

# PSYCHOLOGY AND INTERNATIONAL RELATIONS THEORY

---

J. M. Goldgeier<sup>1</sup> and P. E. Tetlock<sup>2</sup>

<sup>1</sup>*Department of Political Science, George Washington University, 2201 G. Street NW, Washington, DC 20052; e-mail: jimg@gwu.edu;*

<sup>2</sup>*Departments of Psychology and Political Science, Ohio State University, 142 Townshend Hall, 1885 Neil Avenue, Columbus, Ohio 43210; e-mail: tetlock.1@osu.edu*

**Key Words** error and bias, tradeoff reasoning, prospect theory, accountability pressures, internalization

■ **Abstract** Organized around several major theoretical traditions in international relations, this essay suggests which literature in psychology should be of greatest interest to different kinds of international relations scholars. New work in cognitive social psychology and behavioral decision theory simultaneously expands on and qualifies earlier error-and-bias portraits of the foreign policy maker, thereby enriching our understanding of internal divisions within the realist camp. Work on bounded rationality in competitive markets and mixed-motive games, as well as the literature on the power of human emotions to shape judgments of what represents an equitable allocation of scarce resources or a just resolution of conflicts of interest, can inform neo-institutionalist and constructivist theories. Developments in cross-cultural social psychology shed light on constructivist arguments about the creation and maintenance of international social order that typically rest on assumptions about decision making that are qualitatively different from realist and institutionalist approaches to world politics.

## INTRODUCTION

At first glance, the central macro-level theories of international politics appear to rely on minimal assumptions about human cognition and motivation. For realists, states are power or security maximizers; for liberals and neoliberal institutionalists, they are wealth or utility maximizers. For constructivists, human nature is itself a social construction and the appropriate focus is on the intricate webs of normative understanding that shape and are shaped by international actors. Although most scholars in these traditions believe that their models have little or no need to rely on psychological models of individual and group behavior, we argue in this review that they are wrong for a host of interrelated reasons.

They are wrong partly because when we scrutinize what these traditions trumpet as their most distinctive explanatory achievements, we discover that their capacity to explain relevant trends or events hinges on a wider range of implicit psychological assumptions that it is useful to make explicit. In this sense, these macro theorists are already more psychological than they think. And when we shift attention to each tradition's explanatory shortcomings, we believe these can be at least partly corrected by incorporating other psychological assumptions into the conceptual frameworks. In this sense, these macro theories are not as psychological as they should be. We also argue that one of the major benefits of weaving psychological analysis into the fabric of international relations theorizing will be the gradual unveiling of boundary conditions on the applicability of competing frameworks.

The goal of this essay is to demonstrate how developments in psychology can inform what international relations (IR) theorists generally refer to as second-level and third-level arguments about interstate politics (Waltz 1959). We are not arguing that all important regularities in world politics are reducible to psychological laws. Indeed, we suspect that, strictly speaking, none are so reducible. Our guiding philosophy of social science is contextualist, not reductionist. Psychological arguments acquire explanatory force only when they are systematically assimilated into political frameworks that take into account the structural, economic, and cultural conditions within which policy makers work.

This essay considers several major theoretical traditions in IR and suggests which literature in psychology should be of greatest interest to different kinds of IR scholars. Because of space constraints, we focus chiefly on neorealism, neoliberal institutionalism, and constructivism, and within each, we emphasize those aspects of the frameworks that link up in particularly compelling ways with the social psychological literature.

Following the path-breaking work of Jervis (1976) on cognitive constraints on rational decision making within a realist framework, we identify work in cognitive social psychology and behavioral decision theory that simultaneously expands on and qualifies Jervis's error-and-bias portrait of the foreign policy maker and thereby enriches our understanding of internal divisions within the realist camp.

Following developments in behavioral economics, we argue that neoliberal institutionalist and constructivist theories could draw much more effectively than they do from work on bounded rationality in competitive markets and mixed-motive games (Simon 1957, 1982; Kahneman & Tversky 1979, 1984). Also important is the literature on the power of human emotions—often tightly coupled to ideology as well as conceptions of fairness and procedural justice—to shape judgments of what represents an equitable allocation of scarce resources or a just resolution of conflicts of interest.

Following developments in cross-cultural social psychology, we contend that constructivist arguments about the creation and maintenance of international social order typically rest on assumptions about decision making that are qualitatively different from realist and institutionalist approaches to world politics. Experimental

work on tradeoff reasoning sheds light on why some decision makers may view a choice as immoral whereas others are comfortable rationally weighing costs and benefits (e.g. considering the use of taboo biological, chemical, or nuclear weapons systems) or why some decision makers perceive a choice to be moral that others would find irrational (e.g. refusing to enter into a profitable trading or beneficial security relationship with a country that violates human rights internally or fails to protect its environment).

There is no single school of psychological thought from which we draw. Different facets of psychology shed light on the diverse problems addressed by IR theories: identifying and responding to threats to security, recognizing opportunities to achieve more effective economic coordination, and building transnational communities that are not readily reducible to a security or economic calculus. The psychological formulations on which we draw most extensively include the following:

1. the well-known work on cognitive bias and error and the less well-known work on how robust those biases are when decision makers are accountable to skeptical constituencies and/or face the prospect of intellectually sharper adversaries taking advantage of their cognitive foibles;
2. prospect theory and related formulations regarding risk-taking propensities that highlight the importance of the status quo and that specify how risk tolerance varies as a function of whether decision makers frame problems as potential losses or potential gains;
3. theories of procedural and distributive justice that remind us that most people, policy makers included, have strong moral intuitions about how institutions should make decisions and the types of decisions institutions should make (moral intuitions that are often in conflict with the purely instrumental prescriptions of *homo economicus*);
4. theories of cross-cultural psychology that go beyond specifying how people think to advance testable propositions specifying how they relate to one another. In particular, we discuss normative taxonomies of relational schemata that can help constructivists to classify the different kinds of transnational communities that have arisen or might arise and to understand how normative logics change as transformations of these communities occur.

## REALISM

Ironically, given the emphatic rejection of psychology by leading structural realists (Waltz 1959, 1979), it turns out that the incorporation of explicit psychological assumptions is more advanced in the realist—and particularly the neorealist—framework. This is partly because of the seminal influence of Jervis's work on the sources of misperception and partly because misperceptions in the domain of war

and peace are so visibly costly to states and their societies, triggering great interest in the failure of some states to play the international power game well.

Those who believe they have the least need to take account of psychological processes are the strict structural realists (Waltz 1979, Mearsheimer 1990, Layne 1993). Within this orthodox camp, some treat states as security maximizers in an anarchic state system that requires them to practice balance-of-power politics (Waltz 1979, Grieco 1990, Layne 1993); others attribute more grandiose objectives to states and depict them as power maximizers pursuing hegemonic aspirations (Mearsheimer 1994/1995, 2001). Waltz's framework predicts a general tendency for states to make rational, security-maximizing decisions as they are socialized into the international system (and states that are slow to learn the rules disappear). For Mearsheimer, states are under even greater pressure to form accurate representations of the world and respond to the actions of other powers in a timely manner. Either way, states tend to be deeply suspicious and deem it prudent to make worst-case assumptions about the actions of other states in the system, since the penalties for being wrong in a self-help environment are so severe.

Even these pure versions of neorealism implicitly reflect psychological assumptions. The security-maximizing defensive positionalist argument allows for the possibility that states can become satisfied with their position in the system. These status quo powers will tend to be loss averse and disinclined to pursue expansionist policies that could trigger counterbalancing by other states. The power-maximizing offensive variant views states as more gain seeking and assumes they are never satisfied short of hegemony.

Prospect theory—arguably the most influential alternative to subjective expected-utility maximization—can play a role within realism in helping to distinguish conditions under which the Waltz or Mearsheimer model of state motivation applies. Prospect theory posits that under certain conditions, decision makers should be especially willing to take riskier courses of action than would be justified based on calculations of their expected final asset position, and these conditions are as follows: First, they have not made psychological peace with their losses; second, they underweight subjective probabilities of failure by treating small probabilities as functionally equivalent to zero; and third, they overweight subjective probabilities of success by treating large probabilities as equivalent to 1.0 (certainty). By contrast, decision makers should be especially reluctant to take courses of action as risky as those stipulated by expected-value calculations when they have renormalized perceptions of what is rightfully theirs in response to recent gains, when they overweight small probabilities of failure by dwelling on them, and when they underweight large probabilities of success by ruminating over how things could go wrong.

Prospect theory has received extensive support in the experimental literature on choice, and although critical information is often missing (Boettcher 1995), it has been widely applied in the IR field (Levy 1992; Farnham 1994, 1997; McDermott 1998). When states are in the domain of losses (or, like Serbia, have

not psychologically adjusted to ancient losses), they are more likely to take the irredentist approach that Mearsheimer posits (e.g. Germany in 1939, Japan in 1941). When states are in the domain of gain, they are more likely to accept the status quo, as Waltz would predict (examples, arguably, are the US reluctance to incur any casualties in interventions abroad in the 1990s—Haiti, Somalia, Bosnia, and Kosovo—as well as NATO’s hesitation about expanding into parts of Eastern Europe more sensitive to Russia, notwithstanding Russia’s palpable weakness).

Prospect theory also sheds light on the oft-noted differences between deterrence and compellence. It is far more difficult to induce a state to give up something that it already possesses than to prevent it from taking something that it does not possess (Schelling 1967). Indeed, prospect theory—and associated work on the endowment effect, which is the increase in value one places on something once one possesses it (Kahneman et al 1991)—suggests crude quantitative estimates of how much more difficult it is: Prospective gains often need to be roughly twice as large as prospective losses to be of commensurate value.

## The Number of “Poles”

Structural realists argue that errors in statecraft are more likely as the number of great powers increases. With more great powers, uncertainty grows about who will ally with whom. In this view, a bipolar world—one with only two great powers (as existed during the Cold War)—is most stable. For other realists, the focus is not on the static distribution of power at any given moment but rather on the dynamic ebb and flow of military, economic, and technological capabilities of declining and rising powers (Blainey 1973; Gilpin 1981; Wohlforth 1993, 1999; Schweller 1996, 1998), an ebb and flow that inevitably increases uncertainty about the proper course of action for hegemons and challengers alike.

Regardless of the source of the difficulty—the complexity created by multipolarity or the ambiguities created by chance—even these psychologically minimalist variants of realism open the door to considerations of error and bias. The greater complexity of multipolarity increases the difficulty of timely and accurate information-processing regarding the distribution of capabilities. The finite information-processing capacity of states (and implicitly the decision makers within them) can be overwhelmed by environments with unfavorable signal-to-noise ratios. Miscalculation becomes more likely either when the balance is harder to measure or when there is increasing uncertainty about one’s future position in the system. This greater likelihood of miscalculation, in turn, increases the prospects for war.

The psychological literature suggests that decision makers are susceptible to dilution effects. They often lose confidence in the diagnosticity of predictively useful cues when those cues are embedded in arrays of utterly nondiagnostic cues (Nisbett et al 1981, Tetlock & Boettger 1989). From this standpoint, the more unfavorable the signal-to-noise ratio, the greater the risk that decision makers will

be distracted by irrelevancies, which may have been intentionally introduced by adversaries to confuse the real issues at stake.

The dilution literature and the broader body of work on information overload mean that sometimes decision makers are distracted because there are too many balls in the air (multipolarity), sometimes because the ball is moving too fast (periods of hegemonic transition), at other times because, although there is only one ball, there are lots of other things in the air (complex pluralistic polities that send out contradictory cues), and finally because the air is so hazy and the illumination so poor that the ball may be difficult to see (gauging the intentions and capabilities of closed states). Thus, at least four key properties of the information environment interact with cognitive constraints on decision makers in ways that can produce misperception and miscalculation.

## Misperceptions

Some realists have gone much further than the strict structuralists in building psychological constructs into their frameworks. They posit that misperceptions—slippages between reality and decision makers' representations of reality—are not random (the proverbial trembling hand of game theorists) but rather can take systematic forms. Random errors can be explained away easily by strict structuralists and game theorists, but more systematic errors should be a difficult psychological pill to swallow because they call into question the efficiency and thoroughness of the process by which states are socialized into a ruthlessly competitive system in which ultimately only the fittest survive.

One class of misperceptions concerns one's would-be allies. For example, Waltz (1979) has identified two errors that occur in multipolar international environments—states may do too much or they may do too little. First, states can couple themselves too tightly to allies and get dragged into war (World War I model); second, states can free-ride and mistakenly count on others to take care of the balancing against potential aggressors (World War II model). Work seeking to explain these patterns has focused on misperceptions of the offense/defense balance and of the balance of power as a key determinant of which error occurs (Christensen & Snyder 1990, Christensen 1997). When offensive dominance is wrongly thought to prevail—as in 1914—excessive coupling is to be expected; when defensive dominance is wrongly thought to prevail—as in the mid-1930s—free-riding should be common.

This pattern can be amplified by simplistic analogical reasoning that leads states to draw on the wrong historical precedents (states are prone to fighting the last war). The failure to engage in timely Bayesian updating of assessments of changes in military technology and/or intra-alliance behavior leads to systematic lags between perceptions and reality. Despite arguments to the contrary (Gerber & Green 1999), the psychological literature on judgment and choice suggests that most decision makers are not natural Bayesians (Edwards 1962; Tetlock 1998, 1999). People, even experts, are often too slow to change their minds in response to unexpected

events, especially in environments in which causality can be indefinitely contested because no one knows for sure what would have happened in the counterfactual worlds in which alternative policies were pursued.

A second class of misperceptions concerns potential adversaries (Jervis 1976). Again, states must balance the risk of two conflicting perceptual errors. Type I errors involve incorrectly labeling status quo powers (Waltz's and Grieco's defensive positionalists) as expansionist, precipitating a conflict spiral; type II errors involve incorrectly labeling expansionist powers (Mearsheimer's hegemony seekers) as status quo, leading to failures of deterrence. During the Cold War, this debate was central to ideological arguments between liberals and conservatives in both Washington and Moscow over how to deal with the other superpower (Osgood 1981, Garthoff 1994). One implication of Jervis's argument (which relies heavily on experimental social psychology of the 1960s and 1970s) is that type I errors are more common than type II errors as a result of the fundamental attribution error (in which observers are too quick to draw strong dispositional inferences of hostile intent from situationally motivated defensive preparations) and belief perseverance (in which observers are too slow to revise their initial causal inferences in response to unexpected events). Other work on cognitive factors in IR demonstrates that misperceptions occur in a variety of types of international relationships, not just those regarding allies and adversaries (Herrmann & Fischerkeller 1995).

Realists might challenge this argument on one of the following grounds, each compatible with a major line of psychological research on judgment and choice:

1. The state system is overwhelmingly populated by expansionist actors, so the base rate favors an across-the-board inference of expansionist intent.
2. The costs of making a type II error are typically so much more severe than those of making a type I error that making many fundamental attribution errors is a cost worth paying.

One example of this type of adaptive error is the Bush administration's response to Gorbachev's cooperative initiatives. Outsiders criticized both the "strategic pause" announced in 1989 and the general White House attitude that Gorbachev was simply tricking the West into letting its guard down. Many American officials were slow to accept that Soviet policies represented a fundamental redefinition of Soviet interests. Realists would argue, however, that rather than succumbing to error, these officials could not adjust more rapidly because the penalties for being wrong would have been so severe, and in some cases their prior experiences (particularly their involvement in the failed *détente* of the 1970s) added to their concerns (Tetlock & Goldgeier 2000).

## Evolution

The common misconception that realist and psychological theory are inevitably in tension with each other—formally advanced in Waltz's influential 1959 book—should be called into question here (see also Mercer 1995). On the one hand, certain

strands of evolutionary psychology reinforce the realist emphasis on the benefits of being suspicious. On the other hand, realist arguments, far from precluding cognitive bias and error, reinforce psychological analyses that highlight the benefits of making snap judgments of intentionality, especially for potentially threatening conduct, as well as the perils of changing one's mind too soon. For instance, Skyrms (1996) documents that evolutionary theory guarantees only that organisms are in a continual process of maximizing their genetic fit to their ever-changing local environments, not that organisms are perfect Bayesians (truth seekers) or utility maximizers, as normative theories actually require. Indeed, nature might even select for the kind of cognitive biases that US National Security Adviser Brent Scowcroft and others exhibited in 1989. As Stich (1990:25) comments, "A very cautious, risk-averse inferential strategy—one that leaps to the conclusion that danger is present on very slight evidence—will typically lead to false beliefs more often, and true ones less often, than a less hair-trigger one that waits for more evidence before rendering a judgment. Nonetheless, the unreliable, error-prone, risk-averse strategy may well be favored by natural selection. For natural selection does not care about truth; it cares only about reproductive success [read national survival]. And from the point of view of reproductive success, it is often better to be safe and wrong than sorry." Of course, this argument assumes that false positives are evolutionarily inconsequential, an assumption that the spiral theorists, worried about the World War I model, would challenge.

The key point here is that it is possible to grant evolutionary processes of natural selection—whether they have operated on human beings since the Pleistocene or they have operated on nation-states since Westphalia—a central role in shaping mental or organizational mechanisms of decision making (Cosmides & Tooby 1994, Pinker 1997) and still have plenty of room for a psychological research program that demonstrates deviations from rationality defined by perceptual-accuracy criteria.

Revisionist structural realists have begun to explore these attribution errors as they seek to understand how states try to balance against threats rather than power (Walt 1987, 1996). At times, these analyses resemble arguments underlying theories about the democratic peace from a social psychology point of view. For example, Walt (1996) has shown how states undergoing revolutionary transformations change threat perceptions in the international environment and produce systematic errors as others react to the rise of these revolutionary states and exaggerate the threat. Revolutions increase uncertainty for other actors in world politics, since there is a tendency to assume that ideologically alien domestic systems foreshadow threatening international conduct. Walt demonstrates that miscalculations abound owing to reliance on biased information, self-defeating spirals of suspicion occur in numerous cases, and ideology is an impediment to accurate assessments of other states. As psychologists would expect, once a regime is assigned to an ideological outgroup, observers exaggerate and there is reduced ability and motivation to empathize (Brewer & Brown 1998). We feel justified in making nasty inferences about these alien regimes for which we might have weak evidence, and it may prove

difficult to reassess regimes that move away from revolutionary aspirations because of our preference for confirmatory strategies of hypothesis testing (Dawes 1998).

Other realists have attributed self-defeating expansionism to a leader's need to satisfy diverse coalitions in order to stay in power (Snyder 1991). Different groups within the elite place contradictory accountability demands on policy makers, thereby making the overall policy less coherent. For example, Soviet leaders Nikita Khrushchev and Leonid Brezhnev needed to placate, on the one hand, the ideologues and military who wanted to expand into the Third World, and, on the other, the technocrats who wanted the benefits of detente (Anderson 1993, Richter 1994, Snyder 1987/1988). The expansionist behavior was not only costly in itself but also impaired the ability of the Soviet Union to gain economic rewards from the United States. In such cases, the payoff to each specific elite group is rational, but the sum total of payoffs produces irrational, contradictory behavior. Domestic rationality may produce international irrationality (Lebow 1981, Fearon 1998). External observers, however, often underestimate the internal complexity of interactions among factions that produce policy outputs. This underestimation may be the product of cognitive biases toward overcentralization (Jervis 1976, Vertzberger 1990) and entitativity—seeing a collective entity as a single unit (Brewer & Brown 1998)—or a more strategic decision to hold other states tightly accountable for their conduct.

## Moderating Variables

Since Jervis wrote in the mid-1970s, we have learned a lot in experimental social psychology and behavioral economics about the social and market forces that can either amplify or attenuate deviations from rationality. Let us assume, as many experimental researchers do (although not all—see Gigerenzer & Goldstein 1996), that a root cause of error and bias is the tendency of both individual and collective actors to over-rely on simple, easy-to-execute heuristics that often give people unjustifiable confidence in their judgments and decisions. Insofar as this diagnosis is correct, it follows that social, political, and economic systems that encourage actors to engage in more self-critical and reflective forms of information processing should often have the net effect of attenuating bias. Conversely, social, political, and economic systems that encourage mindless conformity, defensiveness, and the perpetuation of shared misconceptions should have the net effect of amplifying bias.

Two important sets of social-institutional moderators of rationality merit mention. First are organizational and domestic accountability pressures. Decision makers virtually never work in social isolation, with the occasional bizarre exceptions of leaders such as Adolf Hitler, Josef Stalin, Kim Il Sung, and Saddam Hussein, who centralized enormous authority in themselves. Decision making more typically unfolds in complex social and political networks of accountability. There are several possible ways of structuring these accountability networks or systems so as to promote more flexible, complex, rigorous, and systematic forms of thinking.

Experimental work indicates that, ideally, decision makers should feel accountable prior to making irrevocable commitments, and they should feel accountable to audiences whose judgment they respect and whose continuing esteem they value. In addition, the simple conformity option—telling people what they want to hear—should be short circuited either by creating normative ambiguity (decision makers should not know what others want to hear) or by creating normative conflict (decision makers know what others want to hear but the others represent diverse interests that want contradictory things). (See Tetlock 1992, Tetlock & Lerner 1999, Lerner & Tetlock 1999.) These are exactly the types of accountability preconditions that characterize multiple-advocacy policy groups (George 1980), many organizational prescriptions for averting groupthink (Janis 1982, Tetlock et al 1992), and arguments in political science about the underlying institutional-political mediators of the democratic-peace effects (Doyle 1983, Russett 1995, Owen 1997, Elman 1997).

Second are competitive market pressures. Skeptics have long suspected that judgmental biases hold up only in single-play situations in which respondents rarely receive feedback concerning the appropriateness or effectiveness of their decisions and respondents have little material incentive to get the answer right anyway. The skeptics overstated their case but there is some truth to this objection. Certain classes of errors and biases—especially breakdowns in consistency and transitivity produced by reliance on simple (lexicographic) decision rules—can be corrected when we move the choice process into open and transparent market settings that provide for repeated play, interaction with attentive competitors, and rapid, unequivocal feedback on the consequences of one's choices (Kagel & Roth 1995, Camerer & Hogarth 1999).

Combining these two lines of argument, it is possible to identify an ideal set of conditions under which the error-and-bias portrait of the decision maker should be of minimal predictive usefulness. Ideally, the inner advisory group deliberates under the norms of multiple advocacy, is accountable to the institutions of a self-correcting democratic polity, and is making decisions in a policy domain in which critical information is readily available; and mistakes are quickly and publicly punished. These conditions are most likely to be satisfied in what is commonly called the zone of peace and prosperity (Goldgeier & McFaul 1992, Russett 1995, Singer & Wildavsky 1996). To be sure, decision makers will still make mistakes in this privileged zone, but the mistakes will be fewer, less systematic, and less serious than elsewhere, and when mistakes do occur, decision makers will correct them more rapidly. An example is the swift response of leaders to what Friedman (1999) calls the electronic herd and its imposition of a "golden straitjacket." The research agenda becomes one of identifying the relative resistance of biases to de-biasing pressures. Holding properties of the environment constant, we suspect the most tenacious biases will prove to be those that are widely shared and relatively subtle, and therefore difficult to identify and exploit. And even if the environment is more conducive to better cognitive performance, emotions associated with stress—such as fear and anxiety—may alter the impact of de-biasing pressures (Crawford 2000).

## The Dangers of Being Too Smart

Psychology can illuminate a long-standing problem identified by prominent realists during the Cold War, namely, the reluctance of thoughtful decision makers to accept the radical strategic implications of the nuclear revolution (Waltz 1979; Jervis 1984, 1989). In this case, the problem was not the typical shortcomings caused by cognitive bias; rather, the error was being “too smart” or hyper-rational. An insightful illustration from behavioral game theory, the “guess-the-number” game first studied by Nagel (1995), illuminates the problem. Contestants guess a number between 0 and 100, with the goal of guessing as close as possible to two thirds of the average number chosen. In a world where all the players are known to be rational, in the sense that they will form expectations about the guesses of others who themselves are carrying out as many levels of deduction as necessary, the equilibrium in this game is 0. In a contest run at Thaler’s (2000) urging by the *Financial Times*, the most popular guesses were 33 (the right guess if everyone else chooses a number at random) and 22 (the right guess if everyone picks 33). The average guess turned out to be 18.91 and the winning guess was 13.

To model how people play this game, we must allow for two kinds of individual differences in sophistication: logical and psychological. First, agents differ in the number of logical levels of processing they work through. A guess of 33 reflects one level (I guess that the average guess will be 50 and  $2/3$  of 50 is 33), 22 is two levels (I infer that you will work through the preceding argument so I will guess  $2/3$  of 33 or 22), and so on. Second, agents differ in their psychological assumptions about other actors and the likelihood of those actors working through the necessary inferences to reach the logically correct answer. Agents who guess 0 are logically sophisticated but psychologically naive. The intriguing implication here is that good judgment requires a game-theoretic appreciation for what the technically correct answer is as well as psychological savvy in appreciating how close people can come to approximating that solution. The lesson for IR is that games can have both logical and psychological equilibria (see Green & Shapiro 1994).

The implication for debates over nuclear doctrine emerges from Jervis’s critique of the so-called countervailing strategy. With perfect logic from a conventional perspective, influential American strategists argued the following: The United States’ current policy posture rests on the threat that if the other side attacks somewhere in a limited fashion, we would launch all-out nuclear war. But this threat is not credible, so we are not deterring a limited attack. We need options at each level of escalation, and we need dominance at each level, so that the other side will know that our potential response is credible and thus they will be deterred from starting even a small-scale attack.

For Jervis, the countervailers were “too rational.” They recognized that the threat of all-out war was incredible, so they worried that it would not be taken seriously and thus would not deter. Stability at the top level of nuclear war led to less stability at lower levels (the stability-instability paradox). Jervis and others, such as Brodie, Waltz, and Bundy, argued that the nature of nuclear weapons

meant that uncertainty was enough. The threat was effective not because the Soviets were certain that the United States would carry it out, but because in war it was a possibility, and the costs of nuclear war were so great that this possibility was sufficient to deter Soviet attack. Given the peculiar psychology of nuclear weapons and the feelings of ultimate dread they inspire, efforts to create limited-war-fighting options were at best unnecessary (wasted resources) and at worst counterproductive by sending the message that it may be permissible to escalate to “low” levels of conflict previously regarded as unthinkable.

## INSTITUTIONALISM

Although realists disagree about many things, there is one matter on which they concur: In the absence of a common threat, sustained interstate cooperation is extremely difficult because (a) states are more concerned about relative than absolute gains, (b) states are always tempted to defect (and there is no higher authority to stop them or others from doing so), and (c) the penalties for being too sanguine about the intentions of others are so severe.

The institutionalist challenge arose in response to this pessimistic appraisal of the prospects for cooperation in the international system. The microeconomic variant of institutionalism articulated by Keohane [building on the earlier work of Coase (1960) and Williamson (1965) in economics and of neofunctionalists such as Haas (1958, 1964)] accepted that the international system is anarchic and populated by unitary, egoistic actors but nonetheless insisted that it is possible for such actors—in principle and in practice—to create institutional frameworks that permit them to secure the gains of cooperation by lowering their fear of being exploited to a tolerable level. Theoretically, transnational institutions should arise whenever it is economically rational to create them (although they might need a hegemon to get them started).

Psychologists are likely to be skeptical and to view such a claim as simultaneously too restrictive and too expansive. The claim is too restrictive because regimes can arise via mechanisms other than transaction-cost calculations—for example, shared norms of fairness or the desire to punish free-riders. The claim is too expansive because actors may fail to form regimes even if all of the economic institutionalists’ preconditions are met; shared mindsets may blind actors to the feasibility or benefits of coordination. [See Thompson (1998) for experimental evidence on the fixed-pie fallacy; see Herrmann et al (2001) for survey evidence of mindset impediments in mass public opinion to movements toward the Pareto frontier in international trade.]

In the neoliberal institutionalist model, actors who recognize that they confront a repeated play situation and that the transaction costs for continuing to improvise ad hoc solutions are high rationally enter into binding compacts that can persist despite changes in the balance of power among the member states. Contrary to many realist views that institutions largely serve the interests of the powerful states

and have minimal independent effect on state interests (Gilpin 1981, Mearsheimer 1994/1995), neoliberal institutionalists believe that international institutions can alter state interests over time as domestic political actors learn the value of continued participation in the international regime (Keohane 1984, Keohane & Martin 1995).

The key issues on which we believe psychology can inform these types of institutional analyses are when and why states seek to create institutions in the first place; when and why states live up to institutional norms and principles; and whether and when, over time, institutions transform the conceptions of state interests held by decision makers.

## Creating Institutions

There are many policy domains in which institutionalists have sought to demonstrate the process by which states, theoretically locked in “prisoner’s dilemma” relationships, successfully escape from the suboptimal mutual defection cell of the payoff matrix and land in the Pareto-optimal mutual cooperation cell. These include the postwar arrangements to facilitate trade by restricting the sovereign authority of states unilaterally to set tariffs and other barriers blocking entry to domestic markets, the rise of the European community to manage a range of economic, environmental, and social issues, and the creation of transnational environmental regimes to cope with perceived externalities of commercial and industrial development (e.g. CFC production, Mediterranean pollution control, restrictions on fishing and whaling, the law of the sea restrictions on resource development).

It should go without saying that ideas—policy makers’ individual or shared mental representations of the problem—shape judgments about whether institutional solutions are feasible and desirable and what forms those institutional solutions should take. Goldstein & Keohane (1993:3) formally recognize this point when they declare, “Ideas influence policy when the principled or causal beliefs they embody provide road maps that increase actors’ clarity about goals or ends-means relationships, when they affect outcomes of strategic situations in which there is no unique equilibrium, and when they become embedded in political institutions.” In our view, the usefulness of adopting a cognitive psychological approach to the role of ideas hinges on the potential for systematic slippage between policy-guiding mental representations of reality and reality itself. Here we invoke the transparency of the environment as a fundamental moderator of the value of resorting to psychological levels of explanation for institution formation. The greater the transparency, the less latitude there is for slippage between reality and mental representations of reality.

“Transparency” has many meanings but two are especially important in this connection. First is the observer’s ability to discern basic facts. Second is the ability to draw sound causal inferences, which can include, for example, the clarity of reward-punishment contingencies in the environment.

Epistemic communities (Haas 1997) can both advance and reflect causal transparency. At one end of the transparency continuum are policies shaped by epistemic communities possessing great authority and prestige by virtue of their technical and scientific achievements in teasing apart cause-effect relationships. This is especially likely in domains where investigators can test theoretical hypotheses in controlled and replicable experiments [and there is minimal reliance on what Tetlock & Belkin (1996) call counterfactual control groups]. Here we have in mind the near consensus among biomedical specialists on the appropriate strategies (or methods of developing strategies) for coping with transnational epidemics, the near consensus among climatologists on the effect of CFCs on the ozone layer of the atmosphere, or the near consensus among physicists and engineers on the prerequisites for nuclear proliferation. Virtually no one in the mainstream of the political elite in the advanced countries is prepared to question advice emanating from these communities on their issues of expertise.

At the other end of the transparency continuum are those policy domains in which the epistemic communities are either deeply divided or relatively undeveloped. Although there is substantial consensus among economists regarding the virtues of free trade, there are still pockets of sharp disagreement on details (strategic trade and infant industry arguments), as well as deeper professional rifts over the advisability of encouraging free flow of capital across national borders. Similarly, scientists differ on the sources, scope, pace, and effects of global warming.

We recognize that all inductive knowledge is tentative, and even well-established epistemic communities can fall prey to belief perseverance and groupthink. Nonetheless, the potential explanatory role of psychological constructs expands rapidly as we move from domains where the design of institutions is guided by “well-known facts” and “solid science” to those where the expert community is deeply divided and there is ample opportunity for cognitive and emotional biases to taint evaluations of evidence and options.

## Compliance and Transformation

IR theorists are deeply divided over the issues of why states comply with the demands of international institutions and whether over time those institutions can change state conceptions of their interests (Simmons 1998). Imagine a social influence continuum. At one end is compliance, where decision makers accede to the rules of international institutions as a result of exclusively utilitarian calculations of the material rewards and punishments of participating in those institutions. Historical prototypes would be military alliances formed to counter direct threats, and the World Trade Organization, which lowers trade barriers for members and provides sanctioning mechanisms to punish norm violators.

Toward the middle of the continuum are decision makers who do what is expected of them because they seek to establish particular social identities in the eyes of certain domestic or international audiences whose opinion they value, and because they may be unwilling to bear the reputation costs of defecting. This

form of social influence—sometimes called identification by social and organizational psychologists—should be less context-specific than mere compliance (which should shift on or off as a function of the cost-benefit calculus). Identification should persist even when it is materially inconvenient, but it still does not represent full-fledged internalization (Kelman 1958).

The self-presentational goals that drive identification do not always lead to “benign” outcomes. Although historical examples include states with dubious civil liberties practices or suspect records on respecting ecosystems signing on to human rights or environmental agreements in order to be part of the community of “modern” and “civilized” states, examples also include moderate Arab states in the 1960s and 1970s paying lip service to pan-Arabism and the destruction of Israel, as well as India’s and Pakistan’s efforts to join the nuclear club in order to achieve the status thought to accompany membership.

The calculations at the identification point of the spectrum are still largely utilitarian, but the benefits are measured not so much in economic and security terms as in social and reputational categories. As we continue toward the internalization end of the continuum, we find states not only doing the right thing but doing so for what they perceive to be the right reasons. The calculation is no longer utilitarian [in March & Olsen’s (1998) terms, following the logic of consequential action] but rather is guided by moral, religious, or ideological ideals (March & Olsen’s rule-bound logic of appropriate action). Historical prototypes might include Scandinavian and Canadian approaches to foreign aid, in which there are minimal assumptions of economic or security *quid pro quos*, and the attitudes of certain revolutionary states seeking to export their ideology to their neighbors.

A contribution of psychologists is to delineate the conditions under which decision makers are especially likely to change their underlying attitudes to bring them into line with initially counterattitudinal behavior. Cognitive dissonance and self-perception research using forced-compliance paradigms suggests, for example, that decision makers are especially likely to internalize attitudes consistent with their behavior when they believe they had some elements of free choice (see Bem 1967, Larson 1985). It is critical here that the external pressure for compliance not be too heavy-handed; otherwise, people will attribute their behavior to external demands and not internal values or attitudes. As good diplomats have long known, heavy-handed pressure tactics often backfire, and even when they do work, the effects are often fleeting.

Most psychologists would probably agree that most political actors (psychopaths excluded) will gradually internalize the norms of fair play implicit in international institutions. These norms can become functionally autonomous from the interests that may once have inspired them. This internalization process should be especially reliable in democracies, in which leaders must justify departures from widely held norms of fair play to a variety of constituencies (Tetlock & Goldgeier 2000).

This does not mean that procedural justice issues are easy to solve. Leading institutionalists have argued that a poorly understood issue is the extent to which

the effectiveness of institutions hinges on whether they are dependent on political decision making or on relatively independent experts or judicial processes (Martin & Simmons 1998, Simmons 1998). The cognitivists have something to say here. Institutions will typically have a hard time establishing that they possess minimal prerequisites of procedural justice (neutrality, equal respect, fairness) in the eyes of participating nation-states. Experimental work reveals that even scrupulous attempts on the part of mediators to achieve neutrality are often misperceived as partisan by both of the contending parties (Ross & Griffin 1991). Domestic political opponents of international institutions can then play on these cognitive biases and mobilize substantial resistance to continued participation in the institution.

Rational choice approaches to norm enforcement stress the danger of free-riding. In the absence of a sovereign, no single party wants to incur the costs of punishing norm violators, so defectors escape sanction (see Coleman 1990). A rapidly growing body of work in experimental social psychology and microeconomics has revealed, however, that people are often willing to make substantial sacrifices to punish cheaters in implicit or explicit social contracts (Fehr & Schmidt 1999). The more deeply internalized the norms and the more egregious the transgression, the sharper the emotional reaction and the more willing people are to go out of their way to punish transgressors and even to punish those who fail to punish transgressors (Axelrod 1984, Crawford 2000). This analysis suggests that international institutions built on shared conceptions of distributive and procedural justice may be able to perpetuate themselves even in the absence of strong central authority.

Finally, just as prospect theory sheds light on the causes of the asymmetry between deterrence and compellence, it also helps us to understand the differential reactions observers often have to the distribution versus the redistribution of scarce resources in both domestic and international economic debates. Because the loss function falls much more rapidly than the gain function rises, prospect theory leads us to expect that even observers trying to be scrupulously neutral will display a preference for upholding the claims of the status quo. It will also be easier to build up psychological momentum to defend the status quo (and avoid imposing painful losses) than it will be to create momentum to overturn the status quo (bestow gains on the have-nots).

## CONSTRUCTIVISM

Whereas realists and institutionalists posit a choice process organized around March & Olsen's logic of consequential action (explicitly utilitarian weighing of costs and benefits), many constructivists posit a choice process heavily informed by the logic of obligatory action, in which people make up their minds in part by matching their conceptions of who they are (social identities) with their assessments of the normative context (what does this situation call for actors of my type to do?).

Insofar as explicit assumptions are made about the motivations of individual actors, these assumptions stress norm-following logics that provide guidelines for legitimate and illegitimate behavior as the actors (*Homo sociologicus*) construct an identity as part of a given social order, itself constructed by the interaction of the actors in the system (Wendt 1999). Examples include the demise of dueling and slave trading (Mueller 1989); the rise of human rights norms, especially the sensitivity to racism (Risse et al 1999); environmentalism (Haas et al 1993); the nonuse of weapons of mass destruction and other constraints on the conduct of war (Legro 1995, Price 1997, Tannenwald 1999); and the changing purpose of military intervention (Finnemore 1996).

At a foundational level, a cognitive psychological analysis of world politics is compatible with the constructivist program. A natural starting point for a cognitive analysis is to consider (a) the nature of the information-processing task that observers confront when they try to draw causal inferences or policy lessons from world politics and (b) the capacity limitations of the human mind. Cognitivists are impressed by the complexity of world politics (the number and variety of alternative explanations that competing schools of thought can typically advance) and the inherently ambiguous feedback that policy makers receive in a path-dependent system that runs once and only once (no one has empirical access to the counterfactual worlds in which alternative policies were pursued). From a cognitivist point of view, all causal inferences and policy lessons are the product of mental constructions of what would, could, or might have happened had a different set of antecedent conditions held or policies been tried. There is, in principle, an infinite number of possible background factors that one could enter as antecedents in one's counterfactual constructions of alternative worlds. In practice, of course, observers must rely on draconian simplifying rules that reduce the number of scenarios to be entertained to a humanly manageable number. These simplifying rules are generally drawn from the shared understandings of epistemic communities.

Notwithstanding this fundamental point of agreement, the efforts to incorporate insights from cognitive science into the constructivist program are only beginning (Finnemore & Sikkink 1998, Wendt 1999). One obvious candidate is the nature of tradeoff reasoning, where a profound disjuncture emerges between rationalists and constructivists (Katzenstein et al 1998, Ruggie 1998). For example, realists expect states rationally to measure the costs and benefits of entering or leaving a military alliance or of taking preemptive military action; institutionalists expect states to examine the costs and benefits of entering or exiting a trading or arms control or environmental regime. By contrast, there are large classes of issues for which constructivists would expect tradeoff reasoning to be extremely difficult. For example, realists may have no trouble believing states will engage in a cost/benefit analysis to determine whether to use a weapon of mass destruction to further their military aims (Sagan 2000), whereas constructivists believe that norms and practices lead decision makers to place boundaries on the thinkable. The constructivists are more open to the possibilities of categorical exclusionary logic than the rationalists, and constructivists have argued that detailed case studies demonstrate that the way

decision makers approach subjects such as the use of weapons of mass destruction supports their way of thinking (Price 1997, Tannenwald 1999).

Tetlock et al (2000) distinguish three types of tradeoffs: routine, taboo (secular versus sacred), and tragic (sacred versus sacred). A routine tradeoff is the type of reasoning one deploys whenever one goes shopping and must compare the relative importance of secular values, which are subject to legal market-pricing in one's social world. (Would I rather have a larger stock portfolio or a Mercedes-Benz?) A tradeoff is taboo if it pits a secular value such as money against a sacred value such as protection of human rights or ecosystems. Taboo tradeoffs are not just cognitively demanding (the familiar incommensurability problem: How much of one value am I willing to give up to achieve an increment on another value?); they are also morally corrosive (the less familiar "constitutive incommensurability" problem, in which merely to think certain thoughts or to make certain comparisons undercuts one's claim to embody shared moral values). Finally, a tradeoff is tragic if it pits two sacred values against each other, e.g. protection of endangered species against protection of indigenous cultures.

These distinctions are psychologically and politically consequential. Decision makers caught making taboo tradeoffs are often ostracized by the moral communities within which they once held leadership roles. Indeed, they may be condemned even if the actual decision they rendered would have been perfectly acceptable had it been semantically framed as a routine or tragic tradeoff. Not surprisingly, decision makers go to great efforts to portray their decision process as free of any taint of taboo tradeoffs, and their adversaries struggle equally tenaciously to convince key constituencies that the boundaries of the unthinkable have been breached.

Should tradeoff reasoning be treated as a defining property of rationality, good judgment, and maturity? For realists, institutionalists, and economic liberals, who argue that decision makers are utilitarians, the answer is yes. Leaders think in terms of how much of  $x$  they are willing to give up for a given increment of  $y$ . For Kantian liberals and constructivists, it is possible to identify large classes of important issues for which decision makers should find compensatory tradeoff reasoning illegitimate. Tragic, but not taboo, tradeoffs are permissible.

For example, regarding weapons systems, constructivists might expect leaders to believe that it is preferable to kill more people with conventional arms than to break a taboo by dropping one small atomic bomb (putting precedent-setting to the side). Realists believe states view actions as mandatory if they promote self-interest (improve one's position in the system). For constructivists, leaders may find actions mandatory in order to preserve international norms, actions that might be unthinkable for a decision maker in a realist mode. Metternich and Kissinger would not think of using military force to serve humanitarian purposes (particularly if these harm self-interest in other areas, e.g. damaging relations with Russia and China by acting in Kosovo), whereas many constructivists posit a world in which upholding norms against genocide is not only thinkable but mandatory.

A second way in which psychological analyses can insinuate themselves into constructivist approaches to world politics derives directly from prospect theory

and work on endowment effects. A major focus of debates that constructivists attempt to explain is fairness: Who owes what to whom? What counts as a just claim on this or that territory or resource? Prospect theory leads us to expect that competing social conceptions of fairness ultimately rest on competing perceptual encodings of the appropriate endowment reference point. Those who challenge the status quo deny the legitimacy of the current allocation of resources and hence downplay the moral seriousness of the pain inflicted on the haves and play up the moral seriousness of the gains that should have gone (in an ideal counterfactual world) to the have-nots. Whether the issue is the forgiveness of loans or the arbitrariness of postcolonial borders in Africa or the Middle East, activists' cognitive constructions of what should be done reflect their perceptual framing of the problem.

Those who take the status quo as their analytical starting point are predisposed to be unsympathetic to redistributive claims advanced by the have-nots. Although the G-7 leaders backed debt forgiveness in 1999 for the poorest of the poor, it is taking them a long time to act, perhaps because many in the G-7 societies ask why international banks or the taxpayers of wealthy countries should subsidize loan defaults by bankrupt and corrupt sub-Saharan countries. Conversely, those who are less fixated on the world that is—and prone to give more weight to counterfactual worlds that could or should have been—are predisposed to be more sympathetic to redistributive claims on behalf of have-nots. Rather than viewing such claims as illegitimate, greedy, and self-serving bids for gain, they see them as just attempts to undo losses imposed by exploitative classes or nations. It should be relatively easy for these activists to imagine worlds that could have been, and might yet be, in which the capricious allocation of resources in this world is rendered more equitable.

Work on emotions in bargaining games reinforces this possibility (Rabin 1993). Economists have traditionally treated human beings as exclusively self-interested and have assumed that when self-interest collides with ethical values, self-interest will generally prevail. But this is not necessarily so—as experimental economists have themselves discovered. The ultimatum game offers a simple demonstration that emotions, coupled with strong, deeply internalized intuitions about fairness, can shape strategic interaction.

In the ultimatum game, one player (the Proposer) is given a sum of money, often \$10, and offers some portion  $x$  to the other player (the Responder). The Responder can either accept the offer, in which case the Responder gets  $x$  and the Proposer gets  $10 - x$ , or reject the offer, in which case both players get nothing. Experimental results consistently reveal that very low offers (less than 20% of the total sum) are often rejected. The Responders appear to react emotionally, indeed indignantly, to the low offers. They do not try to maximize their own payoffs; even in single-play, anonymous games, they turn down low offers and receive zero instead. There is also evidence that, when the veil of anonymity is lifted, players and third-party observers censure those who fail to censure “exorbitant greed.”

For better or for worse, such conduct is far more common than neoclassical economic theorizing would lead us to expect. The emotions activated in the ultimatum game, insofar as they are linked to envy, may sabotage mutually beneficial trade

agreements in which the benefits are shared asymmetrically. But, on the positive side, the same emotions, insofar as they are linked to a deep-rooted justice motive (Welch 1993), may help to deter would-be exploiters from taking what is perceived to be unfair advantage of other parties.

Finally, constructivist analysts can usefully draw on psychology to address what we view as one of the more compelling criticisms of the constructivist program, namely, its inability to specify the conditions under which different groups view different norms as applicable (Checkel 1998). Two lines of research may be especially helpful. First, at the most abstract or process level, experimental work on “natural categories” warns us that such categories are organized not on the principles of classical logic (which are supposed to specify well-defined necessary and sufficient conditions for placing a given instance inside the category) but rather on the principles of fuzzy logic (which stipulate that people often judge whether a given instance falls into a category’s orbit based on the instance’s family resemblance to other category members or to idealized prototypes). Looking for the essential defining features of key normative concepts such as sovereignty or human rights is, from this standpoint, rather futile.

Second, at a more concrete or content level of analysis, recent theoretical advances in cross-cultural psychology suggest taxonomies of norms that can regulate relationships among individuals or groups. These taxonomies, in turn, can serve as platforms for generating hypotheses about when decision makers are more or less likely to rely on various normative rules as guides to action. Coupling neorealism with Fiske’s (1991) model of relational schemata yields five possibilities. The first possibility is the familiar neorealist presocial, anarchic mode of relating, in which no one owes anyone anything, invocations of higher authorities are viewed as meaningless rhetorical flourishes, and there is the omnipresent danger that one’s security will be challenged by other actors. The other four possibilities identify a mutually exclusive and exhaustive set of the forms that normative coordination can take. Starting with the most cognitively complex form of organization—market pricing—actors agree to coordinate their actions by assuming the roles of buyers and sellers in competitive markets, to respect property rights and contractual obligations (the reputation costs of failing to do so are prohibitive), and to maximize expected utility by engaging in compensatory tradeoff reasoning that requires ratio comparisons of competing values. This market-pricing relational template, with supplementary assumptions about transaction costs and monitoring, captures the spirit of neoinstitutionalist approaches to world politics, in which the state system is undergoing a gradual transformation from a presocial Hobbesian anarchy into rule-regulated networks of commerce.

The other three relational schemata within the Fiskean scheme in theory define the constructivist agenda. Norms can take the form of equality matching (e.g. tit-for-tat reciprocity, in which a state calibrates positive or negative responses to the direction and magnitude of other states’ latest moves), authority ranking (e.g. the patron-client or hegemon-satellite relationship, in which the low-status party owes obedience to the high-status party within a certain range of activity),

and communal sharing (e.g. the decision of nation-states to abolish boundaries and merge into a common political entity).

The relational model followed by a given institution or community will lead to different kinds of norm-following logics and thus different implementation rules. The Warsaw Treaty Organization, for example, was an authority ranking system in which the Soviet Union allowed latitude in some highly constrained areas but ruthlessly used force to squelch attempts by local elites to stray from Moscow's norm of one-party control. This logic leads to very different understandings and expectations than the equality-matching norms of the World Trade Organization, which allows states to engage in precisely calibrated tit-for-tat behavior if it deems them the object of unfair trade practices. The Warsaw Treaty Organization was also organized quite differently from its main rival, the North Atlantic Treaty Organization, and thus norms of behavior were different despite the presence of a hegemon in each institution (Ruggie 1993). If constructivism is to explain how actors' identities are mutually constituted with structures, then we need to know which relational schema structures interaction at any given moment in time and how normative logics differ depending on which of the three is dominant.

Interesting cases for the constructivists to explore have been those communities that started with one type of logic and over time developed something quite different. When Haas (1958) began writing about the uniting of Europe in the 1950s, the dominant schema within that nascent community was equality matching. Over time, an institution has developed that increasingly, although still very imperfectly, affirms the norms of communal sharing (all for one and one for all). Enthusiasts consider the abolition of border controls and local currencies only the first steps to a full-scale European entity. By this light, constructivists can compare the development of norm-following logics as relational schemata change within a given set of international relationships that are becoming marked by greater communal sharing, such as the European Union (Moravcsik 1998), or within a security community such as NATO (Adler & Barnett 1998). Developing in a different direction is the relationship between Russia and its former satellites in Europe. The Soviet Union built a bloc in Eastern Europe that was ostensibly based on the ideals of communal sharing but instead followed an authority ranking pattern. Since the collapse of the empire in 1989, the relationship has evolved to one of equality matching, which is what Central and Eastern Europeans want and what some Russians have had a hard time accepting. Shifting relational gears can be painful. Standard assertions of sovereignty can look like deliberate disrespect.

## CONCLUSION

Many IR theorists dismiss psychological arguments, claiming either that they are too reductionist to explain the big patterns we find in world politics or that they are too flimsy or messy to explain anything at all. As this essay suggests, however, psychology's role in identifying boundary conditions is extremely useful in refining arguments within a given theoretical tradition. Prospect theory's identification

of the conditions under which we expect more risk-averse or risk-taking behavior than in an expected-utility model can help explain when a more defensive or offensive realist argument should prevail or when redistributive schemes are likely to have greater or less appeal. Understanding the role of transparency or of domestic/organizational accountability pressures in ameliorating typical errors and biases helps us understand variation in decision-making abilities across types of political systems and might shed light on debates such as those regarding the democratic peace. Work in cross-cultural psychology suggests that there may be a surprisingly small number of basic normative templates from which international organizations can be constituted.

Macrolevel theorists, rather than dismissing psychological theories for presenting obstacles to explaining the broad contours of international behavior, should welcome psychology's help in refining ideas in key debates regarding power, institutions, and norms. Ironically, the buzzwords that dominate recent macrolevel approaches to refine or advance theories are perceptions, ideas, and identity (see Goldstein & Keohane 1993, Katzenstein 1996, Walt 1996). Now is the time for IR theorists to take advantage of systematic arguments about psychological factors to address more explicitly the psychological dimensions of these variables (Goldgeier 1997, Tetlock & Goldgeier 2000). The effort should not be to reduce explanations of behavior to psychological factors but rather to consider how environment and cognition interact in systematic and identifiable patterns to produce the variation we find in world politics.

## ACKNOWLEDGMENTS

We would like to thank Paul Brewer, Dalia Dassa Kaye, Martha Finnemore, Donald Green, Ernst Haas, Richard Herrmann, Robert Jervis, Peter Katzenstein, and Randall Schweller for their comments.

**Visit the Annual Reviews home page at [www.AnnualReviews.org](http://www.AnnualReviews.org)**

## LITERATURE CITED

- Adler E, Barnett MN, eds. 1998. *Security Communities*. New York: Cambridge Univ. Press
- Anderson RD Jr. 1993. *Public Politics in an Authoritarian State: Making Foreign Policy During the Brezhnev Years*. Ithaca, NY: Cornell Univ. Press
- Axelrod RM. 1984. *The Evolution of Cooperation*. New York: Basic Books
- Bem D. 1967. Self-perception: an alternative interpretation of cognitive dissonance phenomenon. *Psychol. Rev.* 74:183–200
- Blainey G. 1973. *The Causes of War*. New York: Free
- Boettcher W. 1995. Context, methods, numbers, and words: prospect theory in international relations. *J. Confl. Resol.* 39:561–83
- Brewer MB, Brown RJ. 1998. Intergroup relations. See Gilbert et al 1998, 2:554–94
- Camerer CF, Hogarth RM. 1999. The effects of financial incentives in experiments: a review and capital-labor-production framework. *J. Risk Uncer.* 19:1–3, 7–42

- Checkel JT. 1998. The constructivist turn in international relations theory. *World Polit.* 50:324–48
- Christensen TJ. 1997. Perceptions and alliances in Europe, 1865–1940. *Int. Organ.* 51:65–97
- Christensen TJ, Snyder J. 1990. Chain gangs and passed bucks: predicting alliance patterns in multipolarity. *Int. Organ.* 44:137–68
- Coase R. 1960. The problem of social cost. *J. Law Econ.* 3:1–44
- Coleman JS. 1990. *Foundations of Social Theory*. Cambridge, MA: Belknap Press of Harvard Univ. Press
- Cosmides L, Tooby J. 1994. Origins of domain specificity: the evolution of functional organization. In *Mapping the Mind: Domain Specificity in Cognition and Culture*, ed. LA Hirschfeld, S Gelman, pp. 85–116. New York: Cambridge Univ. Press
- Crawford NC. 2000. The passion of world politics: propositions on emotion and emotional relationships. *Int. Secur.* 24:116–56
- Dawes RM. 1998. Behavioral decision making and judgment. See Gilbert et al 1998, 2:497–548
- Doyle MW. 1983. Kant, liberal legacies, and foreign affairs. Parts 1 and 2. *Phil. Pub. Aff.* 12:205–54, 323–53
- Edwards W. 1962. Dynamic decision theory and probabilistic information processing. *Human Factors* 4:59–73
- Elman MF, ed. 1997. *Paths to Peace: Is Democracy the Answer?* Cambridge, MA: MIT Press
- Farnham BR, ed. 1994. *Avoiding Losses/Taking Risks: Prospect Theory and International Conflict*. Ann Arbor: Univ. Mich. Press
- Farnham BR. 1997. *Roosevelt and the Munich Crisis: A Study of Political Decision-making*. Princeton, NJ: Princeton Univ. Press
- Fearon JD. 1998. Domestic politics, foreign policy, and theories of international relations. *Annu. Rev. Polit. Sci.* 1:289–313
- Fehr E, Schmidt K. 1999. A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114:817–68
- Finnemore M. 1996. Constructing norms of humanitarian intervention. See Katzenstein 1996, pp. 153–85
- Finnemore M, Sikkink K. 1998. International norm dynamics and political change. *Int. Organ.* 52:887–917
- Fiske AP. 1991. *Structures of Social Life: The Four Elementary Forms of Human Relations*. New York: Free (Macmillan)
- Friedman T. 1999. *The Lexus and the Olive Tree*. New York: Farrar Straus & Giroux
- Garthoff R. 1994. *Détente and Confrontation: American-Soviet Relations from Nixon to Reagan*. Washington, DC: Brookings. Rev. ed.
- George AL. 1980. *Presidential Decisionmaking in Foreign Policy: The Effective Use of Information and Advice*. Boulder, CO: Westview
- Gerber A, Green D. 1999. Misperceptions about perceptual bias. *Annu. Rev. Polit. Sci.* 2:189–210
- Gigerenzer G, Goldstein DG. 1996. Reasoning the fast and frugal way: models of bounded rationality. *Psychol. Rev.* 103:650–69
- Gilbert DT, Fiske ST, Lindzey G, eds. 1998. *The Handbook of Social Psychology*. New York: McGraw Hill. 4th ed.
- Gilpin R. 1981. *War and Change in World Politics*. Cambridge: Cambridge Univ. Press
- Goldgeier JM. 1997. Psychology and security. *Secur. Stud.* 6:137–66
- Goldgeier JM, McFaul M. 1992. A tale of two worlds: core and periphery in the post-cold war era. *Int. Organ.* 46:467–91
- Goldstein J, Keohane RO, eds. 1993. *Ideas and Foreign Policy: Beliefs, Institutions and Political Change*. Ithaca, NY: Cornell Univ. Press
- Green DP, Shapiro I. 1994. *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven, CT: Yale Univ. Press
- Grieco JM. 1990. *Cooperation among Nations: Europe, America, and Non-Tariff Barriers to Trade*. Ithaca, NY: Cornell Univ. Press
- Haas EB. 1958. *The Uniting of Europe*. Stanford, CA: Stanford Univ. Press

- Haas EB. 1964. *Beyond the Nation-State: Functionalism and International Organization*. Stanford, CA: Stanford Univ. Press
- Haas PM, ed. 1997. *Knowledge, Power and International Policy Coordination*. Columbia: Univ. So. Carol. Press
- Haas PM, Keohane RO, Levy MA, eds. 1993. *Institutions for the Earth: Sources of Effective International Environmental Protection*. Cambridge, MA: MIT Press
- Herrmann RK, Fischerkeller MD. 1995. Beyond the enemy image and spiral model: cognitive-strategic research after the cold war. *Int. Organ.* 49:415–50
- Herrmann R, Tetlock PE, Diascro M. 2001. How Americans think about trade: resolving conflicts among power, money, and principles. *Int. Stud. Q.* In press
- Janis IL. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. New York: Houghton Mifflin. 2nd ed.
- Jervis R. 1976. *Perception and Misperception in International Politics*. Princeton, NJ: Princeton Univ. Press
- Jervis R. 1984. *The Illogic of American Nuclear Strategy*. Ithaca, NY: Cornell Univ. Press
- Jervis R. 1989. *The Meaning of the Nuclear Revolution: Statecraft and the Prospect of Armageddon*. Ithaca, NY: Cornell Univ. Press
- Kagel JH, Roth AE. 1995. *The Handbook of Experimental Economics*. Princeton, NJ: Princeton Univ. Press
- Kahneman D, Knetsch JL, Thaler RH. 1991. Anomalies: the endowment effect, loss aversion, and status quo bias. *J. Econ. Persp.* 5:193–206
- Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–91
- Kahneman D, Tversky A. 1984. Choices, values and frames. *Am. Psychol.* 39:341–50
- Katzenstein PJ, ed. 1996. *The Culture of National Security: Norms and Identity in World Politics*. New York: Columbia Univ. Press
- Katzenstein PJ, Keohane RO, Krasner SD. 1998. International organization and the study of world politics. *Int. Organ.* 52:645–85
- Kelman HC. 1958. Compliance, identification, and internalization: three processes of attitude change. *J. Confl. Resol.* 2:51–60
- Keohane RO. 1984. *After Hegemony*. New York: Columbia Univ. Press
- Keohane RO, Martin LL. 1995. The promise of institutionalist theory. *Int. Secur.* 20:39–51
- Larson DW. 1985. *Origins of Containment*. Princeton, NJ: Princeton Univ. Press
- Layne C. 1993. The unipolar illusion: why new great powers will rise. *Int. Secur.* 17:5–51
- Lebow RN. 1981. *Between Peace and War: The Nature of International Crisis*. Baltimore, MD: Johns Hopkins Univ. Press
- Legro JW. 1995. *Cooperation under Fire: Anglo-German Restraint during World War II*. Ithaca, NY: Cornell Univ. Press
- Lerner J, Tetlock PE. 1999. Accounting for the effects of accountability. *Psychol. Bull.* 125:255–75
- Levy J. 1992. Prospect theory and international relations: theoretical applications and analytical problems. *Polit. Psychol.* 13:283–310
- March JG, Olsen JP. 1998. The institutional dynamics of international political orders. *Int. Organ.* 52:943–70
- Martin LL, Simmons BA. 1998. Theories and empirical studies of international institutions. *Int. Organ.* 52:729–58
- McDermott R. 1998. *Risk-taking in International Politics: Prospect Theory in American Foreign Policy*. Ann Arbor: Univ. Mich. Press
- Mearsheimer JJ. 1990. Back to the future: instability in Europe after the cold war. *Int. Secur.* 15:5–56
- Mearsheimer JJ. 1994/1995. The false promise of international institutions. *Int. Secur.* 19:5–49
- Mearsheimer JJ. 2001. *The Tragedy of Great Power Politics*. New York: Norton. In press
- Mercer J. 1995. Anarchy and identity. *Int. Org.* 49:229–52
- Moravcsik A. 1998. *The Choice for Europe: Social Purpose and State Power from*

- Messina to Maastricht*. Ithaca, NY: Cornell Univ. Press
- Mueller J. 1989. *Retreat from Doomsday: The Obsolescence of Major War*. New York: Basic Books
- Nagel R. 1995. Unraveling in guessing games: an experimental study. *Am. Econ. Rev.* 85(5): 1313–26
- Nisbett RE, Zukier H, Lemley RE. 1981. The dilution effect: nondiagnostic information weakens the implications of diagnostic information. *Cogn. Psychol.* 13:248–77
- Osgood RE. 1981. *Containment, Soviet Behavior and Grand Strategy*. Berkeley, CA: Inst. Int. Stud.
- Owen JM. 1997. *Liberal Peace, Liberal War: American Politics and International Security*. Ithaca, NY: Cornell Univ. Press
- Pinker S. 1997. *How the Mind Works*. New York: Norton
- Price R. 1997. *The Chemical Weapons Taboo*. Ithaca, NY: Cornell Univ. Press
- Rabin M. 1993. Incorporating fairness into game theory and economics. *Am. Econ. Rev.* 83:1281–302
- Richter JG. 1994. *Khrushchev's Double Bind: International Pressures and Domestic Coalition Politics*. Baltimore, MD: Johns Hopkins Univ. Press
- Risse T, Ropp SC, Sikkink K, eds. 1999. *The Power of Human Rights: International Norms and Domestic Change*. New York: Cambridge Univ. Press
- Ross L, Griffin D. 1991. Subjective construal, social inference, and human misunderstanding. See Zanna 1992, 24:319–59
- Ruggie JG, ed. 1993. *Multilateralism Matters: The Theory and Praxis of an Institutional Form*. New York: Columbia Univ. Press
- Ruggie JG. 1998. *Constructing the World Polity: Essays on International Institutionalization*. London: Routledge
- Russett B, with W Antholis, CR Ember, M Ember, Z Maoz. 1995. *Grasping the Democratic Peace: Principles for a Post-Cold War World*. Princeton, NJ: Princeton Univ. Press
- Sagan SD. 2000. The commitment trap: why the United States should not use nuclear threats to deter biological and chemical weapons attacks. *Int. Secur.* 24(4):85–115
- Schelling TC. 1967. *Arms and Influence*. New Haven, CT: Yale Univ. Press
- Schweller R. 1996. Neorealism's status-quo bias: what security dilemma? *Sec. Stud.* 5: 90–121
- Schweller R. 1998. *Deadly Imbalances: Tripolarity and Hitler's Strategy of World Conquest*. New York: Columbia Univ. Press
- Simmons BA. 1998. Compliance with international agreements. *Annu. Rev. Polit. Sci.* 1:75–93
- Simon HA. 1957. *Models of Man*. New York: Wiley
- Simon HA. 1982. *Models of Bounded Rationality*. Cambridge, MA: MIT Press
- Singer M, Wildavsky AB. 1996. *The Real World Order: Zones of Peace, Zones of Turmoil*. London: Chatham House. Rev. ed.
- Skyrms B. 1996. *Evolution of the Social Contract*. New York: Cambridge Univ. Press
- Snyder J. 1987/1988. The Gorbachev revolution: a waning of Soviet expansionism? *Int. Secur.* 12:93–131
- Snyder J. 1991. *Myths of Empire: Domestic Politics and International Ambition*. Ithaca, NY: Cornell Univ. Press
- Stich S. 1990. *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. Cambridge, MA: MIT Press
- Tannenwald N. 1999. The nuclear taboo: the United States and the normative basis of nuclear non-use. *Int. Organ.* 53:433–68
- Tetlock PE. 1992. The impact of accountability on judgment and choice: toward a social contingency model. See Zanna 1992, 25:331–76
- Tetlock PE. 1998. Close-call counterfactuals and belief system defenses: I was not almost wrong but I was almost right. *J. Pers. Soc. Psychol.* 75:639–52
- Tetlock PE. 1999. Theory-driven reasoning about possible pasts and probable futures: Are we prisoners of our preconceptions? *Am. J. Polit. Sci.* 43:335–66

- Tetlock PE, Belkin A. 1996. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*. Princeton, NJ: Princeton Univ. Press
- Tetlock PE, Boettger R. 1989. Accountability: a social magnifier of the dilution effect. *J. Pers. Soc. Psychol.* 57:388–98
- Tetlock PE, Goldgeier J. 2000. Human nature and world politics: cognition, influence, and identity. *Int. J. Psychol.* 35:87–96
- Tetlock PE, Kristel OV, Elson SB, Green MC, Lerner JS. 2000. The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *J. Pers. Soc. Psychol.* 78:853–70
- Tetlock PE, Lerner J. 1999. The social contingency model: identifying empirical and normative boundary conditions on the error-and-bias portrait of human nature. In *Dual Process Models in Social Psychology*, ed. S Chaiken, Y Trope, pp. 571–85. New York: Guilford
- Tetlock PE, McGuire C, Peterson R, Feld P, Chang S. 1992. Assessing political group dynamics: a test of the groupthink model. *J. Pers. Soc. Psychol.* 63:402–23
- Thaler RH. 2000. From homo economicus to homo sapiens. *J. Econ. Pers.* 14:133–41
- Thompson L. 1998. *The Mind and Heart of the Negotiator*. Upper Saddle River, NJ: Prentice Hall
- Vertzberger Y. 1990. *The World in Their Minds*. Stanford, CA: Stanford Univ. Press
- Walt SM. 1987. *The Origins of Alliances*. Ithaca, NY: Cornell Univ. Press
- Walt SM. 1996. *Revolution and War*. Ithaca, NY: Cornell Univ. Press
- Waltz KN. 1959. *Man, The State and War*. New York: Columbia Univ. Press
- Waltz KN. 1979. *Theory of International Politics*. Reading, MA: Addison-Wesley
- Welch D. 1993. *Justice and the Genesis of War*. New York: Cambridge Univ. Press
- Wendt A. 1999. *Social Theory of International Politics*. New York: Cambridge Univ. Press
- Williamson O. 1965. A dynamic theory of interfirm behavior. *Q. J. Econ.* 79:579–607
- Wohlforth WC. 1993. *The Elusive Balance: Power and Perceptions During the Cold War*. Ithaca, NY: Cornell Univ. Press
- Wohlforth WC. 1999. The stability of a unipolar world. *Int. Secur.* 24:3–41
- Zanna M, ed. 1992. *Advances in Experimental Social Psychology*. New York: Academic



## CONTENTS

TRANSNATIONAL POLITICS: Contention and Institutions in International Politics, <i>Sidney Tarrow</i>	1
MECHANISMS IN POLITICAL PROCESSES, <i>Charles Tilly</i>	21
DEMOCRATIZATION AND ECONOMIC REFORM, <i>V. Bunce</i>	43
PSYCHOLOGY AND INTERNATIONAL RELATIONS THEORY, <i>J. M. Goldgeier, P. E. Tetlock</i>	67
POLITICAL TRADITIONS AND POLITICAL CHANGE: The Significance of Postwar Japanese Politics for Political Science, <i>Bradley Richardson, Dennis Patterson</i>	93
RELIGION AND COMPARATIVE POLITICS, <i>Anthony Gill</i>	117
TOWARD A FOURTH GENERATION OF REVOLUTIONARY THEORY, <i>Jack A. Goldstone</i>	139
POLITICAL CONSEQUENCES OF MINORITY GROUP FORMATION, <i>M. Hechter, D. Okamoto</i>	189
POLITICAL KNOWLEDGE, POLITICAL ENGAGEMENT, AND CIVIC EDUCATION, <i>William A. Galston</i>	217
THEORIES OF DELEGATION, <i>J. Bendor, A. Glazer, T. Hammond</i>	235
TIME-SERIES-CROSS-SECTION DATA: What Have We Learned in the Past Few Years?, <i>Nathaniel Beck</i>	271
THE ORIGINS AND WAR PRONENESS OF INTERSTATE RIVALRIES, <i>John Vasquez, Christopher S. Leskiw</i>	295
THE POLITICAL ECONOMY OF INTERNATIONAL MONETARY RELATIONS, <i>J. Lawrence Broz, Jeffrey A. Frieden</i>	317
BIOLOGY AND POLITICS: Linking Nature and Nurture, <i>R.D. Masters</i>	345
WOMEN'S MOVEMENTS AT CENTURY'S END: Excavation and Advances in Political Science, <i>Karen Beckwith</i>	371
TAKING STOCK: The Constructivist Research Program in International Relations and Comparative Politics, <i>Martha Finnemore, Kathryn Sikkink</i>	391
VIACRATIC AMERICA: Plessy on Foot v. Brown on Wheels, <i>Douglas W. Rae</i>	417