

Disentangling the Effects of Counterfactual Feedback on Maximization and Risk Preference across Gains and Losses

Ali Shiravand

ali.shiravand@ens.fr

Ecole Normale Supérieure - PSL University <https://orcid.org/0009-0009-8116-8924>

Maëlle Gueguen

Laureate Institute for Brain Research <https://orcid.org/0000-0002-4045-5881>

Sophie Bavard

Paris Brain Institute <https://orcid.org/0000-0002-9283-2976>

Dirk Wulff

Max Planck Institute for Human Development <https://orcid.org/0000-0002-4008-8022>

Julien Bastin

Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, GIN <https://orcid.org/0000-0002-0533-7564>

Stefano Palminteri

ENS <https://orcid.org/0000-0001-5768-6646>

Article

Keywords:

Posted Date: April 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9149770/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Disentangling the Effects of Counterfactual Feedback on Maximization and Risk Preference across Gains and Losses

Ali Shiravand^{a,1}, Maëlle Gueguen^{b,c,d,e}, Sophie Bavard^f, Dirk U. Wulff^{g,h}, Julien Bastin^{d,e}, and Stefano Palminteri^{a,i,1}

This manuscript was compiled on March 17, 2026

Decisions under risk can be characterized by two complementary but distinct behavioral variables: the propensity to choose the expected value-maximizing option and the tendency to select the risky option. Human reinforcement learning (RL) research has traditionally focused more on the former (maximization), providing many insights into how fundamental features of the learning environment influence behavior. For instance, previous studies have shown that counterfactual feedback about unchosen options exerts a strong influence on maximization. However, comparatively little is known about scenarios in which maximization and risk propensity are assessed orthogonally and simultaneously. To address this gap, we designed a series of RL experiments in which we manipulated key factors, including the feedback information regime (partial vs. complete) and the relative expected values of risky and safe options. Contrary to the conventional assumption that “more information is better,” our results clearly show that, in such contexts, the effect of foregone outcomes on risk preference outweighs its influence on maximization. We further compared RL choices with description-based versions of the same decision problems. While replicating established findings on the description–experience gap, we found that complete information had little impact on its expression, suggesting that this gap does not arise primarily from instant feedback or insufficient outcome sampling.

Decision-Making | Reinforcement Learning | Risk Preferences | Description-Experience Gap

1 Introduction

Decision-making under risk is a fundamental cognitive process that has been extensively studied across various disciplines, including psychology, behavioral economics, and neuroscience (1–3). Among the different approaches to studying risky decision-making, a key distinction is made between experience-based and description-based decisions. In experience-based decision-making, individuals learn through direct feedback over time, while in description-based paradigms, they are provided with explicit information about the probabilities and outcomes of different choices (4–7). The majority of real-life decisions fall into the former category, as individuals typically learn about the consequences of their actions through direct interaction with the environment rather than through explicit probabilistic descriptions (8). Within the framework of experience-based decision-making, reinforcement learning (RL) paradigms are prominent (9–11). The reason is that these paradigms, in which feedback is collected on a trial-by-trial basis, are more evolutionarily valid than the alternative sampling paradigm, where individuals gather “free” information about options before making a final choice—a situation that may have become relevant in the digital age but was not characteristic of our evolutionary past, nor is it shared with non-human species (8, 12–15). This study therefore focuses on RL paradigms and investigates how two fundamental aspects of RL performance, maximization and risk preference, are influenced by a crucial factors: the feedback information regime and valence of the decision-problem (6, 16–18).

RL performance can be characterized by two independent dimensions: Maximization (i.e., the tendency to choose the option with the highest expected value (EV), referred to as the M-rate); and Risk Preference (i.e., the tendency to select risky over safe options, independent of their EV, referred to as the R-rate) (19). In human RL studies, the primary focus has traditionally been on maximization (20, 21). Artificial agents and human participants are often assessed based on their ability to select EV-maximizing responses, which are typically considered the “correct” or “optimal” choices. In contrast, risk preference is the primary focus in behavioral economics and judgment and decision-making (JDM) research (4, 22–26). These studies frequently design decision problems with equivalent EVs, making it impossible to assess maximization but allowing for an in-depth exploration of risk attitudes (27, 28). In experience-based decision-making, feedback can be provided in different

Author affiliations: ^aDepartment of Cognitive Studies, École Normale Supérieure-PSL University, Paris, France; ^bLaureate Institute for Brain research, Tulsa, OK, USA; ^cOxley College of Health and Natural Sciences, University of Tulsa, OK, USA; ^dGrenoble Institut Neurosciences, Université Grenoble Alpes, Grenoble, France; ^eInstitut National de la Santé et de la Recherche Médicale (INSERM), U1216, Grenoble, France; ^fMotivation, Brain and Behaviour Laboratory, Paris Brain Institute, Paris, France; ^gCenter for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany; ^hCenter for Cognitive and Decision Sciences, University of Basel, Basel, Switzerland; ⁱInstitut National de la Santé et de la Recherche Médicale (INSERM), U960, Paris, France

¹To whom correspondence should be addressed. E-mail: ali.shiravand@ens.fr and stefano.palminteri@ens.fr

ways (27, 29–32). Partial Feedback is a default condition in which participants only receive information about the outcome of their chosen option. Conversely, Complete Feedback is the condition in which participants receive information about the outcome of both the chosen and unchosen options (33).

The role of feedback in RL has been widely debated, with conflicting findings across different research traditions (31, 34–38). In the RL literature, where the riskiness of options is not typically manipulated, complete feedback has generally been found to increase maximization. This effect is often attributed to the reduction of the uncertainty concerning the unchosen option outcome and the consequent elimination of the exploration/exploitation dilemma, which makes learning more efficient (39, 40). In contrast, in the JDM literature, which focuses on risk preference rather than maximization, complete feedback has been shown to increase risk-taking behavior (33, 41). These differing conclusions bring up a fundamental question: When both maximization and risk preference are examined simultaneously, which effect dominates? This unresolved issue represents the first major knowledge gap that our study aims to address.

Another critical consideration is the well-documented description–experience gap (D/E gap)—the systematic differences observed between decisions made based on described probabilities and those made based on experiential learning (42–46). This phenomenon has significant implications for both theoretical models of decision-making and real-world applications, such as designing behavioral interventions (nudges and boosts) to guide decision-making under uncertainty (47–53). One of the most robust findings in this domain is that subjects behave as if they overweight small probabilities in description-based decisions but underweight in experience-based decisions (4, 43). Beyond probability weighting (which has been extensively studied and is not the focus of the present work; 54), a less explored but equally important aspect of the description–experience gap concerns gain-loss asymmetries. In description-based decisions, the reflection effect is linked to sign-dependent curvature in the utility function, whereas loss aversion refers to greater sensitivity to losses than to gains. In experience-based decisions, the reflection pattern often reverses, with greater risk-taking in gains than in losses (an “inverse” reflection effect) (55, 56). One goal of our study was to replicate this aspect of the D/E gap using a different design and to examine how it relates to the feedback information regime (partial vs. complete). One leading explanation for the D/E gap is that it arises from insufficient sampling. For instance, in RL tasks, losses are generally avoided, leading to their underestimation in experience-based decisions (4, 5, 29, 30, 57, 58). If this hypothesis is correct, then the inverse reflection effect should interact with feedback information, with complete feedback reducing or eliminating the gap by providing additional learning opportunities (thus bringing the RL task closer to the description phase). Testing this hypothesis represents the second major knowledge gap that this study aims to address.

To investigate these research questions, we designed a series of experiments in which we systematically manipulated key variables influencing decision-making under risk. Within each experiment, across different decision problems, we varied option riskiness and EV, which allowed us to create situations where either the risky or the safe option was the EV-maximizing choice. This design enabled us to orthogonalize expected-value maximization and risk preference. Each experiment also manipulated the valence of the decision problems (gain vs. loss), allowing us to compare behavior across the two valence domains and to test for the inverse reflection effect. Across experiments, we further manipulated the feedback regime (partial vs. complete), which enabled us to test the effect of feedback information on both risk propensity and maximization, as well as the outcome probability, to assess the generality of our results. In addition to the RL task, all experiments included equivalent description-based decision contexts, allowing us to empirically examine the D/E gap.

By integrating insights from behavioral economics and human reinforcement learning, this study provides a comprehensive investigation of the effects of feedback information on both maximization and risk preference while also testing its role in shaping the description–experience gap. Specifically, we found that—somewhat surprisingly—feedback information influenced risk preference more strongly than maximization. Furthermore, complete feedback did not eliminate the D/E gap (which, in our study, manifested as an inverse reflection effect), suggesting that this gap does not primarily stem from sampling issues. Our findings will contribute to a better understanding of the cognitive mechanisms underlying experience-based decision-making and have potential applications in fields such as behavioral policy design (59).

Results

Experimental protocol We conducted four experiments (Figure 1) with independent samples of 100 Prolific participants each (total $N = 400$; all participants who completed the task were included; demographics in Table 1).

Table 1. Demographics of the participants. Each experiment varied two between-subject factors: 1) Probability indicates the likelihood of outcomes (P_h and P_l) for the risky option in each context, and 2) Feedback defines whether outcome outcomes were shown for the chosen option only or for both.

Exp.	Probability	Feedback	N	Gender (M/F)	Age
					Mean $\pm$ SD (IQR)
1	80/20	Partial	100	50/50	39.0 $\pm$ 14.1 (27.4–49.0)
2	60/40	Partial	100	50/50	41.9 $\pm$ 13.8 (32.0–53.5)
3	80/20	Complete	100	50/50	41.6 $\pm$ 15.1 (29.5–53.0)
4	60/40	Complete	100	50/50	42.4 $\pm$ 14.7 (30.5–53.6)

The task was designed to examine how people learn to choose between risky and safe options when 1) the valence of outcomes (gain vs. loss) and 2) the amount of feedback (partial vs. complete) vary, and to compare experience-based with

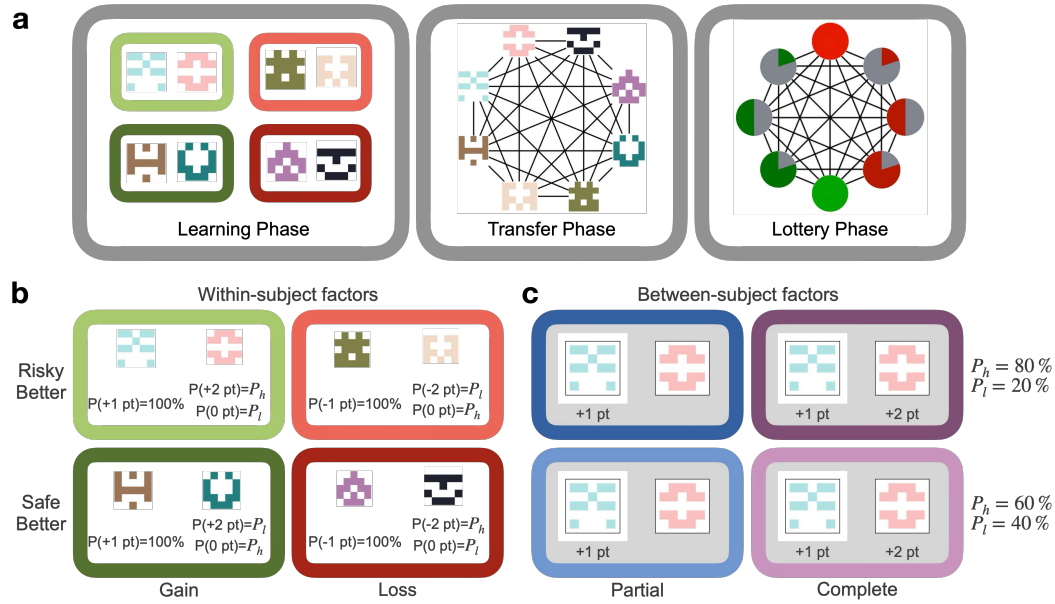


Fig. 1. Experimental paradigm. **a**, The three phases of the task: 1) the learning phase where the choice is made in four predefined contexts (pair of symbols), 2) the transfer phase where the participant chooses among all possible pairs of previously learned symbols without feedback (experience-based decision), and 3) the lottery phase where the participant chooses among all possible pairs of explicitly defined risky decisions (lotteries) with the same risk profile as the learning phase symbols (description-based decision). **b**, Within-subject factors defined by the context structure. In the first row contexts, the risky option has a better expected outcome and in the second row contexts, the safe option has a better expected outcome. In the contexts on the left column, the participant chooses to gain more, and in the right column chooses to lose less. **c**, Between-subject factors. The probability of outcomes and the type of feedback vary between experiments. The outline colors are not visible to participants.

description-based choices. The study comprised three phases: *learning*, *transfer*, and *lottery* (Figure 1a). During learning, two factors varied within participants: 1) valence (gain vs. loss) and 2) which option maximized EV, risky or safe. This yielded four context types: Gain Risky–better (GR), Gain Safe–better (GS), Loss Risky–better (LR), and Loss Safe–better (LS) (Figure 1b). Two factors varied between participants: 1) feedback information: *partial* (only the obtained outcome shown) vs. *complete* (obtained and counterfactual shown), and 2) the risky-option probability schedule P_h and P_l (80/20 or 60/40; Figure 1c).

The learning phase comprised 96 trials (24 per context), with contexts interleaved in random order. Feedback condition (partial or complete) and probability schedule (80/20 or 60/40) were fixed per participant throughout learning. In the transfer phase, all unordered pairs of the eight learned symbols were presented without feedback (28 pairs, each shown twice). This phase measured experience-based choices after learning, free of trial-by-trial feedback influence and by presenting all possible option pairs, including mixed-valence (gain–loss) choice problems. In the lottery (description-based) phase, options were presented as numeric outcomes with pie-chart probabilities that matched the learned symbols. The order of the transfer and lottery phases was counterbalanced across participants.

This protocol provides a clear factorial structure that separates within-participant context effects from between-participant feedback and probability settings, and it delivers two complementary behavioral readouts: EV maximization (M-Rate; choosing the EV-maximizing option) and risk preference (R-Rate; choosing the risky option). It also enables a direct comparison of how learned preferences are expressed in experience-based vs. description-based formats.

Assessing Macroscopic choice patterns during the learning phase To characterize aggregate behavior over the learning phase, we summarized trial-by-trial performance within each context using two complementary measures: the EV-maximization rate (M-Rate; probability of selecting the EV-maximizing option) and the risky choice rate (R-Rate; probability of selecting the risky option).

For the gain/loss panels (Figure 2a,b, left), we contrasted valence within the same participants using paired Wilcoxon signed-rank tests ($N = 400$) and assessed each curve against the chance line at .50 using one-sample Wilcoxon tests. The M-Rate trajectories rose quickly above chance and then stabilized at similar levels for gains and losses, a pattern confirmed statistically by significant one-sample effects in both domains (gain vs. .50: $W = 13,419.0$, $P < 0.001$; loss vs. .50: $W = 8,007.0$, $P < 0.001$) and by a non-significant gain–loss difference (paired Wilcoxon: $W = 34,857.5$, $P = 0.426$). In contrast, valence shaped risk taking: R-Rate settled slightly above .50 for gains and below .50 for losses (gain vs. .50: $W = 29,183.0$, $P = 0.005$; loss vs. .50: $W = 27,064.0$, $P < 0.001$), yielding a reliable gain–loss contrast (paired Wilcoxon: $W = 27,643.0$, $P < 0.001$). Thus, while overall accuracy was comparable across valences, risk preference diverged in the expected direction, with a mild bias toward risk in gains and a robust bias against risk in losses.

For the partial/complete panels (Figure 2a,b, right), we compared feedback regimes across independent experiments using Mann–Whitney tests ($N_{\text{partial}} = N_{\text{complete}} = 200$) and again benchmarked each curve against .50 with one-sample Wilcoxon tests. Complete feedback improved accuracy relative to partial (M-Rate: $U = 17,258.0$, $P = 0.018$), consistent with prior

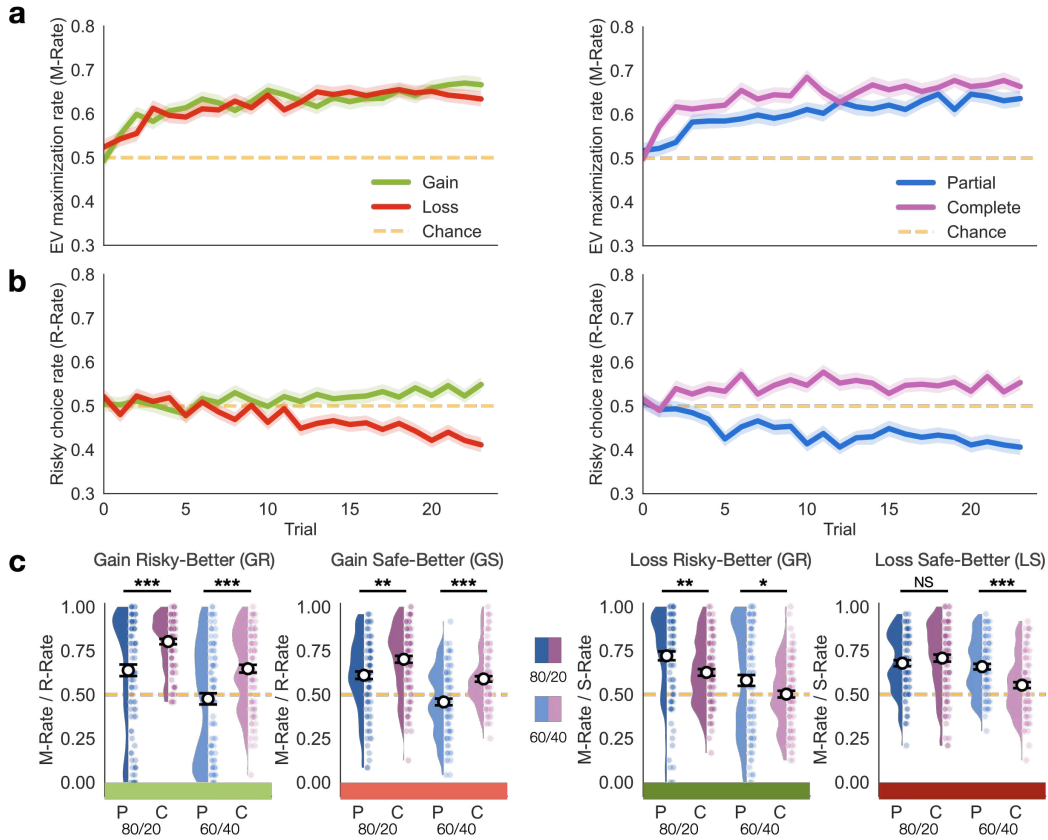


Fig. 2. Choice behavior in the learning phase. **a**, The rate of choosing the higher EV option (M-Rate) over the trials, averaged over all contexts, and **b**, the rate of choosing the risky option (R-Rate) over the trials, averaged over all contexts based on valence (left) and feedback (right). **c**, The choice behavior in the learning phase is separated for each context by the feedback (P: partial, C: complete) and probability type (%80/%20 and %60/%40). S-Rate refers to the safe option choice rate, which is the same as M-Rate in safe-better contexts and R-Rate is the same as M-Rate in risky-better options. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$ Binomial GLM (logit link) with subject-clustered standard errors and Wald χ^2 tests ($N = 100$); N.S., not significant. Error bars and shaded areas on curves represent s.e.m.

evidence that richer outcome information facilitates EV-based learning (29–31); while both regimes were individually above chance (partial vs. .50: $W = 2,941.5$, $P < 0.001$; complete vs. .50: $W = 793.0$, $P < 0.001$). Feedback also produced a marked shift in risk preference: participants were risk averse under partial (R-Rate $< .50$; $W = 5,046.5$, $P < 0.001$) and risk seeking under complete (R-Rate $> .50$; $W = 4,442.5$, $P < 0.001$), and the between-regime difference was large (Mann–Whitney: $U = 10,626.0$, $P < 0.001$).

These observations were corroborated with experiment-specific generalized linear models (GLMs) that quantified how feedback (partial vs. complete) and decision context relate to the two outcome measures (R-Rate and M-Rate). Table 2 summarizes the corresponding GLM contrasts (main effects and interactions), providing a direct comparison of effect sizes across conditions.

The analysis of the learning phase choices revealed two important findings. First, in general, the valence of the decision problem did not affect maximization but did influence risk propensity, with participants being slightly risk-seeking in the gain domain and risk-averse in the loss domain. Second, increasing the amount of feedback information had a moderate impact on maximization and a stronger effect on risk preference, with participants being more risk seeking in the complete feedback situation.

Fine-grained analysis of the effect of feedback information on maximization At the macroscopic level, complete information had a moderate impact on maximization but a strong impact on risk preference. To examine this pattern within each decision problem, we analyzed choice-rate performance by context and probability (Figure 2c). Statistical comparisons use a binomial GLM (logit link) with subject-clustered standard errors and Wald χ^2 tests. For each cell we report the total sample contributing to the contrast.

First, in decision problems where the *risky* option was EV-maximizing, complete feedback increased maximization across valence and probability. In the gain risky context (GR), effects were significant for both probability structures (80/20: $\chi^2(1) = 23.035$, $P < 0.001$, $N = 200$; 60/40: $\chi^2(1) = 19.268$, $P < 0.001$, $N = 200$). In the loss risky context (LR), the same pattern held (80/20: $\chi^2(1) = 9.941$, $P = 0.002$, $N = 200$; 60/40: $\chi^2(1) = 27.536$, $P < 0.001$, $N = 200$).

Table 2. Choice behavior and statistical comparisons across experiments in the learning phase. The table reports mean $\pm$ SEM of risky choice rates (top) and optimal choice rates (EV maximization - bottom) across four experiments. Statistical comparisons include both within-subject factors (*Valence*, *Riskiness/Choice Optimality*) and between-subject factors (*Feedback*, *Probability*), analyzed using GLMs with binomial error distribution and logit link function. Statistical significance was assessed using likelihood ratio tests and Wald tests.

		Risky choice rate				Statistics (within subjects)		
		Gain risky-better	Loss risky-better	Gain safe-better	Loss safe-better	Valence	Riskiness	Interaction
Partial 80/20		0.6383 $\pm$ 0.0331	0.6104 $\pm$ 0.0212	0.2800 $\pm$ 0.0253	0.3221 $\pm$ 0.0175	-0.0421	0.2883***	0.0700
Complete 80/20		0.8021 $\pm$ 0.0155	0.7013 $\pm$ 0.0194	0.3746 $\pm$ 0.0191	0.2929 $\pm$ 0.0200	0.0817**	0.4083***	0.0192
Partial 60/40		0.4758 $\pm$ 0.0320	0.4579 $\pm$ 0.0191	0.4200 $\pm$ 0.0309	0.3417 $\pm$ 0.0160	0.0783*	0.1163**	-0.0604
Complete 60/40		0.6475 $\pm$ 0.0222	0.5892 $\pm$ 0.0159	0.4983 $\pm$ 0.0192	0.4471 $\pm$ 0.0171	0.0512 [†]	0.1421***	0.0071
Statistics (between subjects)	Feedback	0.1717***	0.1312***	0.0783*	0.1054***			
	Probability	0.1625***	0.1525***	-0.1400***	-0.0196			
	Interaction	-0.0079	-0.0404	0.0163	-0.1346***			
		Optimal choice rate (EV Maximization)				Statistics (within subjects)		
		Gain risky-better	Loss risky-better	Gain safe-better	Loss safe better	Valence	Choice Optimality	Interaction
Partial 80/20		0.6383 $\pm$ 0.0331	0.6104 $\pm$ 0.0212	0.7200 $\pm$ 0.0253	0.6779 $\pm$ 0.0175	0.0421	-0.0675 [†]	-0.0142
Complete 80/20		0.8021 $\pm$ 0.0155	0.7013 $\pm$ 0.0194	0.6254 $\pm$ 0.0191	0.7071 $\pm$ 0.0200	-0.0817**	-0.0058	0.1825***
Partial 60/40		0.4758 $\pm$ 0.0320	0.4579 $\pm$ 0.0191	0.5800 $\pm$ 0.0309	0.6583 $\pm$ 0.0160	-0.0783*	-0.2004***	0.0963 [†]
Complete 60/40		0.6475 $\pm$ 0.0222	0.5892 $\pm$ 0.0159	0.5017 $\pm$ 0.0192	0.5529 $\pm$ 0.0171	-0.0513 [†]	0.0363	0.1096**
Statistics (between subjects)	Feedback	0.1717***	0.1312***	-0.0783*	-0.1054***			
	Probability	0.1625***	0.1525***	0.1400***	0.0196			
	Interaction	-0.0079	-0.0404	-0.0163	0.1346***			

[†] $P < 0.06$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

We then turned to decision problems where the *safe* option was EV-maximizing. Here, complete feedback generally reduced maximization. In the gain safe context (GS), maximization decreased under complete feedback for both probability structures (80/20: $\chi^2(1) = 8.353$, $P = 0.004$, $N = 200$; 60/40: $\chi^2(1) = 4.559$, $P = 0.033$, $N = 200$). In the loss safe context (LS), the 60/40 case showed a clear reduction ($\chi^2(1) = 20.057$, $P < 0.001$, $N = 200$), whereas the 80/20 case did not differ reliably from partial feedback ($\chi^2(1) = 1.200$, $P = 0.273$, $N = 200$).

These context-dependent results indicate that complete feedback primarily shifts behavior toward the risky option. It improves EV maximization when the risky option is optimal (GR and LR) and reduces maximization when the safe option is optimal (GS and LS), with LS 80/20 as the only exception (see Supplementary Figure S3 for within context R-Rate dynamics across trials).

Analysing trial-by-trial choice changes Our findings reveal that risk propensity is moderately influenced by the valence of the decision problem (gain and loss), with participants displaying risk aversion in the loss domain, and strongly affected by feedback information. Specifically, we observed a shift from overall risk aversion in the partial feedback condition to overall risk-seeking behavior in the complete feedback condition.

To further investigate trial-by-trial adaptation dynamics, we analyzed choice repetition rates and how they depended on the previous choice and feedback received. In both partial and complete feedback experiments, the probability of repeating a risky choice was strongly modulated by whether the previous risky choice resulted in the best or worst outcome within the given context (Figure 3a - Wilcoxon signed-rank test; partial: $Z = 11356.5$; $P < 0.001$; complete: $Z = 15538.0$; $P < 0.001$). This is consistent with a learning-based explanation and a positive recency effect, where choice that delivered good outcomes tend to be repeated (see Supplementary Figure S2 for the aggregated repetition rates of the safe and risky options irrespective of the risky outcome).

In the complete feedback experiment (Figure 3b), we analyzed how the outcome obtained from the unchosen risky option (i.e., counterfactual feedback) influenced subsequent choices. We found that observing the best possible outcome from the unchosen risky option significantly decreased the likelihood of repeating a safe choice (Wilcoxon signed-rank test; $Z = 12267.0$; $P < 0.001$). This finding supports the idea that participants process forgone outcomes and that the regret associated with comparing the safe choice to the best possible risky outcome drives an increase in risk propensity (40, 60).

In addition, in order to quantify how obtained and counterfactual outcomes shape the tendency to stay with the same symbol on its next presentation within a context, we fit a subject-level GLMM and extracted per-participant coefficients (Figure 3b) for a baseline stay tendency (θ_0), sensitivity to the realized outcome (θ_1), and, only under complete feedback, a counterfactual (regret-like) sensitivity (θ_2 - see Methods for details). Intercepts were positive in both regimes (partial: $\theta_0 = 0.98 \pm 0.04$ SEM; complete: $\theta_0 = 0.65 \pm 0.03$ SEM; one-sample Wilcoxon vs. 0, two-sided, each $P < 0.001$, $N = 200$), and realized outcomes reinforced staying (partial: $\theta_1 = 0.51 \pm 0.02$ SEM; complete: $\theta_1 = 0.62 \pm 0.01$ SEM; each $P < 0.001$, $N = 200$).

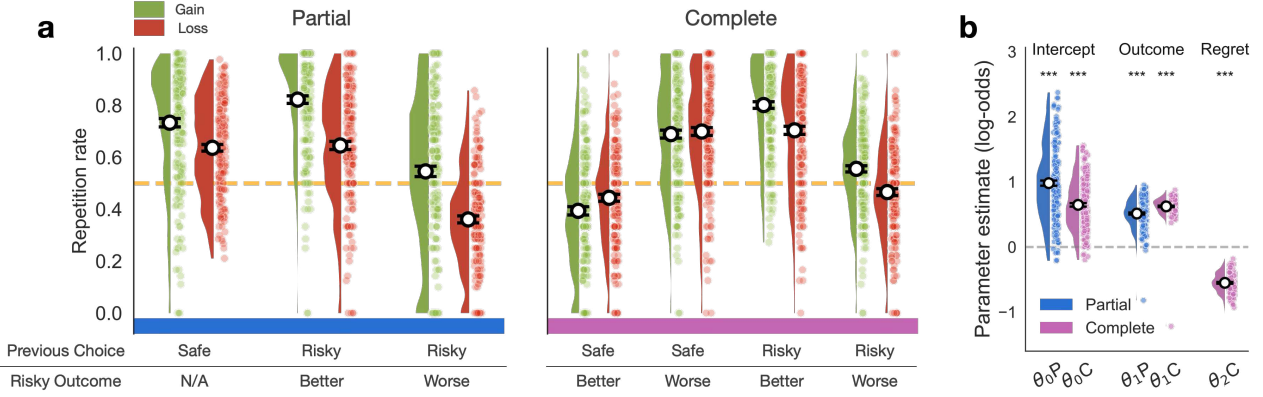


Fig. 3. Choice repetition rates across feedback types. **a**, The repetition ratio during the learning phase for each option type in experiments with partial feedback (left) and complete feedback (right). The repetition ratio represents the probability of selecting the same option on consecutive trials within each choice context, regardless of whether the choice maximizes EV. **b**, Regression coefficients predicting the probability of stay in the next trial. θ_0 : Intercept, θ_1 : chosen outcome, θ_2 : counterfactual (regret) outcome. *** $P < 0.001$, one-sample Wilcoxon against zero ($N = 200$ per condition); Error bars represent s.e.m.

Most importantly, the counterfactual coefficient under complete feedback was negative ($\bar{\theta}_2 = -0.55 \pm 0.01$ SEM; $P < 0.001$, $N = 200$), indicating that counterfactual outcomes for the unchosen option reliably reduce subsequent stay probability.

Assessing the description–experience gap To compare risk taking across description and experience in a controlled way, we extracted from the transfer and lottery phases the same four decision contexts used in learning (GR, GS, LR, LS). For each participant and phase, each context appeared twice, yielding eight binary choices per participant (Figure 4). Risky choice was treated as a binary outcome and analyzed within each experiment ($N = 100$) using generalized linear models with a logit link; statistical significance was assessed with Wald tests on model coefficients. This context-matched approach lets us compare like with like across experience-based choices without feedback (transfer) and description-based choices with explicit probabilities and outcomes (lottery).

In the lottery (description) task, we reproduced the classic reflection effect: risk seeking was higher in losses than in gains (4, 7, 22). The gain–loss contrast was significant in all experiments (GLM with logit link; Wald tests, $N = 100$): 80/20 partial, $\chi^2(1) = 12.700$, $P = 0.012$; 80/20 complete, $\chi^2(1) = 12.100$, $P < 0.001$; 60/40 partial, $\chi^2(1) = 14.000$, $P < 0.001$; 60/40 complete, $\chi^2(1) = 19.500$, $P < 0.001$ (Figure 4a & b).

In the matched transfer (experience) task, the pattern reversed, consistent with prior reports of inverse reflection in experience-based choice (55, 56): risk seeking was higher in gains than in losses in three of four experiments (GLM with logit link; Wald tests, $N = 100$): 80/20 complete, $\chi^2(1) = 9.940$, $P = 0.005$; 60/40 partial, $\chi^2(1) = 17.100$, $P < 0.001$; 60/40 complete, $\chi^2(1) = 12.900$, $P = 0.045$. The 80/20 partial condition was not significant, $\chi^2(1) = 12.500$, $P = 0.105$ (Figure 4a & b).

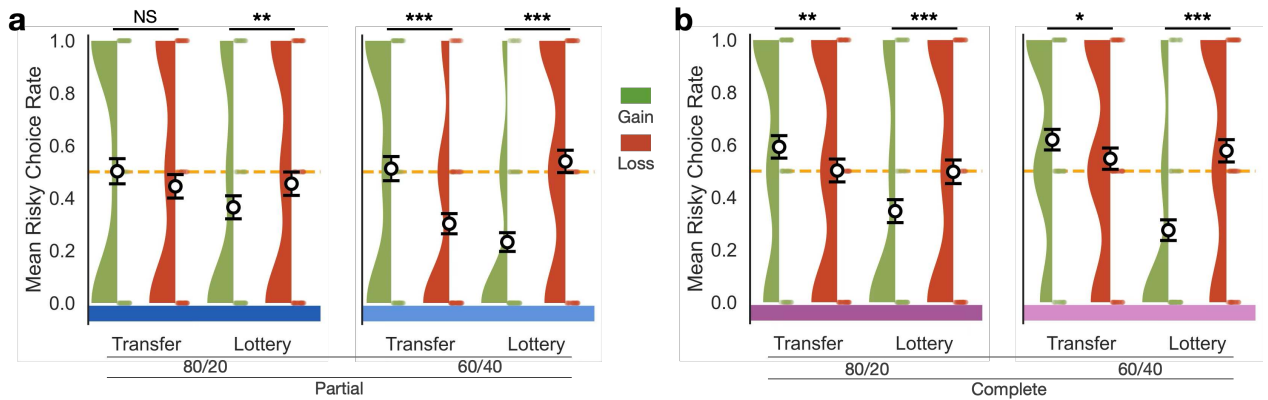


Fig. 4. Inverse reflection effect in experience-based choices. The learning phase choice problems (contexts) were extracted from the transfer and lottery phases. Then the mean risky choice rates (R-Rate) were calculated separately for each participant between gain and loss contexts by the feedback type (**a**, partial vs. **b**, complete) and probability type (80/20 vs. 60/40). *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$ Binomial GLM (logit link) with subject-clustered standard errors and Wald χ^2 tests ($N = 100$); N.S., not significant. Error bars represent s.e.m.

177 These results show a reliable description–experience gap when the same decision problems are evaluated with and without
178 explicit information: description favors greater risk taking in losses relative to gains, whereas experience favors greater risk
179 taking in gains relative to losses. The utility-function modeling that directly contrasts description and experience is provided in
180 Supplementary Figure S1, and comprehensive analyses of transfer and explicit (lottery) choices are provided in Supplementary
181 Figure S4 and S5.

182 Discussion

183 In this study, we leveraged an experimental design in which risk preference was orthogonal to reward maximization, allowing us
184 to dissociate the effects of feedback and valence on learning risky decisions. Our aim was to better understand how feedback
185 and outcome valence shape learning dynamics, risk preferences, and the formation of stable choice patterns.

186 Our first set of results concerned the learning phase. We found that outcome valence had no effect on choice accuracy.
187 This finding is consistent with previous studies and is not readily predicted by simple RL specifications that might anticipate
188 higher performance in the gain domain (the avoidance-learning paradox) (33, 61, 62). Nor is it straightforwardly explained by
189 naïve interpretations of prospect theory (63, 64), which could suggest that losses, being more salient, should lead to improved
190 learning. Indeed, some prior work has reported higher maximization rate under loss-avoidance incentives (e.g., loss-attention
191 accounts) (57, 65). However, a growing body of experience-based RL studies reports no reliable main effect of valence on EV
192 maximization (e.g., equal performance in gain vs loss contexts despite metacognitive or motor biases) (44, 66–68), and other
193 work even finds advantages for gains over losses under some task demands (69). Instead, the similar EV maximization observed
194 across valence domains is traditionally well explained by assuming a form of relative value learning, in which neutral outcomes
195 are subjectively perceived as negative or positive in the loss and gain conditions, respectively (70).

196 In contrast, we observed a clear effect of feedback: complete feedback improved accuracy compared to partial feedback.
197 This replicates prior work (e.g., 33) and can be explained by faster convergence under complete feedback because participants
198 directly observe both chosen and unchosen outcomes, reducing the need for exploration. Under such conditions, complete
199 feedback can shift the exploration–exploitation balance toward exploitation, increasing expected-value–maximizing choices on
200 average. Residual suboptimal choices in this setting likely reflect a mixture of lapses, decision noise, and misestimation rather
201 than systematic learning failures (39, 71, 72).

202 However, complete feedback also led to greater risk taking, even when the risky option was suboptimal, with the influence
203 of feedback information on risk propensity being stronger than its influence on maximization. Since our task dissociates
204 reward maximization from risk preference, this finding cannot be explained as a byproduct of improved learning. A plausible
205 mechanism for risk aversion in the partial feedback condition is the *hot stove effect* (73), whereby unfavorable recent outcomes
206 suppress re-sampling of the risky option and produce apparent risk aversion. This interpretation fits well risk aversion in the
207 loss domain, because unfavorable outcomes are inherently negative and thus readily elicit avoidance. In the gain domain, where
208 outcomes are non-aversive, the hot-stove effect should not be observed if outcomes are encoded on an absolute scale. However,
209 it may still emerge if expectations are defined relative to a local reference point—an assumption that is consistent with the
210 observation of similar levels of maximization in the gain and loss conditions (70, 74–76). This result, along with the hot-stove
211 effect, also aligns well with studies showing that value updating tends to overweight recent negative outcomes. (27).

212 While the hot-stove effect readily explains risk avoidance in partial-feedback contexts (especially in the loss domain), it
213 does not account for risk seeking in complete-feedback contexts. We therefore considered additional mechanisms related to
214 counterfactual outcome processing. Regret avoidance is one such candidate: it requires the comparison between actual and
215 foregone outcomes, which is made explicit when complete feedback is available (77–81). We therefore hypothesised that the
216 increase in risk propensity in the complete-feedback condition was specifically driven by trials in which the comparison between
217 the obtained (safe) outcome and the counterfactual (risky) outcome was unfavorable. To do so, we examined trial-by-trial choice
218 dynamics. As predicted by standard learning models, whether rooted in the RL or Bayesian tradition, participants generally
219 repeated risky choices after being rewarded (positive recency; 5, 82–84). Accordingly, in the partial-feedback condition, when
220 receiving favorable feedback after choosing the risky option, participants were more likely to stay with that choice compared to
221 when they received an unfavorable outcome.

222 When examining the trial-by-trial dynamics in the complete-feedback condition, we found that the probability of repeating
223 a risky option was comparable to that observed in the partial-feedback experiments. Conversely, under complete feedback,
224 the decision to repeat a safe choice was strongly influenced by the outcome of the unchosen alternative. More precisely, and
225 understandably, the probability of staying with the safe option decreased when the counterfactual outcome associated with
226 the unchosen risky option was favorable. This observation is also consistent with the regression analysis showing a negative
227 impact of the value of forgone outcomes on subsequent choices. Overall, the observed pattern suggests that the increase in risk
228 propensity observed in the complete-feedback experiments was primarily driven by a regret-induced tendency to switch away
229 from the safe option when the risky alternative would have yielded a better outcome: a situation that cannot occur in the
230 partial-feedback condition, where the outcome of the risky option remains unknown (60, 85).

231 The analysis of trial-by-trial response changes (specifically the finding that choice repetition rates were higher in the gain
232 condition) is also in line with previous results by Lejarraga and Hertwig, who showed that individuals display less stable
233 preferences in the loss than in the gain domain (18). However, this cannot be explained by an increase in exploratory tendency,
234 as previously proposed, since the effect also persisted in the complete-feedback experiment, where exploration is not required
235 and rather reflects the fact that losses, being more motivationally salient, may hinder the development of stable decision policies
236 (57, 86).

Another robust finding was the gradual divergence of risk preferences across the gain and loss domains during learning. Over time, participants became mildly risk-seeking in the gain domain and more clearly risk-averse in the loss domain. This suggests that the hot stove effect may contribute to risk aversion in the loss domain, but cannot by itself explain the shift toward risk seeking in the gain domain. Other cognitive processes, such as over-weighting of positive outcomes (positivity bias) (70, 87, 88), may lead to an overvaluation of risky options in gains and drive this asymmetric development of preferences.

This pattern is particularly striking, as it represents a clear departure from prospect theory. One of the key behavioral phenomena explained by prospect theory is the reflection effect, according to which individuals are typically risk-averse in the gain domain and risk-seeking in the loss domain. (22, 89). Here, we observed the opposite: participants became more risk seeking in gains and more risk averse in losses. This “inverse reflection effect” has been previously reported (55, 56, 90).

Our study extends previous findings by showing that the effect also persists after learning, in a transfer test where feedback is no longer presented, indicating that its expression does not depend on the immediate delivery of trial-by-trial feedback. The inverse reflection effect was evident both at the behavioral level, but was further supported by comparisons of utility parameters, especially those related to loss sensitivity. In the lottery task, and consistent with prospect theory, loss sensitivity was higher in the loss domain than in the gain domain. However, in the RL tasks, the opposite pattern was found: loss sensitivity was lower in the loss domain. Critically, this pattern held across both feedback conditions (partial and complete), pointing to two important conclusions. First, although there is no feedback during the transfer phase, the nature of the learning experience (i.e., partial or complete) has a durable influence on preference structure. Second, this suggests that the description–experience gap, here instantiated as the inverse reflection effect measured in risk preferences during learning and the feedback-free transfer phase, together with fitted loss sensitivity parameters in the lottery and transfer phases, cannot be attributed solely to limited sampling, insufficient information or instant feedback. If that were the case, the gap should have been smaller in the complete feedback condition, which it was not (6, 91). The absence of loss sensitivity in the RL task is also in line with what was reported by (18). Reanalyzing previous choice-prediction-competition data (84), their model comparison revealed that simple heuristics such as win–stay/lose–shift, rather than value-based models assuming loss aversion, better captured the observed behavior.

It is important to acknowledge some limitations of our experimental setup. First, our study featured relatively small monetary incentives. This may have reduced participant engagement, limiting the depth of their evaluations and risk assessments. However, the observed rate of maximization, as well as our ability to detect robust differences between the learning and lottery tasks, argues against a lack of engagement (29, 57). Second, our study involved only binary choices between a safe and a risky option. Future studies would benefit from including conditions with more than two options and two risky options, which would provide a more sensitive measure of risk attitudes across different domains. Further, mechanistic computational modeling could be used to characterize how feedback shapes belief updating and value comparison mechanisms. This would advance our understanding of how feedback influences learning, beyond what is captured by performance or choice behavior alone.

To conclude, results suggest that the structure of feedback, beyond mere information quantity, plays a critical role in shaping decision strategies. This dissociation has practical implications for behavioral policy design and digital choice architectures in environments where individuals learn from outcomes, such as retail investing platforms, consumer credit interfaces, insurance, and medical decision aids, because the presentation of counterfactual information is a design choice that can durably steer risk preferences. Specifically, our results highlight that providing complete counterfactual feedback may increase risk-taking more than it improves accuracy, implying that “more information” can be behaviorally non-neutral: it may improve learning while simultaneously shifting preferences in ways that could increase exposure to downside risk. From a policy perspective, this suggests that disclosure and feedback interventions should be evaluated not only for their effects on knowledge or accuracy but also for their downstream effects on risk-taking. These findings contribute to a growing body of evidence indicating that experience-based decisions deviate from normative benchmarks not simply due to limited sampling, but also due to distinct cognitive and affective processes triggered by valence, feedback availability and counterfactual information (43, 92).

Methods

Participants A total of 400 participants (100 participants for each of the 4 experiments) were recruited with an equal gender ratio. We recruited participants on the Prolific platform. Each person could join only one of the four experiments, and we blocked repeat sign-ups across experiments. All participants who completed the task are included in the data, and no participant was excluded. Their final reward was calculated based on the sum of outcomes earned across all trials (1 pt = £0.0285). In addition to a fixed Prolific participation fee (£2), participants received an initial bonus of 70 points (equivalent to an additional £2). Depending on their performance, they could either double their points (earning up to £6 in total) or lose the bonus entirely. The study was approved by the INSERM ethics committee (IRB00003888). We obtained online written informed consent before the task, and all data were anonymized.

Behavioral Task Each experiment consisted of 3 phases (Figure 1a). The first is the learning phase, which has 4 types of so-called “contexts” - a pair of symbols that are always presented together. In each context, there is an option with a deterministic outcome (safe option) and a probabilistic outcome (risky option), and the risk probabilities vary between experiments (defined by the parameters P_h and P_l - Figure 1c). The learning phase consists of 2×2 design contexts; one dimension indicates whether the valence of the context is loss or gain, and the other is dedicated to whether the correct option (the one with a higher EV) is risky or safe (Figure 1b). These 4 contexts are called GR (gain risky-better), GS (gain safe-better), LR (loss risky-better), and LS (loss safe-better). Across all contexts, the absolute value of the difference between the expected outcomes

of the options is the same within each trial. Each context is presented 24 times in random order during the learning phase (overall 96 trials). The position of each symbol within the context is also counterbalanced.

During each trial of the learning phase, the participant is presented with a context and is given no time pressure to choose an option. After a confirmation period of 0.5 sec, feedback on the chosen option is presented for 1.5 sec. The unit of the feedback is in *points* (pt) and the range is integers between -2 and +2. At the beginning of the experiment, each participant was given 100 pt and a fixed £2 Prolific credit (Figure 1c). In our experiments, there are two types of feedback: 1) partial: where the participant only sees the result of their chosen option, and 2) complete: where both the results of the chosen and unchosen options are shown, but only the chosen one would count in their collected points. We designed 4 experiments to test all possible combinations based on the feedback type (partial/complete) and the parameter values of P_h and P_l (80%-20% or 60%-40%, Figure 1c).

The second phase is the transfer phase (Figure 1a), where all possible pairs of 8 symbols shown in the previous phase are shown to the participants (28 possible pairs), and each pair is shown 2 times. Thus, in the transfer phase, there is no notion of contexts and no feedback on choice. Participants have to choose between novel pairs based on their experience. This phase represents experience-based decisions over risky choices.

The third phase is called the lottery phase (Figure 1a). In this phase, instead of symbols, the participant is presented with a pair of pie charts, each representing the probability of the indicated outcomes. Without giving the participants any information, the pie charts represent the symbols of the learning phase, but in an explicit way. In addition, in all experiments (except Exp. 1), we added two 50%-50% pie charts (one for the gain domain and one for the loss domain) to elicit more risk preferences from the participants and to have the same number of trials in the learning phase and the lottery phase. Thus, we have 8 unique pie charts with 28 possible combinations, and each combination is represented 2 times (a total of 56 trials for all experiments except Exp. 1 with 30 trials). This phase represents experience-based decisions over risky choices. For half of the participants, the lottery phase came after the learning and transfer phase, and for the rest, it was at the beginning to counterbalance the phase orders. There was a 3-question instruction quiz before the start of both the learning and lottery phases. Participants who failed the quiz could not proceed to the task and were redirected to review the instructions again.

Statistical analysis for Table 2 We tested how feedback information shaped choice rates using compact Gaussian generalized linear models (GLMs) fitted to subject-level averages. Rates were averaged over the 24 repetitions per context (GR, GS, LR, LS). This aggregation yields approximately normal sampling error, so a Gaussian GLM with identity link provides effect estimates directly in probability units. We analyzed two dependent variables in separate passes: risky choice rate (R-Rate) and EV-maximization rate (M-Rate). Cell means are reported as $\bar{y} \pm \text{SEM}$. Each experiment had $N = 100$.

Between-subjects model (per context). To test effects of feedback and probability across experiments we fit, for each context,

$$y = \beta_0 + \beta_F F + \beta_P P + \beta_{FP} FP + \epsilon \quad [1]$$

where $F \in \{0, 1\}$ (Complete=1, Partial=0) and $P \in \{0, 1\}$ (80/20=1, 60/40=0). We report Wald tests (z , two-sided P) for β_F , β_P , and β_{FP} in “Statistics (between subjects)”.

Within-experiment model. To summarize the within-subject structure inside each experiment we fit

$$y = \gamma_0 + \gamma_G G + \gamma_R R + \gamma_{GR} GR + \eta \quad [2]$$

with $G \in \{0, 1\}$ (Gain=1, Loss=0) and $R \in \{0, 1\}$ (risky option is EV-better=1). Wald tests for γ_G , γ_R , and γ_{GR} are shown in the “Statistics (within subjects)” columns. All models were estimated in Python (statsmodels `glm`); two-sided Wald tests.

Statistical analysis of feedback effects on maximization We quantified how feedback information (partial vs complete) affected expected-value (EV) maximization at the level of each decision problem in the learning phase. Analyses were performed separately within each context (GR, GS, LR, LS) and each probability structure (80/20, 60/40). For each trial we defined a binary outcome Success_t that matched the panel semantics: $\text{Success} = 1$ for a risky choice in GR and LR (where the risky option is EV-maximizing) and $\text{Success} = 1$ for a safe choice in GS and LS (where the safe option is EV-maximizing). We then fit a binomial GLM (logit link) at the trial level:

$$\text{logit Pr}(\text{Success}_t = 1) = \beta_0 + \beta_1 \mathbb{I}\{\text{feedback} = \text{comp.}\}. \quad [3]$$

To account for repeated measures, we used subject-clustered (Huber–White) standard errors with Subject as the clustering unit. For each Context $\times$ Probability cell we reported the Wald test for the complete coefficient β_1 as $\chi^2(1)$ together with the corresponding two-sided P value, and we expressed the effect size as an odds ratio $\text{OR} = \exp(\beta_1)$ with a 95% confidence interval based on the cluster-robust variance. Each contrast combined the two matched groups and included $N = 200$ participants. Model fitting and inference were implemented in Python (statsmodels), and figures display group means with standard error of the mean for visualization only.

Probability of stay GLMM and statistical testing For each subject and context, we built trial-to-trial transitions ($i \rightarrow i+1$) and coded stay = 1 if the exact same symbol was chosen again at $i+1$, else 0. Let r_{chosen} be the realized outcome on trial i ; in complete feedback we also defined a counterfactual regressor Δ as the outcome of the unchosen option on trial i . We estimated a binomial GLMM with a logit link and subject-varying intercepts and slopes:

$$\text{logit } P(\text{stay}) = (\theta_0 + u_{0s}) + (\theta_1 + u_{1s}) r_{\text{chosen}} \quad [4]$$

$$\text{logit } P(\text{stay}) = (\theta_0 + u_{0s}) + (\theta_1 + u_{1s}) r_{\text{chosen}} + (\theta_2 + u_{2s}) \Delta \quad [5]$$

where $u_{\cdot s}$ are subject-specific random effects. Subject-level coefficients are $\theta_0 + u_{0s}$, $\theta_1 + u_{1s}$, and, in complete feedback, $\theta_2 + u_{2s}$. To test whether each coefficient distribution differed from zero across subjects, we applied a two-sided one-sample Wilcoxon signed-rank test. We report minimal summaries (means $\bar{\theta}$, p -values, and n) in the Results.

Data and code availability

Anonymized data from all experiments and the analysis codes needed to reproduce the figures and statistical reports are available online on the Open Science Framework (OSF) project at this link: osf.io/gjpyz.

Acknowledgments

S.P. is supported by the European Research Council under the European Union's Horizon 2020 research and innovation program (ERC) (RaReMem: 101043804), and the Agence National de la Recherche (CogFinAgent: ANR-21-CE23-0002-02; RELATIVE: ANR-21-CE37-0008-01; RANGE: ANR-21-CE28-0024-01). The Département d'Etudes Cognitives of the École Normale Supérieure - PSL University is funded by the Agence Nationale pour la Recherche (ANR-17-EURE-0017, ANR-10-IDEX-0001-02). This work has received support under the Major Research Program of PSL Research University "PSL-Neuro" launched by PSL Research University and implemented by ANR (ANR-10-IDEX-0001). A.S. is supported by "Frontiers in Cognition" (ANR-17-EURE-0017).

Competing interests

The authors declare no competing interests.

References

1. RA Poldrack, T Yarkoni, From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annu. review psychology* **67**, 587–612 (2016).
2. B De Martino, CF Camerer, R Adolphs, Amygdala damage eliminates monetary loss aversion. *Proc. Natl. Acad. Sci.* **107**, 3788–3792 (2010).
3. S Mishra, Decision-making under risk: Integrating perspectives from biology, economics, and psychology. *Pers. Soc. Psychol. Rev.* **18**, 280–307 (2014).
4. R Hertwig, I Erev, The description–experience gap in risky choice. *Trends cognitive sciences* **13**, 517–523 (2009).
5. R Hertwig, G Barron, EU Weber, I Erev, Decisions from experience and the effect of rare events in risky choice. *Psychol. science* **15**, 534–539 (2004).
6. DU Wulff, M Mergenthaler-Canseco, R Hertwig, A meta-analytic review of two modes of learning and the description–experience gap. *Psychol. bulletin* **144**, 140 (2018).
7. B Garcia, F Cerrotti, S Palminteri, The description–experience gap: a challenge for the neuroeconomics of decision-making under uncertainty. *Philos. Transactions Royal Soc. B* **376**, 20190665 (2021).
8. R Bhui, SJ Gershman, Decision by sampling implements efficient coding of psychoeconomic functions. *Psychol. review* **125**, 985 (2018).
9. SJ Gershman, Empirical priors for reinforcement learning models. *J. Math. Psychol.* **71**, 1–6 (2016).
10. AG Collins, MJ Frank, How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and neurogenetic analysis. *Eur. J. Neurosci.* **35**, 1024–1035 (2012).
11. SJ Gershman, ND Daw, Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annu. review psychology* **68**, 101–128 (2017).
12. G Gigerenzer, R Hertwig, E Van Den Broek, B Fasolo, KV Katsikopoulos, "a 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis: An Int. J.* **25**, 623–629 (2005).
13. EU Weber, PC Stern, Public understanding of climate change in the united states. *Am. psychologist* **66**, 315 (2011).
14. T Lejarraga, T Pachur, R Frey, R Hertwig, Decisions from experience: From monetary to medical gambles. *J. Behav. Decis. Mak.* **29**, 67–77 (2016).
15. T Lejarraga, JK Woike, R Hertwig, Description and experience: How experimental investors learn about booms and busts affects their financial risk taking. *Cognition* **157**, 365–383 (2016).
16. RS Sutton, AG Barto, et al., *Reinforcement learning: An introduction*. (MIT press Cambridge) Vol. 1, (1998).
17. T Lejarraga, C Gonzalez, Effects of feedback and complexity on repeated decisions from description. *Organ. behavior human decision processes* **116**, 286–295 (2011).
18. T Lejarraga, R Hertwig, How the threat of losses makes people explore more than the promise of gains. *Psychon. bulletin & review* **24**, 708–720 (2017).
19. ML Pedersen, MJ Frank, G Biele, The drift diffusion model as the choice rule in reinforcement learning. *Psychon. bulletin & review* **24**, 1234–1251 (2017).
20. P Dayan, Y Niv, Reinforcement learning: the good, the bad and the ugly. *Curr. opinion neurobiology* **18**, 185–196 (2008).
21. I Erev, AE Roth, Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am. economic review* pp. 848–881 (1998).
22. D Kahneman, A Tversky, Prospect theory: An analysis of decision under risk in *Handbook of the fundamentals of financial decision making: Part I*. (World Scientific), pp. 99–127 (2013).
23. BA Mellers, A Schwartz, AD Cooke, Judgment and decision making. *Annu. review psychology* **49**, 447–477 (1998).
24. J Spicer, TL Mullett, AN Sanborn, Repeated risky choices become more consistent with themselves but not expected value, with no effect of matched trial order. *Judgm. Decis. Mak.* **19**, e2 (2024).
25. I Erev, AE Roth, Maximization, learning, and economic behavior. *Proc. Natl. Acad. Sci.* **111**, 10818–10825 (2014).
26. R Frey, A Pedroni, R Mata, J Rieskamp, R Hertwig, Risk preference shares the psychometric structure of major psychological traits. *Sci. advances* **3**, e1701381 (2017).
27. J Denrell, Adaptive learning and risk taking. *Psychol. review* **114**, 177 (2007).
28. B Rosenberg, E Eldar, The experience–experience gap: Distributional learning is associated with a divergence of preferences from estimations. *Res. Sq.* pp. rs–3 (2025).
29. E Yechiam, JR Busemeyer, The effect of foregone payoffs on underweighting small probability events. *J. Behav. Decis. Mak.* **19**, 1–16 (2006).
30. I Erev, G Barron, On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychol. review* **112**, 912 (2005).
31. L Fontanesi, S Palminteri, M Lebreton, Decomposing the effects of context valence and feedback information on speed and accuracy during reinforcement learning: a meta-analytical approach using diffusion decision modeling. *Cogn. Affect. & Behav. Neurosci.* **19**, 490–502 (2019).
32. E Eldar, C Roth, P Dayan, RJ Dolan, Decodability of reward learning signals predicts mood fluctuations. *Curr. biology* **28**, 1433–1439 (2018).
33. S Palminteri, M Khamassi, M Joffily, G Coricelli, Contextual modulation of value signals in reward and punishment learning. *Nat. communications* **6**, 8096 (2015).
34. S Mousavi, G Gigerenzer, Heuristics are tools for uncertainty. *Homo Oeconomicus* **34**, 361–379 (2017).
35. S Kleinberg, JK Marsh, Less is more: information needs, information wants, and what makes causal models useful. *Cogn. Res. Princ. Implic.* **8**, 57 (2023).
36. AB Zhou, et al., The impact of information presentation on self-other risk decision-making. *Front. Psychol.* **15**, 1357644 (2024).
37. T Rakow, BR Newell, L Wright, Forgone but not forgotten: The effects of partial and full feedback in "harsh" and "kind" environments. *Psychon. bulletin & review* **22**, 1807–1813 (2015).

38. Z Barakchian, AH Vahabie, M Nili Ahmadabadi, Implicit counterfactual effect in partial feedback reinforcement learning: behavioral and modeling approach. *Front. Neurosci.* **16**, 631347 (2022).
39. C Findling, V Skvortsova, R Dromnelle, S Palminteri, V Wyart, Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nat. neuroscience* **22**, 2066–2077 (2019).
40. S Bavard, S Palminteri, Why unchosen options linger in our minds. *Commun. Biol.* **4**, 1271 (2021).
41. O Kopsacheilis, The role of information search and its influence on risk preferences. *Theory Decis.* **84**, 311–339 (2018).
42. R Cubitt, O Kopsacheilis, C Starmer, An inquiry into the nature and causes of the description-experience gap. *J. Risk Uncertain.* **65**, 105–137 (2022).
43. B Garcia, M Lebreton, S Bourgeois-Gironde, S Palminteri, Experiential values are underweighted in decisions involving symbolic options. *Nat. human behaviour* **7**, 611–626 (2023).
44. A Erdman, et al., Experience-based risk taking is primarily shaped by prior learning rather than by decision-making. *Nat. Commun.* **16**, 6310 (2025).
45. K Teodorescu, M Amir, I Erev, The experience–description gap and the role of the inter decision interval. *Prog. brain research* **202**, 99–115 (2013).
46. L Weiss-Cohen, E Konstantinidis, N Harvey, Timing of descriptions shapes experience-based risky choice. *J. Behav. Decis. Mak.* **34**, 66–84 (2021).
47. RH Thaler, CR Sunstein, *Nudge: The final edition*. (Penguin), (2021).
48. R Hertwig, TJ Pleskac, T Pachur, *Taming uncertainty*. (Mit Press), (2019).
49. DU Wulff, TT Hills, R Hertwig, Online product reviews and the description–experience gap. *J. Behav. Decis. Mak.* **28**, 214–223 (2015).
50. R Hertwig, T Grüne-Yanoff, Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychol. Sci.* **12**, 973–986 (2017).
51. SM Herzog, R Hertwig, Boosting: Empowering citizens with behavioral science. *Annu. Rev. Psychol.* **76** (2025).
52. Y Lin, M Osman, R Ashcroft, Nudge: concept, effectiveness, and ethics. *Basic Appl. Soc. Psychol.* **39**, 293–306 (2017).
53. R Hau, TJ Pleskac, J Kiefer, R Hertwig, The description–experience gap in risky choice: The role of sample size and experienced probabilities. *J. Behav. Decis. Mak.* **21**, 493–518 (2008).
54. D Prelec, The probability weighting function. *Econometrica* pp. 497–527 (1998).
55. EA Ludvig, ML Spetch, Of black swans and tossed coins: is the description-experience gap in risky choice limited to rare events? *PLoS one* **6**, e20262 (2011).
56. CR Madan, ML Spetch, FM Machado, A Mason, EA Ludvig, Encoding context determines risky choice. *Psychol. Sci.* **32**, 743–754 (2021).
57. E Yechiam, G Hochman, Loss-aversion or loss-attention: The impact of losses on cognitive performance. *Cogn. Psychol.* **66**, 212–231 (2013).
58. A Mason, EA Ludvig, ML Spetch, CR Madan, Rare and extreme outcomes in risky choice. *Psychon. Bull. & Rev.* **31**, 1301–1308 (2024).
59. S Olschewski, A Luckman, A Mason, EA Ludvig, E Konstantinidis, The future of decisions from experience: Connecting real-world decision problems to cognitive processes. *Perspectives on Psychol. Sci.* **19**, 82–102 (2024).
60. WM Hayes, DH Wedell, Regret in experience-based decisions: The effects of expected value differences and mixed gains and losses. *Decision* **8**, 277 (2021).
61. OH Mowrer, L Solomon, Contiguity vs. drive-reduction in conditioned fear: The proximity and abruptness of drive-reduction. *The Am. J. Psychol.* **67**, 15–25 (1954).
62. O Mowrer, *Learning theory and behavior*. (John Wiley & Sons Inc), (1960).
63. NC Barberis, Thirty years of prospect theory in economics: A review and assessment. *J. economic perspectives* **27**, 173–196 (2013).
64. PP Wakker, *Prospect theory: For risk and ambiguity*. (Cambridge university press), (2010).
65. AC Farinha, TV Maia, People exert more effort to avoid losses than to obtain gains. *J. Exp. Psychol. Gen.* **150**, 1837 (2021).
66. MC Gueguen, et al., Anatomical dissociation of intracerebral signals for reward and punishment prediction errors in humans. *Nat. communications* **12**, 3344 (2021).
67. M Lebreton, K Bacily, S Palminteri, JB Engelmann, Contextual influence on confidence judgments in human reinforcement learning. *PLoS computational biology* **15**, e1006973 (2019).
68. CC Ting, S Palminteri, JB Engelmann, M Lebreton, Robust valence-induced biases on motor response and confidence in human reinforcement learning. *Cogn. Affect. & Behav. Neurosci.* **20**, 1184–1199 (2020).
69. KB Boroujeni, M Watson, T Womelsdorf, Gains and losses affect learning differentially at low and high attentional load. *J. cognitive neuroscience* **34**, 1952–1971 (2022).
70. S Palminteri, M Lebreton, Context-dependent outcome encoding in human reinforcement learning. *Curr. Opin. Behav. Sci.* **41**, 144–151 (2021).
71. M Barretto-Garcia, et al., Individual risk attitudes arise from noise in neurocognitive magnitude representations. *Nat. Hum. Behav.* **7**, 1551–1567 (2023).
72. C Findling, V Wyart, Computation noise in human learning and decision-making: origin, impact, function. *Curr. Opin. Behav. Sci.* **38**, 124–132 (2021).
73. J Denrell, JG March, Adaptation as information restriction: The hot stove effect. *Organ. science* **12**, 523–538 (2001).
74. S Bavard, M Lebreton, M Khamassi, G Coricelli, S Palminteri, Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. *Nat. communications* **9**, 4503 (2018).
75. L Ferguson, M Soukupova, S Bouret, S Palminteri, S Parkes, Reference point-dependent reinforcement learning in humans and rats. *bioRxiv* pp. 2025–04 (2025).
76. S Bavard, S Palminteri, The functional form of value normalization in human reinforcement learning. *Elife* **12**, e83891 (2023).
77. M Zeelenberg, J Beattie, J Van der Pligt, NK De Vries, Consequences of regret aversion: Effects of expected feedback on risky decision making. *Organ. behavior human decision processes* **65**, 148–158 (1996).
78. G Loomes, R Sugden, Regret theory: An alternative theory of rational choice under uncertainty. *The economic journal* **92**, 805–824 (1982).
79. RM Byrne, Mental models and counterfactual thoughts about what might have been. *Trends cognitive sciences* **6**, 426–431 (2002).
80. N Camille, et al., The involvement of the orbitofrontal cortex in the experience of regret. *Science* **304**, 1167–1170 (2004).
81. G Coricelli, et al., Regret and its avoidance: a neuroimaging study of choice behavior. *Nat. neuroscience* **8**, 1255–1262 (2005).
82. G Barron, I Erev, Small feedback-based decisions and their limited correspondence to description-based decisions. *J. behavioral decision making* **16**, 215–233 (2003).
83. G Barron, E Yechiam, The coexistence of overestimation and underweighting of rare events and the contingent recency effect. *Judgm. Decis. making* **4**, 447–460 (2009).
84. I Erev, et al., A choice prediction competition: Choices from experience and from description. *J. Behav. Decis. Mak.* **23**, 15–47 (2010).
85. N Bault, M Joffily, A Rustichini, G Coricelli, Medial prefrontal cortex and striatum mediate the influence of social comparison on the decision process. *Proc. national Acad. sciences* **108**, 16044–16049 (2011).
86. E Yechiam, G Hochman, Losses as modulators of attention: review and analysis of the unique effects of losses over gains. *Psychol. bulletin* **139**, 497 (2013).
87. S Palminteri, M Lebreton, The computational roots of positivity and confirmation biases in reinforcement learning. *Trends cognitive sciences* **26**, 607–621 (2022).
88. A Kappes, AH Harvey, T Lohrenz, PR Montague, T Sharot, Confirmation bias in the utilization of others' opinion strength. *Nat. neuroscience* **23**, 130–137 (2020).
89. K Ruggeri, et al., Replicating patterns of prospect theory for decision under risk. *Nat. human behaviour* **4**, 622–633 (2020).
90. CR Madan, EA Ludvig, ML Spetch, Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychon. bulletin & review* **21**, 629–636 (2014).
91. R Hertwig, DU Wulff, A description–experience framework of the psychology of risk. *Perspectives on psychological science* **17**, 631–651 (2022).
92. A Nasioulas, E Potier, F Cerrotti, M Lebreton, S Palminteri, Feedback-induced attitudinal changes in risk preferences. *Nat. Commun.* (2026).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryRiskRLNatureCommShiravandetal20260317.pdf](#)