PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

Linking compromise and responsibility attribution to risky decision-making in dyadic foraging

Wenning Deng[a], Ketika Garg[a,1,2], and Dean Mobbs[a,b,1,2]

Affiliations are included on p. 10.

Edited by Uta Frith, University College London, London, United Kingdom; received January 22, 2026; accepted June 1, 2026

From hunting to financial planning, many decisions are made under risk and alongside social partners whose decisions and outcomes are coupled with our own. Joint decision-making can create challenges, such as navigating conflicting risk preferences, evaluating each other's actions, and deciding whether to compromise. Yet when and why people compromise, how they evaluate each other's responsibilities during joint decision-making, and the computational and psychological underpinnings of these processes remain unclear. We introduce a dyadic foraging paradigm designed to capture diverse risk preferences, where two participants jointly choose between locations that yield rewards but carry risks and evaluate each other's responsibility for shared outcomes. Across two studies (exploratory N = 250, confirmatory N = 514), people tended to compromise rather than counteract their partner, especially under diverging risk preferences and when compromise was reciprocated. Computationally, compromise was explained by a reinforcement learning model showing that individuals integrate their preferences with that of their partner. Responsibility attributions exhibited egocentric biases—with participants claiming more credit for wins than blame for losses—and these biases were associated with individual differences in compromise behavior. The interplay between individual differences in risk and responsibility attribution further shaped coordination patterns. Finally, we show that compromising improved performance for risk-averse individuals, increased desirability as a social partner, and led to more favorable responsibility attributions, suggesting multiple benefits of compromising. By linking decisions about risk–reward trade-offs to metacognitive judgments about responsibility, our study reveals the social and cognitive processes underlying compromise in risky foraging and, more broadly, in collaborative contexts with conflicting preferences.

risk preferences | social foraging | coordination | metacognition | moral judgements

Significance

People vary in how much risk they tolerate: some prefer safe options with steady rewards, while others pursue risky, high-paying bets. How do people navigate these differences when making decisions together? We developed a dyadic foraging task where pairs jointly managed risk–reward tradeoffs. We found that people frequently compromise, incorporating partners' preferences into their own decisions. Despite frequent compromise, people showed egocentric biases: they claimed more credit for successes than blame for failures, and this bias predicted less compromise. Although compromise could cost rewards, it had social value: frequent compromisers were socially preferred and received less blame. By uncovering computational and psychological mechanisms underlying compromise, our findings illuminate how people balance self-interest with social goals in risky collective decisions.

Our ability to collaborate has played a fundamental role in human evolution and success (1, 2). Many collaborative endeavors involve risk, from cooperative foraging to scientific teams and financial planning, where people pursue shared goals under uncertainty. Yet even when social partners agree on a goal, they often differ in their preferences over how to achieve it (3, 4), which can generate interpersonal tension and divergent perceptions of their own and their partner's contributions to shared outcomes. Joint decisions can be further complicated by the fact that social interactions are inherently dynamic, requiring individuals to continuously adjust to others' actions (5). While prior work has highlighted the benefits of collective action, such as enhanced problem-solving and social learning (1, 5–7), less is known about how people navigate conflicting risk preferences in joint decision-making where outcomes are shared.

One common response to interpersonal differences is to align with others. People tend to align their behavior with others (8) across a variety of domains, including perceptual decisions (9), conversations (10, 11), and value-based judgments (12, 13). Even in risky decision-making, preferences can shift toward those of others (14–16), or move collectively in the same direction (16–19), following group discussion or by simply observing peers. Under divergent preferences, alignment can take the form of compromise, in which individuals adjust their behavior toward others' preferences or decisions. Compromising can facilitate learning by exposing individuals to unexplored options (9, 20) and improve collective performance (20, 21), but it can also be suboptimal (22, 23) and shaped by social norms and signaling (24). In joint risk-taking, decisions about when and how much to compromise might be especially salient in order to navigate both social and environmental uncertainty (25–27). How compromise unfolds across differing risk preferences in jointly controlled decisions, however, remains an open question.

Author contributions: W.D., K.G., and D.M. conceptualized and designed experiments; W.D. performed experiments and collected data; W.D. and K.G. analyzed data; W.D., K.G., and D.M. edited the paper; and W.D. and K.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2026 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]K.G. and D.M. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: kgarg@caltech.edu or dmobbs@caltech.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2602570123/-/DCSupplemental.

Published June 29, 2026.

Downloaded from https://www.pnas.org by CALIFORNIA INSTITUTE OF TECHNOLOGY CALTECH on June 29, 2026 from IP address 83.171.250.77.

While these accounts have characterized compromise in terms of its consequences, investigations into compromise as a process of interaction are lacking, especially under diverging risk preferences in jointly controlled decisions. Work on joint action and coordination has shown how individuals synchronize behavior and exert shared control, but typically in tasks with a single optimal action or aligned goals, thereby minimizing individual preferences and conflict (3). Even paradigms that allow divergent preferences often constrain social decisions to binary choices (28) between self-interest and compromise, producing convergence or turn-taking (29–32). These designs limit behavioral variability and obscure how individuals modulate the degree and direction of their influence over shared outcomes or probe a partner's responsiveness and underlying preferences. As a result, little is known about how people flexibly respond to others' preferences in continuous choice spaces, where strategies can range from compromise to counteracting, and what drives the adoption of one over another.

Alongside the dynamics of compromise, joint decision-making raises a further question: how responsibility for shared outcomes is attributed, and how those attributions shape subsequent decisions. When outcomes depend on the combined actions of multiple individuals, there is inherent ambiguity over who deserves credit or blame. Prior work shows that such judgments are shaped by collaborators' effort and contributions (33–36), and by social values (37–40). At the same time, responsibility attributions are systematically biased, with individuals tending to claim more credit for positive outcomes and less blame for negative ones (41–44). Beyond retrospective evaluation, these judgments can shape social predictions (45, 46) and future behavior. They help individuals interpret outcomes, evaluate their contributions, and adjust subsequent decisions in ways that facilitate learning and coordination (47–49). For example, individuals who perceive themselves as more responsible for risky outcomes tend to regulate their risk-taking accordingly (50–52). Beyond self-evaluation, anticipation of others' evaluations can further promote prosocial behavior and cooperation (53, 54). Together, these findings suggest that responsibility attribution is not only an outcome of joint action but also a metacognitive process that can influence subsequent decisions. However, it remains unclear how responsibility judgments are formed in real time when outcomes depend on joint, interdependent actions, and how such evaluations are associated with the dynamics of compromise and counteraction under risk.

To address these gaps, we conceptualize compromise as a dynamic, graded process through which individuals regulate their influence over a shared outcome, with joint decisions emerging from continuous adjustments to differences in preferences, partner behavior, and shared outcomes. In parallel, we treat responsibility attribution as a metacognitive signal that reflects how individuals interpret their own and their partner's contributions and that may inform their subsequent decisions. We further test whether compromise and responsibility are coupled, and how their association relates to convergence or divergence in joint behavior over time.

To study these processes in risky decision-making, we developed a dyadic risky foraging task in which two individuals jointly select between multiple, graded options varying in risk and reward (55–58), and attribute responsibility for shared outcomes (Fig. 1*A*). Social foraging provides a natural framework for this question, as interacting agents must coordinate under uncertainty while balancing individual preferences and collective outcomes (5, 57). Prior work shows that foraging with others can reduce perceived risk (59), enable social learning (7), and improve decision quality (6, 60, 61), but can also require resolving diverging preferences (62–65). By embedding joint decision-making within a spatially structured risk–reward landscape, our paradigm captures key features of real-world decision environments (such as spatial structure, uncertainty, and interdependent outcomes) within controlled experimental settings (5, 57, 66). This allows precise measurement of how individuals adjust their behavior and assign responsibility under interdependence—and how compromise, counteraction, and responsibility attribution unfold as coupled processes.

Based on exploratory analyses, we predicted that a) individuals would take their partner's preferences into account and compromise more often than they would counteract, b) credit and blame assignment would be based on both trial outcomes and differences between their own and their partner's choices, with a bias toward self-crediting, and c) self-attribution bias would be negatively correlated with the tendency to compromise, such that individuals with stronger self-crediting biases would be less likely to compromise, and that this interplay will shape whether dyads converge or diverge. To test the validity of our task measures, we further predicted that individual traits of intolerance to uncertainty (67) and social value orientation (68) would be linked to risk preferences and responsibility judgments, respectively.

Results

People Compromise with Their Partner in Risk-Based Decisions. First, we asked if individuals adjust their risk preferences based on the social context (i.e., partner's foraging choices). We compared each participant's mean foraging step during the individual phase with their mean foraging step in the dyadic phase, separated by their relative riskiness within the dyad. Relative riskiness was determined by the participant's average foraging step across the last 60 trials of the individual foraging phase (Fig. 1*C*). We found that both risk-averse and risk-prone individuals shifted their foraging steps during the dyadic phase. Specifically, risk-averse individuals made riskier (i.e., more forward) choices (exploratory $t_{rel} = 3.95$, $P < 0.001$; confirmatory $t_{rel} = 6.33$, $P < 0.001$) and risk-prone individuals made less risky (i.e., more backward) choices (exploratory $t_{rel} = -7.58$, $P < 0.001$; confirmatory $t_{rel} = -9.09$, $P < 0.001$) compared to their individual foraging phase. This shift suggests that, overall, people were more likely to compromise with their partner (i.e., move toward the partner) than counteract (i.e., move away from the partner in the direction of one's own initial preference). Risk-taking differences in the dyadic phase also corresponded with stable individual trait differences in uncertainty aversion (*SI Appendix*, Fig. S1*C*), suggesting that they reflect broader dispositional tendencies.

Compromise Facilitated Learning and Increased Rewards for Risk-Averse Individuals. Next, we tested whether the joint decision-making in the social phase was optimal in terms of the rewards obtained by the participants. On average, participants obtained higher rewards in the dyad phase compared to the individual phase (exploratory $t_{rel} = 5.11$, $P < 0.001$; confirmatory $t_{rel} = 5.94$, $P < 0.001$), suggesting an overall benefit of foraging in groups. To rule out the possibility that this increase resulted from participants becoming less sensitive to the predator over repeated exposures, we examined whether reward increased over trials in the individual phase (after 60 trials out of the total 120 in the individual phase) and found no evidence of systematic improvement (*SI Appendix*, Fig. S4*A*).

We further investigated whether reward gains differed by compromise level and risk strategy. We operationalized compromise as the degree to which participants adjusted their foraging location toward their partner's baseline risk preference (defined as partner's average foraging location during the individual phase). We regressed the increase in reward (average reward dyadic phase minus average reward in individual phase) on the degree of

Downloaded from https://www.pnas.org by CALIFORNIA INSTITUTE OF TECHNOLOGY CALTECH on June 29, 2026 from IP address 83.171.250.77.

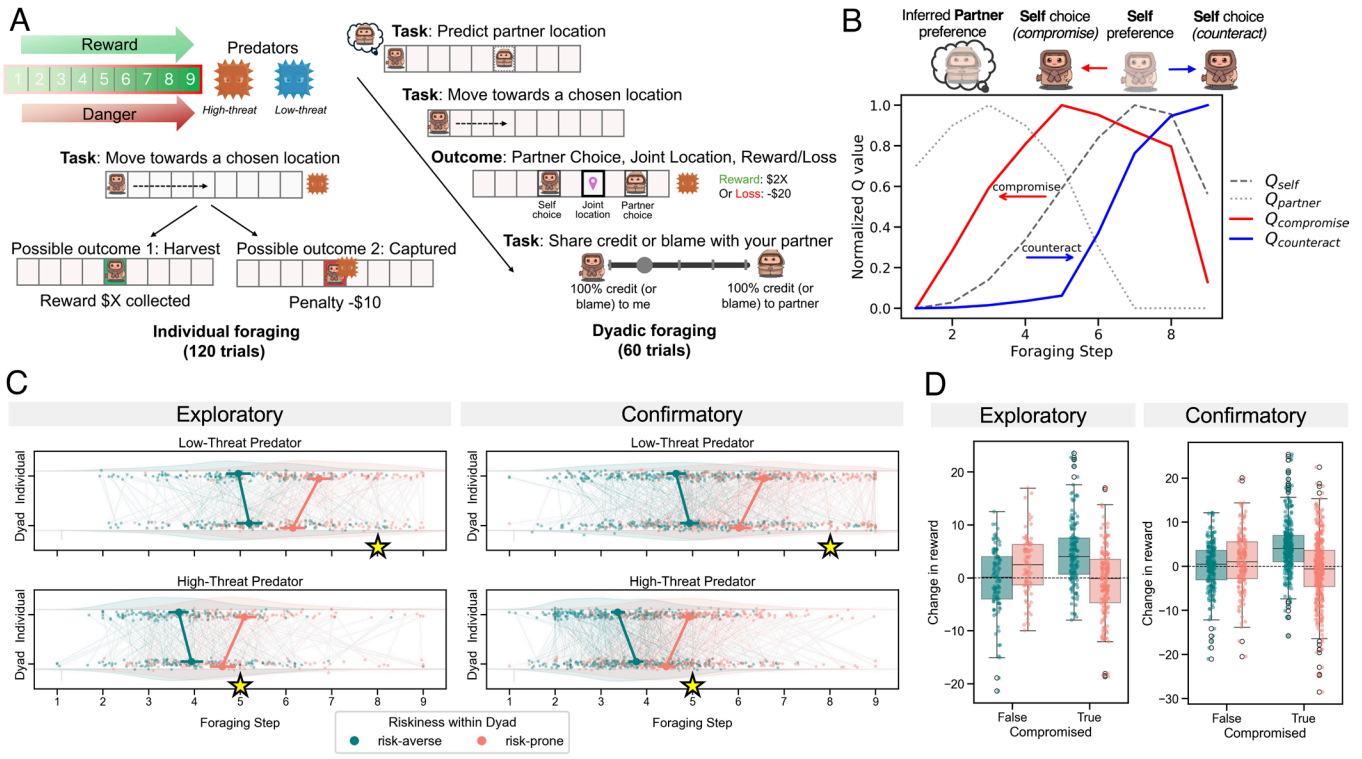

Fig. 1. Overview of the task, computational model, and behavioral results. (*A*) Task paradigm. Squares 1 to 9 represent the nine foraging steps players could choose from, with squares closer to the predator carrying a higher risk than the squares away. Each trial could have two outcomes: a win ("harvest") that yielded rewards or a loss ("captured") that imposed a penalty. Participants first completed 120 rounds of individual foraging (*Left*) and were then paired with another participant for an additional 60 rounds of dyadic foraging (*Right*). In the dyadic phase, the joint choice was the average of the two players' independent choices, and after every trial, they were asked to share credit (after "harvest" trials) or blame (after "captured" trials) with their partner. (*B*) Illustrative example of the Other-regarding Preference Model. This model assumes that players adjust their choices based on their own preferences and their predictions of their partner's preferences. Specifically, the value for each location is given by $Q = wQ_{partner} + (1-w)Q_{self}$ with $w > 0$ indicating compromise and $w < 0$ indicating counteract. The illustration shows $w = 0.4$ for $Q_{compromise}$ and $w = -1$ for $Q_{compensate}$. (*C*) Foraging step choices separated by predator (panels), phase (x-axis), and player's riskiness within the dyad (color). Participants chose more backward (i.e., less risky) steps when encountering the high-threat predator, and they chose more forward (i.e., more risky) steps when encountering the low-threat predator (exploratory $S_{low-threat} = 4.38 \pm 0.09$, $S_{high-threat} = 5.77 \pm 0.10$; confirmatory $S_{low-threat} = 4.15 \pm 0.06$, $S_{high-threat} = 5.59 \pm 0.08$). At the group level, the mean foraging step of dyads resembled that of the individual phase for both predators (exploratory $S_{low-threat} = 5.67 \pm 0.10$, $S_{high-threat} = 4.27 \pm 0.09$; confirmatory $S_{low-threat} = 5.46 \pm 0.07$, $S_{high-threat} = 4.12 \pm 0.06$). Riskiness within the dyad is defined by the average foraging choice of the last 60 rounds of individual foraging. Stars indicate the optimal foraging locations that maximize the expected reward under the risk-neutral assumption (see *SI Appendix*, Fig. S3*E* for calculation). While participants' decisions under the high-threat predator were closely aligned with expected optimal values, their choices under the low-threat predators were significantly more risk-averse than optimal. (*D*) Reward change from individual to dyadic phase by compromise (x-axis) and player's riskiness within the dyad (color). We binarized compromise (split at 0) and riskiness for illustrative purposes. Boxplots show the median and the interquartile range (25th to 75th percentiles).

compromise and self-partner baseline risk difference (a directional measure calculated as the participant's baseline risk preference minus the partner's baseline risk preference) (*SI Appendix*, Table S3 and Fig. 1*D*). We found a positive main effect of compromise (exploratory $\beta_{compromise} = 1.00 \pm 0.21$, $P < 0.001$; confirmatory $\beta_{compromise} = 0.51 \pm 0.15$, $P < 0.001$), a negative main effect of self-partner risk difference (exploratory $\beta_{riskdiff} = -0.36 \pm 0.16$, $P = 0.022$; confirmatory $\beta_{riskdiff} = -0.63 \pm 0.10$, $P = 0.001$), and a negative interaction effect (exploratory $\beta_{interaction} = -0.51 \pm 0.09$, $P < 0.001$; confirmatory $\beta_{interaction} = -0.33 \pm 0.06$, $P < 0.001$). These effects indicate that compromise was beneficial for individuals who were relatively more risk-averse than their partners, and that its benefits decreased for more risk-prone partners. For risk-averse individuals, shifting toward a riskier partner may have facilitated exploration of less frequently visited states and thereby increased their rewards. We also tested whether reward gains and foraging choices in the dyadic phase carry over to individual foraging after dyadic foraging. We found that risk-averse partners increased their foraging location in postdyad individual foraging trials compared to predyad individual foraging phase, suggesting that they socially learned from their partners (*SI Appendix*, Fig. S8). By contrast, for risk-prone individuals who already tended to choose options closer to the reward-maximizing option, compromising conferred no additional reward benefits, and they reverted to riskier locations in the postdyad individual foraging trials (*SI Appendix*, Fig. S8).

The Degree of Compromise Is Based on Risk Differences and the Partner's Tendency to Compromise. We next investigated whether decisions to compromise were driven by their partner's risk preference, partner's compromise, and trial outcomes. We found a negative relationship between an individual's risk change and their partner's risk preference: people who were paired with a riskier partner became riskier, and people who were paired with a less risky partner became less risky in the dyadic phase (Fig. 2*A*), indicating that the magnitude of compromise scaled with the difference in risk preferences between partners.

To test whether compromise was reciprocal and depended on the outcome of a trial, we investigated changes in participant location on a trial-by-trial level (Fig. 2*B*). Specifically, we asked if participants moved toward their partner's location based on their relative positions, the distance between them, the outcome of the previous trial, and their partner's decision to compromise in the previous trial (*SI Appendix*, Table S1).

Downloaded from https://www.pnas.org by CALIFORNIA INSTITUTE OF TECHNOLOGY CALTECH on June 29, 2026 from IP address 83.171.250.77.

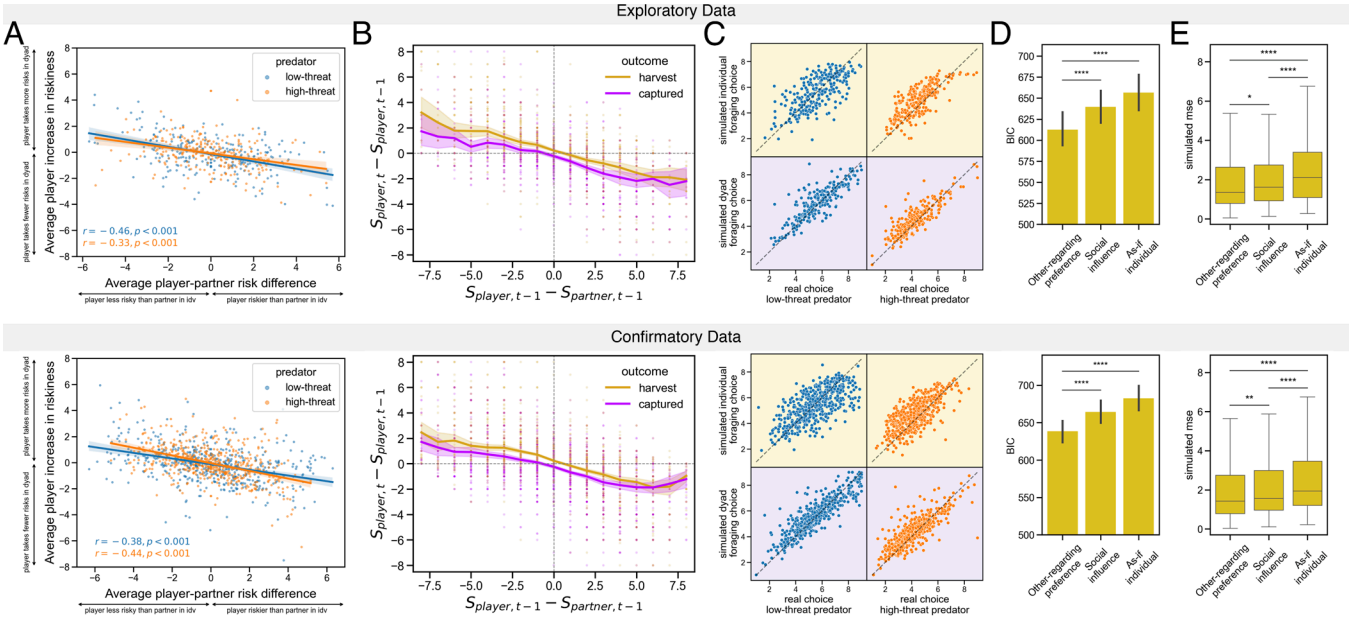

Fig. 2. Trial-by-trial changes in foraging choices and computational modeling of compromise. (*A*) Regression plot showing a negative correlation between the change in player riskiness (player's dyadic phase average step – player's individual phase average step) and the baseline risk differences between player and partner (player's individual-phase average step – partner's individual-phase average step), for both predators. An ANOVA on absolute step change confirmed that this tendency to compromise did not differ significantly across predator types (exploratory $F = 2.88$, $P = 0.09$; confirmatory $F = 0.91$, $P = 0.34$). (*B*) Plot of trial-by-trial changes in player riskiness during the dyadic phase as a function of player-partner choice differences and outcome in the previous round. Error bars indicate 95% credible interval across participants. (*C*) Predictions of each participant's average step based on the best-fitting computational models against their true average, separated by phase and predator. (*D*) Model comparison based on the BIC values of the three models, showing that the other-regarding preferences model provides the best fit Error bars indicate SE across participants. (*E*) Model comparison based on the mean squared error between simulated and observed participant choices of the three models. Box plots indicate median and 25 to 75% IQR.

We regressed each participant's step change ($S_{player,t+1} - S_{player,t}$) on self-partner distance ($S_{player,t} - S_{partner,t}$), partner's previous step change ($S_{partner,t} - S_{partner,t-1}$), and trial outcome (Harvest or Captured). We found that participants' step changes were negatively associated with self-partner distance (exploratory $\beta_{self-partnerdistance} = -0.39 \pm 0.02$, $P < 0.001$; confirmatory $\beta_{self-partner\ distance} = -0.36 \pm 0.03$, $P < 0.001$), and with partner's previous step change (exploratory $\beta_{partnerchanged} = -0.17 \pm 0.01$, $P < 0.001$; confirmatory $\beta_{partnerchanged} = -0.14 \pm 0.03$, $P < 0.001$) (Fig. 2*B*). These effects imply that participants were more likely to move forward (or backward) when they were farther behind their partner (or farther ahead), and when their partner had previously moved backward (or forward), suggesting a reciprocal compromise.

Additionally, participants showed step adjustments based on the trial outcomes, moving forward after wins (exploratory $\beta_{harvest} = 0.20 \pm 0.03$, $P < 0.001$; confirmatory $\beta_{harvest} = 0.16 \pm 0.01$, $P < 0.001$) and backward after losses (exploratory $\beta_{captured} = -0.40 \pm 0.05$, $P < 0.001$; confirmatory $\beta_{captured} = -0.35 \pm 0.02$, $P < 0.001$). We found no evidence that the effect of self-partner distance on step change differed by previous trial outcome (exploratory $\beta_{interaction} = -0.00 \pm 0.01$, $P = 0.885$; confirmatory $\beta_{interaction} = -0.03 \pm 0.02$, $P = 0.292$).

Finally, we examined how accurately participants predicted their partner's next choice and whether discrepancies between their own choices and predicted partner choices were reflected in ongoing decision-making (*SI Appendix*, Table S6). Participants showed low prediction error of the partner's step (absolute difference between partner's step in a trial and participants' prediction; exploratory median = 1.3, IQR = [0.9, 2.0]; confirmatory median = 1.3, IQR = [0.8, 1.9]; *SI Appendix*, Fig. S3*C*), indicating that they could reliably anticipate their partner's choices. To test whether participants were sensitive to their partner's choices on a trial-by-trial basis, we examined how response times varied with the discrepancy between participants' own choices and their predicted partner choices. We regressed log-transformed response times (RT) on the choice discrepancy ($|S_{player} - S_{prediction}|$), while controlling for participant's step choice which affected the time taken to make a given move (higher location value required more presses). While the RT analyses were not preregistered, we report them here given their relevance and note that the findings replicated across both samples. We found a positive association between choice discrepancy and response times (exploratory $\beta_{discrepancy} = 0.026 \pm 0.005$, $P < 0.001$; confirmatory $\beta_{discrepancy} = 0.022 \pm 0.004$, $P < 0.001$; full regression result see *SI Appendix*, Table S4). This effect suggests that larger differences between one's own and predicted partner choices require more strategic deliberation.

Taken together, these results suggest that people were more likely to compromise (i.e., move toward partner) than counteract (i.e., move away from partner), they reliably predicted partner's choices and that the level of compromise was shaped by trial-by-trial risk differences and reciprocity.

Computational Models Show People Integrate Partner's Preferences in Their Choices. We developed computational models to identify the mechanisms underlying convergence in risk preferences. Although participants shifted toward their partner's risk attitudes, multiple processes could produce this pattern: 1) being pulled into previously unvisited states could facilitate better learning of the environment, indirectly producing convergence without social influence; 2) the social context could shift general risk preferences (e.g., through risk dilution or increased

Downloaded from https://www.pnas.org by CALIFORNIA INSTITUTE OF TECHNOLOGY CALTECH on June 29, 2026 from IP address 83.171.250.77.

uncertainty); or 3) participants could actively incorporate their partner's revealed preferences on a trial-by-trial basis. To adjudicate among these possibilities, we developed a reinforcement learning model with risk sensitivity and generalization to capture baseline individual foraging behavior (*Methods*), and we compared three extensions of this model (Fig. 1*B*), each corresponding to a distinct hypothesized process during the dyadic phase. The *as-if individual* model captureds convergence driven solely by improved learning or exploration, without incorporating social information. The *social-influence* model implemented a global shift in risk preference by updating an individual's risk preference toward their partner's preference over time. Finally, the *other-regarding preferences* model captured contingent social integration, in which individuals integrate their own and their partner's risk preferences on a trial-by-trial basis when deciding whether to compromise (Fig. 1*B*).

Model comparison results showed that the *other-regarding preference* model fitted the data best (Fig. 2 *D* and *E*). To validate the model, we simulated data using fitted parameters and found that the model successfully reproduced key features of participants' choices (Fig. 2*C*). We also correlated the model fits with participants' behavior to demonstrate that the parameters of *other-regarding preference* model capture different aspects of observed behavior (*SI Appendix*, Fig. S2). Specifically, the compromise parameter (w), which modulates how much people incorporate others' preferences, correlated with the self-partner distance (*SI Appendix*, Fig. S2*A*).

Response time results (see previous result section and *SI Appendix*, Table S4) further corroborated the validity of the *other-regarding preference* model: response times increased when self-preferred choices deviated more from the partner's predicted choice. Neither the *social-influence* model, which assumed a general shift in risk attitude, nor the *as-if individual* model, which assumed no consideration of partners' choices, can account for response time sensitivity to trial-by-trial choice differences between self and partner. In contrast, the *other-regarding preference* model explains this result by suggesting that participants integrated their partner's trial-by-trial preferences into their own decisions, with greater preference divergence increasing computational load. This implies that compromise in dyads is not simply the outcome of a global shift in preferences but instead emerges from individuals weighing their own and their partner's intentions on each trial.

Finally, we examined whether compromise behavior was strategic or universal by testing whether the compromise parameter (w) varied with individual risk-preferences. If compromise was strategic, we would expect individuals to modulate their compromise based on whether it benefits them: risk-averse individuals should compromise more, while risk-prone individuals should compromise less. However, we found no significant difference in w between riskier and more risk-averse partners (exploratory $t_{\text{ind}} = -0.69$, $P = 0.49$; confirmatory $t_{ind} = -0.75$, $P = 0.45$), even though compromising is beneficial for risk-averse individuals and disadvantageous for risk-prone individuals. We found that the fitted social-weight parameter w was associated with social value orientation in our confirmatory dataset, but this association was not significant in the exploratory dataset (*SI Appendix*, Fig. S1*D*).

People Attribute Credit or Blame Based on Relative Positions and Trial Outcome, but with an Egocentric Bias. Following each trial in the dyadic phase, once the outcome was revealed, we asked participants to attribute responsibility to self and partner. In the case of positive outcomes ("harvest"), they were asked to share credit, and in the case of negative outcomes ("captured"), they were asked to share blame with their partner. The responses were not shown to the other partner, and instead, they served as a measure of self-evaluation, as participants had to decide whether they personally deserved more credit or blame for the result, independent of how the other person might perceive the situation.

We examined how people divide responsibility based on choice differences with their partner and trial outcomes (i.e., whether they share credit and blame differently). Using linear mixed-effect regression, we found a significant effect of self-partner distance, trial outcome, and their interaction on responsibility (Fig. 3*A* and *SI Appendix*, Table S5). People who were more forward tended to take more responsibility (exploratory $\beta_{self-partner\ distance} = 0.04 \pm 0.00$, $P < 0.001$; confirmatory $\beta_{self-partner\ distance} = 0.04 \pm 0.00$, $P < 0.001$). However, their judgments were biased by the outcome.

Participants took more credit (for harvesting rewards) than blame (for being captured) (exploratory $\beta_{captured} = -0.08 \pm 0.01$, $P < 0.001$; confirmatory $\beta_{captured} = -0.05 \pm 0.01$, $P < 0.001$). This egocentric bias was modulated by relative position: participants at safer locations showed stronger egocentric bias, splitting credit equally with their partner for rewards but accepting less blame for attacks (exploratory $\beta_{interaction} = 0.02 \pm 0.00$, $P < 0.001$; confirmatory $\beta_{interaction} = 0.03 \pm 0.00$, $P < 0.001$). We also found greater disagreement between partners after attacks vs. rewards (exploratory $t_{ind} = 3.58$, $P < 0.001$; confirmatory $t_{ind} = 8.27$, $P < 0.001$), suggesting that negative outcomes exacerbate attributional asymmetries.

We also examined if such attribution bias is trait-dependent by correlating participants' self-attributed responsibility with their social value orientation (SVO) score (68). We found that lower SVO score (i.e., more prosocial) was correlated with lower self-attributed blame (exploratory $r = 0.16$, $P = 0.018$; confirmatory $r = 0.14$, $P = 0.002$), and more self-attributed credit (exploratory $r = -0.4$, $P < 0.001$; confirmatory $r = -0.3$, $P < 0.001$; *SI Appendix*, Fig. S1*A*).

Less Egocentric Bias in Responsibility Attribution Correlates with Willingness to Compromise. To assess how the evaluation of relative responsibility for an outcome interacts with decisions to compromise, we first asked whether responsibility attribution predicted subsequent changes in foraging choices at the level of trials. We found that, after getting attacked, participants were more likely to move toward their partner if they blamed themselves more for the loss (Fig. 3*B*). This asymmetry is likely due to the higher sensitivity to self-partner distance in blame assignment than credit assignment. This result also suggests that acceptance of blame after an attack can predict the tendency for participants to adjust their behavior toward greater alignment with their partner.

To investigate whether these trial-by-trial links between attribution and behavioral adjustment scale up to shape participants' overall behavior, we next examined whether a participant's overall egocentric bias (i.e., average credit taken minus average blame taken) predicted their tendency to compromise in the experiment. We found a negative correlation between egocentric bias and participants' weight (w) from the other-regarding preference model, suggesting that stronger self-centered attribution was associated with a reduced willingness to compromise. However, this finding could be confounded by our earlier result showing that players close to each other tend to share responsibility equally. Since participants with greater values of (w) tend to be closer to their partners than those with lower values of (w), their lower bias may simply reflect this closeness.

Downloaded from https://www.pnas.org by CALIFORNIA INSTITUTE OF TECHNOLOGY CALTECH on June 29, 2026 from IP address 83.171.250.77.

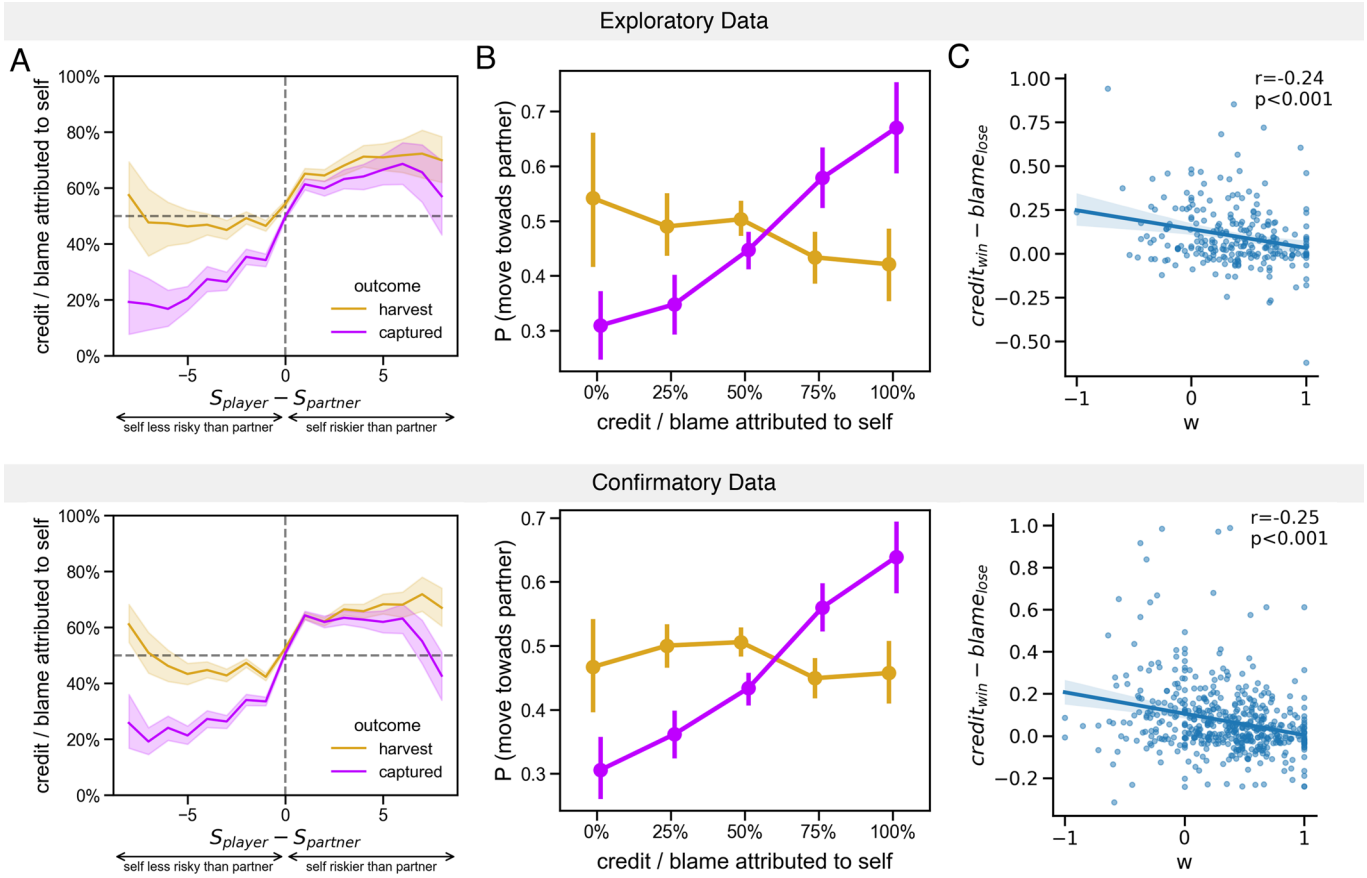

Fig. 3. Judgments of credit/blame and their relationship with compromise. (*A*) Percentage of credit/blame player attributed to themselves, as a function of player-partner step difference (x-axis) and trial outcome (color). Error bars indicate 95% credible interval across participants. (*B*) Trial-by-trial level degree of compromise as a function of credit/blame attributed to oneself in the previous trial and the previous trial's outcome (color). The y-axis was calculated as whether the player moved toward the partner (1 or 0) from trial t-1 to trial t. Error bars indicate 95% credible interval across participants. (*C*) Participants' compromise (*w*) was negatively correlated with their overall egocentric bias (average self-credit – average self-blame) (y-axis). A higher weight on others' preferences was correlated with lower egocentric bias.

To rule out this possibility, we computed a partial correlation between egocentric bias and participants' own compromise parameter (w), controlling for average player-partner distance. Egocentric bias remained significantly correlated with participants' own compromise parameter, (w) after controlling for the average player-partner distance (exploratory $r = -0.17$, $P = 0.008$; confirmatory r = –0.12, $P = 0.006$; *SI Appendix*, Fig. S5*A*). In addition, to rule out the possibility that participants' bias could reflect partner's compromise, we tested whether egocentric bias was related to the partner's compromise. We found no significant correlation between a participant's bias and their partner's compromise parameter (exploratory $r = 0.02$, $P = 0.708$; confirmatory $r = -0.01$, $P = 0.798$; *SI Appendix*, Fig. S5*A* and Table S7). These tests suggest that egocentric bias reflects stable individual differences that are related to decisions to compromise independent of partner's behavior. Overall, our results show that participants who credited themselves more tended to be less willing to adjust their choices to align with their partner, suggesting an association between egocentric bias and reduced compromise.

Risk Preferences and Bias Predict Joint Risk-Adjustment Patterns. Next, we examined whether dyads coordinated in different ways by mutually adjusting their risk-taking, and what influenced these patterns. To classify these patterns, we defined a two-dimensional space capturing changes in risk preferences from individual to dyadic phase for the risk-averse and risk-prone partner in each dyad, separately for each predator (*SI Appendix*, Fig. S6). Splitting each axis at zero yielded four categories: risk-increase (i.e., both moved forward; exploratory N = 52; confirmatory N = 106), risk-decrease (i.e., both moved backward; exploratory N = 74; confirmatory N = 149), risk-converge (i.e., both moved toward each other; exploratory N = 100; confirmatory N = 210), and risk-diverge (i.e., both moved away from each other; exploratory N = 19; confirmatory N = 40). In line with our results so far that show that people tend to compromise, we found that many dyads fell under the risk-converge category. Partners showed asymmetric adjustments even in risk-increase and risk-decrease categories,: in the risk-increase group, risk-averse partners exhibited a larger increase in risk-taking relative to their risk-prone partners; conversely, in the risk-decrease category, risk-prone partners showed a larger decrease in risk-taking.

We found that interindividual differences in baseline risk predicted how the dyads jointly adjusted their risk. First, dyads with higher average risk preferences tend to reduce their risk-taking during the dyadic phase, while those with lower average risk preferences tend to increase it (Fig. 4*B*). Second, dyads with smaller initial differences tended to shift in the same direction (risk-increase or risk-decrease), whereas dyads with larger differences either converged or diverged (Fig. 4*C*). Third, egocentric bias was higher in the divergent dyads (Fig. 4*D*). To explore how risk differences and egocentric biases mechanistically shape these patterns, we designed an agent-based model

Downloaded from https://www.pnas.org by CALIFORNIA INSTITUTE OF TECHNOLOGY CALTECH on June 29, 2026 from IP address 83.171.250.77.

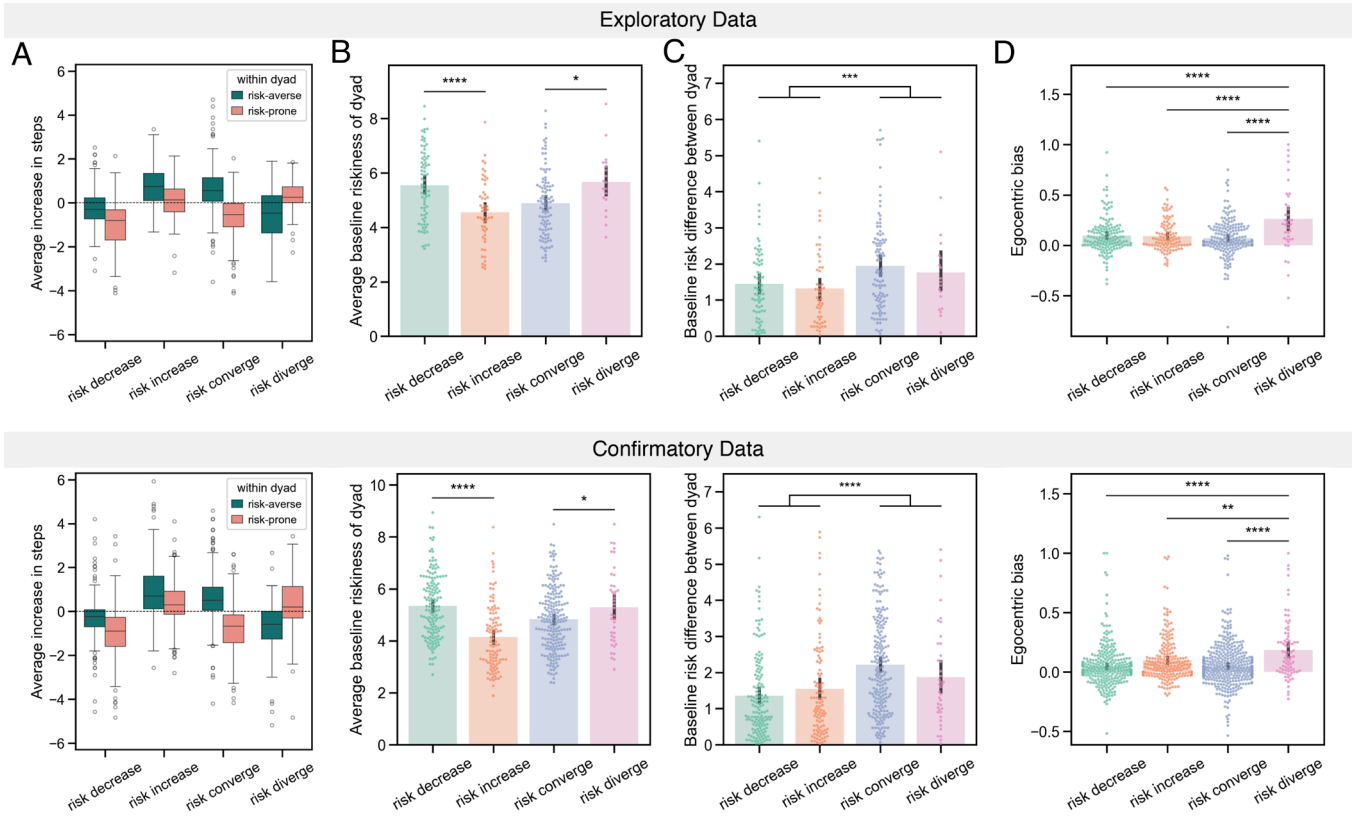

Fig. 4. Dyadic risk-adjustment patterns. (*A*) Distribution of participants' change in riskiness from individual to dyadic phase, separated by the risk-category clusters and relative risk level within the dyad. Except in the *risk-diverge* case, dyads mostly adjust their behavior in opposite directions, narrowing the initial gap in risk preferences. (*B*) Average baseline risk preference of the four risk-category clusters. We calculated the average of the two players' individual foraging step to quantify the dyad's baseline risk preference. Dyads with lower baseline risk increased in risk whereas dyads with higher baseline risk decreased in risk in the dyadic phase (exploratory $t_{ind} = 4.38$, $P < 0.001$; confirmatory $t_{ind} = 7.54$, $P < 0.001$). In addition, dyads that converged in risk (compromised) had lower baseline risk yet dyads that diverged (counteracted) had higher baseline risk (exploratory $t_{ind} = -2.58$, $P = 0.01$; confirmatory $t_{ind} = -2.18$, $P = 0.03$). (*C*) Average baseline risk difference in the four risk-category clusters. We computed the mean individual foraging step difference between the two players in a dyad as a measure of baseline risk difference. Dyads with higher differences in baseline risk were more likely to converge (compromise) or diverge (counteract), while dyads with lower differences in risk were more likely to move in the same direction (exploratory $t_{ind} = -3.47$, $P < 0.001$; confirmatory $t_{ind} = -6.27$, $P < 0.001$). (*D*) Average egocentric bias (difference between self-blame for losses and self-credit for wins) by risk-category cluster. Dyads in the risk-diverge cluster showed higher egocentric bias than dyads in all other clusters. Each dot indicates average egocentric bias per participant per predator.

where agents updated their risk preferences using a win-stay, lose-shift rule modulated by bias (*Methods* and *SI Appendix*, Fig. S7). The model replicated our empirical patterns: partners with similar risk preferences moved in the same direction; but when partners differed greatly, bias determined whether they converged or diverged—low-bias agents converged by updating their positions in a complementary way, whereas high-bias agents resisted updating when outcomes were misaligned with their preferences, leading to divergence. These results imply that egocentric bias can be associated with how shared outcomes are translated into learning signals, thereby influencing emergent coordination dynamics.

People Rate Partners Who Compromise as More Preferable and Blame Them Less. Finally, we asked why people compromise so often, even when it was suboptimal to do so. We hypothesized that compromising might hold an implicit social value that could motivate decisions to compromise. In an exploratory but replicated analysis, we tested whether there is a social benefit to compromising such that people prefer partners who compromise. We found a negative correlation (exploratory $r = -0.38$, $P < 0.01$; confirmatory $r = -0.19$, $P < 0.01$; *SI Appendix*, Fig. S9*A*) between participants' partner preference rating (on a scale of 0 to 10, asked after the task) and self-partner foraging distance in the dyad phase. This effect persisted after controlling for joint reward, suggesting that people enjoyed playing with others whose choices aligned with theirs, independent of their actual earnings. Furthermore, for each participant, we averaged the credit (on successful trials) and blame (on captured trials) that their partner assigned to them and correlated that with the participant's compromise parameter, w. We found a positive correlation (exploratory $r = 0.14$, $P = 0.03$; confirmatory $r = 0.21$, $P < 0.001$; *SI Appendix*, Fig. S9*B*), indicating that participants who compromised more tended to receive more credit and less blame from their partners than the participants who compromised less. These results indicate that compromise carries a social value and may explain why risk-prone individuals compromise even when doing so does not increase their foraging reward.

Discussion

When people make risky decisions together, they must not only navigate the uncertainty about outcomes but also the need to reconcile differing risk preferences, as these differences can create tension between personal motives and collective goals. To understand collaboration under divergent preferences, we investigated two key facets: compromise (i.e., how people adjust decisions toward their partner's preferences) and responsibility attribution (i.e., how they evaluate each other's contributions to shared outcomes). To study these processes, we developed a dyadic foraging task in a risky environment where partners had to make joint decisions while assigning credit or blame for joint outcomes. We found that people often

Downloaded from https://www.pnas.org by CALIFORNIA INSTITUTE OF TECHNOLOGY CALTECH on June 29, 2026 from IP address 83.171.250.77.

compromise to reconcile such differences, and differing preferences drive responsibility (i.e., credit/blame) evaluations, which in turn can influence whether individuals choose to compromise or to counteract. We further showed that individual risk preferences and responsibility attribution biases can predict how dyads coordinate (i.e., their emergent patterns of risk-adjustment). Our findings highlight how individual differences in risk preferences, willingness to compromise, and attributional biases can interact to shape joint decision-making under risk.

We found that differences in risk-preferences did not elicit strategic counteraction; instead, our results revealed a tendency to compromise, with individuals adjusting their choices toward their partner's risk level even in the absence of immediate reward benefits. Critically, compromise was not indiscriminate; rather, it systematically depended on the partner's tendency to compromise (i.e., reciprocity), their risk differences, and trial outcomes. Computational modeling revealed that this pattern emerged from trial-by-trial integration of partner preferences rather than global shifts in risk attitude or improved environmental learning, supporting previous studies showing integration of others' preferences (14, 15, 69). These results also align with previous work showing that humans are biased toward conformity and consensus (8, 14, 15, 70, 71) and guided by norms of reciprocity and fairness (24, 30, 40, 72).

Moreover, our results contextualize these processes beyond binary choice frameworks to a continuous choice space, identifying compromise as a social and cognitive mechanism through which social preferences are flexibly weighted during shared decision-making under uncertainty. Prior work has shown that compromise or aggregation between judgments can increase collective accuracy (21), though such aggregation may still fail to outperform the best individual in dyadic settings, even when outcomes are transparent (22). Our work complements these accounts by treating compromise as a dynamic process and investigating how it emerges. We further show that in learnable environments, compromise can produce asymmetric gains, particularly benefiting individuals with less optimal (e.g., more risk-averse) strategies.

Our results also show how people share responsibility during shared decision-making and its relationship with compromise. We found that people showed systematic biases in how they attributed responsibility for shared outcomes: they claimed more credit after wins than blame after losses. This pattern is consistent with prior work showing that shared responsibility in joint decisions reduces perceived agency and exaggerates self-serving biases (44). We extend this topic by showing that heterogeneity in preferences can amplify these asymmetries, leading to divergence in how partners evaluate the same outcome. Moreover, we found that prosocial individuals [measured via Social Value Orientation (68, 73)] shared both credit and blame more equally. While past studies have documented egocentric biases (43, 74, 75), these studies typically rely on binary judgments ("high credit" vs. "low credit") and are detached from actual actions or outcomes (33, 36). In contrast, our design required participants to make trial-by-trial, graded attributions based on real and observable outcomes during active task performance, allowing us to dynamically link responsibility judgments to participants' and their partner's actions. Our findings highlight that attributional bias persists even when contributions are transparent, and that it is asymmetrically shaped by outcomes and modulated by individual differences.

Critically, our task allowed us to link responsibility attribution to subsequent behavioral adjustment, showing that greater blame-taking after negative outcomes is associated with increased compromise, consistent with the idea that responsibility signals may support learning-related processes (44). Previous research suggests that attributional biases influence learning (76), but their role in coordination and collective behavior has been less studied. Our results show that coordination can be associated with how individuals evaluate their own and their partner's contributions; when these judgments are biased, they can hinder individuals' willingness to compromise. The associations among compromise, bias, and SVO (*SI Appendix*, *Supplementary Results*) raise the possibility that these behavioral signatures reflect a common individual-level disposition in how people weigh self vs. partner during shared decision-making.

Empirical results, supported by agent-based simulations, demonstrate that patterns of risk convergence and divergence can emerge from individuals adjusting their foraging choices based on outcomes, individual differences, and biased credit assignment. While our findings support previous work showing that diversity of initial opinions can increase convergence under uncertainty (77), they also show that in a learnable, value-based task, preference heterogeneity can also lead to persistent divergence, potentially driven by attributional biases, and reciprocal dynamics. Our results are also in line with prior suggestions that causal attributions may serve an important function for future planning and behavioral adjustment (78), extending accounts that link reduced responsibility to diminished outcome processing (44).

In our study, participants had no explicit monetary incentive to compromise—the joint outcome (i.e., the mean of participants' locations) was identical regardless of whether their choices converged or diverged. Nevertheless, people routinely chose to compromise. We identify three potential factors that could have led to such frequent compromise. First, participants attributed more blame to partners who compromised less, whereas individuals who compromised were rated as more desirable partners, indicating a social value to compromising. This aligns with prior work showing that compromise helps individuals diffuse responsibility within groups (38, 41, 44). Second, for suboptimal (e.g., risk-averse) individuals, compromising with more optimal (e.g., risk-prone) individuals conferred learning benefits, which led to an increase in their foraging rewards. Third, our results on reaction times suggest a cognitive benefit of compromising. Larger differences in location preferences required more deliberation, whereas compromising may serve as a low-effort heuristic that reduces the cost of continuous strategizing (79).

Our study has a few limitations. While we establish a link between attributional bias and the tendency to compromise, our design does not experimentally isolate causality. It is possible that people assign responsibility post hoc to justify their decisions to compromise or not rather than to motivate them, and future studies can test this directionality by manipulating partner behavior, inducing perspective-taking, or probing how metacognitive evaluations shape coordination. Further, we defined the joint outcome as the mean of the two participants' choices, providing a simple, transparent mechanism for shared control that allowed graded degrees of compromise, while leaving open the possibility that the observed behavior partly reflects adaptation to the averaging rule itself. In many real-world settings, collective outcomes may depend on other aggregation rules—for example, dominance of the most extreme action, majority rule (44, 80), weighted influence of one individual (81, 82), or coordination mechanisms that penalize disagreement or impose consensus costs (83, 84). Future work could examine such alternative decision rules to test how different forms of joint control shape coordination dynamics.

We also asked participants to predict their partner's choice before deciding. This design choice let us directly measure expectations about the partner, though it may have encouraged perspective-taking and strategic reasoning about their actions (45). This possibility raises an important question for future work: Does perspective-taking

Downloaded from https://www.pnas.org by CALIFORNIA INSTITUTE OF TECHNOLOGY CALTECH on June 29, 2026 from IP address 83.171.250.77.

facilitate compromise or coordination in joint tasks? Further, while our work highlights several social benefits of compromise, we do not directly test to distinguish between them or their relative salience in different contexts. For example, compromise in risk-prone partners could be driven by reciprocal influence, blame avoidance, or a desire to encourage their partners to take greater risks. Further extensions are needed to test the various costs and benefits of compromise (83). Finally, for experimental tractability we focused on dyadic interactions in an abstract task without communication, which can restrict generalizability to larger groups and real-world scenarios. In the real world, discussions guide joint decisions and can inform attributions of blame and credit, serving as feedback that promotes or deters future coordination (85, 86). In larger groups, certain effects can shift. Self-serving biases tend to amplify as the sense of agency weakens and responsibility diffuses across members (38, 44), and compromise itself may depend upon group size, composition, or organization (e.g., leader–follower) (87).

In conclusion, our study makes several important contributions to research on joint decision-making, coordination, and social interaction. Most dyadic decision-making research applies binary choice paradigms in static environments (20, 31, 88), thereby simplifying real-world social dynamics where choices are often continuous and allow for incremental adjustments to others' actions (5, 89–91). Our continuous, spatially correlated grid, combined with solo and social phases, revealed differences in risk preferences and let us study compromise as a dynamic process shaped by both stable preferences and ongoing coordination. Given the correspondence between dyadic-phase risk preferences and independently measured uncertainty aversion, our paradigm also captures individual differences that generalize beyond foraging. By integrating foraging decisions with responsibility attribution, we show how attributional biases emerge from both choices and outcomes, while opening avenues for understanding the cognitive and motivational bases of compromise in collective dynamics. Extending this paradigm also offers a path toward a more integrative understanding of risk-taking (92, 93), social valuation, and norms, and how these processes vary across development (94, 95) and cultural contexts. Ultimately, examining how people navigate collaboration amid conflicting preferences is key to understanding why some collaborations endure while others fracture—dynamics central to team performance, organizational dynamics, and collective action on shared challenges.

Methods

This study was preregistered at https://osf.io/mjcx5 (96). Code and data are available at https://github.com/gracewndeng99/dyadForaging (97). A glossary of all the key measures and variables constructed and used in the analyses is provided in *SI Appendix*.

Experimental Paradigm.

Individual foraging phase. Participants were asked to operate an avatar to forage in a virtual environment with predators. The goal of the task was to maximize rewards while avoiding being captured. To incentivize performance, participants received a bonus payment proportional to their collected rewards. Participants first did the task on their own for 120 trials with two different predators (60 trials each). Both predators were more likely to attack as the participant moved closer (i.e., chose a more forward location), but each predator differed in their overall attack probability (*SI Appendix*, Fig. S3). Rewards increased exponentially as participants got closer to the predator. At each trial, the participant could choose one of the 9 squares to forage at, and each square was associated with a known reward and unknown probability of being captured by the predator. Participant's avatar was spawned at square 1 at the beginning of each trial and they could move their avatar by pressing either the left or right arrow key. After the participant had submitted their choice by hitting the space bar, whether the predator attacks would be revealed. If the predator did not attack, the participant would earn the reward associated with the chosen location. If the predator attacked, the participant would lose a fixed 10 points. Participants needed to learn the attack probability of the two predators and make choices to maximize their reward. The two predators were indicated by two different colors and appeared in a random and intermixed order. Predator-color mappings were counterbalanced across participants. Participants were told that only their choices in the last 60 trials of the individual foraging phase would be counted in their bonus payment.

Dyadic foraging phase. After a participant finished foraging individually, they were randomly paired with another online participant and entered the dyad foraging phase. They foraged in the same environment and encountered the same two predators, one first and then the other, with the order counterbalanced across pairs. With each predator, they played 30 trials. On each trial, each participant was first asked to make a prediction about where their partner would move. They then were asked to move their avatar to their desired location. After both participants had made their choices, their chosen locations were revealed to each other, and their avatars moved to the midpoint of their chosen locations (rounded up when the middle is not an integer). Then, the predator attacked or not based on the probability of attack at the midpoint. The reward structure remained the same as the individual phase, except both rewards and losses were doubled for the group. After each trial outcome was revealed, each participant was asked to assign the responsibility of the outcome between themselves and their partner. In the case of a capture, participants were asked to attribute blame between self and partner on a scale of 5 choices ranging from "completely self" to "completely partner." In the case of a harvest, participants were asked to attribute credit between self and partner on a scale of 5 choices ranging from "completely self" to "completely partner." They were told that their attribution was private and would not affect their or their partner's bonus. Both the participant and their partner were paid a bonus based on their accumulated rewards. To ensure synchrony, participants began each new trial simultaneously once both had completed all questions in the previous trial. A response time limit of 8 s was imposed for each decision; if a participant failed to respond within this window, a response was randomly selected on their behalf.

Data Collection. The study is approved by Caltech IRB #23-1310, and informed consent was obtained from all participants. For the exploratory data, we collected data (after exclusion) from 250 participants (125 groups). For the confirmatory data, we collected one batch from 304 participants (152 groups) and another batch from 210 participants (105 groups). Participants were recruited through the online platform Prolific (www.prolific.com), and each participant conducted a 45-min task plus a 5-min questionnaire at the end. Each participant was paid $10 for completing the task and a bonus depending on their performance in the task.

We collected each participant's foraging location choice and response time in both the individual and dyadic phases. In the dyadic phase, we additionally collected 1) participant's prediction about their partner's choice, and 2) the participant's responsibility attribution of the outcome (see *SI Appendix*, Fig. S3*A* for the distribution of these variables). After the task, we also collected some questionnaire measurements including IUS, STAI-T, SVO, and partner preference rating ("How much did you enjoy playing with your partner?").

In the individual phase, we analyzed data after the first 60 trials to account for a sufficient learning period. We define each participant's baseline riskiness as the average foraging choice over these trials. In the dyadic phase, participants were classified into risk-averse or risk-prone within the dyad based on their baseline riskiness.

In the second batch of the confirmatory dataset, we included an additional 30 rounds of individual foraging after the dyad phase, with results reported in *SI Appendix*. Participants were not informed of this extension. Because the individual and dyadic task phases were identical across the two confirmatory batches, we combined them for the main analyses.

Exclusion Criterion. Pairs of participants were excluded if either member failed to respond on more than one-third of trials in the dyadic phase, resulting in the exclusion of 6 pairs from the exploratory dataset and 13 pairs from the confirmatory dataset. Trials on which one participant failed to respond within the 8-s time limit were excluded from that participant's analyses but retained in their partner's data.

Computational Models. We developed a series of computational models to capture: 1) how people learn and generalize the attack patterns of predators, and 2) how people coordinate with another individual. In these models, participants

Downloaded from https://www.pnas.org by CALIFORNIA INSTITUTE OF TECHNOLOGY CALTECH on June 29, 2026 from IP address 83.171.250.77.

first learn the attack probability of the chosen state based on the outcome, then update the attack probability of the other states, and finally make a choice balancing risk and reward. Each model was fit to the complete behavioral data for each participant, comprising 120 trials from the individual phase and 60 trials from the dyadic phase.

To capture generalization, safety values are updated at every location after each outcome: a prediction error (δ_t) is generated at the chosen location based on the observed outcome (O_t) and propagates to unvisited locations, with propagation strength (γ) decaying with distance from the chosen state (98, 99).

$$\delta_t = O_t - \Pi_{safety}(s,t),$$

$$\Pi_{safety}(s,t+1) = \Pi_{safety}(s,t) + \alpha_t \delta_t,$$

$$\Pi_{safety}(s',t+1) = \Pi_{safety}(s',t) + \alpha_t \left(\delta_t \gamma^{|s'-s|}\right) \quad \forall s' \neq s.$$

Initial safety values were set as a linear gradient across locations, ranging from 0.1 at the square nearest the predator to 1 at the square farthest from it. The model used a decaying learning rate schedule in which the effective learning rate at k^{th} encounter with a predator was defined as $\alpha_t = \alpha/\sqrt{k}$, where α is a free parameter controlling the initial learning rate. This schedule ensures that early observations have greater influence on belief updates, while later trials contribute progressively less, consistent with the notion that beliefs stabilize as evidence accumulates.

To capture individual differences in risk preference, we fit a reward-sensitivity parameter (θ) that controls the curvature of the subjective utility function over rewards and losses (100). The model predicted the participant's choice on each trial as the state with the highest Q-value (*SI Appendix*, Fig. S3 *E* and *F*).

$$Q_{self}(s,t) = \Pi_{safety}(s,t) rewards^{\theta} - \left(1 - \Pi_{safety}(s,t)\right) |loss|^{\theta}.$$

For the dyadic phase, we compared three classes of models. *The as-if-individual model* assumes that individuals do not consider their partner's actions and choose locations to maximize their own utility. *The social influence model* assumes that an individual's internal risk preference is influenced by their partner's risk preference. In this model, Δ captures the additional change in riskiness in the group setting. $\Delta > 0$ indicates an increase in riskiness in the group phase, and vice versa.

$$Q_{self}(s,t) = \Pi_{safety}(s,t) rewards^{(\theta+\Delta)} - \left(1 - \Pi_{safety}(s,t)\right) |loss|^{(\theta+\Delta)}.$$

The other-regarding preference model assumes that an individual values both their own risk preferences and those of their partner. In this model, the group utility is a weighted combination of the individual's preferences and those of their partner. $Q_{partner}\left(\hat{s}_{partner,t}, s\right)$ is constructed as a Gaussian kernel centered on the participant's predicted partner location $\hat{s}_{partner}$, with values being 1 at the predicted location and decreasing smoothly as a function of squared distance from the prediction $(\exp(-\frac{(s-\hat{s}_{partner})^2}{2}))$. When $w = 0$, the individual is insensitive to their partner and relies solely on their own preferences. When $w > 0$, the individual is attracted toward their partner's predicted location, with $w = 1$ indicating full alignment. When $w < 0$, the individual is repelled away from their partner's predicted location–specifically, the individual is attracted to the mirror image of the partner's position reflected around their own preferred location $(2 \cdot \text{argmax}_s Q_{self}(s,t) - \hat{s}_{partner,t})$, with $w = -1$ indicating complete counteralignment. We additionally tested other variations of how partner's preferences are calculated and different learning rate schedules of this model (*SI Appendix*, *Supplementary Methods*).

$$Q(s,t) = \begin{cases} (1-w)\, Q_{self}(s,t) + w Q_{\text{partner}}\left(\hat{s}_{partner,t}, s\right) \text{ if } w \geq 0 \\ (1-w)\, Q_{self}(s,t) + w Q_{\text{partner}}\left(2 * argmax_s Q_{self}(s,t) - \hat{s}_{partner,t}, s\right) \text{ if } w < 0 \end{cases}.$$

To find the best-fitting parameters for each participant, we minimized the negative log-likelihood between model predicted choice and the participant's choice on each trial. We applied a grid generational search procedure with three successive generations. In the first generation, parameters were sampled from a coarse grid over the full search space $(w, \alpha, \gamma, \theta)$, where $w, \alpha, \gamma \in [0, 1]$ and $\theta \in [0, 1.5]$. At each generation, the grid resolution was reduced by a factor of five, and the search was centered on the parameter region that minimized negative log-likelihood in the preceding generation. This iterative refinement allowed us to efficiently locate the parameter set yielding the lowest negative log-likelihood while avoiding exhaustive evaluation of the entire space. For model parameter recovery, we first simulated each participant's choices based on their best-fitting parameters. We then refitted the same model on the simulated data and compared the recovered best-fitting parameters against the parameters that generated the data (*SI Appendix*, Fig. S2*D*). Model comparison was conducted using the Bayesian Information Criterion (BIC). Additionally, we conducted posterior prediction checks by comparing each participants' simulated choices with their actual choices. For each participant, we simulated choices and outcomes on each trial using their best-fitting parameters, repeating each simulation 10 times.

Agent-Based Modeling. The model simulated the 1-D grid of the task and applied the same functions to generate outcomes of win or loss (i.e., reward or attack) based on agents' joint location (or mean location). Each agent had a risk preference as an initial strategy which defined the location they tend to choose. We ran the simulation for all combinations of risk preferences of two agents.

At each time step, each agent picked a location based on their preference. Their joint location was the mean of their individual choices, and the trial outcome was based on the joint location and the predator attack probability. Then, each agent updated their preference using an update rule in a win-stay, lose-shift manner, modulated by their relative position, trial outcome, and egocentric bias. Specifically, each agent has a bias ranging from 0 to 1. If the agent was ahead of their partner, they moved ahead after a win with probability 0.1 [**1**] and moved backward after a loss with probability 1 – bias [**2**]; if the agent was behind the partner, they moved backward after a loss with probability 0.9 [**3**] and moved ahead after a win with probability 1 – bias [**4**]. Magnitude of the movement, whether forward or backward was dependent on the difference between agents' location and scaled by the length of the grid:

$$\Delta x_i \sim Bernoulli(0.1) \times \| \Delta d \|, \text{ if outcome} = \text{win and } \Delta d > 0, \quad [1]$$

$$\Delta x_i \sim Bernoulli(1 - bias) \times \| \Delta d \|, \text{ if outcome} = \text{win and } \Delta d < 0, \quad [2]$$

$$\Delta x_i \sim -Bernoulli(0.9) \times \| \Delta d \|, \text{ if outcome} = \text{lose and } \Delta d < 0, \quad [3]$$

$$\Delta x_i \sim -Bernoulli(1 - bias) \times \| \Delta d \|, \text{ if outcome} = \text{lose and } \Delta d > 0, \quad [4]$$

$$\Delta x_i = 0, \text{ otherwise.} \quad [5]$$

Hierarchical Regressions. All the linear regressions were performed using hierarchical models implemented with the statsmodels package in Python. Random coefficients were included for all parameters and regressions were estimated at the group level. The regression formulas and complete results are provided in *SI Appendix*.

Data, Materials, and Software Availability. Data, code and analyses data have been deposited in Github (https://github.com/gracewndeng99/dyadForaging) (97).

ACKNOWLEDGMENTS. D.M. is supported by National Institute of Mental Health grants R01MH133730-01. We thank Haoxue Fan, Gabriel Ramos-Fernández, Qianying Wu, and Yang Xiang for their thoughtful feedback.

Author affiliations: [a]Department of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125; and [b]Computation and Neural Systems Program at the California Institute of Technology, Pasadena, CA 91125

Downloaded from https://www.pnas.org by CALIFORNIA INSTITUTE OF TECHNOLOGY CALTECH on June 29, 2026 from IP address 83.171.250.77.

1. R. Boyd, P. J. Richerson, J. Henrich, The cultural niche: Why social learning is essential for human adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10918–10925 (2011).
2. M. Tomasello, A. Vaish, Origins of human cooperation and morality. *Annu. Rev. Psychol.* **64**, 231–255 (2013).
3. C. Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton University Press, 2003).
4. D. Kahneman, A. Tversky, Prospect theory: An analysis of decision under risk. *Econometrica* **47**, 263–291 (1979).
5. K. Garg, W. Deng, D. Mobbs, Utilizing social foraging as a framework to study decisions in groups. *iScience* **29**, 114940. (2026).
6. K. Garg, C. T. Kello, P. E. Smaldino, Individual exploration and selective social learning: Balancing exploration–exploitation trade-offs in collective foraging. *J. R. Soc. Interface* **19**, 20210915 (2022).
7. C. M. Wu *et al.*, Adaptive mechanisms of social and asocial learning in immersive collective foraging. *Nat. Commun.* **16**, 3539 (2025).
8. S. G. Shamay-Tsoory, N. Saporta, I. Z. Marton-Alper, H. Z. Gvirts, Herding brains: A core neural mechanism for social alignment. *Trends Cogn. Sci.* **23**, 174–186 (2019).
9. A. Mahmoodi, H. Nili, D. Bang, C. Mehring, B. Bahrami, Distinct neurocomputational mechanisms support informational and socially normative conformity. *PLoS Biol.* **20**, e3001565 (2022).
10. Y. A. Nielsen *et al.*, Conversational alignment as an individual difference. *Proc. R. Soc. B Biol. Sci.* **293**, 20252205 (2026).
11. S. P. H. Speer *et al.*, Hyperscanning shows friends explore and strangers converge in conversation. *Nat. Commun.* **15**, 7781 (2024).
12. D. K. Campbell-Meiklejohn, D. R. Bach, A. Roepstorff, R. J. Dolan, C. D. Frith, How the opinion of others affects our valuation of objects. *Curr. Biol.* **20**, 1165–1170 (2010).
13. V. Klucharev, K. Hytönen, M. Rijpkema, A. Smidts, G. Fernández, Reinforcement learning signal predicts social conformity. *Neuron* **61**, 140–151 (2009).
14. D. Chung, G. I. Christopoulos, B. King-Casas, S. B. Ball, P. H. Chiu, Social signals of safety and risk confer utility and have asymmetric effects on observers' choices. *Nat. Neurosci.* **18**, 912–916 (2015).
15. S. Suzuki, E. L. S. Jensen, P. Bossaerts, J. P. O'Doherty, Behavioral contagion during learning about another agent's risk-preferences acts on the neural representation of decision-risk. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3755–3760 (2016).
16. Y. Jiang, P. Marcowski, A. Ryazanov, P. Winkielman, People conform to social norms when gambling with lives or money. *Sci. Rep.* **13**, 853 (2023).
17. W. Toyokawa, A. Whalen, K. N. Laland, Social learning strategies regulate the wisdom and madness of interactive crowds. *Nat. Hum. Behav.* **3**, 183–193 (2019).
18. J. A. F. Stoner, Risky and cautious shifts in group decisions: The influence of widely held values. *J. Exp. Soc. Psychol.* **4**, 442–459 (1968).
19. S. Moscovici, M. Zavalloni, The group as a polarizer of attitudes. *J. Pers. Soc. Psychol.* **12**, 125–135 (1969).
20. M. Yavuz, E. G. Doğu, B. Bahrami, Exploration-exploitation dilemma in joint value-based decisions is driven by resistance to erratic randomness. PsyArxiv [Preprint] (2025). https://doi.org/10.31234/osf.io/cp8uy_v2 (Accessed 11 June 2026).
21. D. N. Ferreiro, O. Deroy, B. Bahrami, Compromising improves forecasting. *R. Soc. Open Sci.* **10**, 221216 (2023).
22. B. Bahrami *et al.*, Optimally interacting minds. *Science* **329**, 1081–1085 (2010).
23. S. He, J. Wu, Compromise and coordination: An experimental study. *Games Econ. Behav.* **119**, 216–233 (2020).
24. A. Mahmoodi, B. Bahrami, C. Mehring, Reciprocity of social influence. *Nat. Commun.* **9**, 2474 (2018).
25. H. Suganuma, K. Katahira, H. Ohtsuki, T. Kameda, How social learning enhances—or undermines—efficiency and flexibility in collective decision-making under uncertainty. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2516827122 (2025).
26. O. FeldmanHall, A. Shenhav, Resolving uncertainty in a social world. *Nat. Hum. Behav.* **3**, 426–435 (2019).
27. G. Wang, Q. Su, L. Wang, J. B. Plotkin, The evolution of social behaviors and risk preferences in settings with uncertainty. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2406993121 (2024).
28. A. H. H. Rasanan *et al.*, Beyond discrete-choice options. *Trends Cogn. Sci.* **28**, 857–870 (2024).
29. R. Axelrod, W. D. Hamilton, The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
30. R. X. D. Hawkins, R. L. Goldstone, The formation of social conventions in real-time environments. *PLoS One* **11**, e0151670 (2016).
31. S. Moeller *et al.*, Human and macaque pairs employ different coordination strategies in a transparent decision game. *eLife* **12**, e81641 (2023).
32. U. Segal, J. Sobel, Tit for tat: Foundations of preferences for reciprocity in strategic settings. *J. Econ. Theory* **136**, 197–216 (2007).
33. T. Gerstenberg, D. A. Lagnado, Spreading the blame: The allocation of responsibility amongst multiple agents. *Cognition* **115**, 166–171 (2010).
34. A. Mahmoodi, S. Luo, C. Harbison, P. Piray, M. F. S. Rushworth, Human hippocampus and dorsomedial prefrontal cortex infer and update latent causes during social interaction. *Neuron* **112**, 3796–3809.e9 (2024).
35. M. Schäfer, D. B. M. Haun, M. Tomasello, Children's consideration of collaboration and merit when making sharing decisions in private. *J. Exp. Child Psychol.* **228**, 105609 (2023).
36. Y. Xiang, J. Landy, F. A. Cushman, N. Vélez, S. J. Gershman, Actual and counterfactual effort contribute to responsibility attributions in collaborative tasks. *Cognition* **241**, 105609 (2023).
37. A. W. Cappelen, E. Ø. Sørensen, B. Tungodden, Responsibility for what? Fairness and individual responsibility. *Eur. Econ. Rev.* **54**, 429–441 (2010).
38. M. El Zein, B. Bahrami, R. Hertwig, Shared responsibility in collective decisions. *Nat. Hum. Behav.* **3**, 554–559 (2019).
39. E. Fehr, A. Falk, Psychological Foundations of Incentives. SSRN Scholarly Paper (2002). https://doi.org/10.2139/ssrn.294287.
40. E. Fehr, U. Fischbacher, Why social preferences matter–The impact of non-selfish motives on competition, cooperation and incentives. *Econ. J.* **112**, C1–C33 (2002).
41. D. R. Forsyth, L. E. Zyzniewski, C. A. Giammanco, Responsibility diffusion in cooperative collectives. *Pers. Soc. Psychol. Bull.* **28**, 54–65 (2002).
42. D. T. Miller, M. Ross, Self-serving biases in the attribution of causality: Fact or fiction? *Psychol. Bull.* **82**, 213–225 (1975).
43. R. S. Miller, B. R. Schlenker, Egotism in group members: Public and private attributions of responsibility for group performance. *Soc. Psychol. Q.* **48**, 85–89 (1985).
44. M. El Zein, R. J. Dolan, B. Bahrami, Shared responsibility decreases the sense of agency in the human brain. *J. Cogn. Neurosci.* **34**, 2065–2081 (2022).
45. A. R. Todd, D. I. Tamir, Factors that amplify and attenuate egocentric mentalizing. *Nat. Rev. Psychol.* **3**, 164–180 (2024).
46. Z. Zhao, H. Sened, D. I. Tamir, Egocentric projection is a rational strategy for accurate emotion prediction. *J. Exp. Soc. Psychol.* **109**, 104521 (2023).
47. B. Weiner, The development of an attribution-based theory of motivation: A history of ideas. *Educ. Psychol.* **45**, 28–36 (2010).
48. C. D. Frith, The role of metacognition in human social interactions. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 2213–2223 (2012).
49. C. D. Frith, U. Frith, Mechanisms of social cognition. *Annu. Rev. Psychol.* **63**, 287–313 (2012).
50. G. E. Bolton, A. Ockenfels, J. Stauf, Social responsibility promotes conservative risk behavior. *Eur. Econ. Rev.* **74**, 109–127 (2015).
51. I. G. J. Dawson, Taking responsibility: Self-attribution for risk creation and its influence on the motivation to engage in risk management behaviors. *J. Risk Res.* **23**, 1440–1451 (2020).
52. J. Pahlke, S. Strasser, F. M. Vieider, Responsibility effects in decision making under risk. *J. Risk Uncertain.* **51**, 125–146 (2015).
53. J. Andreoni, B. D. Bernheim, Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica* **77**, 1607–1636 (2009).
54. N. Lacetera, M. Macis, Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme. *J. Econ. Behav. Organ.* **76**, 225–237 (2010).
55. N. Kolling, M. Wittmann, M. F. S. Rushworth, Multiple neural mechanisms of decision making and their competition under changing risk pressure. *Neuron* **81**, 1190–1202 (2014).
56. D. Mobbs *et al.*, Foraging under competition: The neural basis of input-matching in humans. *J. Neurosci.* **33**, 9866–9872 (2013).
57. D. Mobbs, P. C. Trimmer, D. T. Blumstein, P. Dayan, Foraging for foundations in decision neuroscience: Insights from ethology. *Nat. Rev. Neurosci.* **19**, 419–427 (2018).
58. L. Real, T. Caraco, Risk and foraging in stochastic environments. *Annu. Rev. Ecol. Evol. Syst.* **17**, 371–390 (1986).
59. B. Silston *et al.*, Neural encoding of perceived patch value during competitive and hazardous virtual foraging. *Nat. Commun.* **12**, 5478 (2021).
60. L. Snijders *et al.*, Causal evidence for the adaptive benefits of social foraging in the wild. *Commun. Biol.* **4**, 94 (2021).
61. A. J. W. Ward, D. J. T. Sumpter, I. D. Couzin, P. J. B. Hart, J. Krause, Quorum decision-making facilitates information transfer in fish shoals. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6948–6953 (2008).
62. L. Conradt, Collective animal decisions: Preference conflict and decision accuracy. *Interface Focus* **3**, 20130029 (2013).
63. C. N. Cook *et al.*, Individual learning phenotypes drive collective behavior. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 17949–17956 (2020).
64. G. H. Davis, M. C. Crofoot, D. R. Farine, Using optimal foraging theory to infer how groups make collective decisions. *Trends Ecol. Evol.* **37**, 942–952 (2022).
65. R. P. Mann, Collective decision-making by rational agents with differing preferences. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10388–10396 (2020).
66. T. Wise, K. Emery, A. Radulescu, Naturalistic reinforcement learning. *Trends Cogn. Sci.* **28**, 144–158 (2024).
67. K. Buhr, M. J. Dugas, The intolerance of uncertainty scale: Psychometric properties of the english version. *Behav. Res. Ther.* **40**, 931–946 (2002).
68. R. O. Murphy, K. A. Ackermann, M. J. J. Handgraaf, Measuring social value orientation. *Judgm. Decis. Mak.* **6**, 771–781 (2011).
69. O. Mayo, S. Shamay-Tsoory, Dynamic mutual predictions during social learning: A computational and interbrain model. *Neurosci. Biobehav. Rev.* **157**, 105513 (2024).
70. M. A. Orloff *et al.*, Social conformity is a heuristic when individual risky decision-making is disrupted. *PLoS Comput. Biol.* **20**, e1012602 (2024).
71. S. Suzuki, R. Adachi, S. Dunne, P. Bossaerts, J. P. O'Doherty, Neural mechanisms underlying human consensus decision-making. *Neuron* **86**, 591–602 (2015).
72. S. Gächter, B. Herrmann, Reciprocity, culture and human cooperation: Previous insights and a new cross-cultural experiment. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 791–806 (2008).
73. D. Balliet, C. Parks, J. Joireman, Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Process. Intergroup Relat.* **12**, 533–547 (2009).
74. M. Ross, F. Sicoly, Egocentric biases in availability and attribution. *J. Pers. Soc. Psychol.* **37**, 322–336 (1979).
75. B. F. Malle, How people explain behavior: A new theoretical framework. *Personal. Soc. Psychol. Rev.* **3**, 23–48 (1999).
76. E. Zamfir, P. Dayan, Interactions between attributions and beliefs at trial-by-trial level: Evidence from a novel computer game task. *PLoS Comput. Biol.* **18**, e1009920 (2022).
77. J. Navajas, O. Armand, R. Moran, B. Bahrami, O. Deroy, Diversity of opinions promotes herding in uncertain crowds. *R. Soc. Open Sci.* **9**, 191497 (2022).
78. A. Morris *et al.*, Looking back to plan ahead: Causal judgments as a sampling approximation for action effectiveness. PsyArxiv [Preprint] (2025). https://doi.org/10.31234/osf.io/nq53z_v2 (Accessed 11 June 2026).
79. A. K. Shah, D. M. Oppenheimer, Heuristics made easy: An effort-reduction framework. *Psychol. Bull.* **134**, 207–222 (2008).
80. R. Hastie, T. Kameda, The robust beauty of majority rules in group decisions. *Psychol. Rev.* **112**, 494–508 (2005).
81. J. A. R. Marshall, G. Brown, A. N. Radford, Individual confidence-weighting and group decision-making. *Trends Ecol. Evol.* **32**, 636–645 (2017).
82. P. Zarnoth, J. A. Sniezek, The social influence of confidence in group decision making. *J. Exp. Soc. Psychol.* **33**, 345–366 (1997).
83. L. Conradt, T. J. Roper, Consensus decision making in animals. *Trends Ecol. Evol.* **20**, 449–456 (2005).
84. J. R. G. Dyer, A. Johansson, D. Helbing, I. D. Couzin, J. Krause, Leadership, consensus decision making and collective behaviour in humans. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 781–789 (2009).
85. B. F. Malle, S. Guglielmo, A. E. Monroe, "Moral, cognitive, and social: The nature of blame" in *Social Thinking and Interpersonal Behavior*, J. Forgas, K. Fiedler, C. Sedikides, Eds. (PA: Psychology Press, Philadelphia), pp. 311–329.
86. B. F. Malle, S. Guglielmo, A. E. Monroe, A theory of blame. *Psychol. Inq.* **25**, 147–186 (2014).

Downloaded from https://www.pnas.org by CALIFORNIA INSTITUTE OF TECHNOLOGY CALTECH on June 29, 2026 from IP address 83.171.250.77.

87. M. Lowe, H. Nikolova, C. J. Miller, S. L. Dommer, Ceding and succeeding: How the altruistic can benefit from the selfish in joint decisions. *J. Consum. Psychol.* **29**, 652–661 (2019).
88. D. Yu, B. Thompson, "People balance joint reward, fairness and complexity to develop social norms in a two-player game" in *Proceedings of the 46th Annual Conference of the Cognitive Science Society*, L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, E. Hazeltine, Eds. (Cognitive Science Society, 2024), pp. 1418–1424.
89. S. Isbaner *et al.*, Dyadic interaction platform: A novel tool to study transparent social interactions. *eLife* **14** (2025). https://doi.org/10.7554/eLife.106757.1.
90. D. Lewen *et al.*, Continuous dynamics of cooperation and competition in social decision-making. *Commun. Psychol.* **3**, 170 (2025).
91. M. A. Pisauro, E. F. Fouragnan, D. H. Arabadzhiyska, M. Apps, M. G. Philiastides, Neural implementation of computational mechanisms underlying the continuous trade-off between cooperation and competition. *Nat. Commun.* **13**, 6873 (2022).
92. E. U. Weber, E. J. Johnson, "Decisions under uncertainty: psychological, economic and neuroeconomic explanations of risk preference," in *Neuroeconomics: Decision Making and the Brain*, P. Glimcher, C. Camerer, E. Fehr, R. Poldrack, Eds. (Elsevier, New York, NY, 2009), pp. 127–144. 10.1016/B978-0-12-374176-9.00010-5.
93. B. Figner, E. U. Weber, Who takes risks when and why?: Determinants of risk taking. *Curr. Dir. Psychol. Sci.* **20**, 211–216 (2011).
94. E. A. Crone, L. Bullens, E. A. A. van der Plas, E. J. Kijkuit, P. D. Zelazo, Developmental changes and individual differences in risk and perspective taking in adolescence. *Dev. Psychopathol.* **20**, 1213–1229 (2008).
95. E. A. Crone, R. E. Dahl, Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nat. Rev. Neurosci.* **13**, 636–650 (2012).
96. W. Deng, K. Garg, D. Mobbs, Pre-registration and preprint from "Linking compromise and responsibility attribution to risky decision-making in dyadic foraging". OSF. https://osf.io/mjcx5. Deposited 12 May 2026.
97. W. Deng, K. Garg, D. Mobbs, Code and data for analyses from "Linking compromise and responsibility attribution to risky decision-making in dyadic foraging". Github. https://github.com/gracewndeng99/dyadForaging. Deposited 8 May 2026.
98. G. E. Wimmer, N. D. Daw, D. Shohamy, Generalization of value in reinforcement learning by humans. *Eur. J. Neurosci.* **35**, 1092–1104 (2012).
99. C. M. Wu, E. Schulz, M. Speekenbrink, J. D. Nelson, B. Meder, Generalization guides human exploration in vast decision spaces. *Nat. Hum. Behav.* **2**, 915–924 (2018).
100. D. Bernoulli, Exposition of a new theory on the measurement of risk. *Econometrica* **22**, 23–36 (1954).

Downloaded from https://www.pnas.org by CALIFORNIA INSTITUTE OF TECHNOLOGY CALTECH on June 29, 2026 from IP address 83.171.250.77.