Cognitive Psychology 77 (2015) 42–76

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

# Do the right thing: The assumption of optimality in lay decision theory and causal judgment

Samuel G.B. Johnson [a,*], Lance J. Rips [b]

[a] Department of Psychology, Yale University, United States
[b] Department of Psychology, Northwestern University, United States

ARTICLE INFO

*Article history:*
Accepted 28 January 2015

*Keywords:*
Lay decision theory
Causal attribution
Rationality
Decision-making
Theory of mind
Behavioral game theory

ABSTRACT

Human decision-making is often characterized as irrational and suboptimal. Here we ask whether people nonetheless assume optimal choices from *other* decision-makers: Are people intuitive classical economists? In seven experiments, we show that an agent's perceived optimality in choice affects attributions of responsibility and causation for the outcomes of their actions. We use this paradigm to examine several issues in lay decision theory, including how responsibility judgments depend on the efficacy of the agent's actual and counterfactual choices (Experiments 1–3), individual differences in responsibility assignment strategies (Experiment 4), and how people conceptualize decisions involving trade-offs among multiple goals (Experiments 5–6). We also find similar results using everyday decision problems (Experiment 7). Taken together, these experiments show that attributions of responsibility depend not only on what decision-makers *do*, but also on the quality of the options they choose *not* to take.



## 1. Introduction

Psychologists, economists, and philosophers are united in their disagreements over the question of human rationality. Some psychologists focus on the fallibility of the heuristics we use and the systematic biases that result (Kahneman & Tversky, 1996), while others are impressed by the excellent

* Corresponding author at: 2 Hillhouse Ave., New Haven, CT 06520, United States.
*E-mail address:* samuel.johnson@yale.edu (S.G.B. Johnson).

http://dx.doi.org/10.1016/j.cogpsych.2015.01.003
0010-0285/

performance of heuristics in the right environment (Gigerenzer & Goldstein, 1996). Economists spar over the appropriateness of rationality assumptions in economic models, with favorable views among classically-oriented economists (Friedman, 1953) and unfavorable views among behavioral theorists (Simon, 1986). Meanwhile, philosophers studying decision theory struggle to characterize what kind of behavior *is* rational, given multifaceted priorities, indeterminate probabilities, and pervasive ignorance (Jeffrey, 1965).

Although decision scientists have debated sophisticated theories of rationality, less is known about people's lay theories of decision-making. Understanding how people predict and make sense of *others'* decision-making has both basic and applied value, just as research on lay theories of biology (e.g., Shtulman, 2006), psychiatry (e.g., Ahn, Proctor, & Flanagan, 2009), and personality (e.g., Haslam, Bastian, & Bissett, 2004) has led to both theoretical and practical progress. The study of lay decision theory can illuminate aspects of our social cognition and reveal the assumptions we make when interacting with others.

In this article, we argue that people use an *optimality* theory in thinking about others' behavior, and we show that this optimality assumption guides the attribution of causal responsibility. In the remainder of this introduction, we first describe game theory research on optimality assumptions, then lay out the connections to causal attribution research. Finally, we derive predictions for several competing theoretical views, and preview our empirical strategy.

### 1.1. Optimality assumptions in strategic interaction

Psychologists are well-versed in the evidence against human rationality (e.g., Shafir & LeBoeuf, 2002; the collected works of Kahneman and Tversky). Nonetheless, optimality assumptions have a venerable pedigree in economics (Friedman, 1953; Muth, 1961; Smith, 1982/1776), and are incorporated into some game-theoretic models. In fact, classical game theory assumes not only first-order optimality (i.e., behaving optimally relative to one's self-interest) but also second-order optimality (assuming that others will behave optimally relative to their own self-interest), third-order optimality (assuming that others will assume that others will behave optimally), and so on *ad infinitum* (von Neumann & Morgenstern, 1944). Understanding the nature of our assumptions about others' decision-making is thus a foundational issue in *behavioral game theory*—the empirical study of strategic interaction (Camerer, 2003; Colman, 2003).

Because people are neither infinitely wise nor infinitely selfish, rational self-interest models of economic behavior break down even in simple experimental settings (Camerer & Fehr, 2006). For example, in the *beauty contest game* (Ho, Camerer, & Weigelt, 1998; Moulin, 1986; Nagel, 1995), a group of players each picks a number between 0 and 100, with the player choosing the number closest to 2/3 of the average winning a fixed monetary payoff. The Nash Equilibrium for this game is that every player chooses 0 (i.e., only if every player chooses 0 is it the case that no player can benefit by changing strategy). If others played the game without any guidance from rationality, choosing randomly, then their mean choice would be 50, so the best response would be around 33. But if others followed that exact reasoning, then their average response would be 33, and the best response to 33 is about 22. Applying this same logic repeatedly leads us to the conclusion that the equilibrium guess should be 0. Yet average guesses are between 20 and 40, depending on the subject pool, with more analytic populations (such as Caltech undergraduates) tending to give lower guesses (Camerer, 2003). Which assumption or assumptions of classical game theory are being violated here? Are people miscalculating the equilibrium? Are they assuming that others will miscalculate, or assuming that others will assume miscalculations from others? Are they making a perspective-taking error, or assuming that others will make perspective-taking errors?

One approach toward answering such questions is to build an econometric model of each player's behavior, interpreting the parameter estimates as evidence concerning the players' underlying psychology (e.g., Camerer, Ho, & Chong, 2004; Stahl & Wilson, 1995). This approach has led to important advances, but the mathematical models often underdetermine the players' thinking, because a variety of mental representations and cognitive failures can often produce identical behavior. In this paper, we approach the problem of what assumptions people make about others' behavior using a different set of tools—those of experimental psychology.