering a microfounded explanation for how context-dependent, socially informed changes in risk acceptability arise, even in the absence of major external shocks. Many HS&E decisions, from health-related behaviors and safety compliance to environmental risk perception and valuation, are known to be shaped by context, social learning, and the observation of others (see the literature surveyed in Section 7). Our model provides a structured mechanism for understanding how such context-dependent and socially informed adjustments in risk attitudes can arise. Future work might apply our model of evolutionary driven preference dynamics to explain specific shifts in health-risk behavior described in the literature (e.g., preventive actions or risky consumption) and workplace safety responses to leadership signals.

Beyond explaining observed behavior, our framework has practical implications for HS&E risk management. It suggests that the effectiveness of risk communication depends not only on the content of information but also on the confidence structure of the audience. Highly confident experts may appropriately discount weak signals but dangerously ignore credible warnings, while less confident decision makers may overreact to salient but uninformative cues. Recognizing this heterogeneity can inform the design of safety reporting systems, expert elicitation procedures, and organizational learn-

ing mechanisms aimed at improving decision quality under uncertainty.

In this way, our evolutionary theory could not only contribute to behavioral decision theory but also to the broader HS&E literature concerned with how individuals adapt their risk preferences across different environments and informational settings.

While HS&E contexts provide a particularly salient domain for applying our framework, given their reliance on thresholds, precaution, and learning from others, the underlying mechanism is more general. Beyond HS&E may be other applications of the model framework developed in this paper. As mentioned, the evolutionary parameter does not need to be *risk aversion* but could also represent degrees of other behavioral factors which affect the prospective number of offspring. Like in the case of risk aversion, a *thrive function* could map such underlying factors into actual, observable preferences.

For example, *patience* plays an important role in conflict situations. If someone demands a change of behavior from someone else, the latter should be given some time to change their ways before punitive measures are taken. Waiting not long enough may lead to an unnecessary escalation of the conflict with negative evolutionary consequences (including violence, injuries, and death). Yet, waiting too long with following-up on a threat or demand that was made will lead to a loss of credibility, yielding disadvantages in future negotiations and conflicts. Like in the case of risk aversion, there is likely an *evolutionarily optimal* level of patience. Analogous trade-offs arise in HS&E contexts, for example, in decisions about when to escalate safety concerns, enforce compliance, or intervene following early warning signs, where acting too quickly or too slowly can both have severe consequences.

Another example is *empathy*. Being empathetic toward nonrelatives can be evolutionarily advantageous because it may, among others, initiate evolutionarily beneficial tit-for-tat relations (see Axelrod and Hamilton 1981, and Chapter 12 in Dawkins 2016). However, when empathy is extended too generously, it can be exploited by others and then *reduce* the expected number of offspring. As before, this reasoning gives rise to the idea that there is an evolutionarily optimal level of empathy. In organizational safety settings, such dynamics may manifest in how managers respond to worker concerns, near-miss reporting, or whistleblowing, where excessive skepticism can suppress valuable information while excessive empathy may undermine enforcement and accountability.

With these initial thoughts, we leave it to future research to explore the applicability of the model framework introduced here to these and other behavioral parameters, both within HS&E risk analysis and in broader decision environments where learning, trust, and context shape adaptive behavior under uncertainty.

Endnotes

- ¹Game theorists have devised formal models to explain how the preference formation proceeds, for example, Samuelson (2001), Heifetz et al.

(2007), Dekel et al. (2007). These models are based on the assumption that, on a fundamental level, genes are “players” (Dawkins 2016), underpinning the extensive application of game theory in theoretical biology (e.g., Alexander 2023; Lewontin 1961; Smith 1982; Weibull 1997).

² This is not the first paper to theorize and speculate about the evolutionary origins of risk preferences. Similar ideas have been developed in different disciplines, notably in philosophy (e.g., Szpiro 1997; Okasha 2007; 2011), theoretical biology (Zhang et al. 2014; Heller and Nehama 2023), and (evolutionary) psychology (Barrett and Fiddick 2000; Slovic et al. 2004; Homma and Takezawa 2024).

³ A strict total relation $>$ is transitive, asymmetric (if $a > b$, then not $b > a$) and connected (if $a \neq b$, then $[a > b \text{ or } b > a]$). A strict total order does not allow for indifference between unequal elements, cutting away a great deal of case distinctions and other hassle without changing the model qualitatively.

⁴ We opt for restricting the range of r because in Section 3, we will model an agent’s *belief* about the correct value of r as a density function. Total ignorance about r would correspond to a uniform distribution, which is not defined on an unbounded support. Moreover, note that r ranges in the *open* interval $(\underline{r}, \bar{r}) = \{r \in \mathbb{R} \mid \underline{r} < r < \bar{r}\}$, meaning that we require r to be an *interior value* of the closed interval $[\underline{r}, \bar{r}] = \{r \in \mathbb{R} \mid \underline{r} \leq r \leq \bar{r}\}$. In the subsequent analysis, this technical assumption helps us to avoid complications that arise from the special cases $r = \underline{r}$ and $r = \bar{r}$. From a practical perspective, the assumption is innocent because our analysis holds for arbitrarily large intervals $(\underline{r}, \bar{r})$. If in a specific application, the cases $r = \underline{r}$ and $r = \bar{r}$ are relevant, it is possible to work with a larger interval $(\underline{r}^*, \bar{r}^*) \supset [\underline{r}, \bar{r}]$ for which $\underline{r}$ and $\bar{r}$ are interior values.

⁵ It is not easy to come up with a meaningful name for a function in a decision-theoretic model. Initially, we used the term *preference function*, but that name could mean anything. The name *thrive function* reflects the idea that if an animal or human adopts the preferences selected by the thrive function, he or she will thrive in an evolutionary sense.

⁶ The reader may notice a similarity between the values defined in Lemma 1 and David Bell’s (1988) concept of *one-switch utility functions*. We are grateful to Peter Wakker for making us aware of this conceptual connection.

⁷ As we are working with strict relations, we could equivalently have written the condition in the set as $l' > l'' \wedge l' \not\geq l''$.

⁸ If one considers a relation to be a subset of $L \times L$ and lines up the $|L|^2$ positions in $L \times L$ in a way that allows to express every relation as a sequence of zeros and ones (i.e., 0 if a pair is not in the relation and 1 if it is), the standard Hamming metric measuring the distance between those sequences would have double the value of the metric defined above applied to the original relations.

⁹ We assumed that the metric $d(\cdot, \cdot)$ had already been used elsewhere but could not find an exact match. Closest may be the *Swaps Index* in Apesteguia and Ballester (2015), which measures how many pairwise comparisons have to be reversed in a preference relation R in order to make it consistent with an empirical observation. As the empirical observation is not a full preference relation but a single choice that is potentially inconsistent with R , their measure does not constitute a metric on the set of preference relations. Halevy et al. (2018) do define a measure of how close two rankings are. As their measure is based on a loss function, and since they are dealing with budget sets, it is not the same as the metric above. We are grateful to Yoram Halevy for these literature pointers.

¹⁰ The definition of linearity may at first sight look asymmetric, but it can be shown that if it is fulfilled, then there are also values r^3, r^4 with $r' \leq r^3 < r^4 \leq r''$ such that for all $r \in (r^3, r^4)$, it holds $d(r', r) = 1$.

¹¹ Some experiments replicating Holt and Laury (2002) are listed in Table 4. Holt and Laury (2002) was recognized as the 5th most cited paper in the *American Economic Review* published between the years

2000 and 2010 (Brice and Montesinos-Yufa 2019). As of January 2025, Google Scholar lists a whopping 8532 citations of the paper.

¹² All CRRA utility functions are positive affine transformations of the function

$$U(x) = \begin{cases} x^{1-r} & \text{for } r < 1 \\ \ln x & \text{for } r = 1 \\ -x^{1-r} & \text{for } r > 1, \end{cases}$$

for which the Arrow–Pratt measure of relative risk aversion equals r (for details, see Wakker 2008).

¹³ If we would use another function from the CRRA family, the r -intervals associated with each switching point would change, but the preference relation associated with the switching point would be the same.

¹⁴ Without even using the specific form of the CRRA utility function, we can also recover most of the preference relation based on first-order stochastic dominance and transitivity (assuming that there is only one switching point). In the above case, we know by first-order stochastic dominance that $A_{10} > A_9 > \dots > A_2 > A_1$ and $B_{10} > B_9 > \dots > B_2 > B_1$. By transitivity then follows from switching point 5 not only that $A_5 > B_5$, but also $A_5 > B_4$, $A_5 > B_3$, and so on. Further such arguments are possible, but some comparisons remain undetermined if the utility function is not used to calculate the expected utility of the lotteries.

¹⁵ Our payoffs are not in dollars but in *Experimental Currency Units* (ECU), and the numerical values are 10 times larger.

¹⁶ This is guaranteed if the CRRA utility function generates a thrive function that is single-crossing and linear.

¹⁷ If r is not risk aversion, the single-crossing property requires a different motivation.

¹⁸ The possibility that the thrive function may be different for different decision makers may be relevant in some contexts, for example, for analyses outside of an evolutionary steady state, or if the Agents A and B come from separate populations that were exposed to different evolutionary pressures.

¹⁹ The existence of multiple medians is an unlikely case if the random variable derives from messy real-world scenarios. Specifically, if the probability mass function $\{p(i)\}_{i=1, \dots, n}$ is itself (partly) determined by random factors, and we can therefore consider it to be an n -dimensional continuous random variable, then the probability of multiple medians is 0.

²⁰ For a screenshot of the menu that was presented to the participants as well as the instructions given to the subjects, see Appendix B.

²¹ From here, one can derive that the decision maker has the risk attitude of risk-seeking in its standard definition, that is, she prefers a lottery over having its expected value as a sure payment.

²² Their finding confirms previous results according to which women typically hold less risky assets compared to men, for example, Barsky et al. (1997), Jianakoplos and Bernasek (1998), Sunden and Surette (1998) (though Schubert et al. 1999 is not in line).

²³ A parameter like λ may also drive the inclination of students to copy other students’ answers in an assessment situation: the more confidence one has in one’s own competence, and the less one trusts others, the lower will be the tendency to peek at the exam scripts of other students (Franklyn-Stokes and Newstead 1995; Khodabakhshzadehs and Shoaahosseini 2021).

²⁴ Some reader may note that the craving for powdered-sugar doughnuts features prominently in the book *Why Buddhism is True* (Wright 2017), which argues that Buddhism is a true religion in the sense that it anticipates central insights from evolutionary psychology.

²⁵ In most decision situations, the cognitive apparatus is balancing and juggling multiple desires, which can be seen as independent behavioral

drives as in the evolutionary psychology version of the *modular theory of mind* (Kurzban 2011). In this context, the rejection of the powdered-sugar doughnut can be interpreted as the overriding of your immediate craving by a more fundamental survival or mating instinct (say, because of the increased risk of heart disease and because of the negative impact of overweight on mating market outcomes), instincts that are likely to be cognitively moderated themselves.

²⁶The need for *sleep*, on the other hand, is highly instinctual, as shown by the fact that it is difficult to override fatigue even if the cognitive apparatus recognizes that in a given situation, the pursuit of this desire is detrimental (there is a body of literature on the interplay between cognition and sleep, for example, Tousignant et al. 2019; Ballot et al. 2021; Cárdenas-Egúisquiza and Berntsen 2022). This is illustrated by the fact that people struggle to resist fatigue even when that is extremely dangerous (e.g., while driving), despite an awareness of the danger and a determination to stay awake (Williamson et al. 2011). At the other end of the spectrum are various preferences that are highly contingent on a cognitive understanding of a situation. For example, genuine emotions of aggression and vengeance, and the preferences derived from those feelings, may depend on a cognitive interpretation of a situation, for example, these emotions may follow if one was insulted publicly but may be absent if the altercation happened in a private setting (Kim et al. 1998; Ijzerman et al. 2007; Torres and Bergner 2010). This difference may result from the potential status loss following from a meek response, which only enters the cognitive calculus in the public case (Santos et al. 2011). Similarly, retaliatory desires only emerge in the presence of a human adversary, not if the offender is a computer (De Melo and Gratch 2015; Adam et al. 2018), a distinction that requires strong cognitive input as it can only be made by a mind who holds a notion of the philosophically nontrivial difference between a person and a machine (Chalmers 1997). The idea that preferences are not only cognitively moderated but even serve *strategic* goals is at the heart of the game-theoretic solution concept of *mental equilibrium* (Winter et al. 2017).

²⁷Lotteries as choice objects may appear to be dull, but this is not always the case: *excitement* is one of the main motivating factors for *gambling* (Lee et al. 2007), which, on an abstract level, is usually not much more than choosing between lotteries.

²⁸This was pointed out both in theoretical and empirical papers, for example, Ciampi et al. (1982), Karni et al. (1983), Mazar et al. (2014).

²⁹In expected utility theory, risk aversion follows from the degree of concaveness of a utility function. If that function is “more concave,” then the risk premium for a given lottery and the Arrow–Pratt measure associated with that utility function have higher values (Arrow 1974). Because a strictly concave utility function of money that is strictly increasing inevitably converges to a linear function (to be precise: a straight horizontal line) when the argument of the function approaches infinity, it follows that for a given lottery, the decision maker who maximizes expected utility will adopt a risk attitude that more and more resembles risk neutrality the higher is the wealth level affected by the lottery outcomes. This point can be illustrated with the following sketch: **art.eps**

³⁰Information sharing and “advice” was identified as one source of peer effects in Lomi et al. (2011), and “knowledge spillovers” have been discussed as an explanation for peer effects among scientists (Waldinger 2010; Azoulay et al. 2010; Borjas and Doran 2012). These articles do not relate to the domain of risk-taking behavior, however, and do not offer an evolutionary explanation.

³¹Some research pursues the idea that risk preferences only change after a long time or when there are exogenous shocks such as the 1929 Great Depression, the 2008 Global Financial Crisis, the 2020 Covid-19 Pandemic, or war (Malmendier and Nagel 2011; Kim and Lee 2014; Guiso et al. 2018; Tsutsui and Tsutsui-Kimura 2022).

References

- Adam, M. T., T. Teubner, and H. Gimpel. 2018. “No Rage Against the Machine: How Computer Agents Mitigate Human Emotional Processes in Electronic Negotiations.” *Group Decision and Negotiation* 27: 543–571.
- Alexander, C., M. Piazza, D. Mekos, and T. Valente. 2001. “Peers, Schools, and Adolescent Cigarette Smoking.” *Journal of Adolescent Health* 29, no. 1: 22–30.
- Alexander, J. M. 2023. *Evolutionary Game Theory*. Cambridge University Press.
- Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström. 2008. “Eliciting Risk and Time Preferences.” *Econometrica* 76, no. 3: 583–618.
- Anderson, L. R., and J. M. Mellor. 2008. “Predicting Health Behaviors With an Experimental Measure of Risk Preference.” *Journal of Health Economics* 27, no. 5: 1260–1274.
- Anderson, L. R., and J. M. Mellor. 2009. “Are Risk Preferences Stable? Comparing an Experimental Measure with a Validated Survey-Based Measure.” *Journal of Risk and Uncertainty* 39, no. 2: 137–160.
- Apesteguia, J., and M. A. Ballester. 2015. “A Measure of Rationality and Welfare.” *Journal of Political Economy* 123, no. 6: 1278–1310.
- Apesteguia, J., M. A. Ballester, and J. Lu. 2017. “Single-Crossing Random Utility Models.” *Econometrica* 85, no. 2: 661–674.
- Aronson, E., T. Wilson, R. Akert, and S. Sommers. 2017. *Social Psychology*, Global Edition ed. Pearson Education.
- Arrow, K. J. 1951. *Social Choice and Individual Values*. John Wiley & Sons.
- Arrow, K. J. 1974. *Essays in the Theory of Risk-Bearing*. North-Holland.
- Ashiya, M., and T. Doi. 2001. “Herd Behavior of Japanese Economists.” *Journal of Economic Behavior & Organization* 46, no. 3: 343–346.
- Axelrod, R., and W. D. Hamilton. 1981. “The Evolution of Cooperation.” *Science* 211, no. 4489: 1390–1396.
- Azoulay, P., J. S. Graff Zivin, and J. Wang. 2010. “Superstar Extinction.” *The Quarterly Journal of Economics* 125, no. 2: 549–589.
- Baker, R. J., S. K. Laury, and A. W. Williams. 2008. “Comparing Small-Group and Individual Behavior in Lottery-Choice Experiments.” *Southern Economic Journal* 75, no. 2: 367–382.
- Ballot, O., Y. Daviaux, E. Sanz-Arigitá, et al. 2021. “Emotion Coping Strategies and dysfunctional Sleep-Related Beliefs Are Associated With Objective Sleep Problems in Young Adults with Insomnia.” *Sleep Medicine* 88: 180–186.
- Barrett, H. C., and L. Fiddick. 2000. “Evolution and Risky Decisions.” *Trends in Cognitive Sciences* 4, no. 7: 251–252.
- Barseghyan, L., F. Molinari, and M. Thirkettle. 2021. “Discrete Choice under Risk With Limited Consideration.” *American Economic Review* 111, no. 6: 1972–2006.
- Barseghyan, L., J. Prince, and J. C. Teitelbaum. 2011. “Are Risk Preferences Stable Across Contexts? Evidence From Insurance Data.” *American Economic Review* 101, no. 2: 591–631.
- Barsky, R. B., F. T. Juster, M. S. Kimball, and M. D. Shapiro. 1997. “Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study.” *The Quarterly Journal of Economics* 112, no. 2: 537–579.
- Bell, D. E. 1988. “One-Switch Utility Functions and a Measure of Risk.” *Management Science* 34, no. 12: 1416–1424.
- Bellemare, C., and B. Shearer. 2010. “Sorting, Incentives and Risk Preferences: Evidence From a Field Experiment.” *Economics Letters* 108, no. 3: 345–348.
- Betto, M. 2024. “Cognition in Preferences and Choice.” Working Paper.
- Borjas, G. J., and K. B. Doran. 2012. “The Collapse of the Soviet Union and the Productivity of American Mathematicians.” *The Quarterly Journal of Economics* 127, no. 3: 1143–1203.

- Brehm, J., and W. Rahn. 1997. "Individual-Level Evidence for the Causes and Consequences of Social Capital." *American Journal of Political Science* 41: 999–1023.
- Breslin, P. A. 2013. "An Evolutionary Perspective on Food and Human Taste." *Current Biology* 23, no. 9: R409–R418.
- Brice, B. D., and H. M. Montesinos-Yufa. 2019. "The Era of Empirical Evidence." Working Paper.
- Cakir, M. S., J. K. Wardman, and A. Trautrim. 2023. "Ethical Leadership Supports Safety Voice by Increasing Risk Perception and Reducing Ethical Ambiguity: Evidence From the COVID-19 Pandemic." *Risk Analysis* 43, no. 9: 1902–1916.
- Calabro, R., E. Kemps, and I. Prichard. 2023. "Socio-cognitive Determinants of Sugar-Sweetened Beverage Consumption Among Young People: A Systematic Review and Meta-Analysis." *Appetite* 180: 106334.
- Cárdenas-Egúisquiza, A. L., and D. Berntsen. 2022. "Sleep Well, Mind Wander Less: A Systematic Review of the Relationship Between Sleep Outcomes and Spontaneous Cognition." *Consciousness and Cognition* 102: 103333.
- Chalmers, D. J. 1997. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford Paperbacks.
- Charness, G., L. Rigotti, and A. Rustichini. 2007. "Individual Behavior and Group Membership." *American Economic Review* 97, no. 4: 1340–1352.
- Chen, Y., and S. X. Li. 2009. "Group Identity and Social Preferences." *American Economic Review* 99, no. 1: 431–457.
- Chiappori, P.-A., B. Salanié, F. Salanié, and A. Gandhi. 2019. "From Aggregate Betting Data to Individual Risk Preferences." *Econometrica* 87, no. 1: 1–36.
- Chmura, T., H. Le, and K. Nguyen. 2022. "Herding With Leading Traders: Evidence From a Laboratory Social Trading Platform." *Journal of Economic Behavior & Organization* 203: 93–106.
- Ciampi, A., M. Silberfeld, and J. E. Till. 1982. "Measurement of Individual Preferences: The Importance of "Situation-Specific" Variables." *Medical Decision Making* 2, no. 4: 483–495.
- Clark, A., and Y. Lohéac. 2007. "It Wasn't Me, It Was Them! Social Influence in Risky Behavior by Adolescents." *Journal of Health Economics* 26: 763–784.
- Cox, S., B. Jones, and D. Collinson. 2006. "Trust Relations in High-Reliability Organizations." *Risk Analysis* 26, no. 5: 1123–1138.
- Dasgupta, U., L. Gangadharan, P. Maitra, and S. Mani. 2017. "Searching for Preference Stability in a State Dependent World." *Journal of Economic Psychology* 62: 17–32.
- Dave, C., C. C. Eckel, C. A. Johnson, and C. Rojas. 2010. "Eliciting Risk Preferences: When Is Simple Better?" *Journal of Risk and Uncertainty* 41, no. 3: 219–243.
- Dawkins, R. 2016. *The Selfish Gene*, 40th Anniversary Edition. Oxford University Press.
- De Melo, C. M., and J. Gratch. 2015. "People Show Envy, Not Guilt, When Making Decisions With Machines." In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 315–321.
- Dean, J. 2014. "Personal Protective Equipment: An Antecedent to Safe Behavior?" *Professional Safety* 59, no. 2: 41–46.
- Decker, S., and H. Schmitz. 2016. "Health Shocks and Risk Aversion." *Journal of Health Economics* 50: 156–170.
- Dekel, E., J. C. Ely, and O. Yilankaya. 2007. "Evolution of Preferences." *The Review of Economic Studies* 74, no. 3: 685–704.
- Dosman, D. M., W. L. Adamowicz, and S. E. Hrudef. 2001. "Socioeconomic Determinants of Health- and Food Safety-Related Risk Perceptions." *Risk Analysis* 21, no. 2: 307–318.
- Duncan, G., J. Boisjoly, M. Kremer, D. Levy, and J. Eccles. 2005. "Peer Effects in Drug Use and Sex Among College Students." *Journal of Abnormal Child Psychology* 33: 375–385.
- Einav, L., A. Finkelstein, I. Pascu, and M. R. Cullen. 2012. "How General Are Risk Preferences? Choices Under Uncertainty in Different Domains." *American Economic Review* 102, no. 6: 2606–2638.
- Fehr, E., U. Fischbacher, and S. Gächter. 2002. "Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms." *Human Nature* 13: 1–25.
- Fei, X., Z. Lou, X. Lv, Z. Ren, and R. Xiao. 2023. "Pollution Threshold Assessment and Risk Area Delineation of Heavy Metals in Soils Through the Finite Mixture Distribution Model and Bayesian Maximum Entropy Theory." *Journal of Hazardous Materials* 452: 131231.
- Fergusson, D. M., N. R. Swain-Campbell, and L. J. Horwood. 2002. "Deviant Peer Affiliations, Crime and Substance Use: A Fixed Effects Regression Analysis." *Journal of Abnormal Child Psychology* 30: 419–430.
- Fischbacher, U. 2007. "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10, no. 2: 171–178.
- Franklyn-Stokes, A., and S. E. Newstead. 1995. "Undergraduate Cheating: Who Does What and Why?" *Studies in Higher Education* 20, no. 2: 159–172.
- Gioia, F. 2017. "Peer Effects on Risk Behaviour: The Importance of Group Identity." *Experimental Economics* 20, no. 1: 100–129.
- Glaeser, E. L., D. I. Laibson, J. A. Scheinkman, and C. L. Soutter. 2000. "Measuring Trust." *The Quarterly Journal of Economics* 115, no. 3: 811–846.
- Goeree, J. K., C. A. Holt, and T. R. Palfrey. 2003. "Risk Averse Behavior in Generalized Matching Pennies Games." *Games and Economic Behavior* 45, no. 1: 97–113.
- Goette, L., D. Huffman, and S. Meier. 2006. "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups." *American Economic Review* 96, no. 2: 212–216.
- Grechuk, B., and M. Zabarankin. 2014. "Risk Averse Decision Making Under Catastrophic Risk." *European Journal of Operational Research* 239, no. 1: 166–176.
- Greiner, B. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments With ORSEE." *Journal of the Economic Science Association* 1, no. 1: 114–125.
- Guignet, D., and A. Alberini. 2015. "Can Property Values Capture Changes in Environmental Health Risks? Evidence From a Stated Preference Study in Italy and the United Kingdom." *Risk Analysis* 35, no. 3: 501–517.
- Guiso, L., P. Sapienza, and L. Zingales. 2018. "Time Varying Risk Aversion." *Journal of Financial Economics* 128, no. 3: 403–421.
- Gupta, P., and P. Goyal. 2024. "Herding the Influencers for Investment Decisions: Millennials Bust the Gender Stereotype." *Journal of Financial Services Marketing* 29, no. 2: 229–241.
- Halevy, Y., D. Persitz, and L. Zrill. 2018. "Parametric Recoverability of Preferences." *Journal of Political Economy* 126: 1558–1593.
- Hamming, R. W. 1950. "Error Detecting and Error Correcting Codes." *The Bell System Technical Journal* 29, no. 2: 147–160.
- Harrison, G. W., E. Johnson, M. M. McInnes, and E. E. Rutström. 2005. "Temporal Stability of Estimates of Risk Aversion." *Applied Financial Economics Letters* 1, no. 1: 31–35.
- He, H., P. Martinsson, and M. Sutter. 2012. "Group Decision Making under Risk: An Experiment With Student Couples." *Economics Letters* 117, no. 3: 691–693.
- Heifetz, A., C. Shannon, and Y. Spiegel. 2007. "The Dynamic Evolution of Preferences." *Economic Theory* 32: 251–286.
- Heller, Y., and I. Nehama. 2023. "Evolutionary Foundation for Heterogeneity in Risk Aversion." *Journal of Economic Theory* 208: 105617.
- Hirsch, R. A., and J. Baxter. 2011. "Context, Cultural Bias, and Health Risk Perception: The "Everyday" Nature of Pesticide Policy Preferences in London, Calgary, and Halifax." *Risk Analysis* 31, no. 5: 847–865.
- Ho, S. S., A. S. Chuah, N. Kim, and E. C. Tandoc Jr. 2022. "Fake News, Real Risks: How Online Discussion and Sources of Fact-Check Influence

- Public Risk Perceptions Toward Nuclear Energy." *Risk Analysis* 42, no. 11: 2569–2583.
- Holt, C. A., and S. K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92, no. 5: 1644–1655.
- Holt, C. A., and S. K. Laury. 2005. "Risk Aversion and Incentive Effects: New Data Without Order Effects." *American Economic Review* 95, no. 3: 902–912.
- Homma, S., and M. Takezawa. 2024. "Risk Preference as an Outcome of Evolutionarily Adaptive Learning Mechanisms: An Evolutionary Simulation Under Diverse Risky Environments." *PLoS One* 19, no. 8: e0307991.
- Ijzerman, H., W. W. van Dijk, and M. Gallucci. 2007. "A Bumpy Train Ride: A Field Experiment on Insult, Honor, and Emotional Reactions." *Emotion* 7, no. 4: 869.
- Jianakoplos, N. A., and A. Bernasek. 1998. "Are Women More Risk Averse?" *Economic Inquiry* 36, no. 4: 620–630.
- Kahneman, D. 2011. "Thinking, Fast and Slow." Farrar, Straus and Giroux.
- Kang, L. 2024. "Describing the Impact of Psychological Safety on Risk Prevention: A Threshold Model Construction." *Work* 79, no. 1: 277–288.
- Karni, E., D. Schmeidler, and K. Vind. 1983. "On State Dependent Preferences and Subjective Probabilities." *Econometrica* 51: 1021–1031.
- Kasperson, R. E., O. Renn, P. Slovic, et al. 1988. "The Social Amplification of Risk: A Conceptual Framework." *Risk Analysis* 8, no. 2: 177–187.
- Khodabakhshzadehs, H., and R. Shoahosseini. 2021. "The Relationships Among Attitudes Towards Cheating, Academic Self-Confidence, and General Language Ability Among Iranian EFL Learners." *International Journal of Language Testing* 11, no. 2: 1–12.
- Kim, S. H., R. H. Smith, and N. L. Brigham. 1998. "Effects of Power Imbalance and the Presence of Third Parties on Reactions to Harm: Upward and Downward Revenge." *Personality and Social Psychology Bulletin* 24, no. 4: 353–361.
- Kim, Y.-I., and J. Lee. 2014. "The Long-Run Impact of a Traumatic Experience on Risk Aversion." *Journal of Economic Behavior & Organization* 108: 174–186.
- Krebs, J. R. 2009. "The Gourmet Ape: Evolution and Human Food Preferences." *The American Journal of Clinical Nutrition* 90, no. 3: 707S–711S.
- Kurzban, R. 2011. *Why Everyone (Else) Is a Hypocrite: Evolution and the Modular Mind*. Princeton University Press.
- LaJoie, A. S., J. C. Kerr, R. D. Clover, and D. M. Harper. 2018. "Influencers and Preference Predictors of HPV Vaccine Uptake among US Male and Female Young Adult College Students." *Papillomavirus Research* 5: 114–121.
- Lamont, O. A. 2002. "Macroeconomic Forecasts and Microeconomic Forecasters." *Journal of Economic Behavior & Organization* 48, no. 3: 265–280.
- Lande, R., and S. J. Arnold. 1985. "Evolution of Mating Preference and Sexual Dimorphism." *Journal of Theoretical Biology* 117, no. 4: 651–664.
- Lee, H.-P., P. K. Chae, H.-S. Lee, and Y.-K. Kim. 2007. "The Five-Factor Gambling Motivation Model." *Psychiatry Research* 150, no. 1: 21–32.
- Lewontin, R. C. 1961. "Evolution and the Theory of Games." *Journal of Theoretical Biology* 1, no. 3: 382–403.
- Lichtenstein, S., and P. Slovic. 1971. "Reversals of Preference between Bids and Choices in Gambling Decisions." *Journal of Experimental Psychology* 89, no. 1: 46.
- Lin, X., and B. A. Weinberg. 2014. "Unrequited Friendship? How Reciprocity mediates Adolescent Peer Effects." *Regional Science and Urban Economics* 48: 144–153.
- Lindell, M. K., and S. N. Hwang. 2008. "Households' Perceived Personal Risk and Responses in a Multihazard Environment." *Risk Analysis* 28, no. 2: 539–556.
- Lindhjem, H., S. Navrud, N. A. Braathen, and V. Biaisque. 2011. "Valuing Mortality Risk Reductions from Environmental, Transport, and Health Policies: A Global Meta-Analysis of Stated Preference Studies." *Risk Analysis* 31, no. 9: 1381–1407.
- Lomi, A., T. A. Snijders, C. E. Steglich, and V. J. Torló. 2011. "Why Are Some More Peer Than Others? Evidence From a Longitudinal Study of Social Networks and Individual Academic Performance." *Social Science Research* 40, no. 6: 1506–1520.
- Lundborg, P. 2006. "Having the Wrong Friends? Peer Effects in Adolescent Substance Use." *Journal of Health Economics* 25, no. 2: 214–233.
- Lusk, J. L., and K. H. Coble. 2005. "Risk Perceptions, Risk Preference, and Acceptance of Risky Food." *American Journal of Agricultural Economics* 87, no. 2: 393–405.
- Maglio, M. A., C. Scott, A. L. Davis, J. Allen, and J. A. Taylor. 2016. "Situational Pressures That Influence Firefighters' Decision Making About Personal Protective Equipment: A Qualitative Analysis." *American Journal of Health Behavior* 40, no. 5: 555–567.
- Malmendier, U., and S. Nagel. 2011. "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *The Quarterly Journal of Economics* 126, no. 1: 373–416.
- Masclat, D., N. Colombier, L. Denant-Boemont, and Y. Loheac. 2009. "Group and Individual Risk Preferences: A Lottery-Choice Experiment With Self-Employed and Salaried Workers." *Journal of Economic Behavior & Organization* 70, no. 370: 470–484.
- Maslow, A. 1943. "A Theory of Human Motivation." *Psychological Review* 2: 21–28.
- Mazar, N., B. Koszegi, and D. Ariely. 2014. "True Context-Dependent Preferences? The Causes of Market-Dependent Valuations." *Journal of Behavioral Decision Making* 27, no. 3: 200–208.
- McCready, J. L., B. Nichol, M. Steen, J. Unsworth, D. Comparcini, and M. Tomietto. 2023. "Understanding the Barriers and Facilitators of Vaccine Hesitancy Towards the COVID-19 Vaccine in Healthcare Workers and Healthcare Students Worldwide: An Umbrella Review." *PLoS One* 18, no. 4: 1–41.
- Meng, G., Q. Li, X. Yuan, et al. 2023. "The Roles of Risk Perception, Negative Emotions and Perceived Efficacy in the Association Between COVID-19 Infection Cues and Preventive Behaviors: A Moderated Mediation Model." *BMC Public Health* 23, no. 1: 109.
- Mittelhammer, R. C. 2012. *Mathematical Statistics for Economics and Business*. Springer.
- Morgano, G. P., L. Mbuagbaw, N. Santesso, et al. 2022. "Defining Decision Thresholds for Judgments on Health Benefits and Harms Using The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence to Decision (EtD) Frameworks: A Protocol for a Randomised Methodological Study (GRADE-THRESHOLD)." *BMJ Open* 12, no. 3: e053246.
- Mourato, S. 2006. *Cost-Benefit Analysis and the Environment*. OECD.
- Neumann, I., E. Quiñelen, P. Nahuelhual, et al. 2022. "Using Explicit Thresholds Were Valuable for Judging Benefits and Harms in Partially Contextualized GRADE Guidelines." *Journal of Clinical Epidemiology* 147: 69–75.
- Okasha, S. 2007. "Rational Choice, Risk Aversion, and Evolution." *The Journal of Philosophy* 104, no. 5: 217–235.
- Okasha, S. 2011. "Optimal Choice in the Face of Risk: Decision Theory Meets Evolution." *Philosophy of Science* 78, no. 1: 83–104.
- Patelaraou, A., A. Saliag, P. Galanis, et al. 2022. "Predictors of Nurses' Intention to Accept COVID-19 Vaccination: A Cross-Sectional Study in Five European Countries." *Journal of Clinical Nursing* 31, no. 9–10: 1258–1266.
- Pawar, N. M., and N. R. Velaga. 2022. "Analyzing the Impact of Time Pressure on Drivers' Safety by Assessing Gap-Acceptance Behavior at Un-Signalized Intersections." *Safety science* 147: 105582.

- Phimister, J. R., U. Oktem, P. R. Kleindorfer, and H. Kunreuther. 2003. "Near-Miss Incident Management in the Chemical Process Industry." *Risk Analysis: An International Journal* 23, no. 3: 445–459.
- Pooladvand, S., and S. Hasanzadeh. 2022. "Neurophysiological Evaluation of Workers' Decision Dynamics Under Time Pressure and Increased Mental Demand." *Automation in Construction* 141: 104437.
- Portera, M., and L. Bartalesi. 2016. *Aesthetic Preferences: An Evolutionary Approach*, 375–388. Springer International Publishing.
- Powell, L. M., J. A. Tauras, and H. Ross. 2005. "The Importance of Peer Effects, Cigarette Prices and Tobacco Control Policies for Youth Smoking Behavior." *Journal of Health Economics* 24, no. 5: 950–968.
- Reason, J. 1998. "Achieving a Safe Culture: Theory and Practice." *Work & Stress* 12, no. 3: 293–306.
- Rieger, M. O., M. Wang, and T. Hens. 2015. "Risk Preferences Around the World." *Management Science* 61, no. 3: 637–648.
- Samuelson, L. 2001. "Introduction to the Evolution of Preferences." *Journal of Economic Theory* 97, no. 2: 225–230.
- Santos, M. D., D. J. Rankin, and C. Wedekind. 2011. "The Evolution of Punishment Through Reputation." *Proceedings of the Royal Society B: Biological Sciences* 278, no. 1704: 371–377.
- Schildberg-Hörisch, H. 2018. "Are Risk Preferences Stable?" *Journal of Economic Perspectives* 32, no. 2: 135–154.
- Schkade, D. A., and E. J. Johnson. 1989. "Cognitive Processes in Preference Reversals." *Organizational Behavior and Human Decision Processes* 44, no. 2: 203–231.
- Schubert, R., M. Brown, M. Gysler, and H. W. Brachinger. 1999. "Financial Decision-Making: Are Women Really More Risk Averse?" *American Economic Review* 89, no. 2: 381–385.
- Shanmugam, R., and R. Chattamvelli. 2015. *Statistics for Scientists and Engineers*. John Wiley & Sons.
- Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor. 2004. "Risk as Analysis and Risk as Feelings: Some Thoughts About Affect, Reason, Risk, and Rationality." *Risk Analysis* 24, no. 2: 311–322.
- Smith, J. M. 1982. *Evolution and the Theory of Games*. Cambridge University Press.
- Stigler, G. J., and G. S. Becker. 1977. "De Gustibus Non Est Disputandum." *The American Economic Review* 67, no. 2: 76–90.
- Sunden, A. E., and B. J. Surette. 1998. "Gender Differences in the Allocation of Assets in Retirement Savings Plans." *The American economic review* 88, no. 2: 207–211.
- Szpiro, G. G. 1997. "The Emergence of Risk Aversion." *Complexity* 2, no. 4: 31–39.
- Tajfel, H., M. G. Billig, R. P. Bundy, and C. Flament. 1971. "Social Categorization and Intergroup Behaviour." *European Journal of Social Psychology* 1, no. 2: 149–178.
- Talpsepp, T., and A.-L. Tänav. 2021. "Do Gender, Age and Education Affect Herding in the Real Estate Market?" *Journal of Behavioral and Experimental Finance* 32: 100571.
- Torres, W. J., and R. M. Bergner. 2010. "Humiliation: Its Nature and Consequences." *Journal of the American Academy of Psychiatry and the Law Online* 38, no. 2: 195–204.
- Tousignant, O. H., N. D. Taylor, M. K. Suvak, and G. D. Fireman. 2019. "Effects of Rumination and Worry on Sleep." *Behavior Therapy* 50, no. 3: 558–570.
- Tsutsui, Y., and I. Tsutsui-Kimura. 2022. "How Does Risk Preference Change Under the Stress of COVID-19? Evidence From Japan." *Journal of Risk and Uncertainty* 64, no. 2: 191–212.
- Tversky, A., and D. Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211, no. 4481: 453–458.
- Tversky, A., P. Slovic, and D. Kahneman. 1990. "The Causes of Preference Reversal." *The American Economic Review* 204–217.
- Van der Pol, M., and M. Ruggeri. 2008. "Is Risk Attitude Outcome Specific within the Health Domain?" *Journal of Health Economics* 27, no. 3: 706–717.
- Vaquera, E., and G. Kao. 2008. "Do You Like Me As Much As I like You? Friendship Reciprocity and Its Effects on School Outcomes Among Adolescents." *Social Science Research* 37, no. 1: 55–72.
- Vaughan, D. 1996. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. University of Chicago Press.
- Wakefield, J. G., M.-L. McLaws, M. Whitby, and L. Patton. 2010. "Patient Safety Culture: Factors That Influence Clinician Involvement in Patient Safety Behaviours." *BMJ Quality & Safety* 19, no. 6: 585–591.
- Wakker, P. P. 2008. "Explaining the Characteristics of the Power (CRRA) Utility Family." *Health Economics* 17, no. 12: 1329–1344.
- Waldinger, F. 2010. "Quality Matters: The Expulsion of Professors and the Consequences for PhD Student Outcomes in Nazi Germany." *Journal of Political Economy* 118, no. 4: 787–831.
- Wallach, M. A., N. Kogan, and D. J. Bem. 1962. "Group Influence On Individual Risk Taking." *The Journal of Abnormal and Social Psychology* 65, no. 2: 75–86.
- Weibull, J. W. 1997. *Evolutionary Game Theory*. MIT Press.
- White, T. L., T. Thomas-Danguin, J. K. Olofsson, G. M. Zucco, and J. Prescott. 2020. "Thought for Food: Cognitive Influences on Chemosensory Perceptions and Preferences." *Food Quality and Preference* 79: 103776.
- Williamson, A., D. A. Lombardi, S. Folkard, J. Stutts, T. K. Courtney, and J. L. Connor. 2011. "The Link between Fatigue and Safety." *Accident Analysis & Prevention* 43, no. 2: 498–515.
- Winter, E., L. Méndez-Naya, and I. García-Jurado. 2017. "Mental Equilibrium and Strategic Emotions." *Management Science* 63, no. 5: 1302–1317.
- Wolf, C., and L. Pohlman. 1983. "The Recovery of Risk Preferences from Actual Choices." *Econometrica: Journal of the Econometric Society* 843–850.
- Wong, T. K. M., S. S. Man, and A. H. S. Chan. 2020. "Critical Factors for the Use or Non-Use of Personal Protective Equipment Amongst Construction Workers." *Safety Science* 126: 104663.
- Wright, R. 2017. *Why Buddhism Is True: The Science and Philosophy of Meditation and Enlightenment*. Simon and Schuster.
- Yang, M., L. S. Roope, J. Buchanan, et al. 2022. "Eliciting Risk Preferences That Predict Risky Health Behavior: A Comparison of Two Approaches." *Health Economics* 31, no. 5: 836–858.
- Zajonc, R. B., and H. Markus. 1982. "Affective and Cognitive Factors in Preferences." *Journal of Consumer Research* 9, no. 2: 123–131.
- Zhang, R., T. J. Brennan, and A. W. Lo. 2014. "The Origin of Risk Aversion." *Proceedings of the National Academy of Sciences* 111, no. 50: 17777–17782.
- Zheng, Z., K. Tang, Y. Liu, and J. M. Guo. 2021. "Gender and Herding." *Journal of Empirical Finance* 64: 379–400.
- Zmerli, S., and K. Newton. 2008. "Social Trust and Attitudes Toward Democracy." *Public Opinion Quarterly* 72, no. 4: 706–724.

Appendix A: Proofs

For the reader's convenience, we restate the results here.

Lemma 1. Assume that M is single-crossing and there are no dominated lotteries in L . Then, for $l', l'' \in L$ with $l' \neq l''$, exactly one of the following statements is true:

1. There exists a value $r(l', l'') \in (\underline{r}, \bar{r})$ such that $l' \succ_r l''$ if $r < r(l', l'')$ and $l'' \succ_r l'$ if $r > r(l', l'')$.
2. There exists a value $r(l'', l') \in (\underline{r}, \bar{r})$ such that $l'' \succ_r l'$ if $r < r(l'', l')$ and $l' \succ_r l''$ if $r > r(l'', l')$.

Proof. Consider the sets

$$A := \{r \in (\underline{r}, \bar{r}) \mid l' \succ_r l''\} \text{ and } B := \{r \in (\underline{r}, \bar{r}) \mid l'' \succ_r l'\}.$$

If A was empty, then lottery l' was dominated, and if B was empty, then l'' was dominated. Therefore, both sets are nonempty. Because $\succ_r$ is a complete order for all r , it holds $A \cup B = (\underline{r}, \bar{r})$. From this follows that $\underline{r} = \inf A$ or $\underline{r} = \inf B$. If neither was the case, we would have $\inf A, \inf B > \underline{r}$, and then, by definition of the infimum, any value r with $\underline{r} < r < \min[\inf A, \inf B]$ would be neither in A nor in B , contradicting $A \cup B = (\underline{r}, \bar{r})$.

Assume $\underline{r} = \inf A$ (this does not lose generality as our argument can be made analogously for $\underline{r} = \inf B$). Pick $r' \in A$ and $r'' \in B$. $r'' < r'$ is impossible because the single-crossing property would require that for all $r < r''$ holds $r \in B$, but then $\inf A \geq r'' > \underline{r}$, contrary to our assumption. Therefore, we have $r' < r''$. As we chose r', r'' arbitrarily from A and B , this implies that $\sup A \leq \inf B$. However, there exists no value r with $\sup A < r < \inf B$ because such r would be neither in A nor in B . The case $\sup A < \inf B$ is thus impossible, and it follows that we must have $\sup A = \inf B$. Then, $r(l', l'') := \sup A = \inf B$ has the desired properties by the definition of the supremum, the infimum, and the sets A and B .

Our argument was based on the assumption $\underline{r} = \inf A$. If we would assume $\underline{r} = \inf B$ instead, the argument can be made analogously and it follows that $\sup B = \inf A$, and $r(l'', l') := \sup B = \inf A$ has the desired properties. It remains to be shown that exactly one of the two cases, either $\underline{r} = \inf A$ or $\underline{r} = \inf B$, is true. Pick r', r'' such that $r' \in A$ and $r'' \in B$. Trivially, either $r' > r''$ or $r'' > r'$ is true. If $r' > r''$, the single-crossing property requires that for every $r \leq r''$, it holds $r \in B$; hence, $\underline{r} \neq \inf A$. Analogously, from $r'' > r'$ follows $\underline{r} \neq \inf B$. $\square$

Lemma 2. The function $d : L \times L \rightarrow \mathbb{N} \cup \{0\}$ defined by (1) is a metric on L .

Proof. It is clear that $d(\succ', \succ') = 0$ and $d(\succ', \succ'') = d(\succ'', \succ') > 0$ for $\succ', \succ'' \in \mathcal{P}$ with $\succ' \neq \succ''$. Only the Triangle inequality

$$d(\succ', \succ''') \leq d(\succ', \succ'') + d(\succ'', \succ''') \quad (\text{A.1})$$

requires an argument. Here, we note that for each two unequal lotteries that are ranked differently by $\succ'$ and $\succ'''$, the relation $\succ''$ either agrees with $\succ'$ or with $\succ'''$ because there are three relations but, in the absence of indifference, only two ways how to rank the two elements. Specifically, consider a pair $(l', l'') \in L \times L$ that is counted for the value of $d(\succ', \succ''')$, that is, for the left-hand side of (A.1). Then, if $\succ''$ agrees with $\succ'$ on $\{l', l''\}$, the pair is not counted for $d(\succ', \succ'')$. However, it is then counted for $d(\succ'', \succ''')$. Thus, this pair is also counted once for the right-hand side of (A.1). Alternatively, if $\succ''$ agrees with $\succ'''$ on $\{l', l''\}$, the pair is counted for $d(\succ', \succ'')$ but not for $d(\succ'', \succ''')$. Again, the pair is also counted once for the right-hand side of (A.1). Hence, if we restrict the relations to the subset of pairs in $L \times L$ which are counted for $d(\succ', \succ''')$, the Triangle inequality is fulfilled with equality. Among those pairs that are ranked equally by $\succ'$ and $\succ'''$, that is, which are not counted for $d(\succ', \succ''')$, some may be ranked differently by $\succ''$ and by $\succ'$ and $\succ'''$. These would be counted for both $d(\succ', \succ'')$ and $d(\succ'', \succ''')$, raising the right-hand side of (A.1) above its left-hand side. In the presence of such pairs, the Triangle inequality is fulfilled strictly. $\square$

Lemma 3. Let $\{l^1, l^2, l^3, l^4\} \in L$ with $\{l^1, l^2\} \neq \{l^3, l^4\}$. If the thrive function M is single-crossing and linear, then $r(l^1, l^2) \neq r(l^3, l^4)$.

Proof. Let $R := \{r(l', l'') \in (\underline{r}, \bar{r}) \mid l', l'' \in L\}$ be the set of values whose existence was proved in Lemma 1. By way of contradiction, assume that there are lotteries $\{l^1, l^2\} \neq \{l^3, l^4\}$ with $\hat{r} := r(l^1, l^2) = r(l^3, l^4)$. Set

$$a := \max\{x \in \{r\} \cup R \mid x < \hat{r}\} \text{ and } b := \min\{x \in \{\bar{r}\} \cup R \mid x > \hat{r}\}.$$

Select $r' \in (a, \hat{r})$ and $r'' \in (\hat{r}, b)$. Then, $d(\succ_{r'}, \succ_{r''}) > 2$ because, with $r' < \hat{r}$ and $r'' > \hat{r}$, the relations $\succ_{r'}$ and $\succ_{r''}$ rank (at least) the lotteries l^1, l^2 and l^3, l^4 differently. Because we assume the thrive function to be linear, there must be values r^1, r^2 with $r' \leq r^1 < r^2 \leq r''$ such that for all $r \in (r^1, r^2)$, it holds $d(\succ_r, \succ_{r''}) = 1$. We will derive a contradiction from here. If there are such values r^1, r^2 , then $\hat{r} \leq r^1$ because left of $\hat{r}$ all $\succ_r$ differ from $\succ_{r''}$ in at least two comparisons. However, for all $r \in (\hat{r}, r^2)$, it holds $d(\succ_r, \succ_{r''}) = 0$, that is, the relations $\succ_r$ and $\succ_{r''}$ are the same. To show this, assume again by way of contradiction that there is $r \in (\hat{r}, r^2)$ and two lotteries $l', l'' \in L$ that are ranked differently by $\succ_r$ and $\succ_{r''}$. W.l.o.g., assume $l' \succ_r l''$ and $l'' \succ_{r''} l'$. By Lemma 1, there exists a value $r(l', l'')$ or a value $r(l'', l')$ with properties as specified in the lemma. Again w.l.o.g., assume the existing value is $r(l', l'')$. If $r(l', l'') < r < r''$, then, by definition of $r(l', l'')$, we must have $l'' \succ_r l'$, contrary to our assumption. If, on the other hand, we have $r < r'' < r(l', l'')$, then we must have $l' \succ_{r''} l''$, again contradicting our assumption. It therefore follows that $r(l', l'') \in [r, r'']$. Yet, $r(l', l'')$ is in R , and r'' was picked such that $r'' < b = \min\{x \in \{\bar{r}\} \cup R \mid x > \hat{r}\}$, rendering the case $r(l', l'') \in [r, r''] \subseteq (\hat{r}, b)$ impossible. This contradiction proves that the existence of lotteries $\{l^1, l^2\} \neq \{l^3, l^4\}$ with $\hat{r} := r(l^1, l^2) = r(l^3, l^4)$ is incompatible with linearity. $\square$

Proposition 1. If the function M is single-crossing and linear, then $(\underline{r}, \bar{r})$ can be partitioned into n intervals $I_1, \dots, I_n$ (each of which may be open, closed, or semiclosed) such that:

1. $\succ_{r'} = \succ_{r''}$ if and only if $r', r'' \in I_j$ for one $j \in \{1, \dots, n\}$.
2. The n intervals can be labeled in a way that fulfils:

$$r' \in I_i, r'' \in I_j \Rightarrow d(\succ_{r'}, \succ_{r''}) = |i - j|.$$

Proof. We will carry out the proof by constructing a sequence of labeled intervals and then showing that the intervals and the labeling have the two properties stated above. As before, let $R := \{r(l', l'') \in (\underline{r}, \bar{r}) \mid l', l'' \in L\}$ be the set that contains the values specified in Lemma 1. Define

$$\rho(\hat{r}) := \begin{cases} \min\{r \in R \mid r > \hat{r}\} & \text{if } \{r \in R \mid r > \hat{r}\} \neq \emptyset \\ \bar{r} & \text{otherwise.} \end{cases}$$

With this notation, we devise the following algorithm:

Initialization:

Set

$$I_1 := \begin{cases} (\underline{r}, \min R] & \text{if } \succ_r = \succ_{\min R} \text{ for } r := \frac{1}{2}\underline{r} + \frac{1}{2}\min R \\ (\underline{r}, \min R) & \text{otherwise.} \end{cases}$$

Set $j := 2$.

Main iteration:

Step 1:

If $\rho(\sup I_{j-1}) \neq \bar{r}$, go to Step 2. Otherwise, set

$$I_j := (\underline{r}, \bar{r}) \setminus \left\{ \bigcup_{k=1, \dots, j-1} I_k \right\},$$

terminate the algorithm, and let the output be the intervals $I_1, \dots, I_n$ with $n = j$.

Step 2:

If I_{j-1} is a right open interval, set

$$I_j := \begin{cases} [\sup I_{j-1}, \rho(\sup I_{j-1})] & \text{if } \succ_r = \succ_{\rho(\sup I_{j-1})} \text{ for } r := \frac{1}{2} \sup I_{j-1} + \frac{1}{2} \rho(\sup I_{j-1}) \\ [\sup I_{j-1}, \rho(\sup I_{j-1}) & \text{otherwise.} \end{cases}$$

If I_{j-1} is a right closed interval, we set

$$I_j := \begin{cases} (\sup I_{j-1}, \rho(\sup I_{j-1})] & \text{if } \succ_r = \succ_{\rho(\sup I_{j-1})} \text{ for } r := \frac{1}{2} \sup I_{j-1} + \frac{1}{2} \rho(\sup I_{j-1}) \\ (\sup I_{j-1}, \rho(\sup I_{j-1})) & \text{otherwise.} \end{cases}$$

Set $j := j + 1$. Return to Step 1.

As there are finitely many elements in L , there are finitely many elements in R . Therefore, there are finitely many steps until $\rho(\sup I_{j-1}) = \bar{r}$ is reached and the algorithm ends. The first property of the partition, namely, $\succ_{r'} = \succ_{r''}$ if and only if $r', r'' \in I_j$, results from the fact that each interval border (except $\bar{r}$ and $\bar{r}$ of the extreme intervals) is an element of R . For the “if”-part, consider by way of contradiction that there was an index j with $r', r'' \in I_j$ such that $r' < r''$, and two lotteries l', l'' with $l' \succ_{r'} l''$ and $l'' \succ_{r''} l'$. By Lemma 1, there must exist a value $r(l', l'') \in R$ or a value $r(l'', l') \in R$. W.l.o.g., assume $r(l', l'')$ exists. By definition of $r(l', l'')$, if $r(l', l'') < r'$, then $l' \succ_{r'} l''$ is impossible. Likewise, if $r(l', l'') > r''$, then $l'' \succ_{r''} l'$ is impossible. Therefore, we must have $r' \leq r(l', l'') \leq r''$. As $r(l', l'') \in R$, it is in itself an interval border; hence, if $r' < r(l', l'') < r''$, the values r' and r'' could not be from the same interval, contrary to our assumption. If, on the other hand, $r(l', l'') = r'$ or $r(l', l'') = r''$, then r' or r'' are themselves boundaries of an interval, and as $r', r'' \in I_j$, that interval must be I_j . W.l.o.g., assume $r(l', l'') = r'$ is the left boundary of I_j . That means that I_j is left-closed, which implies I_{j-1} is right-open. By construction, this is only the case if for the midpoint $\bar{r}$ of I_{j-1} holds $\succ_{\bar{r}} \neq \succ_{r'}$. However, by Lemma 3, the only way in which the three relations $\succ_{\bar{r}}$, $\succ_{r'}$, and $\succ_{r''}$ can differ is by the ranking of l' and l'' (because $r(l', l'') = r'$, but for any other $l^1, l^2 \in L$, we have $r(l^1, l^2) \neq r'$). Thus, as we assume that $\succ_{r'}$ and $\succ_{r''}$ rank l' and l'' differently, it must be the case that $\succ_{\bar{r}}$ and $\succ_{r'}$ rank l' and l'' in the same way, contradicting that we had already derived $\succ_{\bar{r}} \neq \succ_{r'}$.

It remains to be shown that if $r' \in I_i, r'' \in I_j \Rightarrow d(\succ_{r'}, \succ_{r'') = |i - j|$. We can restrict our attention to the case that both r' and r'' are interior values in their respective intervals because from the first part of the proof follows that the boundaries of an interval, if they are included in the interval, will be mapped into the same relation as an interior value. Then, considering Lemma 3 and the first part of this proposition, $d(\succ_{r'}, \succ_{r'') equals the number of interval boundaries that are strictly between r' and r'' . By the way the intervals were constructed and labeled, that is, by defining a sequence of increasing interval boundaries that are labeled consecutively, the number of interval boundaries between the two interior values $r' \in I_i$ and $r'' \in I_j$ is $|i - j|$. $\square$$

Lemma 4 (Median minimizes expected absolute error). *If $m(p)$ is the unique median of the probability mass function p , then $i = m(p)$ minimizes the expression*

$$\sum_{k=1, \dots, \infty} k(p(i+k) + p(i-k)).$$

Proof. See, for example, Shanmugam and Chattamvelli (2015), Theorem 2.3 (p. 55). $\square$

Lemma 5. *Let p and q be probability mass functions defined on $\mathbb{Z}$, supported on a finite set $\{1, \dots, n\} \subset \mathbb{N}$, and let $s_\lambda = \{\lambda q(i) + (1 - \lambda)p(i)\}_{i \in \mathbb{Z}}$*

with $\lambda \in [0, 1]$ be a linear combination of p and q . If $m(q), m(p), m(s_\lambda)$ are the unique medians of the three functions and $m(p) \leq m(q)$, then $m(p) \leq m(s_\lambda) \leq m(q)$.

Proof. By definition, $m(s_\lambda)$ is the smallest index $\bar{m}$ for which

$$\lambda \sum_{k=1, \dots, \bar{m}} p(k) + (1 - \lambda) \sum_{j=1, \dots, \bar{m}} q(j) \geq 0.5 \quad (\text{A.2})$$

If $\bar{m} < m(p)$, we have $\sum_{k=1, \dots, \bar{m}} p(k) < 0.5$ and, because of $m(p) \leq m(q)$, $\sum_{j=1, \dots, \bar{m}} q(j) < 0.5$, and thus,

$$\lambda \underbrace{\sum_{k=1, \dots, \bar{m}} p(k)}_{< 0.5} + (1 - \lambda) \underbrace{\sum_{j=1, \dots, \bar{m}} q(j)}_{< 0.5} < 0.5.$$

Likewise, for $\bar{m} \geq m(p)$, we have and $\sum_{j=1, \dots, \bar{m}} q(j) \geq 0.5$ and, because of $m(p) \leq m(q)$, also $\sum_{k=1, \dots, \bar{m}} p(k) \geq 0.5$, and thus,

$$\lambda \underbrace{\sum_{k=1, \dots, \bar{m}} p(k)}_{\geq 0.5} + (1 - \lambda) \underbrace{\sum_{j=1, \dots, \bar{m}} q(j)}_{\geq 0.5} \geq 0.5.$$

Therefore, $\bar{m} > m(p)$ would not be the smallest index to fulfill (A.2). It follows that we must have $m(p) \leq m(s_\lambda) \leq m(q)$. $\square$

Proposition 2. *Let $\lambda \in [0, 1]$. Then,*

$$d(\succ^{m(s_\lambda)}, \succ^{m(q)}) \leq d(\succ^{m(p)}, \succ^{m(q)}).$$

Proof. If $m(p) = m(q)$, Lemma 5 ensures $m(s_\lambda) = m(p) = m(q)$, and thus, in view of Proposition 1 (Statement 1), we have $\succ^{m(s_\lambda)} = \succ^{m(p)} = \succ^{m(q)}$ so that the result is trivially true. W.l.o.g. assume $m(p) < m(q)$. By Proposition 1 (Statement 2), Proposition 2 is true if

$$|m(q) - m(s_\lambda)| \leq m(q) - m(p) \quad (\text{A.3})$$

(as $m(q), m(p), m(s_\lambda) \in \{1, \dots, n\}$ and thus $m(q), m(p), m(s_\lambda) > 0$, we have dropped the absolute value function on the right-hand side). Because of Lemma 5, there exists $\delta \in [0, 1]$ such that

$$m(s_\lambda) = \delta m(q) + (1 - \delta)m(p). \quad (\text{A.4})$$

Plugging (A.4) into (A.3) leads to

$$\begin{aligned} |m(q) - \delta m(q) - (1 - \delta)m(p)| &\leq m(q) - m(p) \Leftrightarrow \\ (1 - \delta)(m(q) - m(p)) &\leq m(q) - m(p) \end{aligned}$$

which is obviously true for $\delta \in [0, 1]$ and $0 < m(p) < m(q)$. $\square$