

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Computational and neurobiological foundations of leadership decisions

Micah G. Edelson*, Rafael Polania, Christian C. Ruff, Ernst Fehr*, Todd A. Hare*

INTRODUCTION: Decisions as diverse as committing soldiers to the battlefield or picking a school for your child share a basic attribute: assuming responsibility for the outcome of others. This responsibility is inherent in the roles of prime ministers and generals, as well as in the more quotidian roles of firm managers, schoolteachers, and parents. Here we identify the underlying behavioral, computational, and neurobiological mechanisms that determine the choice to assume responsibility over others.

METHODS: We developed a decision paradigm in which an individual can delegate decision-making power about a choice between a risky and a safe option to their group or keep the right to decide: In the “self” trials, only the individual’s payoff is at stake, whereas in the “group” trials, each group member’s payoff is affected. We combined models from perceptual and value-based decision-making to estimate each individual’s personal utility for every available action in order to tease apart potential motivations for choosing to “lead” or “follow.” We also used brain imaging to examine the neurobiological basis of leadership choices.

RESULTS: The large majority of the subjects display responsibility aversion (see figure, left panel), that is, their willingness to choose between the risky and the safe option is lower in the group trials relative to the self trials, independent of basic preferences toward risk, losses, ambiguity, social preferences, or intrinsic valuations of decision rights. Furthermore, our findings indicate that responsibility aversion is not associated with the overall frequency of keeping or delegating decision-making power. Rather, responsibility aversion is driven by a second-order cognitive process reflecting an increase in the demand for certainty about what constitutes the best choice when others’ welfare is affected. Individuals who are less responsibility averse have higher questionnaire-based and real-life leadership scores. The center panel of the figure shows the correlation between predicted and observed leadership scores in a new, independent sample. Our analyses of the dynamic interactions between brain regions demonstrate the importance of information flow between brain regions involved in computing separate components of the choice to understanding

leadership decisions and individual differences in responsibility aversion.

DISCUSSION: The driving forces behind people’s choices to lead or follow are very important but largely unknown. We identify responsibility aversion as a key determinant of the willingness to lead. Moreover, it is predictive of both survey-based and real-life leadership scores. These results suggest that many people associate a psychological cost with assuming responsibility for others’ outcomes. Individual differences in the perception of, and willingness to bear, responsibility as the price of leadership may determine who will strive toward leadership roles and, moreover, are associated with how well they perform as leaders.

ON OUR WEBSITE
Read the full article
at <http://dx.doi.org/10.1126/science.aat0036>

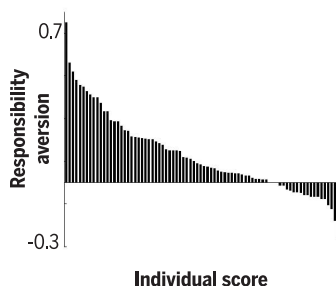
Our computational model provides a conceptual framework for the decision to assume responsibility for others’ outcomes as well as insights into the cognitive and neural mechanisms driving this choice process. This framework applies to many different leadership types, including authoritarian leaders, who make most decisions themselves, and egalitarian leaders, who frequently seek a group consensus. We believe that such a theoretical foundation is critical for a precise understanding of the nature and consequences of leadership. ■

The list of author affiliations is available in the full article online.

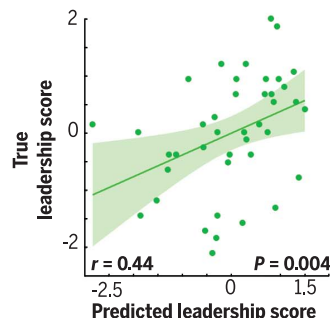
*Corresponding author. Email: micah.edelson@econ.uzh.ch (M.G.E.); todd.hare@econ.uzh.ch (T.A.H.); ernst.fehr@econ.uzh.ch (E.F.)

Cite this article as M. G. Edelson *et al.*, *Science* **361**, eaat0036 (2018). DOI: 10.1126/science.aat0036

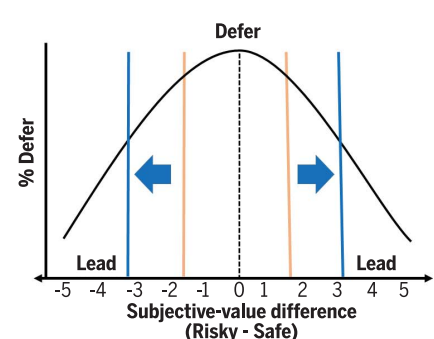
The large majority of subjects display responsibility aversion (each bar represents one subject)



Predictions of leadership scores based on responsibility aversion are correlated with true leadership scores



Delegation of decision-making based on subjective-value differences between the choice options



Frequency, out-of-sample predictive power, and computational foundations of responsibility aversion. (Left) Responsibility aversion differs widely across individuals. (Center) These individual differences in responsibility aversion can be used to predict leadership scores in a new, independent sample. (Right) The lead-versus-defer decision process is illustrated. The black curve shows the proportion of defer choices increasing when the subjective-value difference between actions approaches zero

(dashed line). This pattern holds in both self and group trials. What changes is where people set deferral thresholds (orange, self; blue, group), which determine when they are most likely to defer. More responsibility-averse individuals show a larger shift in the deferral thresholds, which our computational model links to increased demand for certainty about the best course of action when faced with assuming responsibility for others. *r*, Spearman rank correlation coefficient.

RESEARCH ARTICLE

NEUROSCIENCE

Computational and neurobiological foundations of leadership decisions

Micah G. Edelson^{1*}, Rafael Polania^{1,2}, Christian C. Ruff¹, Ernst Fehr^{1*}, Todd A. Hare^{1*}

Leaders must take responsibility for others and thus affect the well-being of individuals, organizations, and nations. We identify the effects of responsibility on leaders' choices at the behavioral and neurobiological levels and document the widespread existence of responsibility aversion, that is, a reduced willingness to make decisions if the welfare of others is at stake. In mechanistic terms, basic preferences toward risk, loss, and ambiguity do not explain responsibility aversion, which, instead, is driven by a second-order cognitive process reflecting an increased demand for certainty about the best choice when others' welfare is affected. Finally, models estimating levels of information flow between brain regions that process separate choice components provide the first step in understanding the neurobiological basis of individual variability in responsibility aversion and leadership scores.

Leadership decisions pervade every level of society, from the basic family unit up to global organizations and political institutions. Parents, teachers, CEOs, and heads of state all lead their respective groups and make decisions that have widespread and lasting consequences for themselves and others (1). Thus, a key aspect of leadership is the acceptance of responsibility for others. We developed a behavioral task that, together with computational modeling and neuroimaging (2–4), allows us to determine the cognitive and neural mechanisms driving the choice to assume or forgo the responsibility of leading a group.

There are some key features of leadership choices that are potential drivers of decisions to lead. For example, a position of leadership is associated with the right to make decisions that affect one's own and others' welfare. Therefore, the choice to lead a group may be taken particularly often by those who put a high value on decision rights or who are driven by a desire to determine and control others' outcomes (5, 6). Alternatively, leadership might be perceived as a burden, and those who are most willing to shoulder this responsibility may be most likely to choose to lead. Furthermore, the decision to lead could be predicated on the willingness to accept losses or potential failures for oneself or others or to act under conditions of high uncertainty and ambiguity. Finally, because leaders' decisions often have far-reaching consequences that require careful forethought, those who are most competent in the task at hand (for example, make more accurate and objective assessments

of probabilities) may be more likely to make decisions to lead.

We designed an experiment to allow us to distinguish between the hypotheses that decisions to lead others are related to (changes in) basic preferences over risk, loss, or ambiguity and the possibility that responsibility affects choices through a separate mechanism. Participants were initially divided into groups of four. After a group induction phase designed to enhance inter-individual affiliation (7) (see supplementary methods 2.1.1), each participant completed a “baseline choice task” independently of the other group members. In this task, participants decided in each trial whether to accept or reject a gamble that involved probabilities of gains and losses (Fig. 1A and appendix S1). As the exact probability of success is rarely known in realistic choice situations, the task included many trials with ambiguous probabilities of gains and losses. However, to distinguish individuals' attitudes toward pure risk versus ambiguity, the task also contained trials in which the exact probabilities were known.

In the “delegation task” (Fig. 1B), the participants faced the same gambles as in the baseline task, but now they had the option to make the decisions themselves (i.e., to lead) or to defer and follow the decision of the group. If a participant deferred, the action implemented (risky or safe) was the one chosen by the majority of the other group members in response to the exact same gamble in the baseline task. The delegation task had two types of trials, the “self” trials and the “group” trials, which were matched on all features except who received the outcome (Fig. 1B). In the self trials, only the payoff of the deciding participant is at stake and the payoffs of the other group members were not affected. By contrast, in the group trials, the decision outcome affected the payoff of every group member equally.

In real-life decisions, individual group members, even though they may objectively face the same situation, often possess unique information or perspectives (8). Our task incorporated this aspect by ensuring that, for every matched baseline and delegation trial, no two group members saw the exact same segment of the probability space (Fig. 1C). Consequently, the group, as a whole, always had more information about the probabilities with which gains or losses occurred than any single individual in the group.

All participants were explicitly informed about the nature of the group-level informational advantage before the delegation task (see supplementary methods 2.2.1 and appendix S2 for task instructions). This group advantage increased with the level of ambiguity, resulting in an identical parametric manipulation of the incentive to defer in both the self and group trials (fig. S1). Although in all trials, deferring to the majority meant taking a better-informed action, it also meant the loss of the individual's decision rights or power to determine the choice (see fig. S1 and supplementary results 1). Thus, participants always had to weigh both of these aspects—the subjective value they put on their decision right versus the value of a better-informed decision—when choosing to lead or defer.

We collected and analyzed choice data from two independent samples of participants: an initial dataset examining only choice behavior and a second dataset in which we replicated the behavioral experiment but also collected neuroimaging data. For brevity, we discuss the behavioral results across all subjects and, in the main text, only report those results that replicated within each dataset independently (for results of each group separately, see the supplementary materials).

Baseline preferences and leadership scores

We initially measured individuals' leadership scores with two widely used scales (9, 10) that predict leadership positions and ability in numerous domains, including politics, athletics, and business (1, 11–13), and later supplemented these questionnaire measures with data on actual leadership roles (see supplementary methods 2.3). On the basis of these measures, we examined whether risk, loss, and ambiguity preferences in the baseline task were associated with leadership scores. None of these preference measures was consistently correlated with leadership scores across both independent samples (table S1 and fig. S2). Moreover, sensitivity to the informational advantage, response times, and choice consistency were not reliably associated with leadership scores (table S1 and supplementary results 1, 2, and 7).

The role of preferences for decision rights and control

Every decision in the delegation task, across both self and group conditions, requires the participant to choose whether or not she will make the decision herself or give up the right to make the choice and follow the other group members'

¹Zürich Center for Neuroeconomics, Department of Economics, University of Zürich, 8006 Zürich, Switzerland.

²Decision Neuroscience Lab, Department of Health Sciences and Technology, ETH Zürich, 8057 Zürich, Switzerland.

*Corresponding author. Email: micah.edelson@econ.uzh.ch (M.G.E.); todd.hare@econ.uzh.ch (T.A.H.); ernst.fehr@econ.uzh.ch (E.F.)

collective judgment. Individuals who put a high value on maintaining their private decision rights should display a relatively lower deferral rate in the self trials when compared to individuals who do not value their private decision rights as highly.

Consistent with the view (5, 6) that decision rights are generally valued positively, participants preferred, on average, to maintain control over their own outcomes in the self trials and were willing to forgo the informational advantage available when deferring to the majority in most trials (mean = 62.7%; Wilcoxon signed-rank test versus a random-choice null hypothesis, z score = 6.0, $P = 2 \times 10^{-9}$). However, the proportion of control-taking choices in the self condition was not related to individual leadership scores (Fig. 2A; Spearman rank correlation coefficient (r) = -0.03, $P = 0.84$).

The driver behind leadership might not be the desire to control only one's own outcome but rather to exert decision rights with broad implications for whole groups. This would imply that the frequency of keeping control in the group trials is informative about real-life leadership measures. Just as in the self trials, on average, participants preferred to maintain control over group outcomes despite the informational advantage of deferring. However, again there is no evidence for an association between the strength of the preference for control in group trials and leadership scores (Fig. 2B; $r = 0.13$, $P = 0.33$; see also supplementary results 1). Thus, preferences in favor of decision rights and control over self or others did not explain individual differences in leadership scores, suggesting that different motivational forces are at work.

Leadership and responsibility aversion

If it is not the aforementioned preferences that distinguish high- from low-scoring leaders, then perhaps a dynamic change to the decision process between individual versus group choices holds the key. A critical difference between group and self trials is the potential responsibility for others' welfare in group trials. Relatively little is known about how responsibility for others' outcomes influences decision-making. Indeed, we do not even know yet whether the average person prefers to seek or avoid responsibility, much less how responsibility preferences might relate to leadership.

The majority of participants preferred to avoid responsibility, that is, participants deferred more on group than self trials. Thus, we term this preference responsibility aversion. The mean percent increase in deferral rate from self to group trials was 17.3% (Wilcoxon signed-rank test, z score = 5.4, $P = 5 \times 10^{-8}$). However, there was substantial variability in the level of responsibility aversion across individuals (SD = $\pm 43\%$). Critically, individuals who showed less responsibility aversion had higher leadership scores (Fig. 2C; $r = -0.46$, $P = 2 \times 10^{-4}$). This variability in responsibility aversion was not significantly correlated with baseline preferences over risk, ambiguity, or loss, nor did it correlate with personality traits from

the "five-factor model" (table S1 and supplementary results 4: for risk, loss, and ambiguity preferences, all $P > 0.66$; for the five-factor model, all $P > 0.2$).

To assess the ecological validity of this association between responsibility aversion and leadership scores, we collected real-life expressions of leadership behavior from our participants (rank obtained during mandatory military service and leadership experience in scouts organizations, supplementary methods 2.3.4). Responsibility aversion was the only measure that significantly correlated with these real-life expressions of leadership (Fig. 2D, $r = -0.49$, $P = 0.02$).

This relationship between responsibility aversion and leadership is also robust. First, all results presented above and in the upcoming sections on computational modeling were initially obtained in the behavioral group and then independently replicated in the functional magnetic resonance imaging (fMRI) group (see the supplementary materials). Second, we computed out-of-sample predictions of the leadership scores for the fMRI sample based on parameter es-

timates computed on the basis of the original behavior-only sample. The predicted leadership scores for the fMRI sample were, indeed, significantly correlated with the empirically observed leadership scores from those participants (Fig. 2E, $r = 0.44$, $P = 0.004$; supplementary results 3).

Taken together, these results suggest that responsibility aversion, an as yet mechanistically undetermined behavioral preference, is a robust and ecologically valid predictor of leadership. Critically, these results hint that some key latent factor(s) in the decision process must change when individuals are faced with the choice to lead others versus making the same choice for themselves alone. What are the underlying cognitive computations and neural mechanisms?

What is responsibility aversion, and why does it arise?

Responsibility aversion, as an interpersonal phenomenon, might be related to social preferences, that is, a concern for others' payoffs. We therefore examined several measures of social preferences as well as feelings of group affiliation and democratic tendencies. We also performed a

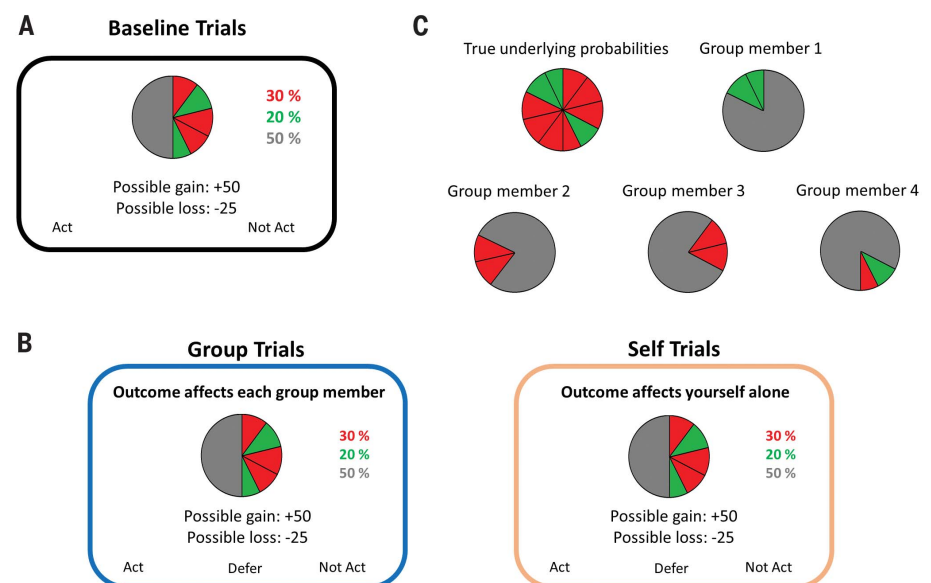


Fig. 1. Experimental design. (A) Baseline task. Individuals needed to select a risky option ("act") or safe option ("not act") on the basis of the probability of success of the risky option and the possible gain or loss if that option was chosen. The probability of success and failure was indicated by the proportion of green or red slices, respectively, in the probability pie and by adjacent text. In each trial, a varying amount of the probability information was obscured by a gray cover. If the individuals preferred the safe choice, they received a sure outcome of 0 for that trial. (B) Delegation task. Two days later, individuals were faced with the same choices but had the additional option to "defer" to the majority opinion of their group and gain access to the group's informational advantage. This task involved two conditions, group (where the participant's action affected the payoff of all group members) and self (where the participant's action affected only herself). (C) Informational advantage for the group. Shown is one example of potential observable probabilities seen by each of the four individuals in the group as well as the true underlying probability pie, which was not displayed to the participants. The position of the obscuring gray cover changed for each individual, resulting in the exposure of a different part of the probability information. Consequently, in our task, the group, as a whole, had more information than each individual alone. For a full description, see supplementary results 1 and fig. S1. The informational advantage and optimal choice, in terms of expected monetary payoff, were identical for each matched group and self trial (see also supplementary results 8).

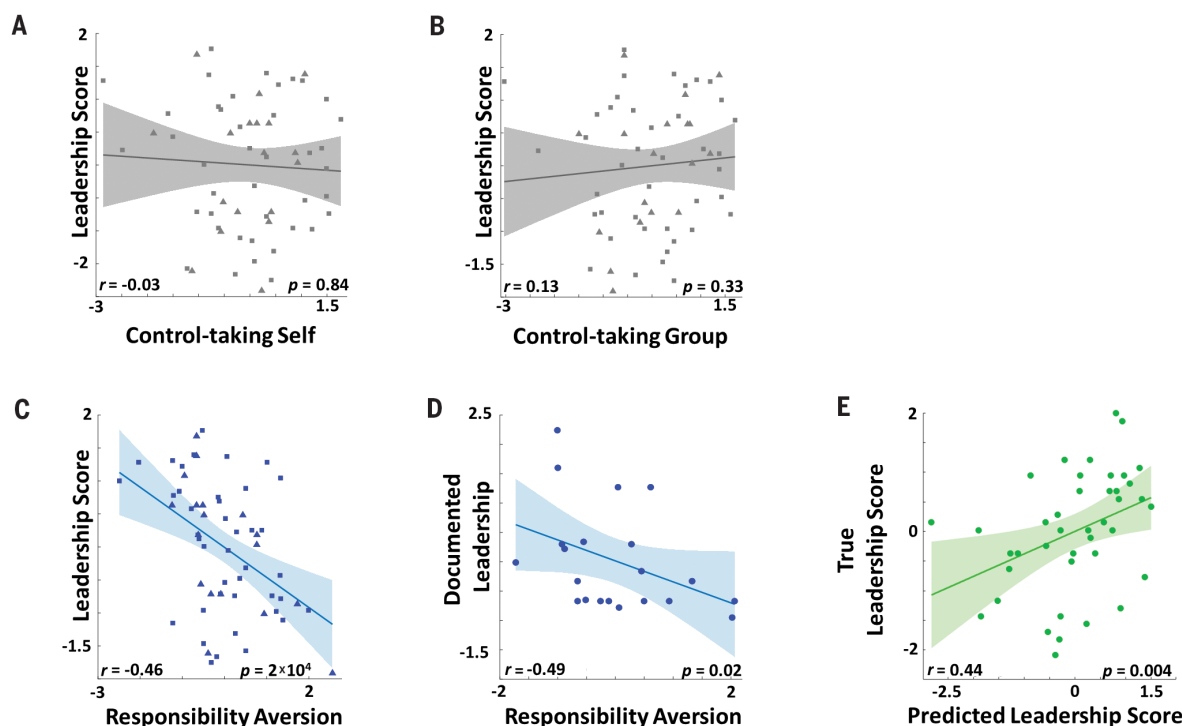


Fig. 2. Behavioral evidence for responsibility aversion. (A and B) Leadership scores as a function of control-taking in self (A) and group (B) trials. The scatter plots and the associated regression line show the (lack of) association between normalized leadership scores and a basic preference for controlling one's own or common outcomes. (C) Responsibility aversion scores correlated negatively with leadership questionnaire scores ($r = -0.46$, $P = 2 \times 10^{-4}$). For (A) to (C), each marker (triangles for the original behavioral group and squares for the fMRI replication group) represents one participant. (D) Responsibility aversion scores (normalized) correlated negatively with real-life manifestation of leadership behavior (such as military rank, $r = -0.49$, $P = 0.02$, data obtainable for $n = 21$). (E) Out-of-sample

prediction of leadership scores for individuals in the fMRI sample. This prediction is based on the parameter coefficients estimated using participants in the original, behavior-only dataset and then applied to each individual in the independent fMRI dataset to predict leadership scores (for full details, see supplementary results 3). The correlation between the observed leadership score and the predicted leadership scores is $r = 0.44$ ($P = 0.004$). For all scatter plots, the solid line is the best-fit regression line, and shaded areas indicate a 95% prediction interval for fit lines estimated from new out-of-sample data points. The correlation coefficients and P values were calculated by using the nonparametric Spearman rank correlation.

control experiment to identify the potential impact of regret, blame, or guilt on responsibility aversion. However, none of these measures was correlated with responsibility aversion (table S1 and supplementary results 5, 6, and 8). Moreover, the association between leadership scores and responsibility aversion remained significant after controlling for such measures in a multiple regression analysis (table S1). Thus, responsibility aversion is distinct from other trait-level preference categories. This raises the questions of why and how it affects decision processes—questions that can only be answered by identifying the underlying computational mechanism—and how the brain implements these processes.

One possibility is that responsibility aversion is driven by a tendency to become more conservative in terms of risk, loss, or ambiguity when making choices that can affect others. Alternatively, responsibility aversion could be driven by an as yet uncharacterized cognitive process. Therefore, we analyzed participants' behavior by developing a computational model that allowed us to determine the mechanism underlying responsibility aversion.

To convey the logic of our computational modeling approach, we first describe the choice

behavior that participants demonstrated in the baseline and self trials, in which responsibility can play no role, and then explain how this inspired our efforts to formally model the mechanisms generating the observed changes in behavior for the matched group trials. The patterns of deferral choices (Fig. 3A) and reaction times (Fig. 3B) provide an initial clue as to how deferral decisions are made and the type of computational process that might underlie these choices. We estimated subjects' preference parameters (i.e., attitudes toward risk, loss, and ambiguity and probability weights), using a prospect theory model (supplementary methods 3.1; see also supplementary results 9), and used these parameters to compute the subjective-value differences between accepting and rejecting the gamble in each trial. Fig. 3A depicts the proportion of deferral choices during self trials as a function of these subjective-value differences.

The figure shows an inverted U-shaped pattern. For large subjective-value differences, the probability of deferral is close to zero, whereas for small differences, average deferral rates reach almost 60%. Low subjective-value differences mean that the values of the two options are difficult to distinguish, that is, the discrimina-

bility between the options is low, whereas high subjective-value differences imply high discriminability between the options. This interpretation is also supported by reaction-time data (Fig. 3B), which show that response times are highest when subjective-value differences are low. Thus, when there is little doubt that accepting (or rejecting) the gamble is the superior option in a given trial, participants generally make the decision themselves rather than letting the group decide. However, when standard preferences toward loss, risk, and ambiguity provide little guidance about what constitutes the best choice because the subjective-value difference is small, participants defer more often to the group.

We thus postulated that responsibility aversion might be due to changes in the demand for certainty about what constitutes the best choice when also deciding for others instead of only for oneself. According to this hypothesis, the subjective value of the gamble and the uncertainty about what is the best choice do not change between the self and the group trials. Rather, it is the required level of certainty about the best response to the gamble that changes when individuals are responsible for others. In mechanistic terms, the demand for certainty in a given

Fig. 3. Patterns of deferral

behavior. (A) Percentage of choices to defer for self trials as a function of the subjective-value difference between the safe and risky options (10 bins; negative values indicate a relative advantage for the safe option, whereas positive values indicate an advantage for the risky option; values calculated independently in the baseline task by using a prospect theory model, see supplementary methods 3.1). Bins in the middle (−1 and 1) of the x axis are those in which the subjective values of the risky and safe choices are most similar. For bins on the extreme right of the x axis (5), risky options are strongly preferred, whereas safe options are strongly preferred at the extreme left (−5). **(B)** Reaction times (RTs, measured in milliseconds) as a function of subjective-value difference in baseline trials, in which deferring was not an option. Thus, we measure the RT specific to the risky or safe choices in every trial. In line with a large amount of literature on perceptual and value-based decision-making (36), one would predict that low discriminability (higher choice difficulty) corresponds to longer RTs, whereas high discriminability is associated with shorter RTs. **(C)** Illustration of the hypothesized mechanism involving a shift in a deferral threshold. In the self condition, values more extreme than the deferral threshold (orange lines) indicate that the participant feels certain enough to make the choice herself, in most cases. A shift in this deferral threshold toward the extremes of the distribution in the group condition (blue lines) would result in less

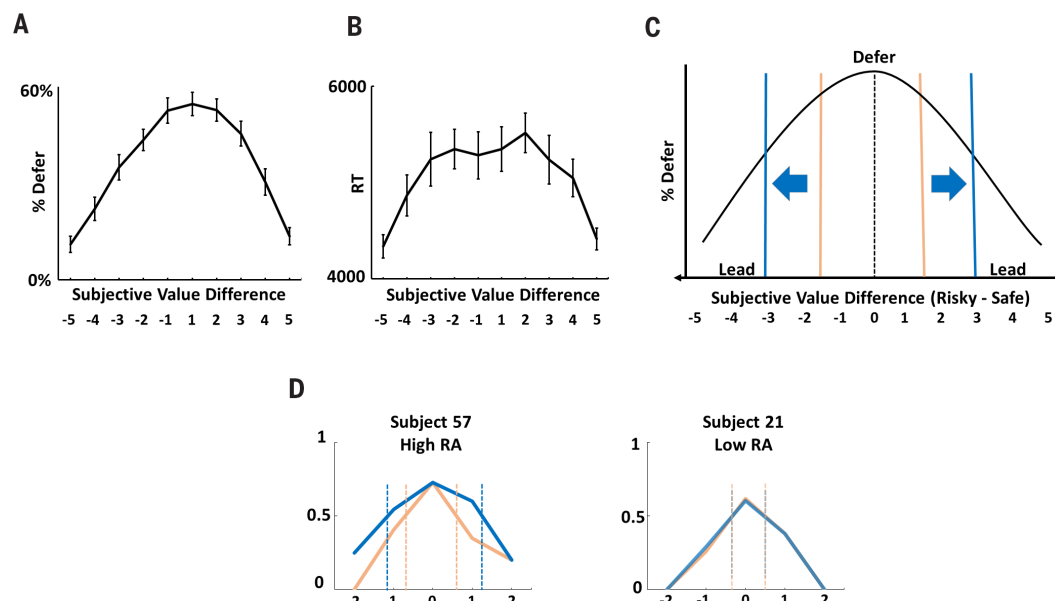
choice condition can be represented by deferral thresholds. A deferral threshold is defined by the critical subjective-value difference between accepting and rejecting the gamble (i.e., the vertical lines in Fig. 3C) at which the subject switches between preferring to lead, on average, versus deferring. Naturally, there will be a critical subjective-value difference (deferral threshold) for switching between deferring and leading in both the negative (i.e., when the safe option is preferred) and positive (i.e., when the risky option is preferred) domains. The thresholds define a critical range of subjective-value differences within which the participant prefers to defer to the group and beyond which the participant prefers to make the decision herself (Fig. 3C). The optimal deferral thresholds are determined by the size and precision of the subjective-value difference (i.e., certainty) and the potential leader's prior beliefs about the utility of leading and the utility of deferring as a function of subjective-value differences (supplementary methods 3). If, for example, the demand for certainty increases in one condition relative to another, then the deferral thresholds become wider and the potential leader will defer more often.

Thus, a responsibility-averse individual could potentially be characterized as someone who demands higher certainty about what is the best

choice in the group trials compared to self trials, which is tantamount to wider deferral thresholds in the group trials (Fig. 3, C and D). It is critical to note that we are proposing that this mechanism involves a change in the level of certainty required to take the choice when faced with potential responsibility for others rather than an overall high or low demand for certainty.

A change in the demand for certainty about the best choice represents an alternative mechanism to the hypothesis that changes in the subjective-value construction process via preferences over risk, loss, and ambiguity or probability weighting across self and group trials lead to responsibility aversion (14, 15). This “shift-in-standard-preferences hypothesis” can, in principle, account for the higher willingness to defer in the group trials while maintaining a constant threshold across trial types (see fig. S3). For example, if a subject becomes more loss averse in the group trials, then the subjective-value difference between accepting and rejecting becomes smaller in many trials. Therefore, a subject may prefer to keep the decision right for a given lottery in the self trials (because the subjective-value difference is outside the fixed critical range) but defer the decision right in the group trials (because the subjective-value difference shrinks and is now within the fixed critical range). Thus,

trials crossing this threshold and a reduced tendency to lead. The dashed black line indicates the zero point in the difference between the subjective values of the safe and risky options. **(D)** Shifts in deferral thresholds at the individual level. The choice patterns for two example participants with either high or low responsibility aversion (29 versus 0% increase in deferral frequency in the group trials). The point of indifference between deferring and leading shifts in the strongly responsibility-averse individual (subject 57) but remains constant in the low responsibility-averse participant (subject 21). Note that we use 5, instead of 10, levels of subjective-value difference in the individual plots because there are fewer trials at the individual level. For (A) and (B), error bars represent SEM.



it is not clear a priori which potential mechanism is more consistent with the leadership decisions we observed.

A mechanistic explanation of responsibility aversion and leadership behavior

We specified a computational model in which individuals' preference parameters and their deferral thresholds are simultaneously estimated on the basis of their behavior in the self and group trials. This model constitutes an implicit horse race between the shift-in-standard-preferences hypothesis and an explanation of responsibility aversion in terms of differences in deferral thresholds across conditions. If standard preferences vary substantially between the conditions while deferral thresholds remain constant, responsibility aversion is best explained in terms of changes in conventional preferences. If, however, conventional preference estimates remain constant across group and self trials while the deferral thresholds vary, then responsibility aversion can be attributed to changes in deferral thresholds and the beliefs about the relative utility of deferring that they signify.

Our computational model combines aspects of optimal categorization (16, 17), which enable the empirical identification of individuals' deferral

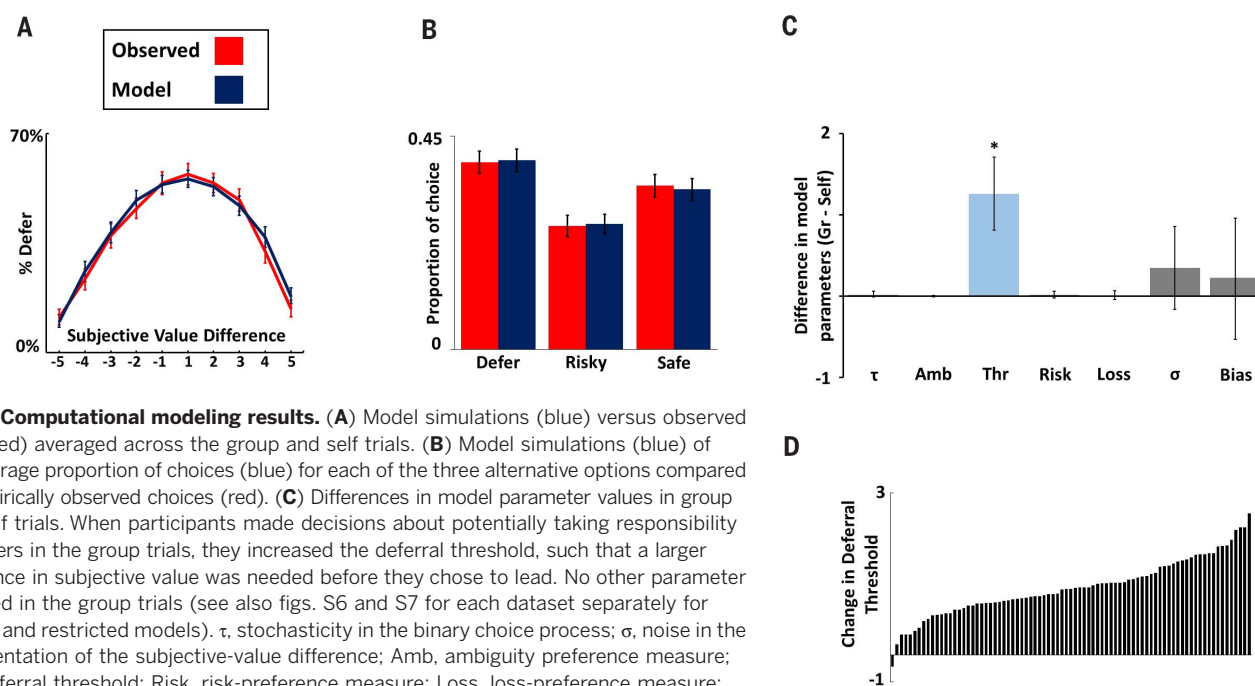


Fig. 4. Computational modeling results. (A) Model simulations (blue) versus observed data (red) averaged across the group and self trials. (B) Model simulations (blue) of the average proportion of choices (blue) for each of the three alternative options compared to empirically observed choices (red). (C) Differences in model parameter values in group and self trials. When participants made decisions about potentially taking responsibility for others in the group trials, they increased the deferral threshold, such that a larger difference in subjective value was needed before they chose to lead. No other parameter changed in the group trials (see also figs. S6 and S7 for each dataset separately for the full and restricted models). τ , stochasticity in the binary choice process; σ , noise in the representation of the subjective-value difference; Amb, ambiguity preference measure; Thr, deferral threshold; Risk, risk-preference measure; Loss, loss-preference measure; Bias, measure of left or right asymmetry in deferral thresholds. *The posterior probability of a difference between the conditions is >0.999 . The blue and gray shading highlight significant and nonsignificant changes across conditions, respectively. (D) The change in the deferral threshold, measured in subjective-value units, between the group and self conditions. Each bar represents one individual. For (A) and (B), error bars represent SEM; for (C), errors bars represent 95% credible intervals because they are obtained from a posterior distribution on the population level (see supplementary methods 3).

thresholds, with prospect theory (18), which enables the empirical identification of individuals preference parameters for risk, loss, and ambiguity and probability weights (see supplementary methods 3 and supplementary results 9). The model simultaneously estimates a condition-specific (group or self) deferral threshold and condition-specific preference parameters from each individual's pattern of choices. The probability of deferring is jointly determined by the subjective value of the gamble and the deferral thresholds. The probability of choosing the risky versus safe action conditional on leading is determined for each decision problem on the basis of the subjective value of the risky relative to the safe option.

Our computational model accurately captures the patterns of choice behavior (Fig. 4, A and B; see also model comparison results in table S2 and parameter recovery exercise in table S7). This allowed us to use it in determining which of the underlying components of the decision process are affected by responsibility for others' welfare. Direct tests of model parameters between conditions showed that the group trials led to a specific increase in the deferral threshold [mean change (\pm SD) is 1.26 (\pm 0.23); posterior probability of a difference between the conditions is >0.999] but did not influence any other model parameter (Fig. 4C). Thus, being responsible for others did not change the way participants processed key decision-relevant information such as reward magnitude, risk, or ambiguity but rather induced a shift in the deferral threshold, indicating a higher demand

for certainty about the best choice in the group trials. Critically, the σ parameter quantifying the noise in the subjective-value difference representation, and partially determining the threshold values, does not change, suggesting that changes in prior beliefs about the utility of leading and the utility of deferring as a function of the subjective-value difference drive responsibility aversion.

Almost all individuals increased their deferral threshold in the group trials relative to the self trials (Fig. 4D). Moreover, these individual-level changes in the deferral threshold were correlated with leadership scores ($r = -0.46$, $P = 3 \times 10^{-4}$). More stable thresholds across conditions were associated with higher leadership scores.

Our results suggest the following theoretical conceptualization of the choice to lead or to defer: Depending on their demand for certainty about the best choice, the subjects establish boundaries in subjective-value space (i.e., deferral thresholds) that are used to determine whether leading or deferring is the best course of action. In each lead or defer decision, the subjective values of the available options are constructed from underlying basic preferences over risk, loss, ambiguity, decision rights, and so on. Only once these values are constructed can they be compared to the deferral threshold. Therefore, responsibility aversion is fundamentally different from basic preferences over risk, loss, and ambiguity or probability weights. Although these preferences play a role in determining the subjective value of the gambles, they are stable

across self and group trials and therefore cannot explain the existence of responsibility aversion. Instead, changes in beliefs about the utility of leading and deferring when potentially deciding for others underlie responsibility aversion. The resulting change in the demand for certainty for group trials relative to self trials suggests that a form of second-order introspection or metacognitive processing (2, 19, 20) is involved in responsibility aversion.

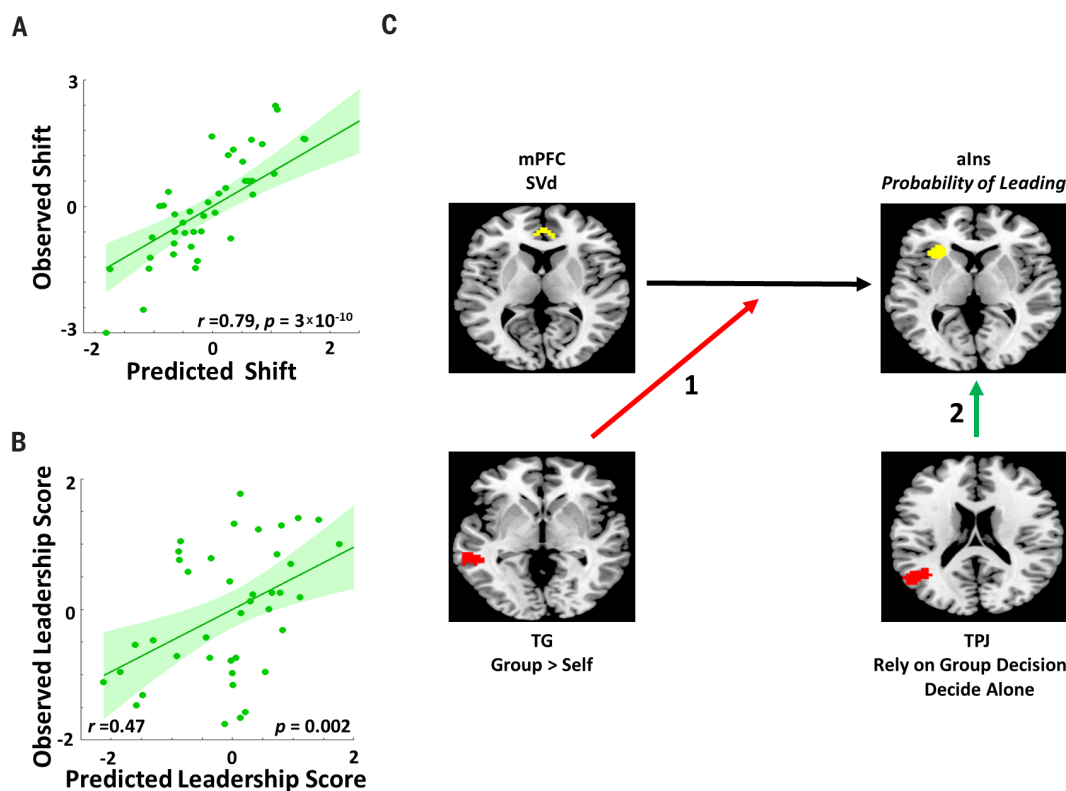
Although high-scoring leaders can vary substantially in terms of underlying preferences (e.g., risk, loss, and control preferences), the unifying element is that they calibrate their prior beliefs about the utility of leading and deferring similarly across group and self trials. This characterization of the choice to lead is compatible with many different leadership styles or leadership types (see fig. S4) (11, 21–26). Consider, for example, an “authoritarian” leader with a strong preference for control and thus a very narrow deferral threshold in both the group and self trials. Compare her with a “democratic” leader with a strong preference for consensus who displays a rather broad deferral threshold in both group and self trials. Both leadership types are consistent with our conceptualization of leadership choice, and our theory predicts that both will have a high score for goal-oriented leadership because the key mechanism underlying the choice to lead is the similarity in the deferral thresholds across group and self trials. Thus, the choice process we describe can serve as a unifying mechanism across the variety

Fig. 5. Predictions about responsibility aversion and leadership choices from a minimal neural network model. (A) The

scatter plot shows the correlation between the out-of-sample predicted shift in deferral thresholds, which are based on individuals' connectivity parameters in the neural network, and individuals' observed scores computed from their choices ($r = 0.79$, $P = 3 \times 10^{-10}$).

(B) The scatter plot shows the correlation between the out-of-sample predicted leadership scores, which are based on individuals' connectivity parameters in the neural network and the preference measures in table S1, and individuals' observed leadership scores ($r = 0.47$, $P = 0.002$).

(C) Schematic representation of a subset of the neural network parameters, specifically those most closely linked to individual differences in the modification of the deferral threshold (see fig. S8 and table S6 for all DCM and regression weights, respectively). Individuals who shift their deferral threshold showed a reduced influence of mPFC activity on the aIns. The degree of this reduction was proportional to activity in the TG, which is higher in group relative to self trials (arrow 1). The reduced influence of mPFC on aIns and the impact of TG activity on this reduction suggest that the influence of the subjective-value difference on choices is modulated under responsibility. Participants with a larger



shift in deferral thresholds also show a stronger negative effect of the TPJ input on the aIns (arrow 2). This TPJ activity had a stronger effect on the aIns among participants who showed a larger shift in deferral thresholds. Yellow-colored regions represent parametric correlations with trial-wise regressors obtained from our computational model. Red-colored regions represent simple binary contrasts. For (A) and (B), shaded areas indicate a 95% prediction interval for fit lines estimated from new out-of-sample data points.

of traits and characteristics associated with leadership (1, 17).

Neural mechanisms of responsibility aversion

We next turned to neural data to further understand the latent determinants of this process and how they are implemented in the brain. In our computational model, the key factor determining whether an individual will assume responsibility in any given trial is whether the current subjective-value difference exceeds the deferral threshold. Consequently, we can test the hypothesis that individual differences in responsibility aversion will manifest as differences in this comparison process at the neural level.

How might such a comparison process be implemented in the brain? Higher-order cognitive functions, such as leadership decisions, are most likely supported by interactions between both local and anatomically distinct pools of neurons (27). Therefore, we constructed a minimal model of the neural processing nodes that can incorporate the different choice aspects related to assuming responsibility and used this

minimal network to test manifestations of individual differences at the neural level.

We first used fMRI data from participants who made decisions in the delegation task to identify brain regions (i.e., potential network nodes) where activity correlated with the four key aspects of our task: (i) the trial type (group versus self), (ii) relying on the group's decision (defer rather than lead), (iii) the subjective-value difference, and (iv) the estimated probability of leading, $p(l)$ in each trial. Our goal here was not an exhaustive characterization of all brain activity patterns underlying leadership decisions. Rather, we aimed to test if activity patterns, centered on the time of choice, in a minimalistic brain network, can further uncover unobservable aspects of the internal decision process and test the mechanism for choosing the responsibility of leadership derived through computational modeling of the choice data.

First, we identified activation that correlated with the four aforementioned variables in our leadership decision task at the time of choice (see tables S3 to S5). The basic contrast testing for differential activity as a function of choice condition (group versus self) revealed increased activity in the middle-superior temporal gyrus

(TG) when participants were potentially responsible for the welfare of others. The temporal parietal junction (TPJ) (i) was more active when participants deferred their decision right to the group and (ii) also increased as a function of the informational advantage (i.e., potential benefit) available by deferring and taking advantage of the other group members' knowledge regardless of the decision outcome (see supplementary results 10.2).

We also used the model-derived, trial-wise estimates of the subjective-value difference and the probability of leading, $p(l)$ as parametric regressors in our fMRI analyses. These two parametric contrasts revealed that the subjective-value difference was associated with activity in several brain regions, including the medial prefrontal cortex (mPFC), whereas the probability of leading was most strongly reflected in the activity of the anterior insula (aIns; for additional details and full results of all univariate analyses, see supplementary methods 5, supplementary results 10, and tables S3 to S5).

Having identified regional activity (TG, TPJ, mPFC, and aIns) that correlated with the four critical components of our leadership task, we

next quantified the levels of functional interaction between these four network components. We fit a stochastic dynamic causal model (DCM) (28) to estimate the context-dependent changes in functional coupling within our network on the group and self choices (for full details, see supplementary methods 5.4). Once we obtained the parameters representing the levels of local activity and functional coupling within our brain network model on group relative to self trials, we tested whether these measures can be used to predict and, ultimately, better understand individual patterns of leadership choices.

Individual differences in the parameters of our brain network model were indeed predictive of individual differences in the shift in deferral thresholds and leadership scores (Fig. 5, A and B). A model including only the neural network parameters yielded accurate out-of-sample predictions for each participant's shift in the deferral threshold (median split classification accuracy = 91%, $P = 2 \times 10^{-11}$).

We also tested if these neural parameters explained variation in leadership scores over and above the behavioral measures listed in table S1 (including responsibility aversion). Model comparison demonstrated that including the parameters of the DCM along with the behavioral measures provided a better fit to the data (Akaike information criterion and Bayesian information criterion differences are equal to 186.6 and 119.8, respectively). Once again, this combined model made accurate out-of-sample classifications of the participants' leadership scores (median split classification accuracy = 71%, $P = 0.006$).

Next, we turned to the question of which brain network parameters best explained individual differences in behavior. In our computational model of behavior, the deferral thresholds are compared to the subjective-value difference to determine whether it is best to lead or defer in each trial, and these thresholds generally increase with responsibility for others (Figs. 3D and 4C). This widening of the deferral thresholds signifies a change in the association between the subjective-value difference and deferral-choice probabilities, and this change is greater in highly responsibility-averse individuals because the deferral threshold moves further out. Therefore, if mPFC activity is associated with the subjective-value difference and aIns activity is associated with the probability of leading, then we should see a differential impact of mPFC activity on the aIns in participants with larger responsibility aversion (i.e., greater widening of the thresholds) in the group trials.

This pattern of results was indeed observed (Fig. 5C and table S6) and was conditional on activity in the TG. Recall that TG activity was higher in the group trials compared to the self trials. Increased TG activity was associated with a lower or inhibited influence of mPFC on aIns at the neural level. Leaders show less of this inhibition. This provides a potential neural mechanism for the change in deferral thresholds. These findings further support the conclusion

that responsibility aversion is the result of a second-order process operating on the results of subjective-value computations generally thought to be related to mPFC activity (29, 30).

We also found that, during the group trials (relative to the self trials), there was a stronger influence of TPJ activity on the aIns in individuals who showed a larger shift in deferral thresholds. Activity in the TPJ reflected, in our task, the potential informational advantage available by deferring to the decisions of the other group members, consistent with theories on the role of the TPJ in mentalizing (31). We speculate that stronger signaling from the TPJ to the aIns in group trials may be one means through which the deferral threshold is increased, thus producing the observed responsibility-averse choices.

Conclusion

Being a leader requires making choices that will determine others' welfare. Decisions as diverse as committing soldiers to the battlefield or picking a school for your child share a basic attribute: assuming responsibility for the outcome of others. Thus, although the motivations driving one to lead a country, business, or classroom are many and varied (and domain-specific attitudes most likely play an important role), a willingness to shoulder responsibility is present in all who choose to lead, shaping every level of society for better or worse.

Our results provide a behavioral, computational, and neurobiological microfoundation of the processes underlying the decision to lead. Although early conceptual leadership research emphasized the importance, and speculated on the nature, of internal decision-making processes (32), the necessary empirical and analytic tools to directly address these questions were not available at the time. We identify low responsibility aversion as an important determinant of the decision to lead and demonstrate, empirically and computationally, that it is based on a multilevel evaluation of the subjective evidence in favor of one potential action over another in the light of prior beliefs about the utility of maintaining control (33), gaining information, and taking responsibility for others' outcomes.

We provide both a precise empirical measure and a theoretical foundation of responsibility aversion that make it possible to further explore its implications for social and economic phenomena (34). There could be a psychological cost for assuming responsibility for others' outcomes, which may require extra compensation. It may explain why "responsibility" is often used to justify pay differentials in hierarchical organizations (35), as well as why organizations may want to economize on these costs and preferentially choose individuals with low responsibility aversion for leadership positions and why individuals with low responsibility aversion are more likely to self-select into such positions (see Fig. 2D). These conjectures and our characterization of the leadership choice process raise many future research opportunities, and we hope that the empirical and theoretical concepts developed

in this paper will prove useful in providing a more thorough understanding of these issues.

Methods summary

A full description of the materials and methods is provided in the supplementary materials. Briefly, we collected choice data from 40 participants on a decision paradigm in which an individual could delegate decision-making power about a choice between a risky and a safe option to their group or keep the right to decide. In the main task, the participants made 140 different choices under two conditions: in the self trials, only the individual's payoff is at stake, whereas in the group trials, each group member's payoff is affected. We combined computational modeling approaches from the perceptual and value-based decision-making domains to estimate each individual's personal utility for every available action in order to tease apart potential motivations for choosing to "lead" or "follow." In a separate sample of 44 participants, we collected choice data using the same decision paradigm in conjunction with fMRI. The fMRI data were analyzed with effective functional connectivity modeling techniques to examine the neurobiological basis of leadership choices.

REFERENCES AND NOTES

1. B. Bass, R. Bass, *The Bass Handbook of Leadership: Theory, Research, and Managerial Applications* (Free Press, New York, 2009).
2. S. M. Fleming, N. D. Daw, Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychol. Rev.* **124**, 91–114 (2017). doi: [10.1037/rev0000045](https://doi.org/10.1037/rev0000045); pmid: [28004960](https://pubmed.ncbi.nlm.nih.gov/28004960/)
3. C. Summerfield, K. Tsetos, Do humans make good decisions? *Trends Cogn. Sci.* **19**, 27–34 (2015). doi: [10.1016/j.tics.2014.11.005](https://doi.org/10.1016/j.tics.2014.11.005); pmid: [25488076](https://pubmed.ncbi.nlm.nih.gov/25488076/)
4. B. De Martino, S. Bobadilla-Suarez, T. Nouguchi, T. Sharot, B. C. Love, Social information is integrated into value and confidence judgments according to its reliability. *J. Neurosci.* **37**, 6066–6074 (2017). doi: [10.1523/JNEUROSCI.3880-16.2017](https://doi.org/10.1523/JNEUROSCI.3880-16.2017); pmid: [28566360](https://pubmed.ncbi.nlm.nih.gov/28566360/)
5. B. Bartling, E. Fehr, H. Herz, The intrinsic value of decision rights. *Econometrica* **82**, 2005–2039 (2014). doi: [10.3982/ECTA11573](https://doi.org/10.3982/ECTA11573)
6. E. Fehr, H. Herz, T. Wilkening, The lure of authority: Motivation and incentive effects of power. *Am. Econ. Rev.* **103**, 1325–1359 (2013). doi: [10.1257/aer.103.4.1325](https://doi.org/10.1257/aer.103.4.1325)
7. C. Eckel, P. Grossman, Managing diversity by creating team identity. *J. Econ. Behav. Organ.* **58**, 371–392 (2005). doi: [10.1016/j.jebo.2004.01.003](https://doi.org/10.1016/j.jebo.2004.01.003)
8. J. Faria, J. Dyer, C. Tosh, J. Krause, Leadership and social information use in human crowds. *Anim. Behav.* **79**, 895–901 (2010). doi: [10.1016/j.anbehav.2009.12.039](https://doi.org/10.1016/j.anbehav.2009.12.039)
9. A. W. Halpin, B. J. Winer, in *Leader Behavior: Its Description and Measurement*, R. M. Stogdill, A. E. Coons, Eds. (Ohio State Univ., 1957).
10. R. Blake, J. Mouton, A. Bidwell, Managerial grid. *Adv. Manag.-Office Exec.* **1**, 12–15 (1962).
11. E. A. Fleishman et al., Taxonomic efforts in the description of leader behavior: A synthesis and functional interpretation. *Leadersh. Q.* **2**, 245–287 (1992). doi: [10.1016/1048-9843\(91\)90016-U](https://doi.org/10.1016/1048-9843(91)90016-U)
12. C. S. Burke et al., What type of leadership behaviors are functional in teams? A meta-analysis. *Leadersh. Q.* **17**, 288–307 (2006). doi: [10.1016/j.leaqua.2006.02.007](https://doi.org/10.1016/j.leaqua.2006.02.007)
13. T. A. Judge, R. F. Piccolo, R. Ilies, The forgotten ones? The validity of consideration and initiating structure in leadership research. *J. Appl. Psychol.* **89**, 36–51 (2004). doi: [10.1037/0021-9010.89.1.36](https://doi.org/10.1037/0021-9010.89.1.36); pmid: [14769119](https://pubmed.ncbi.nlm.nih.gov/14769119/)
14. G. Charness, M. Jackson, The role of responsibility in strategic risk-taking. *J. Econ. Behav. Organ.* **69**, 241–247 (2009). doi: [10.1016/j.jebo.2008.10.006](https://doi.org/10.1016/j.jebo.2008.10.006)

15. J. Pahlke, S. Strasser, F. Vieider, Responsibility effects in decision making under risk. *J. Risk Uncertain.* **51**, 125–146 (2015). doi: [10.1007/s11166-015-9223-6](https://doi.org/10.1007/s11166-015-9223-6)
16. D. M. Green, J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, 1966).
17. A. T. Qamar et al., Trial-to-trial, uncertainty-based adjustment of decision boundaries in visual categorization. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20332–20337 (2013). doi: [10.1073/pnas.1219756110](https://doi.org/10.1073/pnas.1219756110); pmid: [24272938](https://pubmed.ncbi.nlm.nih.gov/24272938/)
18. A. Tversky, D. Kahneman, Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertain.* **5**, 297–323 (1992). doi: [10.1007/BF00122574](https://doi.org/10.1007/BF00122574)
19. C. D. Frith, U. Frith, Mechanisms of social cognition. *Annu. Rev. Psychol.* **63**, 287–313 (2012). doi: [10.1146/annurev-psych-120710-100449](https://doi.org/10.1146/annurev-psych-120710-100449); pmid: [21838544](https://pubmed.ncbi.nlm.nih.gov/21838544/)
20. Y.-L. Boureau, P. Sokol-Hessner, N. D. Daw, Deciding how to decide: Self-control and meta-decision making. *Trends Cogn. Sci.* **19**, 700–710 (2015). doi: [10.1016/j.tics.2015.08.013](https://doi.org/10.1016/j.tics.2015.08.013); pmid: [26483151](https://pubmed.ncbi.nlm.nih.gov/26483151/)
21. B. J. Avolio, B. M. Bass, D. I. Jung, Re-examining the components of transformational and transactional leadership using the Multifactor Leadership. *J. Occup. Organ. Psychol.* **72**, 441–462 (1999). doi: [10.1348/096317999166789](https://doi.org/10.1348/096317999166789)
22. P. Balthazard, D. Waldman, R. W. Thatcher, S. T. Hannah, Differentiating transformational and non-transformational leaders on the basis of neurological imaging. *Leadersh. Q.* **23**, 244–258 (2012). doi: [10.1016/j.leaqua.2011.08.002](https://doi.org/10.1016/j.leaqua.2011.08.002)
23. S. Einarsen, M. S. Aasland, A. Skogstad, Destructive leadership behaviour: A definition and conceptual model. *Leadersh. Q.* **18**, 207–216 (2007). doi: [10.1016/j.leaqua.2007.03.002](https://doi.org/10.1016/j.leaqua.2007.03.002)
24. J. Antonakis, D. V. Day, B. Schyns, Leadership and individual differences: At the cusp of a renaissance. *Leadersh. Q.* **23**, 643–650 (2012). doi: [10.1016/j.leaqua.2012.05.002](https://doi.org/10.1016/j.leaqua.2012.05.002)
25. R. E. Boyatzis, K. Rochford, A. I. Jack, Antagonistic neural networks underlying differentiated leadership roles. *Front. Hum. Neurosci.* **8**, 114 (2014). doi: [10.3389/fnhum.2014.00114](https://doi.org/10.3389/fnhum.2014.00114); pmid: [24624074](https://pubmed.ncbi.nlm.nih.gov/24624074/)
26. J. Jiang et al., Leader emergence through interpersonal neural synchronization. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 4274–4279 (2015). doi: [10.1073/pnas.1422930112](https://doi.org/10.1073/pnas.1422930112); pmid: [25831535](https://pubmed.ncbi.nlm.nih.gov/25831535/)
27. S. L. Bressler, V. Menon, Large-scale brain networks in cognition: Emerging methods and principles. *Trends Cogn. Sci.* **14**, 277–290 (2010). doi: [10.1016/j.tics.2010.04.004](https://doi.org/10.1016/j.tics.2010.04.004); pmid: [20493761](https://pubmed.ncbi.nlm.nih.gov/20493761/)
28. B. Li et al., Generalised filtering and stochastic DCM for fMRI. *Neuroimage* **58**, 442–457 (2011). doi: [10.1016/j.neuroimage.2011.01.085](https://doi.org/10.1016/j.neuroimage.2011.01.085); pmid: [21310247](https://pubmed.ncbi.nlm.nih.gov/21310247/)
29. O. Bartra, J. T. McGuire, J. W. Kable, The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* **76**, 412–427 (2013). doi: [10.1016/j.neuroimage.2013.02.063](https://doi.org/10.1016/j.neuroimage.2013.02.063); pmid: [23507394](https://pubmed.ncbi.nlm.nih.gov/23507394/)
30. J. A. Clithero, A. Rangel, Informatic parcellation of the network involved in the computation of subjective value. *Soc. Cogn. Affect. Neurosci.* **9**, 1289–1302 (2014). doi: [10.1093/scan/nst106](https://doi.org/10.1093/scan/nst106); pmid: [23887811](https://pubmed.ncbi.nlm.nih.gov/23887811/)
31. R. M. Carter, S. A. Huettel, A nexus model of the temporal-parietal junction. *Trends Cogn. Sci.* **17**, 328–336 (2013). doi: [10.1016/j.tics.2013.05.007](https://doi.org/10.1016/j.tics.2013.05.007); pmid: [23790322](https://pubmed.ncbi.nlm.nih.gov/23790322/)
32. V. Vroom, P. Yetton, *Leadership and Decision-Making* (Univ. Pittsburgh Press, 1973).
33. L. A. Leotti, M. R. Delgado, The value of exercising control over monetary gains and losses. *Psychol. Sci.* **25**, 596–604 (2014). doi: [10.1177/0956797613514589](https://doi.org/10.1177/0956797613514589); pmid: [24390827](https://pubmed.ncbi.nlm.nih.gov/24390827/)
34. Consider the analogy with the intuitive notion of risk aversion, which lacked a precise theoretical underpinning before Arrow and Pratt precisely defined it in terms of the concavity of an individual's utility function, thereby opening the door for theoretical modeling and the precise interpretation of empirical measures of risk aversion. Without these foundations, progress in understanding the concept of risk aversion would have been seriously impeded.
35. J. Dickinson, Employees' preferences for the bases of pay differentials. *Employee Relat.* **28**, 164–183 (2006). doi: [10.1108/01425450610639383](https://doi.org/10.1108/01425450610639383)
36. I. Krajbich, B. Bartling, T. Hare, E. Fehr, Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nat. Commun.* **6**, 7455 (2015). doi: [10.1038/ncomms8455](https://doi.org/10.1038/ncomms8455); pmid: [26135809](https://pubmed.ncbi.nlm.nih.gov/26135809/)

ACKNOWLEDGMENTS

We thank Y. Berson, T. Fitzgerald, M. Grueschow, and T. Sharot for helpful feedback; L. Kasper and K. Treiber for technical assistance; and S. Gobbi for error-proofing scripts. **Funding:** E.F. was supported by the European Research Council (Advanced Grant on the Foundations of Economic Preferences). M.G.E., R.P., C.C.R., E.F., and T.A.H. were supported by the Swiss National Science Foundation (grant numbers 100014_140277, 320030_143443, and 105314_152891 and Sinergia grant CRSI3_141965). C.C.R. was supported by the European Research Council (BRAINCODES). **Author contributions:** M.G.E. conceived the idea. M.G.E., T.A.H., and E.F. designed experiments with contributions from C.C.R. M.G.E. conducted the experiments. M.G.E., R.P., and T.A.H. performed the analyses and computational modeling with contributions from E.F. M.G.E., T.A.H., and E.F. wrote the paper with contributions from C.C.R. and R.P. All authors discussed the results and implications and commented on the manuscript at all stages. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** Data and analyses codes are available at https://econgilt.uzh.ch/thare/Edelson_Polania_Ruff_Fehr_Hare_2018.git.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6401/eaat0036/suppl/DC1
Materials and Methods
Supplementary Results
Fig. S1 to S8
Tables S1 to S7
References (37–122)
Appendices S1 and S2

14 January 2018; accepted 6 July 2018
10.1126/science.aat0036



Computational and neurobiological foundations of leadership decisions

Micah G. Edelson, Rafael Polania, Christian C. Ruff, Ernst Fehr, and Todd A. Hare

Science **361** (6401), eaat0036. DOI: 10.1126/science.aat0036

Leadership and responsibility

Leadership of groups is of paramount importance and pervades almost every aspect of society. Leadership research has rarely used computational modeling or neuroimaging techniques to examine mechanistic or neurobiological underpinnings of leadership choices. Edelson *et al.* found empirically and theoretically that the choice to lead rests on a metacognitive process (see the Perspective by Fleming and Bang). Individuals who showed less “responsibility aversion” had higher leadership scores. A computational model combining signal detection theory with prospect theory provided a mechanistic understanding of this preference. Neuroimaging experiments showed how the key theoretical concepts are encoded in the activity and connectivity of a brain network that comprises the medial prefrontal cortex, the superior temporal gyrus, the temporal parietal junction, and the anterior insula.

Science, this issue p. 467; see also p. eaat0036

View the article online

<https://www.science.org/doi/10.1126/science.aat0036>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2018, American Association for the Advancement of Science