Good Decisions, Good Causes: Optimality as a Constraint on Attribution of Causal Responsibility

Samuel G. B. Johnson (samuel.johnson@yale.edu)
Department of Psychology, Yale University
2 Hillhouse Ave., New Haven, CT 06520 USA

Lance J. Rips (rips@northwestern.edu)
Department of Psychology, Northwestern University
2029 Sheridan Road, Evanston, IL 60208 USA

Abstract

How do we assign causal responsibility for others' decisions? The present experiments examine the possibility that an optimality constraint is used in these attributions, with agents considered less responsible for outcomes when the decisions that led to those outcomes were suboptimal. Our first two experiments investigate scenarios in which agents are choosing among multiple options, varying the efficacy of the forsaken alternatives to examine the role of optimality in attributing responsibility. Experiment 3 tests whether optimality considerations also play a role in attribution of causality more generally. Taken together, these studies indicate that optimality constraints are used in lay decision theory and in causal judgment.

Keywords: Causal attribution; decision-making; theory of mind; responsibility; lay decision theory.

Introduction

Many of the decisions we make on a daily basis are thoroughly mediocre. This conclusion has been the joint product of the philosophical discipline of *Decision Theory*, which aims to characterize the decisions we ought to make given our knowledge and priorities (e.g., Jeffrey, 1965), and the psychological discipline of *Judgment and Decision-Making* (JDM), which aims to characterize actual decision-making behavior (e.g., Tversky & Kahneman, 1974). JDM research has shown that normative decision theory largely fails as a descriptive theory of human decision-making, documenting a plethora of ways in which our actual decision-making does not live up to normative standards.

Less is known, however, about how people conceptualize and evaluate the decisions of others—a research question one might term *lay decision theory*. The present research begins to examine this question, investigating how people assign causal responsibility to agents for the outcomes of their decisions.

In these experiments, we consider situations in which an agent made a decision that led to an outcome with probability P_{ACT} (always 50%), but could have made an alternative decision that would have led to that outcome with probability P_{ALT} (which was varied between 10% and 90% across conditions). For example:

Angie has a shrub, and wants the shrub's flowers to turn red. She is considering two brands of fertilizer to apply:

If she applies Formula PTY, there is a 50% chance that the flowers will turn red.

If she applies Formula NRW, there is a 10% chance that the flowers will turn red.

Angie chooses Formula PTY, and the flowers turn red.

Thus, Angie's actual choice of Formula PTY led to the outcome with probability P_{ACT} = 50%, and her alternative choice of Formula NRW led to the outcome with P_{ALT} = 10%. Because $P_{ACT} > P_{ALT}$, Angie's choice was optimal. However, if Formula NRW had led to the outcome with P_{ALT} = 90%, then $P_{ALT} > P_{ACT}$, and Angie's choice would have been suboptimal. Finally, if P_{ALT} had been 50%, then $P_{ACT} = P_{ALT}$, and there would have been no uniquely optimal decision.

There are at least three possible predictions one could make about how judgments of Angie's responsibility for the outcome would vary as a function of the counterfactual alternatives that Angie forsook.

First, according to the *Counterfactual Stability* view, counterfactual alternatives are irrelevant to computing causal responsibility in situations like Angie's. This seems plausible, since P_{ACT} is fixed across all conditions at 50%, and the likelihood of goal completion does not depend on the efficacy of the alternative. Moreover, it is stipulated in the vignette that the agent achieves her goal, eliminating uncertainty about the outcome. This result would be obtained if people are permissive of suboptimal decision-making when computing causal responsibility. In actual decision-making practice, after all, computational limitations prevent us from analyzing every possible course of action, so we often settle for an option that is satisfactory even if not optimal (Simon, 1956).

Second, according to the *Difference-Making* view, judgments of responsibility are a linear function of the difference made by the actual choice, relative to the alternative choice. On this view, responsibility judgments would be proportional to $[P_{ACT} - P_{ALT}]$, known as ΔP in the causal learning literature. On this view, one is most responsible for an outcome when the quality of the alternative options is very low, because the choice made a

large difference, while one is viewed as less responsible as the size of this gap decreases. For suboptimal choices (i.e., when $\Delta P < 0$), one's responsibility could further decrease (or be viewed as preventive) as the forsaken alternatives become increasingly efficacious and ΔP becomes increasingly negative. Though prior research on causal attribution (e.g., Cheng & Novick, 1992; Spellman, 1997) does not directly predict this result, this pattern would be most consistent with those previous findings.

Finally, according to the *Optimality* view, judgments of responsibility would be higher for optimal decisions than for suboptimal decisions, without consideration for *how much* better that decision is, compared to its alternatives. As we know humans to be satisficers and heuristic decision-makers, it may seem unlikely on the surface that we should require optimal behavior from others when assigning causal responsibility. However, theoretical considerations make this view seem less far-fetched.

Dennett (1987; see also Davidson, 1967) proposed that mental state inferences can often be accomplished by invoking a *well-formedness rule* called the *Principle of Rationality*. Just as we can solve the equation '$X + Y = Z$' if given the values of two of the three variables, so can we make inferences about agents' *actions*, *goals*, and *situational constraints* by using the principle that agents act *rationally* to satisfy their goals, given situational constraints. These inputs to the rationality 'formula' can be either states of the world (when reasoning *teleologically*), or mental states (i.e., beliefs, desires, and intentions, when reasoning *mentalistically*). In either case, the Principle of Rationality produces a unique prediction for one element given the other two, just as the facts of arithmetic yield a unique solution for '$2 + Y = 5$'. Actions conforming to the Principle of Rationality are optimal, relative to the agent's goals and situational constraints.

Previous research has shown that both adults and infants often make inferences afforded by the Principle of Rationality. In a series of experiments using a violation-of-expectation paradigm, Csibra et al. (1999) presented young infants with simple visual displays of geometric figures. The infants successfully used teleological constraints to predict these figures' actions from their goals and situational constraints. More recently, Baker, Saxe, and Tenenbaum (2009) developed a computational model to capture adults' inferences about goals from a display of the agent's movements in a simple maze.

The Principle of Rationality could lead people to discount the causal efficacy of suboptimal decision-makers in two ways. First, because actions are assumed to follow the Principle of Rationality, apparently suboptimal actions are often reinterpreted as optimal actions under different assumptions—for example, that the agent was acting under a different goal, or that the agent's beliefs were incomplete or erroneous (Baker et al., 2009; Buchsbaum et al., 2011). Although the action is optimal under such a reinterpretation, the assumptions made about the agent (such as ignorance) to rationalize the action may undermine the agent's perceived causal role in producing the outcome. Second, Csibra et al. (1999) have suggested that conformity to the Principle of Rationality is used as a principle for determining which entities are treated as agents, that is, as subject to folk-psychological principles. A decreased belief in the decision-maker's status as an agent could lead to a decreased attribution of causation. Indeed, such reasoning may not be restricted only to human agents. Kelemen and Rosset (2009) found that even adults apply teleological principles 'promiscuously' to inanimate objects. If people use efficiency cues to classify entities as agents, they may assign greater causal responsibility for objects that fulfill their causal affordances in the most efficient manner.

If the Optimality view is correct, responsibility judgments would be higher when $P_{ACT} > P_{ALT}$ (an optimal choice) than when $P_{ACT} < P_{ALT}$ (a suboptimal choice). However, the size of $[P_{ACT} - P_{ALT}]$ would not affect judgments, since optimality is a qualitative property of a choice and does not depend on the magnitude of this difference. This view does not make a specific prediction about judgments when $P_{ACT} = P_{ALT}$, because there is no uniquely optimal choice in such situations.

Three experiments distinguished among these possibilities. First, we examined whether an agent's perceived responsibility (Experiment 1A) or causal contribution (Experiment 1B) for the outcome of a decision depends on the efficacy of an alternative, forsaken option. Second, we replicated and extended these findings using situations in which agents decide among three options (Experiment 2A) or in which the base rate of the outcome is specified (Experiment 2B), ruling out an alternative interpretation of Experiment 1. Finally, we explored the possibility that optimality constraints are used in assessing causation more generally, even for inanimate causes (Experiment 3).

Experiments 1A and 1B

In Experiment 1, we examined whether the predictions of the Counterfactual Stability, Difference-Making, or Optimality view best capture judgments about vignettes such as those presented in the introduction. Additionally, to assess the consistency of these effects across measures, we included questions both about responsibility (in Experiment 1A) and about causation (in Experiment 1B).

Method

Participants Fifty participants (56% female) were recruited from Amazon Mechanical Turk to participate in Experiment 1A, and a different group of 50 participants (44% female) to participate in Experiment 1B.

Materials and Procedure Participants read five vignettes similar to the text given above, with five different cover stories. In these vignettes, P_{ACT} was fixed at 50%, while P_{ALT} was varied across cover stories (at 10%, 30%, 50%, 70%, and 90%) using a Latin square. The corresponding values of ΔP ($P_{ACT} - P_{ALT}$) are thus 40%, 20%, 0%,

−20%, and −40%, respectively. Participants were asked to rate their agreement with either a responsibility statement (e.g., "Angie is responsible for the flowers turning red") in Experiment 1A, or with a causal statement ("Angie caused the flowers to turn red") in Experiment 1B, on an 11-point scale (0: 'disagree'; 5: 'neither agree nor disagree'; 10: 'agree'). Manipulation check questions were included in this and all subsequent studies to monitor comprehension of the vignettes; however, these questions are not discussed further because no participants were eliminated from the analysis for these or any subsequent experiments.

Results

As shown in Table 1, judgments of responsibility and causation were higher when $P_{ACT} > P_{ALT}$ than when $P_{ACT} < P_{ALT}$, while judgments were intermediate when $P_{ACT} = P_{ALT}$. Yet, the magnitude of the difference between P_{ACT} and P_{ALT} had no effect on judgments beyond the direction of the difference, consistent with the Optimality view.

This pattern was confirmed with a mixed-model ANOVA on judgments, with P_{ALT} (10%, 30%, 50%, 70%, or 90%) as a within-subjects factor, and Experiment (responsibility question or causal question) as a between-subjects factor. This revealed a significant main effect of P_{ALT}, $F(4,392) = 13.97$, $MSE = 3.34$, $p < .001$, $\eta_p^2 = .13$, and a main effect of Experiment, $F(1,98) = 10.54$, $MSE = 15.41$, $p = .002$, $\eta_p^2 = .10$, with responsibility ratings higher overall than causal ratings ($M = 6.09$, $SD = 1.67$ vs. $M = 4.95$, $SD = 1.83$). This main effect may have occurred because the word 'cause' triggered a deterministic causal concept at odds with the probabilistic character of the decision problem. There was no interaction between P_{ALT} and Experiment, $F(4,392) = 0.87$, $MSE = 3.34$, $p = .48$, $\eta_p^2 < .01$, indicating no reliable difference in the effect of P_{ALT} between experiments.

To explore the main effect of P_{ALT}, pairwise planned comparisons were conducted on adjacent P_{ALT} conditions (means from the combined experiments are presented in the bottom row of Table 1). The 10% and 30% conditions did not differ, $t(99) = 1.09$, $SEM = 0.21$, $p = .28$, $d = 0.11$, nor did the 70% and 90% conditions, $t(99) = -0.21$, $SEM = 0.24$, $p = .84$, $d = -0.02$. However, the 50% condition differed significantly from both the 30% condition, $t(99) = -3.13$, $SEM = 0.25$, $p = .002$, $d = -0.31$, and the 70% condition, $t(99) = 2.22$, $SEM = 0.21$, $p = .029$, $d = 0.22$.

Table 1: Results of Experiment 1 (SDs in parentheses).

P_{ALT}	10%	30%	50%	70%	90%
Exp. 1A	6.74 (2.24)	6.60 (2.10)	5.82 (2.35)	5.68 (2.34)	5.60 (2.49)
Exp. 1B	5.96 (2.34)	5.64 (2.29)	4.84 (2.47)	4.06 (2.73)	4.24 (2.58)
Mean	6.35 (2.31)	6.12 (2.24)	5.33 (2.45)	4.87 (2.66)	4.92 (2.61)

Discussion

These results are most consistent with the Optimality view. Although there was an effect of P_{ALT} on responsibility ratings, this occurred only because judgments were dependent on the *sign* of ΔP: Judgments were higher when $P_{ACT} > P_{ALT}$ than when $P_{ACT} = P_{ALT}$, and higher when $P_{ACT} = P_{ALT}$ than when $P_{ACT} < P_{ALT}$, but the *magnitude* of ΔP did not affect judgments. This finding cannot be explained by either the Counterfactual Stability view, according to which responsibility judgments would be invariant over different values of P_{ALT}, or by the Difference-Making view, according to which responsibility judgments would be proportional to ΔP ($P_{ACT} - P_{ALT}$). Moreover, the lack of magnitude-dependence held for both attributions of responsibility and of causation, indicating that these results are not due to idiosyncratic properties of either phrasing.

Although the Counterfactual Stability and Difference-Making views cannot explain the results of Experiment 1, these results do not uniquely entail the Optimality view, because participants could have made these responses on the basis of whether $\Delta P > 0$. In more complex decision problems, it is possible for multiple options to have positive ΔP, yet for only one option to be uniquely optimal. Experiment 2 investigated whether people would still be sensitive to optimality in more complex scenarios.

Experiments 2A and 2B

To examine whether the response pattern in Experiment 1 was based on optimality or simply on whether $\Delta P > 0$, Experiment 2 used vignettes in which agents faced three choices, of which (*A*) had a low probability of leading to the goal, (*B*) had a moderate probability of leading to the goal, and (*C*) had the highest probability of leading to the goal. Thus, *C* is the optimal choice, but both *B* and *C* have positive ΔP relative to *A*. If judgments are based on optimality, then an agent choosing *C* should be rated more responsible than an agent choosing *B*, since *C* is optimal but *A* and *B* are not. However, if people are merely sensitive to ΔP being positive, they should rate agents choosing *B* and *C* equally highly, since $\Delta P > 0$ for both.

Experiment 2 employed two different framings of the "least optimal" alternative. In Experiment 2A, all three options were described as alternative choices with varying probabilities of success. In Experiment 2B, the "least optimal" alternative was described as a base rate—the probability of the goal occurring in the absence of any action. This second phrasing makes the fact that $\Delta P > 0$ for both of the other alternatives more salient, providing a stronger test against the Optimality hypothesis.

Method

Participants One hundred participants (52% female) were recruited from Amazon Mechanical Turk to participate in Experiment 2A, and a different group of 100 participants (49% female) participated in Experiment 2B.

Each experiment was conducted as part of a session that included additional experiments not reported here; the order of the experiments was counterbalanced.

Materials and Procedure In Experiment 2A, participants read two vignettes from Experiment 1, modified to read:

Angie has a shrub, and wants the shrub's flowers to turn red. She is thinking about applying a fertilizer, and has three options:

If she applies Formula LPN, there is a 10% chance that the flowers will turn red.

If she applies Formula PTY, there is a 50% chance that the flowers will turn red.

If she applies Formula NRW, there is a [30/70]% chance that the flowers will turn red.

Angie chooses Formula PTY, and the flowers turn red.

For Experiment 2B, the phrase "if she applies Formula LPN" was replaced by the phrase "if she applies nothing," to make the 10% base rate more salient. Whether Formula NRW had a 30% or 70% chance of leading to the goal (P_{ALT}) was manipulated within-subjects (in the former case, the actual choice was optimal, while in the latter case, the actual choice was suboptimal), with the assignment of P_{ALT} to vignette counterbalanced. Participants rated the agent's responsibility for the outcome on the same 11-point scale as Experiment 1A.

Results and Discussion

As shown in Figure 1, agents were viewed as less responsible when their choice was suboptimal, whether the "least optimal" option was described as an alternative (Experiment 2A) or as a base rate (Experiment 2B). This occurred even though ΔP was positive for both choices.

An ANOVA was conducted on responsibility judgments, with P_{ALT} (30% or 70%) as a within-subjects factor and Experiment (2A or 2B) as a between-subjects factor. There was a main effect of P_{ALT}, $F(1,198) = 18.57$, $MSE = 1.91$, $p < .001$, $\eta_p^2 = .09$, with responsibility rated higher when P_{ALT} = 30% ($M = 6.83$, $SD = 2.01$) than when P_{ALT} = 70% ($M = 6.24$, $SD = 2.14$). Thus, responsibility ratings were higher for optimal decisions.

There was also a main effect of Experiment, $F(1,198) = 12.57$, $MSE = 6.37$, $p < .001$, $\eta_p^2 = .06$, with judgments higher in Experiment 2B ($M = 6.98$, $SD = 1.53$) than in Experiment 2A ($M = 6.09$, $SD = 2.01$). This may have occurred because Experiment 2B described the least effective option as an omission rather than as an action, creating a qualitative difference among the options. However, there was no interaction between Experiment and P_{ALT}, $F(1,198) = 0.30$, $MSE = 1.91$, $p = .59$, $\eta_p^2 < .01$.

Although the Difference-Making account would make the same predictions as the Optimality account for these cases, the results of Experiment 2 show that the responses in Experiment 1 are unlikely to have been based merely on considerations of whether ΔP > 0, since this was the case for both the optimal and the suboptimal case in Experiment 2. The Optimality view is the only account that is consistent with the results of both experiments.

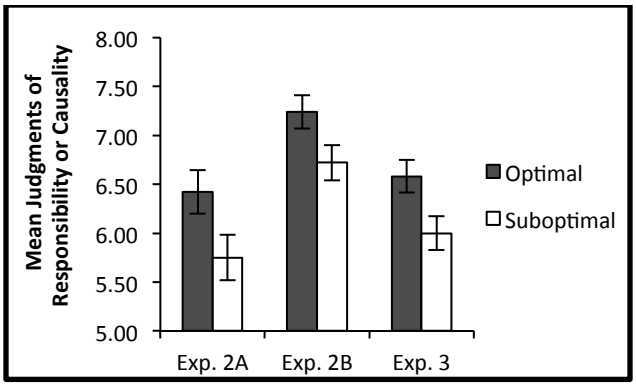

Figure 1: Mean responsibility judgments (Experiment 2) and causal judgments (Experiment 3) on 11-point scales.

Experiment 3

One potential explanation for the results of Experiments 1 and 2 is that participants were reinterpreting apparently suboptimal actions as guided by a different set of assumptions about the agent's knowledge or goals (e.g., Baker et al., 2009; Buchsbaum et al., 2011). Angie's objectively suboptimal decision to use Formula PTY may have led participants to view her as ignorant of the choice situation, or as having some other goal in mind other than making the flowers turn red, and participants may have accordingly downgraded her responsibility.

Another possibility, however, is that these results reflect principles used to designate entities as subject to our folk-psychological theorizing in the first place. Csibra et al. (1999) suggested that the Principle of Rationality is used for this purpose; indeed, efficiency may even be detected at the perceptual level (Gao & Scholl, 2011). If individuals failing to behave optimally are not conceptualized as agents to the same extent as those behaving optimally, this could lead people to discount their role in causally producing the outcome.

If our earlier effects were obtained at least in part because optimality is used as a principle for designating agents, optimality considerations might be used more generally in causal reasoning, outside the social realm, because people may reason about efficient causes as though they were endowed with agent-like properties. This prediction, while counterintuitive, is bolstered by findings of 'promiscuous' teleological reasoning (Kelemen & Rosset, 2009), with children and even adults under time pressure treating natural kinds as though they were artifacts endowed with purposes. Experiment 3 examined this prediction by testing whether people treat event types as more causal when they lead optimally to their effect, relative to other possible causes.

Method

Participants One hundred participants (44% female) were recruited from Amazon Mechanical Turk to participate in Experiment 3. This experiment was

conducted as part of a session that included an additional experiment not reported here; the order of the experiments was counterbalanced.

Materials and Procedure Participants read two vignettes adapted from Experiment 2B so that they no longer referred to a choice made by a human agent, but instead to the probability of an effect occurring given two different (non-human) causes:

There is a certain shrub that has flowers which sometimes turn red. There are two brands of fertilizer: Formula PTY and Formula NRW.

When nothing is applied, there is a 10% chance that the flowers turn red.

When Formula PTY is applied, there is a 50% chance that the flowers turn red.

When Formula NRW is applied, there is a [30/70]% chance that the flowers turn red.

Each participant was then asked to what extent they agreed with the statement that "Formula PTY causes the flowers to turn red" on the same 11-point scale used in previous experiments. In both conditions, participants judged the strength of a cause with 50% efficacy (i.e., P_{ACT} = 50%), while the efficacy of the other cause (P_{ALT} = 30% or P_{ALT} = 70%) was manipulated within-subjects and counterbalanced with vignette.

Results and Discussion

As shown in Figure 1, Formula PTY was judged more causal when it was optimal than when it was suboptimal, even though it always had a 50% chance of leading to the effect. A paired-sample *t*-test revealed that causal ratings were higher in the P_{ALT} = 30% condition (M = 6.58, SD= 1.67) than in the P_{ALT} = 70% condition (M = 6.00, SD = 1.72), $t(99) = 4.50$, $SEM = 0.13$, $p < .001$, $d = 0.45$.

This result suggests that the mechanism underlying the optimality effect in lay decision theory is not specific to human agents, but can be extended 'promiscuously' to other entities, with people discounting a cause's efficacy in the face of a superior alternative cause.

One concern about this result might be potential scale or contrast effects. For example, suppose that participants implicitly judge each cause in each vignette (including the alternative cause that was not asked about), and always give the 'best' cause for each vignette the same rating. Then, a participant in the optimal condition might have assigned the better (actual) cause a rating of '7' and implicitly assigned the worse (alternative) cause a rating of '5', and in the suboptimal condition implicitly assigned the better (alternative) cause a rating of '7' and the worse (actual) cause a rating of '5', leading to our effect. Similarly, a contrast effect could have occurred if the psychological weight of P_{ACT} differed between conditions. P_{ACT} (50%) could have felt like a larger magnitude when compared to P_{ALT} = 30% than to P_{ALT} = 70%, leading to higher ratings in the optimal condition.

Although these possibilities can only be ruled out definitively with future study, our within-subjects design renders these explanations unlikely. The vignettes were read and judged consecutively, which both calls attention to the consistency of P_{ACT} across conditions, and creates pressure to give identical responses across conditions. Nonetheless, converging evidence from other tasks will be of use in ruling out these possibilities more directly.

The present result should be distinguished from the superficially similar phenomena of discounting (e.g., Khemlani & Oppenheimer, 2011) and cue competition (e.g., Waldmann & Holyoak, 1992). *Discounting* occurs when one has a prior causal schema in which two causes (e.g., Formulas PTY and NRW) are each sufficient for an effect (the flowers turning red). If one cause (Formula NRW) is known to occur on some particular occasion, this makes the other cause (Formula PTY) less likely to be present on that occasion, because the known presence of Formula NRW "explains away" the effect and removes any reason to posit Formula PTY. Thus, the discounting phenomenon involves prior knowledge of causal types influencing subsequent inferences about causal tokens.

In the related phenomenon of *cue competition*, token-level observational data affect subsequent formation of type-causal schemas. For example, in *backward blocking*, two candidate causes are first paired with the effect (i.e., Formulas PTY and NRW are both applied for several trials on which the flowers turn red), then one of the candidates alone is paired with the effect (i.e., only Formula PTY is applied for several trials on which the flowers turn red). Observing that the alternative cue (Formula PTY) produces the effect by itself reduces the belief that Formula NRW causes the effect in general.

However, the logics underlying these phenomena do not apply to our experimental situation. The input to the discounting process is type-causal schemas, and the output token-causal inferences; the input to cue competition is token-causal observations, and the output type-causal schemas. In our task, in contrast, participants made type-causal judgments from knowledge about statistical relationships at the type level. Thus, the present phenomenon is conceptually distinct.

Although Experiment 3 suggests that the optimality effect in lay decision theory occurs at least in part because conformity to the Principle of Rationality is used to designate entities as subject to folk-psychological principles, this does not preclude the possibility that some participants in Experiments 1 and 2 were additionally re-interpreting the agents' actions as optimal under a different set of assumptions. Nonetheless, Experiment 3 shows that re-interpretation cannot be a full explanation. Indeed, the effect in Experiment 3, which cannot not be explained in terms of re-interpretation, was of similar magnitude to that in Experiment 2.

General Discussion

The present studies examined whether optimality is used as a cue for assigning causal responsibility. In Experiment 1, agents were judged more responsible for, and more

causal in, producing an outcome when their decision was the optimal choice for obtaining the outcome, but the magnitude of the difference between the efficacy of the optimal and suboptimal choices did not affect judgments. Experiment 2 showed that perceived responsibility is greater when a decision is optimal than when suboptimal, even when the suboptimal option is superior to a worst option or to the base rate of the outcome. Finally, Experiment 3 demonstrated an optimality effect in reasoning about causation for inanimate causes, suggesting that the optimality effect occurs at least in part because entities acting optimally are more likely to be designated as agents subject to our folk psychology.

Our results suggest several potentially promising avenues for future research. We are seldom confronted in real life with decisions for which we know the exact probabilities, more often entertaining a range of probabilities as potentially valid (Levi, 1985). A more ecologically valid test of our optimality hypothesis would specify realistic decision alternatives for which participants had a range of prior beliefs about the efficacy for achieving an outcome, rather than a single probability value. We chose to instead specify the probabilities so as to maximize experimental control. However, replicating the current results with more naturalistic stimuli would both enhance the generality of our findings and allow for exploration of boundary conditions.

Little appears to be known concerning folk beliefs about decision-making, what we term *lay decision theory*. In addition to shedding light on our theory of mind abilities, understanding the principles of lay decision theory may have practical implications for behavioral game theory, in which people must model others' behavior in order to make their own decisions. The present research addresses only a small fraction of the questions that might be asked: for example, how these beliefs are used in explaining and predicting behavior, whether (and when) people conceptualize decisions in terms of mental states or as states of the world (i.e., with mentalistic or teleological representations), how people conceptualize more complex decision problems in which multiple goals must be balanced against one another, and whether optimality constraints are applied equally to our own behavior as to the behavior of others. We are currently conducting research to probe these questions.

Conclusion

Rationality constraints are commonplace heuristics for making inferences about human actions. The present research shows that such constraints also play a role in the evaluation of decisions, affecting how causal responsibility is assigned for an outcome.

As suboptimal decision-makers ourselves, it may appear hypocritical for us to hold others less responsible for the outcomes of their decisions when they decide suboptimally. Yet, we may have little choice—our rationality is, after all, bounded.

Acknowledgments

This research was partially supported by funds awarded to the first author by the Yale University Department of Psychology. We thank Fabrizio Cariani, Winston Chang, Angie Johnston, Doug Medin, Emily Morson, Axel Mueller, Eyal Sagi, and Laurie Santos for their insightful suggestions, and the city of Nashville, TN for its inspirational, sunny November weather.

References

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.

Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition, 120*(3), 331–340.

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*(2), 365–382.

Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of 'pure reason' in infancy. *Cognition, 72*(3), 237–267.

Davidson, D. (1967). Truth and meaning. *Synthese, 17*(3), 304–323.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Gao, T., & Scholl, B. J. (2011). Chasing vs. stalking: Interrupting the perception of animacy. *Journal of Experimental Psychology: Human Perception and Performance, 37*(3), 669–684.

Jeffrey, R. C. (1965). *The logic of decision*. New York: McGraw-Hill.

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition*, *111*(1), 138–143.

Khemlani, S. S., Oppenheimer, D. M. (2011). When one model casts doubt on another: A levels-of-analysis approach to causal discounting. *Psychological Bulletin, 137*(2), 195–210.

Levi, I. (1985). Imprecision and indeterminacy in probability judgment. *Philosophy of Science, 52*(3), 390–409.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138.

Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General, 126*(4), 323–348.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121*(2), 222–236.