
Rationality and Psychology in International Politics

Jonathan Mercer

Abstract The ubiquitous yet inaccurate belief in international relations scholarship that cognitive biases and emotion cause only mistakes distorts the field's understanding of the relationship between rationality and psychology in three ways. If psychology explains only mistakes (or deviations from rationality), then (1) rationality must be free of psychology; (2) psychological explanations require rational baselines; and (3) psychology cannot explain accurate judgments. This view of the relationship between rationality and psychology is coherent and logical, but wrong. Although undermining one of these three beliefs is sufficient to undermine the others, I address each belief—or myth—in turn. The point is not that psychological models should replace rational models, but that no single approach has a lock on understanding rationality. In some important contexts (such as in strategic choice) or when using certain concepts (such as trust, identity, justice, or reputation), an explicitly psychological approach to rationality may beat a rationalist one.

Rational choice theorists and political psychologists agree that psychology explains only deviations from rationality. The primary task of political scientists, according to Downs, is to study “rational political behavior, not psychology or the psychology of political behavior.”¹ According to this view, rational explanations should avoid psychology because psychology explains only mistakes. As Satz and Ferejohn note, “mental entities need not figure in the best rational-choice explanations of human action.”² Harsanyi suggests that when behavior appears to be irrational, then as a last resort analysts might turn to “specific assumptions about the psychological mechanisms underlying human behavior.”³ Yet even in this case, psychology can help only after the establishment of a rational baseline.⁴ Knowing irrational behavior depends on knowing rational behavior. For example, psychology can

I thank Deborah Avant, James Caporaso, James Davis, Bryan Jones, Margaret Levi, Peter Liberman, Lisa Martin, Susan Peterson, Jason Scheideman, Jack Snyder, Michael Taylor, two anonymous reviewers, and especially Robert Jervis and Elizabeth Kier for their thoughtful comments and critiques. Jason Scheideman also helped with research assistance.

1. Downs 1957, 9–10.
2. Satz and Ferejohn 1994, 76.
3. Harsanyi 1986, 83.
4. See Verba 1961; and Kahler 1998.

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009120685.003>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009123291.003>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009123291.003>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009123291.003>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009123291.003>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009123291.003>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009123291.003>

Psychology cannot explain accurate judgments. Although recognizing the first two beliefs are myths means psychology does more than explain mistakes, I address the “psychology of mistakes” myth by focusing on emotion. The last section discusses how emotion is necessary to both rationality and to basic international relations concepts such as trust, identity, and justice. For example, I use the emotion in trust to solve collective action problems. Rethinking the relationship between rationality and psychology should help rationalists overcome what one economist describes as an “aggressive uncuriosity” about psychology,⁸ and it should help political psychologists overcome their disinterest in accurate judgments.

The belief that psychology explains mistakes implies that rationality is free of psychology: because rational models cannot be built on an irrational (or psychological) foundation, each must be distinct from the other. I argue that rational models are not distinct from, but rather depend on, psychological assumptions. I use Elster's definition of rationality as using the best means to achieve a given end.⁹ I use accurate outcomes (or judgments) to mean obtaining a given end and mistaken outcomes to mean not obtaining a given end. I also distinguish "process" from "outcome," which makes it possible for nonrational (or nonnormative) decision-making processes to yield accurate judgments.¹⁰ In other words, just as a psychological process can result in either accurate or inaccurate judgments, a rational process can do so as well. Not only can actors make mistakes by failing to adhere to a rational process, but analysts can also make mistakes (or cause policymakers to make mistakes) by presenting as rational a process that actually undermines rationality.¹¹ Rational choice theorists and political psychologists can agree on Elster's

10. Funder 1987.

11. To avoid the terminological confusion that may result from the observation that psychological processes can be rational and that rational (choice) processes can be irrational, I adhere to the conventional view that models, processes, and baselines are rational to the extent that they conform to a rational choice (or “coherence”) standard of rationality. When necessary, I substitute “rationalist” for “rational choice” to distinguish it from “rational.”

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009100000.003>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009100000.003>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009100000.003>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009100000.003>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009100000.003>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009100000.003>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9781009100000.003>

Whereas rationalists emphasize how a rational process, *ceteris paribus*, produces accurate judgments, psychologists emphasize how nonnormative processes can result in accurate judgments. Thus the key difference between rationalists and psychologists is not over what they explain, but over how they explain it. Rationalists rely on deduction, statistics, and probability theory, whereas psychologists rely on induction and description of how the mind actually works. Neither focuses primarily on mistakes, though each emphasizes a different way of obtaining the same desired outcome.

Behaviorists and neoclassical economists attempted to eliminate the “mind” from causal explanations of human behavior. It was not that mental phenomena caused mistakes, but that it was impossible to know what role mental phenomena played. Behaviorists and economists rejected the metaphysical in favor of the observable and material. They aimed to overthrow “folk psychology” and replace it with science. Although behaviorists did not address rationality directly, their attempt to substitute observable stimuli for mental phenomena is central to contemporary rational choice theory. Methodological reasons also drove economists to eliminate mental phenomena from their explanations. Their apparent success, coupled with a coherence epistemology, led them to believe that psychology is useful only to explain deviations from rationality. These two traditions explain rational choice theorists’ belief that psychology explains mistakes.

17. See Hammond 1996; Funder 1987; Krueger 1998; Gigerenzer et al. 1999; and Rosen 2004.

Eliminating the mind from explanations of behavior required overthrowing folk psychology. The folk psychology of human action, also known as intuitive psychology, is the commonsense belief that one's desires and beliefs explain one's actions.¹⁸ For example, why did Bob go to the opera? Because he desires to hear an opera and believes he will hear one at the opera house. Using Bob's beliefs and desires inferred from his trip to the opera to explain that trip is teleological: the behavior explains the purpose for that behavior. If Bob desires to hear jazz, but goes to the opera, then analysts must either view him as irrational or "fix" Bob's desires and beliefs to make them consistent with his action. Because beliefs and desires constitute action and cause action, folk psychology is unfalsifiable. The theory has a few other problems. For example, it is indeterminate (without numerous *ceteris paribus* conditions) and it hinges on access to mental states that are difficult to know. These (and other) problems make folk psychology problematic and unscientific. Understandably, behaviorists, economists, and rational choice theorists sought to replace folk psychology with a causal theory of human action.

Behaviorism emerged in the early twentieth century as a response to the problematic and unscientific nature of folk psychology.¹⁹ Behaviorists study behavior, not the mind. People respond to incentives, and to understand incentives is to understand behavior. In what became the behaviorist manifesto, Watson declared: "Psychology as the behaviorist views it is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behavior. Introspection forms no essential part of its methods, nor is the scientific value of its data dependent on the readiness with which they lend themselves to interpretation in terms of consciousness."²⁰ Psychologists should study behavior, not mental processes. As Skinner observed: "We do not need to try to discover what personalities, states of mind, feelings, traits of character, plans, purposes, intentions, or other perquisites of autonomous man really are in order to get on with a scientific analysis of behavior."²¹ Analysts do not need to study personality because, as Skinner discovered, "Thinking is behaving. The mistake is in allocating the behavior to the mind."²² Behaviorists accepted that private mental states exist, but rejected their causal power.²³

Despite behaviorism's thirty-year dominance in psychology, it failed to liberate psychology from the mind.²⁴ One can teach a food-deprived pigeon a lot—behaviorists recently trained pigeons to discriminate between Monet and Picasso—

18. See Rosenberg 1988; Egan 1995; and Pinker 2002.

19. In contrast, behavioralism emerged in the 1950s from the efforts of political scientists to apply scientific methods to a range of issues, including cognition and affect. See Merkl 1969; and Dahl 1969.

20. Watson 1913, 158.

21. Skinner 1972, 12–13.

22. Skinner 1974, 104.

23. See Hayes et al. 1999, 163; and Moore 1995, 81.

24. For behaviorism's dominance, see Abelson 1994; and Hunt 1993.

Even if researchers control the environment and provide only one stimulus, how subjects respond to that stimulus depends not on its physical attributes but on the subjective understanding (or construal) of the stimulus.³¹ Researchers can reliably predict that a food-deprived chicken will respond to a lever that gives it food, or that a person will respond to a \$10 bill on the ground by picking it up, but prediction becomes unreliable in slightly more complex settings. How students respond to very low (but passing) grades differs dramatically: some students work harder, some blame the exam, and some pump their fist and say "Yes!" Despite behaviorists' attempts to rely only on behavior, they nonetheless relied on folk psychology. Behaviorists eliminated from their explanations neither desires (I want that corn pellet) nor beliefs (at the end of the third maze on the left is food). Without knowing desires and beliefs, one cannot know what "works" as an incentive.

31. For construal, see Aronson et al. 1994, 16. For the importance of subjective understanding, see Mischel 1999; Ross and Nisbett 1991; Ross 1990; and Jervis 1976.

As with their colleagues in psychology, economists coveted scientific progress, which meant developing a causal theory of human behavior. Although marginal utility theory began by expressing in mathematical terms Bentham's view that the source of all human behavior is avoiding pain and seeking pleasure, economists believed that measuring hedonic states—my pleasure in consumption is two and a half times your pleasure—was an illusion.³² Progress required abandoning the metaphysical and focusing on the material and observable. Switching to a simple ranking of preferences (ordinal utility) did not address the problem; economists continued to assume individual cardinal utility, which requires introspection to assign precise numbers to experienced utility. Samuelson's introduction of revealed preference theory appeared to solve the psychological problem.³³ Instead of ranking preferences to explain behavior, Samuelson used behavior to rank preferences. Like behaviorists, economists denied the causal power of the mind without denying the existence of the mind.

Inferring preferences from behavior resulted in empirical and theoretical problems. Empirically, economists have found so little support for the axiom of revealed preference that it is "empirically useless."³⁴ Theoretically, reliance on revealed preferences meant that economists abandoned causal theory. Revealed preferences explain action by inferring preferences from that action, just as folk psychology explains action by inferring desires and beliefs from that action.³⁵ Both are teleological arguments. The problems of revealed preferences encouraged economists to junk both behaviorism and any claim about the psychological plausibility of their assumptions.³⁶ Whereas behaviorists believed that relying on mental phenomena was a mistake, economists believed that mental phenomena caused mistakes. Because psychology explains deviations from rationality, and given the assumption that psychological biases are idiosyncratic rather than systematic, psychology becomes unimportant with a large enough sample. Markets are free of psychology even if individuals are not. Friedman's famous "as if" assumption gave economists license to stop worrying about psychology and learn to love aggregate data.³⁷ When explaining market behavior, psychology happily, perfectly, and conveniently washes out.³⁸ The emergence of behavioral (or psychological) economics reinforces the view that psychology explains mistakes. Behavioral economists attack neoclassical approaches for ignoring psychology, and typically reintroduce psychology to explain irrational behavior.³⁹ The market's imperfections reflect individuals' psychological limitations.

32. See Kahneman 2000; Lewin 1996, 1297–1308; and Rosenberg 1988, 66–68. Psychologists now study and measure hedonic notions of utility. See Mercer forthcoming.

33. Samuelson 1938.

34. Lewin 1996, 1316. See also Sen 1973, 243.

35. Rosenberg 1988, 72.

36. Lewin 1996, 1318.

37. Friedman 1953.

38. See Rabin 2002, 678–79; and Rosenberg 1988, 75.

39. See Shiller 2000; and Tirole 2002. For an exception, see Rabin and Thaler 2001.

Because rationalists reject the mind as causing behavior, the actor's environment carries the explanatory burden. Hardin advises that analysts should study the incentives that cause ethnic conflict, not the psychology of the participants.⁴⁶ Achen and Snidal note that rational deterrence theory never refers to mental calculations because decision makers need not calculate: "If they simply respond to incentives in certain natural ways, their behavior will be describable by utility functions."⁴⁷ Incentives explain the behavior of actors engaged in ethnic cleansing, the behavior of decision makers in a nuclear crisis, or even, as Satz and Ferejohn note, the behavior of pigeons in a Skinner box: "Studies have shown that pigeons do reasonably well in conforming to the axioms of rational-choice theory."⁴⁸ Whether rational choice explains pigeon behavior, or vice versa, is beside the point.⁴⁹ Even if people are pigeons and incentives explain behavior, analysts cannot know what counts as an incentive without reference to the beliefs and desires of the animal.

49. Operant conditioning provides many of the same basic predictions as microeconomics; see Alhadeff 1982.

Rational deterrence rests on the manipulation of incentives, which are properties of individuals, not environments. Thus what is rational for me to do depends on what inferences I believe you will make from my actions, and this cannot be specified in the abstract. How you respond to my threats or promises depends on your image of me and of yourself.⁵¹ Treating beliefs as exogenous can be sensible but does not solve the problem of inference: What will you infer from my behavior? Relying on Bayesian updating works well when the new information is objective and subject to one interpretation: I'll recognize a poker chip as red even if I was expecting to see a green chip and will update my belief about the probable distribution of red and green chips in a bag. However, in most cases that interest students of international politics, new information is ambiguous, subject to multiple interpretations, and interpreted through one's beliefs, expectations, and desires. As Jervis has emphasized, people do not simply revise their preexisting beliefs based on new evidence, but they use their preexisting beliefs to interpret that evidence.⁵² Unless one is in the lab predicting the color of poker chips in a bag, to make a rational choice one needs to understand beliefs, desires, and their influence on the interpretation of evidence.

50. Schnaitter 1999.

51. Davis 2000.

52. Jervis 2002, 307.

58. Funder 1992. For other critiques, see Gigerenzer 1991; Gigerenzer et al. 1999; Krueger 1998; Funder 1987; and Krueger and Funder forthcoming. See the response by Kahneman and Tversky 1996; and the open peer commentary to Todd and Gigerenzer 2000.

63. Jervis 1976, 119.

Psychology and Rational Baselines

For example, rational reputation models advise decision makers that reputations form predictably and systematically, and are worth defending.⁶⁶ Common sense, formal models, and logic tell rationalists how reputations form and thus how rational actors should behave. What is striking about the rational reputation literature is that it should even exist. Dictionaries define a reputation as a judgment about someone's character, yet character or disposition is psychology's home turf. Riker understood that it makes "no sense in terms of rational choice theory to attempt to attribute consistent character traits to actors."⁶⁷ Driving the cause of

67. Riker 1995, 37.

war and peace back into the personality of decision makers smacks of “mentalism,” which rationalists and behaviorists gave up long ago in the name of science. Behaviorists would say that food, not an internal state of hunger, causes a pigeon to peck a bar. Or, brittle glass breaks because it is hit by a stone, not because of an inherent quality of brittleness.⁶⁸ Rationalists praised behaviorists for liberating psychology from the mind and creating a true science.⁶⁹ In the spirit of the behaviorists, rationalists improved the concept of reputation by making it rational, which meant eliminating the psychology. Once purged of all things mental, rationalists can use reputation to address everything from market failure to collective action and commitment problems.⁷⁰ Insisting that a rational baseline precede a psychological explanation makes sense if that baseline first, is free of psychology, and second, provides the best means to a desired end. Rational reputation arguments fall short on both counts.

First, eliminating psychology from a psychological concept also eliminates its explanatory power. For example, some rationalists define reputation as a judgment about someone’s situation.⁷¹ In this case, a reputation is a function not of mental properties but of the costs and benefits of a situation. If a “mindless” cost-benefit calculus explains behavior, then the situation, not a reputation, explains behavior. “Situational reputation” is an oxymoron. Reputation’s explanatory power depends on explaining either why an actor behaves similarly in different situations or why different actors respond differently to the same situation. When situational incentives govern—such as when people flee a burning building—reputation is irrelevant. When situational incentives do not govern, or when an actor behaves contrary to compelling situational incentives—such as when someone stays inside a burning building—then an actor’s beliefs, desires, subjective calculations of costs and benefits, or reputation may be relevant.⁷² Draining psychology from a psychological concept guts its explanatory power.

Second, an analyst’s reliance on unacknowledged psychological assumptions can cause mistakes. Just as behaviorists used terms such as “drive” to replace ordinary terms such as “desire,” rationalists use elaborate definitions of reputation to evade psychology. For example, a reputation becomes a belief about the probability that another is a particular “type.” A “type” is a player’s private information about its expected utility for particular outcomes. Its expected utility is its payoff for a particular outcome, and observers infer probable payoffs from behavior.⁷³ In order for a revealed reputation argument to explain or predict behavior, payoffs must be

68. Moore 1999, 64–65. This behaviorist explanation works only if one takes the characteristics of glass for granted; a rock dropped on a sheet of steel has a different effect. Jervis, personal communication.

69. For example, Riker 1962, 7.

70. See respectively, Keohane 1984; Ostrom 1998; and Fearon 1995.

71. See Sartori 2002, 136; and Downs and Jones 2002, 101.

72. Jervis 1976.

73. Morrow 1994, 55, 219, 242.

The two observations above—that psychological concepts drained of psychology lose explanatory punch and that neglect of psychological assumptions can result in mistakes—reinforce a final point: no reason exists to privilege rationalist over psychological explanations. A correspondence approach and a coherence approach to rationality can yield competing means to the same desired end, and no principled reason exists to prefer one over the other. For example, Mercer argues that because reputations for resolve form rarely, rational actors should not fight wars over reputation.⁷⁸ Mercer uses attribution theory to understand when people will use dispositional explanations to account for someone's behavior. The only way to know whether reputations should form, or to understand when reputations do form, is by understanding how people think. In the case of reputation formation, or more generally in problems of strategy (such as finance or nuclear deterrence), knowing how people tend to reason and tend to behave is crucial to making accurate judgments.⁷⁹ Predicting reputations is different than predicting the distribution of heads or tails in a coin flip. Coins are mindless; people (generally) are not. The point is not that rational reputation arguments must be wrong, but that psychological expla-

79. For psychological approaches to finance, see Shleifer 2000; for deterrence, see Jervis 1984, and Lebow 1987.

Even on grounds of parsimony and generalizability, no reason exists to privilege rationalist explanations over psychological ones.⁸⁰ Rabin points out that psychological approaches can be “simpler, more tractable, and more useful (less post hoc) explanations than existing models. . . . [T]here is no inherent negative correlation between psychological realism on the one hand, and taste for tractability, formalism, parsimony, and simplification on the other hand.”⁸¹ For example, loss aversion and mental accounting may serve as a simpler and superior model to expected utility theory, and a simple attribution theory may beat complex rational reputation models.⁸² Parsimony and generalizability are attributes of theories, not of disciplines. The belief that rationalist models necessarily provide the best means to a given end, that they are free of psychology, and so should precede psychological explanations, is wrong. Coherence models can be powerful and revealing and helpful, but so too can correspondence models, and the use of one does not depend on the existence of the other.

The belief that psychology explains mistakes implies that psychology cannot explain accurate judgments. Scholars typically view cognitive biases and emotion as undermining rationality, which is why rationality is thought to be free of psychology and why psychology is thought to require a rational baseline. By eliminating the mind, analysts can determine what constitutes rational behavior. This view is mistaken: rational models typically rely on psychological assumptions; rational baselines need not precede psychological explanations; and, as I argue in this section, nonnormative (or nonrationalist) decision-making processes can provide the best means to a given end. I use emotion to demonstrate the value of a correspondence approach to rationality.

80. See Verba 1961; and Kahler 1998.

81. Rabin 2002, 673–74.

82. For the logical but absurd consequences of adhering to expected utility theory, see Rabin 2000. For further discussion and an alternative, see Rabin and Thaler 2001; and Camerer and Thaler 2003.

83. For exceptions, see McDermott *forthcoming*; Marcus 2003; and Crawford 2000.

84. Skinner 1976, 92.

Damasio's research on the brain is not a parlor trick. People without emotion may know they should be ethical, and may know they should be influenced by norms, and may know that they should not make disastrous financial decisions, but this knowledge is abstract and inert and does not weigh on their decisions. They do not care about themselves or about others, and they neither try to avoid making mistakes nor are they capable of learning from their mistakes. They "know"

90. Ibid., 193.

If rationality depends on emotion, then correspondence models of rationality may have an advantage over coherence models. In this case, idealized (and emotionless) models of decision making may be a source of mistakes. For example, the rational choice of defection in a PD game, or in a collective action problem, is normatively self-defeating (because it reliably fails to maximize expected utility) and positively wrong (because cooperation is common). Since the mid-1970s, experimentation on PD and strategically equivalent games (such as commons dilemmas) has invariably found what one researcher calls “rampant cooperation.”⁹² People who irrationally cooperate in PD commonly do better than those who rationally defect. In a variety of games of strategic interaction, the rational choice is typically the worst choice.⁹³ Although neurologists and neuroeconomists suggest that humans may be hardwired to cooperate, I focus on how emotion solves problems of strategic interaction.⁹⁴

91. Ibid., 45.

92. Colman 2003, 147.

93. See Colman's discussion of PD, and of Matching, Centipede, and Ultimatum games. Colman 2003.

94. For the neuroscience of cooperation, see Rilling et al. 2002. For a review of evolutionary models of cooperation and reputation, see Fehr and Fischbacher 2003. For neuroscience and political psychology, see Cacioppo and Visser 2003. For the use of neuroscience to examine decision making and war termination, see Rosen 2004.

95. Dawes 1980.

96. See Dawes 1980; and Taylor 1987.

97. Taylor 1987.

Rationalists do to trust what they do to reputation.¹⁰³ The belief that psychology explains mistakes, and thus the fear of “going mental,” leads rationalists to drain psychology from psychological concepts. As the philosopher Becker noted rational accounts of trust appear to eliminate that which they claim to describe.¹⁰⁴ Rationalists drain the psychology from trust by turning it into a consequence of incentives. Emphasizing incentives as the basis for trust eliminates both the need for trust and the opportunity to trust. If trust depends on external evidence, transparency, iteration, or incentives, then trust adds nothing to the explanation.¹⁰⁵ Equally importantly, emphasizing incentives as the basis for trust eliminates the opportunity for trust to form. Trust is a feeling toward someone, not toward an inanimate object or toward a situation. I trust my friends but rely on my car.¹⁰⁶ If observers attribute cooperation to the environment rather than the person, then trust cannot—and need not—develop. My argument is not that incentives are unimportant—though Chomsky’s observation that incentives depend on beliefs and desires merits repeating—but that incentive-based behavior is not a substitute for trust-based behavior.

98. See Jones 1996; Becker 1996. See also Tyler 1998; Tyler and Degoeey 1996; and Uslander 1998.

100. Oatley 2000, 90.

102. For feelings as evidence, see Clore and Gasper 2000, 25.

104. Becker 1996, 47.

105. See Williamson 1994; and Granovetter 1985.

106. Jones 1996, 14.

leave their group, they do not.¹¹⁷ The emotion in identity makes trust available to solve collective action problems.

Because identity demands discrimination, relying on identity and trust to solve collective action problems is a double-edged sword. Tajfel singled out the esteem emotions as the driving force behind competition. It is the feeling of self-esteem, or pride, that makes people in groups view their group as different and better than other groups. In general, the more positively one feels about one's group, the more negatively one feels about rivalrous groups.¹¹⁸ In-group trust does not require out-group distrust—which is a feeling of pessimism about another's goodwill and competence—but it does require one to distinguish between trusting one's group and not trusting an out-group.¹¹⁹ Whether discrimination is a consequence of animus or an unintended consequence of preference for one's group, the result for the out-group is similar. After slaughtering 12,000 heathen in a day, Joshua gave thanks to God by carving the commandment, "Thou shall not kill." Joshua was not being hypocritical, for the prohibition applied only to the in-group, just as Moses preached to "love thy neighbor," as long as the neighbor was a member of the tribe.¹²⁰ Preferential treatment of one's own group is a source of peace and cooperation within the group, and often of distrust and hostility between groups.¹²¹

Emotion drives in-group cooperation and out-group discrimination. If this perspective is correct, then an emphasis on material incentives, abstract cognitive beliefs, or incomplete information is mistaken.¹²² The emotion in identity explains why group members may trust each other and why they may distrust out-group members. For example, should analysts view alliances between states as instrumental, or should they sometimes view them as reflecting a security community of shared values, beliefs, and trust? Both realists and constructivists eliminate emotion from their assessments. Constructivists, for example, view Deutsch's "we feeling" (which provides the basis for a security community) as indicating purely cognitive beliefs free of affect.¹²³ Because political scientists typically view emotion as irrational, it is not surprising that constructivists, no less than rationalists, are eager to purge it from their explanations.¹²⁴ Instead of running from emotion, recognizing emotion's role in trust and in identity may help analysts better understand how alliances might work and how security communities might form. Emo-

117. Orbell et al. 1988. For emotion's influence on out-group discrimination in minimal group settings, see DeSteno et al. 2004.

118. Reynolds et al. 2000.

119. For distrust, see Jones 1996. For trust and out-groups, see Brewer 1999; and Brewer and Brown 1998, 575.

120. See Hartung 1995; Ridley 1997; and Wilson 2002.

121. Brewer 1997, 201.

122. Stuart Kaufman reaches similar conclusions in his study of ethnic conflict. Kaufman 2001.

123. See Adler 1997, 255, 263; and Wendt 1999, 319.

124. Exceptions include Kier 2004; Finnemore 2003; Crawford 2000; and Finnemore and Sikkink 1998.

A correspondence approach to rationality ought to be better than a coherence approach when the explanation for a phenomenon rests on psychological concepts such as trust, identity, or reputation. For example, “justice” depends on emotion. As Solomon argued, “The idea that justice requires emotional detachment, a kind of purity suited ultimately only to angels, ideal observers, and the original founders of society has blinded us to the fact that justice arises from and requires such feelings as resentment, jealousy, and envy as well as empathy and compassion.”¹²⁵ If emotion subverts rationality, then any conception of justice that depends on emotion must capture irrational behavior. Rationalists drain emotion from justice and assume self-interested actors that have preferences over outcomes rather than preferences over strategies. Rational people care only about the realization of their preferences and are indifferent to how an outcome is obtained; they care about their final payoffs but not the payoffs of others.¹²⁶ Yet preferences over process appear to be just as important. How a decision is made, what the other’s motivation appears to be, and whether one is treated with respect, consideration, and fairness often matter more than the outcome.¹²⁷ Although rationalists can use non-standard hedonic values in their utility functions, doing so risks creating unfalsifiable arguments and, even worse, acknowledging a reliance on psychology. While political psychologists recognize that justice depends on emotion, this dependence leads them to view a concern for justice as a deviation from rationality.¹²⁸ It is as if a rational person must be greedy but indifferent to justice.

125. Solomon 1990, 34.

127. See Tyler and Degeoy 1996; Rabin 2002; Camerer and Thaler 2003; and Kier 2004. See also, for experimental work on the Ultimatum game, Colman 2003; and Henrich et al. 2001.

129. Philosophers in the early seventeenth century began to view greed as so powerful that, when harnessed with reason, it might tame the other passions. Thus began the semantic shift from the emotion "greed" into emotionless (and virtuous) "interest." Hirschman 1977.

Elster suggests that because emotion undermines rationality, the only problems emotion solves are those it causes in the first place.¹³⁰ Emotion can cause problems (such as out-group discrimination), but emotion can also solve problems that rational models cause (such as the rational strategy of defection in PD). The concepts “identity” and “trust” have causal power only when analysts recognize their dependence on emotion. Identification without feeling implies a cold, neutral, bloodless observation; it inspires no action. Trust without emotion implies an expectation of trustworthiness based on incentives; trust adds nothing if incentives explain cooperation. Emotion makes identity, trust, and solutions to collective action problems possible. Political scientists should view emotion not only as a problem to be overcome, but also as a solution that can help individuals and groups overcome problems. Damasio has shown that rationality depends on emotion, and I have discussed how emotion may solve collective action problems and may contribute to understanding alliance formation and justice. Psychology does more than explain mistakes. It also helps prevent them.

Conclusion

Apprehending rationality is hard, and it is made harder by the belief that psychology explains only mistakes. This belief results in three myths: that rational explanations are free of psychology, that psychological explanations demand a rational baseline, and that psychology cannot explain accurate judgments. To the contrary, rational models usually rest on psychological assumptions, a correspondence model does not depend on the existence of a coherence model, and psychology can explain accurate judgments. For example, psychological economists use loss aversion and mental accounting to prevent mistakes that expected utility theory causes, neuroscientists use emotion to explain rational behavior, political psychologists use attribution theory to explain reputation formation, and I use emotion in trust and identity to solve collective action problems. The point is not that correspondence models should replace coherence models, but that no single approach has a lock on understanding rationality. In some important contexts (such as in strategic choice) or when using certain concepts (such as reputation, trust, identity, or justice), a correspondence approach to rationality may beat a coherence approach.

Rejecting the belief that psychology explains only mistakes should help rationalists overcome their fear of “going mental” and thus encourage them to pay attention to the psychological assumptions that often drive their explanations. Rejecting the belief should also encourage political psychologists to stop conceding rationality to the rationalists. Political psychology is—or at least should be—as much about accurate judgments as inaccurate ones. Finally, rejecting these beliefs

130. Elster 1999, 291.

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/S0000013718000000>

Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/S0000013718000000>

- Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/S0000013718000000>

- Clare, Gerald L., and Karen Gasper. 2000. Feeling Is Believing: Some Affective Influences on Belief. In *Emotions and Beliefs: How Feelings Influence Thoughts*, edited by Nico H. Frijda, Antony S. R. Manstead, and Sacha Bem, 10–44. Cambridge: Cambridge University Press.
- Colman, Andrew M. 2003. Cooperation, Psychological Game Theory, and Limitations of Rationality in Social Interaction. *Behavioral and Brain Sciences* 26 (2):139–53.
- Crawford, Neta C. 2000. The Passion of World Politics: Propositions on Emotion and Emotional Relationships. *International Security* 24 (4):116–56.
- Dahl, Robert A. 1969. The Behavioral Approach in Political Science: Epitaph for a Monument to a Successful Protest. In *Behavioralism in Political Science*, edited by Heinz Eulau, 68–92. New York: Atherton Press.
- Damasio, Antonio R. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam.
- Davis, James W. 2000. *Threats and Promises: The Pursuit of International Influence*. Baltimore, Md.: Johns Hopkins University Press.
- Dawes, Robyn M. 1980. Social Dilemmas. *Annual Review of Psychology* 31:169–93.
- . 1998. Behavioral Decision Making and Judgment. In *The Handbook of Social Psychology*, 4th ed., Vol. 1., edited by Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey, 497–548. Boston: McGraw-Hill.
- Dawes, Robyn M., and Richard H. Thaler. 1988. Anomalies: Cooperation. *Journal of Economic Perspectives* 2 (3):187–97.
- DeSteno, David, Nilanjana Dasgupta, Monica Y. Bartlett, and Aida Caidric. 2004. Prejudice from Thin Air: The Effect of Emotion on Automatic Intergroup Attitudes. *Psychological Science* 15 (5):319–24.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper & Row.
- Downs, George W., and Michael A. Jones. 2002. Reputation, Compliance, and International Law. *Journal of Legal Studies* 31 (1):S95–S114.
- Egan, Frances. 1995. Folk Psychology and Cognitive Architecture. *Philosophy of Science* 62 (2):179–96.
- Elster, Jon. 1986. Introduction. In *Rational Choice*, edited by Jon Elster, 1–33. New York: New York University Press.
- . 1989. *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.
- . 1999. *Alchemistries of the Mind: Rationality and the Emotions*. Cambridge: Cambridge University Press.
- Fearon, James D. 1995. Rationalist Explanations for War. *International Organization* 49 (3):379–414.
- Fehr, Ernst, and Urs Fischbacher. 2003. The Nature of Human Altruism. *Nature* 425 (23):785–91.
- Finnemore, Martha. 2003. *The Purpose of Intervention: Changing Beliefs About the Use of Force*. Ithaca, N.Y.: Cornell University Press.
- Finnemore, Martha, and Kathryn Sikkink. 1998. International Norm Dynamics and Political Change. *International Organization* 52 (4):887–917.
- Frank, Robert H. 1988. *Passions Within Reason: The Strategic Role of the Emotions*. New York: Norton.
- Frieden, Jeffrey A. 1999. Actors and Preferences in International Relations. *Strategic Choice and International Relations*, edited by David A. Lake and Robert Powell, 39–76. Princeton, N.J.: Princeton University Press.
- Friedman, Milton. 1953. The Methodology of Positive Economics. In *Essays in Positive Economics*, 3–43. Chicago: University of Chicago Press.
- Frijda, Nico H., Antony S. R. Manstead, and Sacha Bem. 2000. The Influence of Emotions on Beliefs. In *Emotions and Beliefs: How Feelings Influence Thoughts*, edited by Nico H. Frijda, Antony S. R. Manstead, and Sacha Bem, 1–9. Cambridge: Cambridge University Press.
- Frijda, Nico H., and Batja Mesquita. 2000. Beliefs Through Emotions. In *Emotions and Beliefs: How Feelings Influence Thoughts*, edited by Nico H. Frijda, Antony S. R. Manstead, and Sacha Bem, 45–77. Cambridge: Cambridge University Press.
- Funder, David C. 1987. Errors and Mistakes: Evaluating the Accuracy of Social Judgment. *Psychological Bulletin* 101 (1):75–90.
- . 1992. Everything You Know Is Wrong. *Contemporary Psychology* 37 (4):319–20.

- . 2001. Personality. *Annual Review of Psychology* 52:197–221.
- George, Alexander L., and Juliette L. George. 1964. *Woodrow Wilson and Colonel House: A Personality Study*. New York: Dover.
- Gigerenzer, Gerd. 1991. From Tools to Theories: A Heuristic of Discovery in Cognitive Psychology. *Psychological Review* 98 (2):254–67.
- Gigerenzer, Gerd, Peter M. Todd, and the ABC Research Group. 1999. *Simple Heuristics that Make Us Smart*. New York: Oxford University Press.
- Glimcher, Paul W. 2003. *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*. Cambridge, Mass.: MIT Press.
- Goemans, Hein E. 2000. *War and Punishment: The Cause of War Termination and the First World War*. Princeton, N.J.: Princeton University Press.
- Goldgeier, James M. 1994. *Leadership Style and Soviet Foreign Policy: Stalin, Khrushchev, Brezhnev, Gorbachev*. Baltimore, Md.: Johns Hopkins University Press.
- Goldgeier, James M., and Philip E. Tetlock. 2001. Psychology and International Relations Theory. *Annual Review of Political Science* 4:67–92.
- Granovetter, Mark. 1985. Economic Action and Social Structures: The Problem of Embeddedness. *American Journal of Sociology* 91 (3):481–510.
- Gries, Peter Hays. 2004. *China's New Nationalism: Pride, Politics, and Diplomacy*. Berkeley: University of California Press.
- Hammond, Kenneth R. 1996. *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. New York: Oxford University Press.
- Hardin, Russell. 1995. *One for All: The Logic of Group Conflict*. Princeton, N.J.: Princeton University Press.
- Harsanyi, John C. 1986. Advances in Understanding Rational Behavior. In *Rational Choice*, edited by Jon Elster, 82–107. New York: New York University Press.
- Hartung, John. 1995. Love Thy Neighbour: The Evolution of In-Group Morality. *Skeptic* 3 (4):86–99.
- Hayes, Steven C., Kelly G. Wilson, and Elizabeth V. Gifford. 1999. Consciousness and Private Events. In *The Philosophical Legacy of Behaviorism*, edited by Bruce A. Thyer, 153–87. Boston: Kluwer Academic.
- Heider, Fritz. 1958. *The Psychology of Interpersonal Relations*. London: Lawrence Erlbaum.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. *American Economic Review* 91 (2):73–78.
- Hirschman, Albert O. 1977. *The Passions and the Interests: Political Arguments for Capitalism Before Its Triumph*. Princeton, N.J.: Princeton University Press.
- Hirshleifer, Jack. 1993. The Affections and the Passions: Their Economic Logic. *Rationality and Society* 5 (2):185–202.
- Hogg, Michael A., and Dominic Abrams. 1988. *Social Identification: A Social Psychology of Inter-group Relations and Group Processes*. New York: Routledge and Kegan Paul.
- Hopf, Ted. 1994. *Peripheral Visions: Deterrence Theory and American Foreign Policy in the Third World, 1965–1990*. Ann Arbor: University of Michigan Press.
- Houghton, David P. 2001. *U.S. Foreign Policy and the Iran Hostage Crisis*. Cambridge: Cambridge University Press.
- Hunt, Morton M. 1993. *The Story of Psychology*. New York: Doubleday.
- Innis, Nancy K. 1999. Edward C. Tolman's Purposive Behaviorism. In *Handbook of Behaviorism*, edited by William O'Donohue and Richard Kitchener, 97–117. New York: Academic Press.
- Janis, Irving L. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascos*. Boston: Houghton Mifflin.
- Jervis, Robert. [1970] 1989. *The Logic of Images in International Relations*. Reprint, with a new preface. New York: Columbia University Press.
- . 1976. *Perception and Misperception in International Politics*. Princeton, N.J.: Princeton University Press.

- Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9780521875886.008>

- Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/S0000000000000000>

- Downloaded from <https://www.cambridge.org/core>. Dartmouth College, on 02 Jan 2018 at 01:29:14, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/9780521875885.008>

- Tyler, Tom R., and Robyn M. Dawes. 1993. Fairness in Groups: Comparing the Self-Interest and Social Identity Perspectives. In *Psychological Perspectives on Justice: Theory and Applications*, edited by Barbara A. Mellers and Jonathan Baron, 87–108. Cambridge: Cambridge University Press.
- Tyler, Tom R., and Peter Degoe. 1996. Trust in Organizational Authorities: The Influences of Motive Attributions on Willingness to Accept Decisions. In *Trust in Organizations: Frontiers of Theory and Research*, edited by Roderick M. Kramer and Tom R. Tyler, 331–56. Thousand Oaks, Calif.: Sage Publications.
- Tyler, Tom R., and Roderick M. Kramer. 1996. Whither Trust? In *Trust in Organizations: Frontiers of Theory and Research*, edited by Roderick M. Kramer and Tom R. Tyler, 1–15. Thousand Oaks, Calif.: Sage Publications.
- Uslaner, Eric M. 1998. Social Capital, Television, and the 'Mean World': Trust, Optimism, and Civic Participation. *Political Psychology* 19 (3):441–67.
- Verba, Sidney. 1961. Assumptions of Rationality and Non-Rationality in Models of the International System. *World Politics* 14 (1):93–117.
- Watanabe, Shigeru, Junko Sakamoto, and Masumi Wakita. 1995. Pigeons' Discrimination of Paintings by Monet and Picasso. *Journal of the Experimental Analysis of Behavior* 63 (2):165–74.
- Watson, John B. 1913. Psychology as the Behaviorist Views It. *Psychological Review* 20 (2):158–77.
- Welch, David A. 1993. *Justice and the Genesis of War*. Cambridge: Cambridge University Press.
- Wendt, Alexander. 1999. *Social Theory of International Politics*. Cambridge: Cambridge University Press.
- Williamson, Oliver E. 1994. Transaction Cost Economics and Organizational Theory. In *The Handbook of Economic Sociology*, edited by Neil J. Smelser and Richard Swedberg, 77–107. Princeton, N.J.: Princeton University Press.
- Wilson, David S. 2002. *Darwin's Cathedral: Evolution, Religion, and the Nature of Society*. Chicago: University of Chicago Press.