

HANDBOOK
of
INTERNATIONAL RELATIONS

Edited by
WALTER CARLSNAES, THOMAS RISSE
and BETH A. SIMMONS



SAGE Publications
London • Thousand Oaks • New Delhi

Rationalism v. Constructivism: A Skeptical View

JAMES FEARON AND ALEXANDER WENDT

In the introduction to the fiftieth anniversary issue of *International Organization*, Peter Katzenstein, Robert Keohane and Stephen Krasner (1998) suggest that the main axis of debate in the field of international relations (IR) in the coming years is likely to be rationalism versus constructivism.¹ In at least one important respect, this would be a remarkable development. For whatever they are, rationalism and constructivism are not in the first instances theories of international politics. Rather, rationalism seems to refer to a methodological approach that may imply a philosophical position on what social explanation is and how it ought to work, the nature of which is debated. And constructivism seems to refer to a set of arguments about social explanation that may imply preferences over specific questions and methods of social inquiry, the nature of which are debated. If the field does focus on rationalism versus constructivism, then the central debate in IR will not be about international relations but rather about how to study international relations.

To be sure, the concern in IR with questions of method and philosophy of social science has precedents, in recent arguments over positivism and post-positivism and in the earlier debate over behavioralism versus historical traditionalism.² The terms of the emerging debate between rationalism and constructivism are different than these, but the concern with second- rather than first-order issues is similar.

One may reasonably ask whether progress in understanding international relations and improving human (and planetary) welfare is best served by structuring the field of IR in this way, as a battle of analytical paradigms. At the very least it can encourage scholars to be method-driven rather than problem-driven in their research, which may result

in important questions or answers being ignored if they are not amenable to the preferred paradigmatic fashion. For this chapter, however, we leave this important question aside. Supposing that 'rationalism v. constructivism' does orient some debate in IR in the coming years, we ask what the contrast amounts to. What are the 'isms' referred to? And do the differences between them provide grounds for a war of paradigms?

We answer the last question mainly in the negative. Although there are some important differences between the two approaches, we argue that there are also substantial areas of agreement, and where genuine differences exist they are as often complementarities as contradictions. Our objective is not to suggest that there is no need for discussion, or that rationalism and constructivism should or could be synthesized into one perspective. It is to suggest, rather, that the most interesting research is likely to be work that ignores zero-sum interpretations of their relationship and instead directly engages questions that cut cross the rationalist/constructivist boundary as it is commonly understood.³

A key argument towards this conclusion is that, in our view, rationalism and constructivism are most fruitfully viewed *pragmatically* as analytical tools, rather than as metaphysical positions or empirical descriptions of the world.⁴ Since the ontological and empirical interpretations of the debate seem more common in the literature and lead to more zero-sum pictures, it may be useful to explain briefly why we resist them.

The ontological reading treats rationalism and constructivism as sets of assumptions about what social life is made of and what kinds of relationships exist among these elements. For example,

from this perspective rationalism is usually seen as assuming an individualist ontology, in which wholes are reducible to interacting parts, and constructivism as assuming a holist ontology, in which parts exist only in relation to wholes. In each case, certain empirical arguments and analytical tools are prescribed or proscribed a priori as legitimate or illegitimate, scientific or unscientific, and thus the stage is set for a genuine war of paradigms.

It is important to understand these ontological issues, since failure to do so can lead to analytical tools or frameworks becoming *tacit* ontologies (Ruggie, 1983: 285), foreclosing potentially interesting lines of argument without justification. However, we do not believe this framing of the rationalist-constructivist debate is the most useful, for three reasons. First, the issues are by definition philosophical, and as such not likely to be settled soon, if ever, and almost certainly not by IR scholars. Second, although some rationalists and constructivists may in fact have strong ontological commitments, others may not, since there is no inherent *need* to commit to an ontology to work in these traditions. Just as quantum physicists can do their work without any idea how to interpret its ontological implications, social scientists too can proceed pragmatically, remaining agnostic about what society is 'really' made of. Finally, it seems doubtful that as a discipline we know so much about international life that we should rule out certain arguments a priori on purely philosophical grounds. Thus, while recognizing the role that ontological issues play in structuring the rationalist-constructivist debate, in this chapter we will largely avoid them, adopting a stance of ontological pluralism instead.⁵

A second way to frame the debate is in empirical terms, as a disagreement about substantive issues in the world like how often actors follow a logic of consequences or logic of appropriateness, or whether preferences really are exogenous or endogenous to a given social interaction. We explore some of these questions below, and find some genuine differences, but this too is not our preferred way to proceed. First, in their purest, most stripped-down forms, neither approach makes many interesting empirical predictions about the world. To a large extent it is only with the addition of auxiliary assumptions – a particular theory of preferences, for example – that such predictions emerge. Moreover, although one *can* interpret an assumption of, say, exogenous preferences, as a factual claim about a certain social system, there is no *need* to do so. It is perfectly legitimate to view it instead as merely a methodological convenience necessitated by the fact that one cannot study everything at once. As in the ontology case, there is always a danger here that analytical assumptions will become tacit empirical ones, but given sufficient methodological self-consciousness this problem can be avoided.

This brings us to the pragmatic interpretation of rationalism and constructivism, as analytical tools or lenses with which to theorize about world politics. Analytical lenses do not in themselves force the researcher to make ontological or empirical commitments. What makes a comparison of them interesting none the less is that they view society from opposing vantage points – roughly speaking, rationalism from the 'bottom-up' and constructivism from the 'top-down'. As a result they tend in practice to ask somewhat different questions and so bring different aspects of social life into focus. It would be surprising if this did not lead to different pictures of world politics, and thus to 'paradigmatic' debate about what world politics is really like. Emphasizing these differences would have been one way to write this chapter. Yet, in IR today there is ample perception already of conflict between rationalism and constructivism, in our view much of it unnecessary or ill-founded, based either on treating them in ontological or empirical terms or on misunderstandings about what they entail. Moreover, we are also struck by two areas of potential convergence that are insufficiently appreciated. First, the two approaches often yield *similar*, or at least complementary, accounts of international life. This redundancy may arise because in the end they are studying the same underlying reality. Second, even though their respective vantage points tend in practice to highlight some questions and not others, in many cases there may be much to be gained by using the tools of one to try to answer questions that tend to be asked primarily by the other. Such a cross-paradigmatic exchange of characteristic questions and answers is in our view the most fruitful way to advance not only these two research agendas, but more importantly, our understanding of world politics. With these considerations in mind, we shall write this chapter with a view toward deconstructing some of the supposed contradictions between the two approaches, and highlighting the convergences. Again, this is not to suggest that there are no differences, many of which we discuss below. But once viewed in an analytical, tool-kit fashion, we believe many putative disputes lose much of their force.

From this pragmatic stance, we seek to clarify what each approach brings to the table and how they relate. In standard representations of the debate, two issues seem most at stake: (1) whether and how ideas matter in world politics, and (2) the relationship between international actors and the structures in which they are embedded. These are large issues and we do not attempt a full discussion of both here. We will argue, however, that on the first issue there is considerably less difference than is often thought. Rationalism is sometimes portrayed as emphasizing material as opposed to ideational factors, but this misunderstands what is entailed by the approach.

This allows us to focus most of our attention on the second issue. Here it seems useful to begin by dividing the debate into two issues, conceptions of structure and conceptions of agency. In the literature it has become increasingly common to assume that one approach is agent-centric and one structure-centric,⁶ but this can be misleading insofar as it suggests that one is only about agents and the other only about structures. In fact, both have an agentic and a structural aspect.

Although we believe that there is much useful work to be done thinking through the structural side of the debate, given space constraints we shall set this issue aside in favor of an approach to the problem through the agency side.⁷ We do so in part because there seems to be more interest in contemporary IR in agency than in structure. But this focus also makes sense given that constructivism entered the field in part by criticizing the rationalist view of agency as being exogenous to structure, and constructivism is now in turn being criticized for lacking a theory of agency.⁸ As such, even though what follows neglects some important issues in the rationalist/constructivist debate, we do hope it will speak to one of its major concerns so far.

The next section provides brief overviews of the positions typically denoted by 'rationalism' and 'constructivism'. The rest of the chapter examines five common ways of characterizing the debate with respect to the agent side of the agent-structure problem: material versus ideational; logics of consequences versus logics of appropriateness; norms as useful versus norms as right; exogenous versus endogenous actors, when this is understood in causal terms; and exogenous versus endogenous actors, when understood in constitutive terms.

This list by no means exhausts the possibilities for discussion. A particularly interesting one that we take up only in passing is the role of 'performativity' in the constitution of actors. Although associated with postmodernism and thus constructivist in a broad sense, a divide has emerged between postmodernism and constructivism as this term has come to be defined in IR, and as such performativity has not figured in the rationalist/constructivist debate *per se*.⁹ We hope the implications of performativity for this debate will be addressed soon.

OVERVIEW OF RATIONALISM AND CONSTRUCTIVISM

Rationalism

As used in IR contexts, 'rationalism' seems to refer variously to formal and informal applications of rational choice theory to IR questions, to any work drawing on the tradition of microeconomic theory from Alfred Marshall to recent developments in

evolutionary game theory, or most broadly to any 'positivist' exercise in explaining foreign policy by reference to goal-seeking behavior. In the first two senses, rationalism can be characterized as a method, that is, as a cookbook or recipe for how to explain actions (and especially actions taken in a strategic or multi-actor context). The recipe may be summarized as follows.

- 1 One starts with an action or pattern of actions to be explained.¹⁰ Some international relations examples would include decisions to send troops into battle, the formation of a balance of power, an arms race, currency devaluations, the imposition of a tariff, or protests at the meetings of an international organization.
- 2 One posits a set of actors with the capacity to take the actions in question, and probably others who can take actions that may bear on their considerations. Especially in IR, this step typically involves simplification and abstraction (for example, states or international organizations as actors instead of individuals).
- 3 One proposes a structure of interaction, a sequence of choices for the actors identified in (2), that embeds the pattern of actions to be explained in a larger universe of possibilities. For instance, to explain an arms race one needs a structure of interaction that allows, in principle, for actions that would not produce an arms race. In any structure of interaction some elements will be taken as exogenous for the purpose of the analysis. Exogenous elements of a rationalist model or argument are features not explained within the argument, such as (mostly) preferences over outcomes, the beliefs that actors hold at the start of the posited interaction, or technological capabilities (such as the time it takes to move X troops Y miles given Z terrain). Endogenous features are things explained within the model/argument, such as actor preferences over actions, and beliefs about other actors in light of their actions. The variation of exogenous elements – for instance the relative size of two armies or a state's current account balance in a specific model – allows statements about when the (endogenous) outcome to be explained is more or less likely.
- 4 Either (a) one makes arguments about the actors' preferences over the universe of possible outcomes identified in (3). Or (b), in evolutionary game-theoretic models/arguments, one may posit rules for how different outcomes associate with the differential reproduction of actors and their 'programs' or decision rules in subsequent rounds. In addition, in some models/arguments, one may posit a structure of initial beliefs held by the actors in question.
- 5 One shows how or under what conditions the outcome or pattern of actions in question would

emerge if the actors were choosing rationally in light of their beliefs and the other actors' choices, or as the result of the longer-term selection of decision rules or habits in more evolutionary models.¹¹ In game-theoretic models, beliefs are themselves subject to criteria of rational revision, both through Bayesian and more boundedly rational forms of learning.

Steps 2, 3 and 4 (which need not occur in any particular order) are all both empirical and theoretical. They are empirical in that each is open to criticism on the grounds of being consequentially unrealistic. For instance, a critic might reasonably say 'it is empirically implausible that in your argument state A has no opportunity to concede the territory at stake rather than fight, and this restriction drives your conclusions'.¹² They are theoretical in that each involves the creative simplification and re-presentation of a complex reality in schematic form, which may or may not yield valuable insights and clarity. In other words, this is art as much or more than science. Following the recipe does not guarantee tasty or filling results.

Two Common Misunderstandings about Rationalism

Scientism and the status of formal models The recipe for explanation described above may be pursued informally, in ordinary language, or formally, in the language of game or decision theory. In the latter case, many scholars infer from the appearance of mathematics and symbols that the 'rationalist' must believe that there are no fundamental differences between social and natural science, and that social science can and should aspire to be the same as, say, theoretical physics. While there may be scholars who believe this, the position is definitely not entailed by the use of formal models in the microeconomic tradition. Quite the contrary, the rationalist recipe described above embraces intentionality and the explanation of actions in terms of beliefs, desires, reasons and meanings.¹³ Models in physics purport to describe invariant laws governing a world of inanimate objects to which we have no immediate, intuitive access. Microeconomic models purport to show how initially puzzling patterns of action may emerge from individual choices that make sense (are comprehensible) to us in light of the beliefs, desires and constraints they face. They are a form of ethnography more than an effort to find equations that govern putative laws of behavior.

Here is another way to make this point: a formal model is just an argument. Models in the rationalist tradition are arguments that formalize and explore the collective consequences of a fundamental principle of folk psychology – that actions are

explained by showing how they make sense in light of particular beliefs and desires.¹⁴ Some users of game models present them instead as if they were magic boxes. Assumptions go in at one end and (presto!) out come hypotheses and results at the other end, with little or no attention paid to explaining what is happening in between to make the connection. In our view, this is bad practice. It should always be possible to translate the action within a model of intentional decision-making into readily comprehensible, ordinary language terms, and it is incumbent on users of such models to do so.

'Rationalism' is a moving target Rationalist analyses in IR have drawn and continue to draw heavily on the results of the evolving program of microeconomic theory.¹⁵ It is worth pointing out that the makers of this program, mainly economic theorists, have spent little time reflecting on the defining features or philosophical foundations of their approach and its results. Instead, efforts to delimit 'rationalism' (or under other names) almost invariably come from without, from philosophers, sociologists, psychologists and political scientists. One consequence is that these efforts sometimes become dated as microeconomic theorists progress from one set of problems to another. It is difficult and hazardous to try to define rationalism by picking out a set of core assumptions said to characterize all past and all future 'rationalist' work.

Two examples are useful to develop the point. First, in the late 1950s Herbert Simon persuasively delimited rationalism by the assumption of perfectly rational agents, whom he characterized as possessing (1) perfect information and (2) a perfect ability to perform calculations. Beginning (more or less) with Akerlof's work on the 'lemon's problem' in 1970 and continuing through the development of incomplete information game theory in the 1980s, microeconomic theory developed a powerful set of tools for explaining action in contexts where people lack knowledge about some important aspect of the situation they face. These developments helped clarify that imperfect information did not entail 'bounded rationality', as Simon had assumed, but rather that bounded rationality should be identified with an imperfect ability to perform calculations, to remember or envision states of affairs.

Second, it is striking that for roughly a decade the cutting edge of microeconomic theory has been devoted *almost entirely* to models in which the actors are less than fully rational in this second, more narrow sense.¹⁶ The 1980s had seen an explosion of work exploring the meaning and consequences of the rationality assumption in models with imperfectly informed agents. By the early 1990s this research had either answered its questions, pushed them as far as they could reasonably go, or had made older questions (such as the problem of equilibrium selection) more pressing.

This effective completion of a research program in non-cooperative game theory with rational agents led directly into work on evolutionary models with agents who use fixed decision rules or adjust their actions in some myopic, boundedly rational way.¹⁷

Notably, no economic theorist has decried this shift away from models about fully rational agents as a betrayal of the premises of 'rationalism' (so far as we know). Rather, the spirit of the research has been, as before, 'what happens in a model of this problem if one makes such-and-such assumptions?' In addition, many of the main results from evolutionary and bounded rationality models show how less-than-full rationality assumptions often yield similar or identical aggregate implications as did the earlier models with full rationality. If the shift reveals anything about the philosophical commitments of the current microeconomic program, it makes clear that the core commitment is to the formalization of arguments to explore the consequences of different assumptions rather than to any particular assumption of actor rationality.

This second example raises the question of what exactly rationality has to do with 'rationalism'. Certainly it would be odd to define rationalism as whatever it is that the users of microeconomic theory do, and certainly the program's point of departure has long been the explanation of action by reference to optimality in light of beliefs and desires.¹⁸ But perhaps a more plausible candidate for a constitutive feature of rationalism is a commitment to explaining macro-social phenomena in terms of more micro-level phenomena – as Thomas Schelling (1978) suggested, going from 'micromotives' to 'macrobehavior'.¹⁹ Note that in the rationalist recipe posed above, one might employ all sorts of rationality assumptions, or perhaps none at all, as in a model that explained an aggregate outcome such as a balance of power by reference to the differential selection of culturally or otherwise given state 'programs'. By contrast, the recipe is more fundamentally characterized by an effort to explain a whole – an outcome, or pattern of actions – in terms of component parts.

This does *not* imply that 'rationalism' lacks any account of how macro-level phenomena, such as social structures, impinge on and even 'socially construct' individual actors. Indeed, the point of equilibrium analysis is to elaborate how certain 'macro' structures select for or create incentives for individual actions that in turn constitute the structure. We briefly take up such accounts and their relationship to constructivism in the penultimate section of the chapter. But it is still fair to say that as a provisional starting point, rationalist analyses begin from the micro level and try to work to the macro level.

It is not clear whether 'from micro to macro' is or needs to be an expression of an ontological commitment to some form of methodological individualism, or just a pragmatic proposal about a potentially

productive way to pursue social explanation. Some rationalists seem inclined toward the former position,²⁰ in which case the rationalist/constructivist debate is pushed into the domain of metaphysics and the game becomes more winner-take-all. Such 'ontologizing' of rationalism, whether tacit or explicit, worries many constructivists (like Wendt), and this concern underlies some of their critical reaction to the rationalist research program. However, many users of rationalist methods (like Fearon) see no need to make broad metaphysical claims about this approach, and want simply to explore its implications for social explanation. This pragmatic stance does not rule out constructivist approaches to social explanation a priori and as such does much to deflate any notion of a new 'Great Debate'.

Constructivism

Like rationalism, constructivism can be seen in either ontological, empirical, or analytical terms. In either case, however, it is not a substantive theory of world politics. This is important to note because constructivism has sometimes been identified with the latter, and then compared to bona fide theories of world politics like realism or liberalism.²¹ This is problematic. As in rationalist IR scholarship, in its constructivist counterpart one can find state-centric and non-state-centric theories, second-image and third-image theories, pessimistic and optimistic theories, and so on. As such, there is a great deal of variation on substantive issues within constructivist IR, and indeed, given the often self-consciously political character of constructivist scholarship, these issues are if anything even more intense sources of disagreement than they are among rationalists. Fortunately, we do not need to address these substantive variations here, but let there be no mistake up front that when it comes to the content and nature of international politics, constructivism is not a 'theory' at all, any more than is rationalism.

Even narrowing the focus in this way, however, there seems to be considerably less agreement among constructivists than among rationalists. Thus whatever the risks of stipulating a single 'recipe' for rationalist research above, it is even more difficult to do so here. This is an important part because whereas rationalists generally agree on questions of epistemology, the debate over the nature of knowledge and truth claims is very much alive within constructivist IR. Indeed, since rationalists have tended not to have deep epistemological qualms about social science and thus not to see the point of debating epistemology in the first place, it seems fair to say that it is primarily within constructivism that these questions are being argued.²²

In particular, constructivist IR scholarship is currently divided on at least two epistemological

questions: (1) whether knowledge claims about social life can be given any warrant other than the discursive power of the putative knower (call this the relativism issue); and (2) whether causal explanations are appropriate in social inquiry (the naturalism issue). These questions are partially independent, which has allowed three distinct epistemological positions to emerge within constructivism: a 'positivist' position that answers yes to both questions; an 'interpretivist' position that answers yes and no respectively; and a 'postmodern' position that (seems to?) answer no to both. These epistemological differences are deep and sufficiently contentious as to raise the question of whether one can speak of 'constructivism' in the singular at all. Moreover, in the eyes of many constructivists epistemological positions have implications for the ontological and analytical questions they believe are at stake in arguments over rationalism and constructivism, making it impossible or illegitimate to separate them. So on the constructivist side at least, the basic parameters of the 'rationalism v. constructivism' issue are essentially contested.

This leaves the authors two options in handling constructivism: we can either show how epistemological and ontological issues are or are not intertwined, or we can bracket that nexus and focus just on the latter. Ultimately we believe that the first option needs to be pursued. Some of the issues in the rationalist/constructivist contrast do involve questions of what it means to know something and how this is possible. However, in our view this does not mean that the epistemological differences are the whole of the matter – the rationalist/constructivist contrast does not reduce to positivism versus post-positivism. Given its less developed state, therefore, as well as the limitations of time and space, we adopt the second option above, although some differences among constructivists with respect to rationalism arising from epistemological disputes emerge below. In bracketing epistemological questions in this chapter, however, we emphasize that the result is only a partial and debatable interpretation of 'constructivism'.

So, what generalizations can be made about constructivism? To start with the obvious, constructivists are interested in how the objects and practices of social life are 'constructed', and especially those that societies or researchers take for granted as given or natural. Naturalization is problematic because it obscures the ways in which social objects and practices depend for their existence on ongoing choices, and as such it can be oppressive and a barrier to social change. However, while the general purpose of de-naturalizing a previously unquestioned object or practice – for example, power politics, ethnic identity, or sovereignty – is therefore to open up possibilities for progressive transformation, it need not have that effect. In some cases actors may decide that a practice should not

be changed, but if so at least its acceptance would then be more self-conscious and democratically accountable.

One can identify at least four characteristic and inter-related features of constructivist thinking about the construction of social objects and practices. First, constructivism is centrally concerned with the role of ideas in constructing social life. These ideas will often be shared by many people, and in order to have social relevance they need to be instantiated in practices, which on both counts means they may have considerable objectivity, facticity, or 'materiality'. Constructivism is not subjectivism or pure idealism. Instead, the emphasis on ideas is meant to oppose arguments about social life which emphasize the role of brute material conditions like biology, geography and technology. This is not to say that these have no role whatsoever, but rather that their impact is always mediated by the ideas that give them meaning.

Second, constructivism is concerned with showing the socially constructed nature of agents or subjects. Rather than taking agents as givens or primitives in social explanation, as rationalists tend to do (though see below), constructivists are interested in problematizing them, in making them a 'dependent variable'. This concern operates on two levels. On the more superficial level the focus is on the causal processes of socialization by which particular agents acquire their identities and interests. On a deeper level, constructivists are concerned with the constitutive conditions of possibility for certain modes of subjectivity in the first place. Some of these conditions are historical, in the sense that understandings of what it means to be an agent may change over time, and thus are culturally relative rather than reducible to universal features of human beings' biological constitution. In modern liberal society, for example, we often take it for granted that agents are 'individuals' with powers of reason, autonomy and responsibility. But as John Meyer and Ron Jepperson, among others, have shown, this is very much a culturally specific way of thinking about subjectivity.²³ Other conditions of possibility for subjectivity are synchronic, in the sense that the ideas that actors have in their heads about what they want to do depend for their content or meaning on discursive structures shared with other actors, so that one cannot be a certain kind of subject – say, a witch-doctor – unless others in the society make that possible. At stake in all this is partly an explanatory question of whether social forms can be adequately understood by starting with given agents, but also ultimately a political question of whether society can be normatively grounded on the liberal conception of the individual as some kind of natural baseline.

Third, constructivism is based on a research strategy of methodological holism rather than methodological individualism. In a strict form,

methodological individualism requires that explanations in social science be reducible in the last analysis to 'micro-foundations', which is to say statements about ontologically primitive individuals and/or their interactions. For various reasons holists argue that this effort must ultimately fail, and so we need to make social wholes and internal relations rather than individuals the primitives in social scientific explanation.²⁴ The commitment to holism, while related to the second point above about endogenizing the given individual, is not the same thing. As will become clear below, it is possible to explain certain aspects of agents' subjectivity in ways that do not violate the individualist requirement of reducibility. (Note that this is not to suggest that rationalism is necessarily individualistic. Whether rationalism is committed to a reducibility requirement depends on how it is interpreted: if as an ontology then probably yes, if only as a pragmatically useful strategy of social explanation, then no.)

Finally, what ties the three foregoing points together is a concern with constitutive as opposed to just causal explanations.²⁵ Causal theorizing seeks to establish the necessary and sufficient conditions relating a pre-existing cause to a subsequent effect in a more or less mechanistic way. An assumption of such theorizing, therefore, is that cause and effect are independently existing phenomena. Constitutive theorizing, in contrast, seeks to establish conditions of possibility for objects or events by showing what they are made of and how they are organized. As such, the object or event in question is an 'effect' of the conditions that make it possible, but it does not exist independent of them. A common example illustrating this point is the master-slave relationship. The nature and meaning of 'master' and 'slave' as modes of subjectivity are constituted by their relationship in the sense they cannot *be* 'masters' and 'slaves' except in relation to the other. This highlights the way in which social relations can be a primitive in analysis, or irreducible to propositions strictly about pre-existing individuals. This is not to say that constructivists, particularly on the 'positivist' wing, are uninterested in causal explanations. After all, masters and slaves are also effects of shared ideas in the causal sense that their identities and interests are generated and sustained by the interaction between them. But the constitutive aspect of constructivist scholarship is more distinctive.²⁶

BONES OF CONTENTION?

We turn now to the substance of the rationalist-constructivist debate. The discussion takes place on many fronts simultaneously and continually evolves, and as such it is difficult to know what all

the issues are, much less cover them all in any detail. However, as we indicated at the outset, one way of organizing the terrain and slicing off a manageable piece for this chapter is to distinguish questions about agents from questions about structure. In this chapter we address only the former. Here there seem to be at least five ways of characterizing what 'rationalism v. constructivism' is about. Although judging from the literature it may seem that they are equally divisive, we argue that this is not the case. Some involve genuine rival hypotheses about what is going on in social contexts; others involve differences in emphasis or research question, and as such more complementarity than contradiction; and still others seem to involve hardly any difference at all.

Material versus Ideational

It is not uncommon in the literature to see the rationalist-constructivist divide characterized in terms of the former being about material factors and the latter being about ideas. Put into the frame of this chapter's concern with conceptualizations of agency, this often translates into the proposition that rationalists believe that people are always acting on material self-interest, and constructivists believe that people are always acting on the basis of norms or values. Whatever the relative merits of self-interest versus non-self-interest descriptions of actor motivation, we think that seeing this as an issue of material conditions versus ideas is not very useful. The problem here lies more in the perception of rationalism than of constructivism.

Constructivism is correctly seen as defined in part by opposition to materialism. The character of this opposition depends on how materialism is understood, which we shall not get into here, and it should not be over-stated. Constructivism does not imply a radical, 'ideas *all* the way down' idealism which denies any role whatsoever to material considerations. As John Searle points out, brute (material) facts are logically prior to institutional facts (Searle, 1995: 34-5). And neither does constructivism imply that the ideational structures of social life are not objective or real. 'Material' is not the same thing as 'objective'. But given those qualifications, John Ruggie is correct to say that constructivism emphasizes the role of consciousness in social life, Emanuel Adler to stress its focus on Popper's World 3 of shared understandings, and so on (Adler, 1997; Ruggie, 1998: 856). Material factors matter at the limit, but *how* they matter depends on ideas.²⁷

The picture is more complicated on the rationalist side. On the one hand, proponents and critics alike have sometimes associated rationalism either tacitly or explicitly with materialism. In their influential treatment of the role of ideas in foreign

policy, for example, Judith Goldstein and Robert Keohane argue that explanations emphasizing ideas are 'rivals' to the 'rationalist' concern with explanations emphasizing preferences (Goldstein and Keohane, 1993: 4). This putative rivalry suggests that preferences or interests²⁸ are not themselves ideas and thus, presumably, material.²⁹ Similarly, Ruggie's point about constructivism's focus on ideas is meant as a contrast with rationalism, which he argues does not encompass 'normative factors' and treats ideas either not at all or only 'secondarily' (Ruggie, 1998: 864, *passim*).

These associations of rationalism with materialism may stem from the sociology of knowledge of how rational choice theory entered IR. Early in the 1960s, it was seen as a useful way of exploring the logic of nuclear deterrence and military strategy more generally.³⁰ Since these intellectual enterprises were influenced by political realism, and realism gives pride of place to material power in international politics, it was perhaps natural for 'rationalism' to acquire a materialist connotation. This may have been reinforced by the publication in 1979 of Kenneth Waltz's *Theory of International Politics*, whose neorealism combined an implicitly materialist definition of system structure with microeconomic analogies for thinking about the logic of anarchy (Waltz, 1979). (We say 'implicitly' because Waltz does not actually defend materialism or argue that ideas do not matter. Rather, he suggests that due to evolutionary pressures in a self-help system, perceptions will tend to reflect the reality of who has the material power to hurt whom, which leads to his equating international structure with the distribution of material capabilities.) Finally, facing what they saw as a disciplinary hegemony of rationalist realists, postmodern and constructivist critics in the 1980s and early 1990s failed to disentangle the two strands, reinforcing the perceived materialist bent of rational choice theory.³¹

It is true that in the hands of rationalists who were influenced by materialist conceptions of politics the explanatory role of ideas has tended to be ignored or minimized. But this should be seen as a function of materialist commitments, not rationalism. At least three considerations bear on this conclusion.

First, rationalist explanations are a species of intentional explanation, the basic structure of which is the formula, 'Desire + Belief = Action'.³² This means that at their core – the level of individual choice – ideas are an essential, not just secondary, element of rationalist explanation. Second, at the level of social interaction, game theory typically explains aggregate outcomes by reference to 'equilibria', which are *made up of* patterns or structures of beliefs that satisfy various stability properties.³³

Finally, it is not clear in what sense even desires are necessarily material. Some desires may be material in the sense of having a biological basis,

like the desire for food or sleep. But in what sense is, say, a desire to get tenure material? It is in the sense that getting tenure will result in pecuniary rewards. On the other hand, the fact that one sees oneself (and is seen by others) as a professor, such that one could plausibly have a preference for tenure in the first place, seems more a fact about ideas than biology. At some level there is always a material base to desire because human beings are physical creatures, but in most cases this base is 'directionless' in the absence of ideas that give it content (Howe, 1994). This does not violate the 'Desire + Belief' model, since there is nothing in the model which requires that 'Desire' be material. It *may* be material, but then again it may not; rational choice theory, as a theory of choice given desires and beliefs, is strictly speaking agnostic on this question.

Two lessons can be drawn from this discussion. One is that there is little difference between rationalism and constructivism on the issue of *whether* ideas 'matter'. Constructivists might criticize rationalists for the way in which they study ideas, for example by imputing the content of actors' consciousness on the basis of some deductive theory rather than proceeding inductively from what kind of ideas actors really do have, and rationalists might criticize constructivists for failing to explicate whether and how a given pattern of actions and a system of beliefs are mutually reinforcing.³⁴ But that is a different point. The logic of both approaches depends crucially on actors making choices on the basis of their beliefs. That there should be no fundamental difference here makes sense if we pause to consider that rationalism and constructivism can both trace roots to Weber. The differences between 'intentional' and 'interpretive' explanations notwithstanding, they also have a lot in common.³⁵

The other lesson follows from the first, which is that when rationalist models *do* seem to downplay ideas (and sometimes they do) this is a function of materialism rather than rationalism. John Ferejohn's distinction between 'thin' and 'thick' rationalist models is useful here (Ferejohn, 1991). The thin model is simply the logic of intentional explanation referred to above, which says nothing about the content of desires or beliefs. The thick model adds assumptions about the content of desires and beliefs (for example, 'self-interest' and 'complete information about preferences') in a given case. Where those assumptions come from, however, is a question about which rationalism in the thin sense is agnostic. There are materialist rationalisms and idealist rationalisms, and as such if we want to debate the relative importance of ideas in social life it makes more sense to focus on thick theories of interest than it does on rationalism and constructivism.

All of this is not to suggest that there are no significant differences between rationalism and constructivism with respect to *how* they think ideas

matter. In particular, as we noted above, rationalists tend to draw a clear distinction between ideas/beliefs and desires or preferences. This may be related to a more basic feature of rationalist thinking about ideas, which one of the authors here thinks is to treat their explanatory role in more causal than constitutive terms. Ideas are a causal mechanism like any other, existing independently of other causal mechanisms and explaining some portion of the variance in actors' behavior. Constructivists, on the other hand, tend to emphasize the constitutive role of ideas, the ways in which ideas give other factors the explanatory role that they have by investing them with meaning and content. From this perspective ideas permeate social life rather than form a distinct variable whose explanatory force can be isolated. This may be an important difference, to which we return below. But it is a difference *within* a largely shared agreement that ideas 'matter'.

The Logic of Consequences versus the Logic of Appropriateness

Another typical way of interpreting 'rationalism v. constructivism' in IR is in terms of the contrast between *homo economicus* and *homo sociologicus*. The former is a calculating machine who carefully assesses different courses of actions, choosing whichever provides the most efficient means to her ends. The latter is a rule-follower who acts out of habit or decides what to do by posing the question 'how is a person in my role (or with my identity) supposed to act in this circumstance?' In March's terms, *homo economicus* is said to follow a 'logic of consequences' in her mode of decision-making, while her sociological counterpart follows a 'logic of appropriateness'.³⁶

Partisans face powerful temptations to reduce one logic to an instance of the other. Economists are apt to see the logic-of-appropriateness as consequentialist – it is simply that the consequence of concern is conformity of one's actions with a set of norms or an identity. Economists stress that nothing in their approach prevents taking desires (or 'interests' or 'preferences') as being informed by or based on norms. On the other side, sociologists are apt to see the logic of consequences as simply rule-following in settings where it is regarded as socially appropriate to be calculating about the choice of efficient means to given ends.

As blanket statements – ontological claims about the nature of decision-making – we agree with March and Olsen's skepticism about either reduction being a good idea (March and Olsen, 1998: 953–4). Surely the distinction between the two logics points to an empirically interesting phenomenon. Sometimes actors do decide by attempting

to calculate consequences. On the other side, some choices seem so tightly constrained by webs of norms and roles that they scarcely seem like 'choices'. And even if not tightly constrained, in some settings the problem of figuring out what to do seems to entail primarily the interpretation and application of conflicting normative claims, rather than estimating the likelihood that such-and-such action will lead to such-and-such result. Even if it were possible to subsume one logic theoretically as an instance of the other, if we are not to obscure these empirical differences we would then simply need to introduce a new linguistic or conceptual distinction to capture them.

There are also good reasons to think that both rationalist and constructivist analyses as commonly practiced may have a comparative advantage in analyzing settings where one or the other mode of decision-making is predominantly at issue. Arguably, decisions with great importance for international politics have often proceeded from a person's or group's interpretation of the internal logic of a complicated ideological or religious system (for example, Marxist-Leninism, Islam, or liberalism). Rationalist methods in their present form are ill-suited to provide insightful analyses of the ideational logics embedded in such systems, or their consequences for debates and actions. Constructivists, by contrast, are in their element here. On the other hand, rationalists have developed a powerful set of tools for thinking about the choice of means to diverse ends in multi-actor settings. Even a constructivist committed to 'reduction' in favor of the logic of appropriateness might concede some value to such analyses in domains where norms permit consequentialism.

But this division-of-labor framing should not be pushed too hard. There is no reason to rule out, a priori, the possibility that a rationalist (constructivist) analysis might yield valuable insights applied to a domain where the logic of appropriateness (consequences) predominates. For example, suppose that for some recurrent pattern of behavior the truth is that people or states are purely acting out norms, or out of habit, or both. An analysis that assumes the agents are acting as calculators-scheming-consequences might none the less be valuable for revealing how the observed pattern can be stable and self-reproducing over time. To be stable, a social pattern of habitual or norm-based actions still needs to be robust against 'entry' by agents espousing alternative norms and against agents who experiment with new (non-habitual) actions. Game-theoretic models are well-suited for analyzing this sort of robustness in social settings. On the other side, IR research that problematizes logic-of-consequences thinking by challenging its empirical universality or theoretical necessity may yield valuable insights as well.³⁷

Norms as Useful versus Norms as Right

A closely related framing of 'rationalism *v.* constructivism' sees the two approaches as differing in their understanding of social norms and the reasons thought to explain norm compliance. Here the issue is not so much 'Do people follow a logic of consequences or appropriateness in their behavior?' as 'Why do people follow norms? What motivates them to do so?' Some see rationalists as arguing that people follow norms only because (and when) it is useful to do so, whereas constructivists allow that people can be motivated to follow norms simply because they think it the right or legitimate thing to do.³⁸

Of course, the idea that actors may desire to follow norms for their own sake rather than just because it is useful is perfectly compatible with the 'thin theory' of rational choice (Ferejohn, 1991). The latter, after all, is agnostic between different 'thick theories' about the content of preferences, and as such does not rule out actors having a preference to follow a norm for its own sake. Indeed, interpreting norms as preferences probably represents the first way that rationalists tried to conceptualize norms.³⁹ Constructivists might point out that the idea of a 'preference for a norm' could refer simply to a 'taste', like for chocolate, and as such does not capture their interest in the perceived normative or obligatory force of norms.⁴⁰ But the logical ability of rationalism to accommodate a preference for rule-following nevertheless does seem to take much of the wind out of the sails of at least this one framing of 'rationalism *v.* constructivism'.

In order to generate a real debate on this issue, therefore, it is necessary to arbitrarily restrict the rationalist position to a particular thick theory of actor preferences, namely one in which actors do *not* have an intrinsic preference to follow norms. This amounts to saying that actors' attitudes toward norms is a 'realist' one of 'self-interest', since the norms would not be seen as having intrinsic worth or being ends in themselves.⁴¹ This move in effect limits the role of norms to the Belief side of the intentional action equation, rather than allowing them to appear as arguments in actors' utility functions (Desire).

The constructivist position then becomes equivalent to a different, 'non-Realist', thick theory of preferences, namely one in which actors *do* have an intrinsic desire to follow norms, perhaps rooted in a belief that this is the right or obligatory thing to do given a certain identity. This seems to imply that actors possess non-selfish or collective interests toward norms, which is to say that they identify with or make them part of their conception of self, and as such make the group's interest in upholding norms their own individual interest as well (Wendt, 1999: 337).⁴² Norms here figure as arguments in

actors' utility functions, rather than being limited to beliefs about the environment.

Having imposed such a constraint on the 'rationalist' view, rationalists and constructivists would now have a genuine empirical disagreement. On one level, the issue at stake is about actor motivations. But it may also be seen as about the degree to which norms are internalized. Saying that norms have become desires with perceived obligatory force implies deeper internalization than saying that norms are only beliefs about the environment to which actors relate instrumentally. As such, there may be something to the common characterization – given the arbitrary domain restriction we have imposed – that in rationalism the main explanatory role of norms is 'regulative' of the behavior of exogenously given agents, whereas in constructivism norms are 'constitutive' of actors' identities and interests in the first place.⁴³

These differences might have observable implications for both discourse and behavior. On the discursive side, actors might justify their actions differently under the two logics, the one by appeal to instrumental considerations, the other by appeal to normative ones. For example, is a norm of reciprocity in trade or other international negotiations defended on the grounds of intrinsic fairness or as a useful means to an end? Such evidence is not decisive, since actors may talk publicly in normative terms even if they are motivated primarily by instrumental and selfish concerns.⁴⁴ But unless we are prepared to dismiss all talk as cheap then the kinds of discourse that actors use should count for something. On the behavioral side, in turn, we might expect to see observable differences at both the individual and aggregate levels in rates of norm compliance. If actors are motivated to follow norms for their own sake then they should be more inclined to observe them (other things being equal).

So in the end there does seem to be something at stake, both theoretically and empirically, in the distinction between the two motivations for norm compliance. Yet, beyond the difficulty already noted that rationalists need not exile norms from preferences, there are at least three further reasons not to treat the rivalry between these two views of motivation in zero-sum, let alone paradigmatic, terms.

First, there is little reason to think that human behavior toward norms is either *always* self-interested or *always* a function of perceived legitimacy. Different people may vary in the extent to which they have internalized a given norm, and the same person may vary in the extent to which she has internalized different norms. The theoretical challenge is therefore one of identifying the conditions under which each hypothesis holds, rather than showing that one is always right or wrong. In the larger scheme of things, both hypotheses are probably true.

A second problem is that empirically it may be impossible to discriminate between the views, especially when both predict – as they may often do – the same outcome. This will be particularly problematic if we have access only to behavioral evidence, whether because of data unavailability or because no conscious or explicit ‘choice’ was made to follow a norm in the first place. Why did Germany not annex Denmark last year? Both hypotheses offer plausible explanations: because the consequences would have been too grave, and because the German leadership believed that this would have been wrong. Which is the ‘real’ reason, and how would we know, given that Germany probably never made a conscious decision not to invade last year in any case? Moreover, what is to stop someone from saying that he obeyed a norm for *both* reasons, perhaps with one in the foreground and the other in the back? Clearly, the two hypotheses need not be mutually exclusive for a given case. These empirical problems will be less debilitating in situations where the two hypotheses generate different behavioral predictions (like compliance versus violation), since we can then substitute observable behavior for assessments of subjective intention. But it is unclear what percentage of cases this will be (or even how we could find this percentage out), nor what to do with the remaining cases where predictions are the same, nor what kind of metric one could develop to assess the relative importance of normative versus non-normative motivations in a useful way.

Finally, there is the problem that the two motivations for norm compliance – fear of bad consequences and desire to do right – may interact with each other over time, in either direction. On the one hand, if in a given situation ‘second-’ or ‘third-party enforcement’ (punishment by society or state) is consistent enough that actors repeatedly comply with a norm, then over time they may internalize it to the point of acquiring a preference to comply for the sake of doing right or acting appropriately. Indeed, that seems a fair description of the socialization process we all go through as children. First we get punished for doing bad things, later we learn to see doing bad things as morally wrong.

But on the other hand, desires to do right may also decay over time if there is not enough enforcement against norm violators. Consider someone who stops at a red light at 3a.m. on an empty road, which is plausibly taken as evidence of internalization. If it happens that police enforcement suddenly plummets and traffic violations increase, then this person is more likely to ask ‘Why should I follow the rules when no one else does?’⁴⁵ Thus, third-party enforcement of norms may sometimes undergird ‘first-party enforcement’, the desire to comply because one believes it the right thing to do. These potential interactions suggest a developmental division of

labor between first- and third-party enforcement. When norms are new, we might expect the fear of bad consequences for violation to dominate. Over time, with internalization, the logic of appropriateness may take over, but may still depend in part on social or legal institutions that use the threat of punishments to prevent the entry and proliferation of ‘exploiters’.⁴⁶

In sum, although under at least one important formulation there are some differences between the ‘rationalist’ and constructivist explanations for why people follow norms, there are also good reasons not to make too much of them. In the aggregate the two explanations are complementary rather than mutually exclusive, may be hard to distinguish empirically, and in some cases there might not even be any fact of the matter to distinguish at all. At the extremes or ‘tails’ of the case distribution the rivalry between the two claims is clear, but in the middle there may be no deciding between them. This muddy empirical situation may encourage partisans to privilege their favored hypothesis on purely theoretical grounds, trying to subsume the other as a special case, but that seems unlikely to advance our understanding of the world.

Problematizing Actors I: Preference Formation

Perhaps the most widely cited issue thought to divide rationalist and constructivist scholarship concerns what the dependent variable or explanandum should be, in particular whether to take actors as ‘exogenously given’ and focus on explaining their actions, or to ‘problematize’ or ‘endogenize’ actors themselves. Rationalism is usually seen as doing the former and constructivism the latter, although we shall argue that this difference is difficult to sustain in a hard and fast form. Since the issue is many-sided and thus fraught with potential confusion, however, we should say a few words about how we see the analytical terrain going into the discussion. In particular, it is useful to make two distinctions, one between different ways in which actors might be problematized or explained, the other between different kinds of actor properties which could be at stake in such a process.

There are two broad senses in which one might try to ‘endogenize’ actors, causal and constitutive.⁴⁷ The causal approach asks where actors came from, or came to have the qualities that they have today. Hendrik Spruyt’s explanation of how over the centuries states became the dominant actors in world politics by driving out competitors like city-states and city-leagues, or Rodney Hall’s account of how changes in the domestic organization of states from dynastic to national foundations transformed inter-state relations, are good examples of what can

be learned when we problematize state actors rather than take them as given (Hall, 1999; Spruyt, 1994; see also Reus-Smit, 1999). More generally, a number of constructivists in IR have advanced variations on the causal argument that state identities and interests have evolved over time through interaction with other states and NGOs.⁴⁸

In contrast, the constitutive approach asks not where actors or their properties come from, in an historical or process-tracing sense, but about their social conditions of possibility at a given moment. What is it about Costa Rica that enables it to participate as an equal in the UN? The recognition by other states of its sovereignty as a right. What is it that, for a time in the early 1990s, made Iraq a 'rogue' state with which most other states refused to have contact, when in the past its aggressiveness might have been evaluated differently? Shared understandings that determine the boundaries of acceptable foreign policy practice. Constitutive explanations of actors 'explain' in the sense of telling us what actors are made of, or how their properties are made meaningful or possible by the society in which they are embedded. As we argue below, the causal approach to endogenizing actors is not that much at odds with rationalism, whereas the constitutive approach may be more difficult to reconcile with it.

Because these two ways of thinking about problematizing actors are different, and have different implications for the debate between rationalism and constructivism, we address them in different sections. In this section we take up the causal issue, in the next the constitutive one.

Turn, then, to the second analytical distinction, between the kinds of actor properties that might be at stake. Whether approaching actors from a causal or constitutive standpoint, we can take three different things about them as given or not: their bodies, beliefs, or desires. These should be kept distinct in discussions about exogenous versus endogenous actors, since they vary in the extent to which they are a source of disagreement between rationalists and constructivists.

A *body* is the platform on which actorhood is constructed. The social position and meaning of bodies will vary, but before this variation can occur bodies must be constituted by an internal organizational structure and process that enables them to move, act and acquire meaning in the first place. In the case of individuals this internal organizational structure is given by biology. In the case of corporate actors like states it is constituted by biologically given people engaging in ongoing collective action enabled by the structure of the organization.⁴⁹ Interestingly, on the question of whether to take bodies as given the main theoretical cleavage is not between rationalists and constructivists, but between rationalist and constructivist 'moderns'

who both see themselves as part of the Enlightenment, liberal project in which the individual or agent is granted a privileged status, and 'postmoderns' who reject that project and want to deconstruct the individual or agent all the way down (this is one place where the performativity argument referred to in the introduction comes into play). Thus, like rationalists, modern constructivists have been largely content to take as 'exogenously given' that they were dealing with *some* kind of actor, be it a state, transnational social movement, international organization or whatever. As such, the constructivist concern with identity-formation has typically focused on the construction of variation within a given actor class (type or role identities), rather than on explaining how organizational actors come into being in the first place (corporate identities).⁵⁰

Actors also have *beliefs*. Here too there is little disagreement between rationalists and constructivists (of any stripe), this time because rationalists have been perfectly willing to try to explain beliefs and changes in beliefs. All non-cooperative game-theoretic solution concepts (such as Nash equilibrium) amount to proposals about what sort of patterns of beliefs one would expect to arise in different social settings. Further, dynamic games with incomplete information can explicitly model the evolution of actor beliefs about others' preferences and beliefs (including, for instance, beliefs about others' beliefs about oneself, which figure prominently in discussions of 'identity'). And it is perfectly possible within such an analysis for a person's beliefs even about their *own* preferences to change and evolve as a result of acquiring new information.

To be sure, there is debate about how *deeply* rationalism can explain beliefs – for example, whether it can handle the 'complex' learning involved in preference formation, or the ways in which individuals' beliefs may be constituted by social collectives. But it is clear that at least with respect to 'simple' learning about an external environment rationalism is itself very much in the business, with constructivism, of problematizing actors.⁵¹ Since we address preference formation as a separate point in a moment, and there is otherwise relatively little to disagree about, we shall not address this aspect of the exogenous actor problem in this chapter.

That leaves *preferences*, desires, or in constructivist parlance, 'identities and interests'. This is where most of the debate has occurred, with rationalists tending to treat preferences as given, and constructivists trying to endogenize them. Our view is that while there are characteristic differences here, they are not as fundamental as is sometimes supposed.

Before turning to that argument, however, we emphasize up front that the choice of whether to treat preferences as given is an important one in

social inquiry, for both theoretical and political reasons. It matters theoretically because to assume exogeneity is implicitly to make an empirical claim about the world, namely that what actors want is constant within the context of the study in question. If this claim is not accurate, then we are led to question the subsequent causal story being told about behavior. And an exogeneity assumption can matter politically because if what actors want is not stable, or could be made not stable, then policies based on an assumption of stability may not have the desired effect or may understate the potential for social change. The latter is of course the main reason why constructivists are concerned to endogenize identities and interests. If it can be shown that these are produced and reproduced by social interaction then the possibilities for change may be greater than if, say, a 'prisoners' dilemma' is treated as an unchangeable fact about some aspect of world politics.

Yet there are also at least three reasons for caution about making it the basis for a deep, paradigmatic divide. One is that the choice of exogenous versus endogenous preferences can be treated as purely analytical, rather than as a substantive claim about the nature of the world.⁵² By this we mean that, on one level, the choice can be about nothing more serious than what question, or dependent variable, researchers are personally interested in. After all, there *are* two questions here – 'what are the causes of *X* behavior?' (an action in the world) and 'what are the causes of *X* preference?' (a state of mind) – and it is not obvious that we have to answer one in order to answer the other. It is perfectly legitimate to answer the former while holding preferences constant, and to answer the latter while bracketing the causes of behavior. Jeffrey Legro has summarized this overall situation with a dance metaphor, the 'two-step': first we explain preferences, then we explain behavior (Legro, 1996).

This is not to say that there is no risk in separating the two questions. As we noted above, the assumption that preferences are given brings baggage with it, an implicit assumption of stability. Probably few rationalists are committed to this assumption as a matter of principle, that is, to the proposition that preferences *really are* stable, for all time.⁵³ For most it is merely a 'methodological bet', an analytical convenience that allows them to answer the question that interests them, which is the effect of context on choice behavior. But there is nevertheless a danger that, through a process of forgetting what we are doing, what starts out as merely an analytical convenience can become something more than that, a tacit assumption about what the world is really like which limits our theoretical and/or political horizons. The assumption that states are self-interested, for example, is harmless when made as an analytical convenience, but if turned into a tacit universal claim it can lead us to conclude, mistakenly, that anarchic systems are necessarily

self-help worlds rather than contingently so in particular historical circumstances. This transformation of harmless analytical assumptions into tacit ontologies seems particularly likely to happen in a 'division-of-labor' approach to the two-step, where constructivists and rationalists address their respective questions in isolation from each other. The best way to keep the two-step honest is to make sure that the partners are coordinated, rather than go their separate ways.

A second reason for not putting too much weight on whether preferences are taken as given is that the boundary between preferences and action, on which the debate over this issue inherently turns, is relative and unstable. One researcher's preference over outcomes is another researcher's preference over actions. Consider the Cuban Missile Crisis. A 'rationalist' might ask, why did US decision-makers, given their preference that the Soviets remove their missiles from Cuba but an even stronger preference to avoid nuclear war, choose a strategy of naval blockade? And why did this convince the Soviets, given their preferences and available actions, to retreat? On the other hand, one could also ask the 'constructivist' question, as Jutta Weldes has done, of how US policy-makers constructed the removal of Soviet missiles as their interest in the first place, since such an interpretation was not absolutely necessary (Weldes, 1999). Weldes answers by reference to a national security discourse and its associated Cold War identity, that constituted US interests in a certain way. Yet, the collective agreement within the Kennedy administration on this assessment of 'US interests' can also be seen as the outcome of strategic behavior in an intra-administration game of talk and access, conducted in the shadow of expectations about likely public and elite reactions to different courses of action and outcomes. Insofar as the determination of collective interests is itself the result of a set of choices, we could then ask a new 'rationalist' question in which *that* was the 'action' to be explained, and the problem was to show how that choice emerged in light of higher or 'meta'-preferences (for example, 'security of the United States as a capitalist, democratic state'), and 'lower' preferences concerning re-election and relative power within the administration.

We have, in other words, a potentially endless means–ends chain in which any given end can be seen as a means to some other ends depending on what question is being asked. As such, the decision to call something an outcome over which preferences are assumed, or an action (or set of actions) to be explained is not a statement about the world but rather an analytical move by the investigator. This absence of a fixed boundary between action and interests may help explain their occasional conflation in constructivist critiques of rationalist models, in which the claim to explain interests sometimes turns out to be difficult to separate from

an explanation of action.⁵⁴ And in any case, if 'rationalists' can be turned into 'constructivists', or *vice versa*, simply by pushing the research question one step up (or down), whether or not preferences are taken as given seems like a slippery foundation for a paradigm war.

A final issue is that it is not even clear that the 'two-step' accurately describes the division of labor between rationalists and constructivists, on either side. Thus, on the one hand, some 'rationalists' do not take preferences as exogenously given. In IR, for example, Andrew Moravcsik accepts the logic of the two-step but nevertheless seeks to explain foreign policy-makers' preferences by reference to domestic politics (Moravcsik, 1997). That rationalists could do this is not surprising in light of the dependence of what counts as a preference on what question is being asked: a preference (end) on one level of theory can be a choice (means) on another. In evolutionary game theory some rationalists have gone even farther, building models in which actors acquire preferences either as a result of differential reproduction or a process of imitation or adaptation; thus actor preferences are explained endogenously by some kind of selection or evolutionary stability (equilibrium) argument.⁵⁵ As such, Ruggie goes too far in suggesting that rationalism cannot accommodate complex learning, unless we restrict the label 'rationalist' arbitrarily to models that do not address it (Ruggie, 1998: 868).

If 'rationalists' can justifiably claim to offer some insight into the formation of preferences, then 'constructivists' in turn can justifiably claim to offer some insight into the second half of the two-step, the choice of action. As we discussed above, constructivism has become associated with the hypothesis that much of human behavior is driven by a normative logic of appropriateness rather than an instrumentalist logic of consequences, and as such invades the turf of the 'rationalist' step of the supposed division of labor.

In sum, the decision about whether to causally explain preferences does not seem like an occasion for a profound or divisive debate. The boundaries between preferences and choice of action are in important part question-relative and thus unstable, and even less do they have to constitute significant epistemological or ontological cleavages. This is not to say that in every case it will be useful to explain preferences over outcomes by translating them into means to higher ends. And certainly it remains reasonable to criticize an argument or model that draws policy conclusions without exploring the possibility of preference endogeneity, just as it may be warranted to criticize an analysis that does not treat identity and interest-formation as the consequences of choices that are potentially amenable to rationalist explanation. But these are relatively concrete issues that can be handled without implicating 'paradigmatic' sensibilities.

Problematizing Actors II: On the Constitution of Subjectivity

One of the most persistent and at least superficially plausible ways of characterizing 'rationalism *v.* constructivism' in IR is by reference to the divide between methodological individualism and holism in the philosophy of social science. We already gave credence to this view by picturing rationalism as an approach that tries to explain macro-level phenomena (such as a 'balance of power system') by reference to more micro-level phenomena (such as state motivations and capacities). Holists in a range of fields have expressed serious doubts about whether many central features of social life, in international politics and elsewhere, can be adequately or at all understood by somehow resolving them to component parts. In practical terms, constructivists in IR argue for understanding parts, such as states, in terms of wholes like international systems or reigning ideas, rather than exclusively the other way around.

Another way of expressing this opposition is by contrasting causal and constitutive forms of explanation. Causal explanations, which refer to the action of pre-existing, temporally prior causes that produce the effects to be explained, would seem to have an affinity with the micro-to-macro program of rationalism. Constitutive explanations, which characterize systems of beliefs and practices that in effect create or define social objects and actors – such as master and slave, or states, for instance – would seem to illustrate holism in action.

Even here, though, we encounter difficult issues that caution against drawing too sharp a line. Perhaps the main question concerns whether rationalism has the conceptual resources for a defensible and useful account of how structures constitute agents. We lack the space here to provide anything like a full investigation of rationalist/constructivist differences on questions about social structure (as opposed to agency). But we do want to suggest how some standard formulations in the literature may be 'too quick'.

In a broad sense, rationalist studies do typically involve efforts to explain wholes in terms of the actions and interactions of parts. Contrary to a common misconception in IR theory, however, this does *not* imply that rationalists have no account of how macro-level phenomena, such as 'social structures', impinge on individual actors. An example is useful to illustrate how this account works.

Consider the following application of the rationalist recipe to explaining why, in the United States, people drive their cars almost exclusively on the right side of two-lane roads. Imagine a model/argument in which the actors are a large number of individuals who must choose simultaneously whether to drive on roads or not, and if on a road on the left, in the middle, or on the right side. Individuals are assumed to

desire to arrive at their destination quickly but without damage to body or car. This is a coordination problem – a problem in which one's optimal choice depends on how others choose and in which some patterns of choice are better for all than some others. The observed pattern in which more or less everyone drives on the right side of the road is explained as an equilibrium pattern of optimal choices. That is, given that everyone else is expected to drive on the right, driving on the right is an efficient means to reach one's destination quickly but without harm.⁵⁶

Notice that there are *two* sorts of 'structure' implicit in this story, exogenous and endogenous. Individuals in the argument face exogenous structure in the sense of physical constraints. If you drive off the road, your car is likely to be damaged and it may be impossible to get where you want to go.⁵⁷ But in an equilibrium, they also face a social structural constraint that derives from the fact that everyone expects everyone else to drive on the right. This is endogenous structure in that it is *mutually constituted* by the beliefs and attendant actions of all individuals in the model, and it is explained within the model rather than postulated. From the vantage point of any one actor, this endogenous structure of beliefs and attendant actions is just as objective and real as the trees on the side of the road, even though the actor's own actions contribute to making the reality and it could be made otherwise. Note also that this social structure is not determined by material conditions; the convention could just as well be to drive on the left.

If this is a *causal* explanation, it is not causal in the most straightforward sense of pre-existing causes that reliably produce subsequent effects. Actions are explained in part by reference to beliefs in this account, but at the same time beliefs are explained as correct perceptions of actions (in an equilibrium). Explanation by reference to an equilibrium pattern of beliefs and behaviors answers a 'how is this possible?' question more than it does a 'what caused this to come about?' question.⁵⁸ In this respect it appears closely akin to the constitutive form of explanation associated with holism.

The convention of driving on the right is regulative rather than constitutive; what it means to drive is not constituted by this convention.⁵⁹ But a convention account can also be offered for the constitutive rules that define the meanings of words and actions or, say, actor identities. The mappings from behaviors such as speech sounds and gestures to meanings (which mappings constitute 'actions') are obviously also matters of coordination within a culture. In the United States a wave with a smile generally means 'Hi!' and not 'I want to kill you', though it could be otherwise.⁶⁰

Now consider a question of social identity, such as the master-slave dichotomy. In contrast to a

property such as being six feet tall, the social identity 'master' or 'slave' cannot be defined solely by reference to facts about an individual. There are no masters without slaves and vice versa. Instead, to be a master (for example) is to be accorded certain powers with respect to certain other individuals, *by social convention*. Just as in the case of driving conventions, the coordinated actions and beliefs that constitute a system of slavery could be otherwise, but none the less have for any one individual an objective reality posed by the beliefs and expected actions of others in various contingencies. In this approach, an actor's identity, a complex of beliefs about self, others and relations between them, would be endogenously explained as an equilibrium in a coordination game rather than posited as an exogenously given fact about an individual.⁶¹

It is true that, in practice, emphasizing the ways in which actor identities are constituted by social conventions is not how rationalist arguments usually proceed. Rather, the rationalist strategy is usually to build in or presuppose some social structures and the identities they constitute, and then to explain from the 'bottom-up' a pattern of choices and the structures they imply. Constructivists have objected to this building in of structurally constituted identities, since they are interested in how these are constituted in the first place.⁶² So in showing how identities can be seen as constituted by an equilibrium in a coordination game we are going beyond the typical rationalist story, in effect using a rationalist approach to answer a question normally asked only by constructivists.⁶³ The value of such a move is both rhetorical and substantive. By highlighting the flexibility of rationalism to accommodate 'constructivist' insights it suggests there may be less opposition here than is often thought.⁶⁴ And by emphasizing the mutual determination of social structure and individual choice in equilibrium, it highlights an aspect of the micro-foundations of actor constitution that constructivists have sometimes neglected.

Even so, constructivists may wonder if a rationalist approach to conventions is really up to the task of comprehending the constitution of social identities and structures of meaning more broadly. One question is whether such an account can explain the constitution of one identity without presupposing some other (or others). If not, then the convention account may be missing an important aspect of the constructivist position.⁶⁵ In practical terms, this might imply that the rationalist approach to conventions would sometimes be useful for 'cutting into' a network of social identities to understand how one particular identity is sustained given others, but less useful for gaining insight into the bigger picture.

A second issue is whether a rationalist approach necessarily implies that conventions are aggregates of, and thus ontologically reducible to, pre-existing

beliefs and meanings. If so, this would conflict with the holist argument that the contents of the actor beliefs that sustain social conventions do not exist apart from those conventions. But in themselves, the equilibrium explanations just given of the driving and master–slave conventions carry no such implication of reducibility. The meanings that constitute and sustain these conventions may be pre-existing in actors' heads or they may not. If rationalism is viewed in analytical rather than ontological terms it can be agnostic on this question, and thus be compatible with holism. And well that is, since it would be unfortunate if constructivists could not avail themselves of the insights provided by equilibrium arguments in their own work just because they are associated with 'rationalism'.

We are not suggesting that the ontological debate between holism and individualism is thereby settled or unimportant. Philosophers have been arguing about their relative merits for decades.⁶⁶ The eventual 'solution', if there is one, may matter in a broad sense for IR insofar as it speaks to the question of whether or not rationalism and constructivism reduce to the same vantage point on international life, and thus to what the search for 'micro-foundations' in IR can mean. We are in no position to settle this dispute. However, if the rationalist–constructivist debate in IR is understood in methodological rather than ontological terms, as we recommended in the introduction, then it is not clear that IR scholars *need* to settle it to do their work. A lesson to take away from this discussion is that there seem to be at least two ways of telling stories about the constitution of actor subjectivity, which may or may not on close examination turn out to be the same. Although constructivists have tried to make this issue exclusively their own – an effort abetted by relative neglect from rationalists – the rationalist approach appears to have the conceptual resources for an endogenous account of actor identities, both constitutive and, as we saw in the previous section, causal. This account may or may not ultimately capture the essence of the constructivist argument, and it may or may not yield insightful analyses of the phenomenon in question. This remains to be seen. But in the meantime, there is a strong pragmatic case for treating the two stories as complementary at the least. This will encourage IR scholars to pursue questions about the constitution of actors in whatever way seems to yield insights, and to think creatively about how they might be combined.

CONCLUSION

The idea of a battle royal or 'Great Debate' between rationalism and constructivism is appealingly dramatic, but properly understood many of the

issues dissolve upon close inspection. Although often framed as an argument about ontology or empirical descriptions, we have argued for a pragmatic interpretation in which these are two approaches to answering questions about international politics, rather than two competing *Weltanschauungen*.

If the debate is defined as a matter of ontology, then it approaches zero-sum and a great deal rides on who wins. Yet it is not clear how much this would tell us about world politics. Knowing that international reality consists 'ultimately' of wholes or parts, for example, tells us little about how states, non-governmental organizations or multinational corporations affect international politics; about the conditions under which world politics is more conflictual or cooperative; about whether and how anarchy can be transcended – in short, about most of the political questions of concern to IR scholars. To answer such questions we need to make further assumptions that go beyond those supplied by rationalism and constructivism.

Again, this is not to say that ontological (and for that matter epistemological) issues should not be engaged by IR scholars, or that doing so will have no benefits. Indeed, some benefits are already apparent. Of the empirically oriented sub-fields of political science, IR is probably the most philosophically self-conscious and informed, the most interested in the continuous examination of fundamental questions about what social inquiry is supposed to be and do. On the whole, this engagement makes it harder for scholars to lapse into an unthinking 'normal science' or 'normal postmodernism' that cannot defend or think through its standard practices. But we should also be conscious of the limits of philosophical debates for making sense of international politics. If 'rationalism v. constructivism' is to be another 'Great Debate' in IR, then let it not be constructed as an argument about ontology.

If the debate is viewed in more empirical terms then the relationship between the two approaches is more complex. In some cases they offer rival hypotheses, in others they seem complementary, in others they are redundant. Discussion about issues like these is likely to provide more insight into world politics than will ontology, but here too it is problematic to see the question as 'rationalism v. constructivism'. Neither perspective necessarily commits the researcher to a claim about the world like 'preferences *really are* exogenous (endogenous) to interaction'. It is equally valid to treat rationalism and constructivism as merely analytical statements about what the researcher is interested in, which can never be everything at once. Moreover, even if we do choose to test the claim that, for example, in context *X* preferences 'really are' exogenous (endogenous), this may tell us little about *Y* or *Z*, where the

opposite claim might be true. Even when defined as an empirical dispute, in other words, the eventual result of a 'Great Debate' would probably be that both approaches are true some of the time. This does not mean we should never try to adjudicate between 'rationalist' and 'constructivist' hypotheses in those cases where they can be made to generate rival empirical predictions. But we should be clear about what those cases are, and about what does and does not follow 'paradigmatically' from one hypothesis prevailing over the other.

In short, we believe the most fruitful framing of 'rationalism v. constructivism' is a pragmatic one, treating them as analytical lenses for looking at social reality. It is common in articles of this sort to try to delimit (or legislate) the types of problems for which each lens works best. Although we have offered some suggestions of this kind, on the whole we have argued that the standard ways of drawing lines between the two 'isms' and their presumed competences are on shaky grounds. Thus, even the question of what lens to use for a particular research question should be left open and not fixed by a priori, methodological or theoretical considerations.

Our discussion has focused on the 'agency' side of the problem. We did not explore the structural side of the equation except indirectly in our discussion of the constitution of agents. Some additional 'bones of contention' between rationalism and constructivism might be found in their approaches to structure, involving issues such as micro versus macro structure, common versus collective knowledge, external versus internal relations, and the reality of 'deep' structures. An exploration of these issues would be worthwhile, but we suspect the result would parallel our conclusion here, that the relationship of the two approaches, when understood pragmatically, is largely either complementary or overlapping.

It should be stressed that in advocating a pragmatic view we are not endorsing method-driven social science. Too much research in international relations chooses problems or things to be explained with a view to whether the analysis will provide support for one or another methodological 'ism'. But the point of IR scholarship should be to answer questions about international politics that are of great normative concern, not to validate methods. Methods are means, not ends in themselves. As a matter of personal scholarly choice it may be reasonable to stick with one method and see how far it takes us. But since we do not know how far that is, if the goal of the discipline is insight into world politics then it makes little sense to rule out one or the other approach on a priori grounds. In that case a method indeed becomes a tacit ontology, which may lead to neglect of whatever problems it is poorly suited to address. Being conscious about these choices is why it is important to distinguish between the ontological, empirical, and pragmatic levels of

the rationalist-constructivist debate. We favor the pragmatic approach on heuristic grounds, but we certainly believe a conversation should continue on all three levels.

This prompts a concluding suggestion: that the rationalism-constructivism issue be seen not as a debate but as a conversation. The connotation of 'debate' is of a zero-sum conflict between two sides with firm substantive commitments about what the world is like. This might have been appropriate in the first 'Great Debate' between realists and idealists, who disagreed about the essential nature of world politics. But all of the subsequent 'debates', including this one, have been more about method than substance, and on that level considerably less is at stake. Rationalists and constructivists approach international life from different analytical standpoints, which has led them to ask characteristically different questions and develop characteristically different answers. Rather than a dialogue of the deaf in which each side tries to marginalize or subsume the other in the name of methodological fundamentalism, the challenge now should be to combine insights, cross boundaries and, if possible, synthesize specific arguments in hope of gaining more compelling answers and a better picture of reality.

We have tried to contribute to such a conversation by working through a number of commonly perceived points of conflict and disagreement. Our own experience going back and forth on ten (!) drafts of this chapter might suggest that a conversation across these 'isms' is too difficult to pursue or sustain. But interestingly, most of our difficulties arose not from clear substantive disagreements but rather from matters of presentation and efforts to gain clarity about just how 'the other side's' argument works on specific points. In any event, we believe that the blind men's best hope of progress in understanding international politics lies in a conversation of truth-seekers rather than lawyerly debate.

Notes

The authors gratefully acknowledge comments on earlier drafts by Walter Carlsnaes, Thomas Risse, Joel Westra, the participants at a PIPES workshop and the Stanford Reading Group on International Relations, and one anonymous reviewer.

1 As if confirming this suggestion, the editors of the present volume have organized the theory section in these terms as well.

2 Which have since become known as, respectively, the third and second 'Great Debates' (see Lapid, 1989). By this reckoning, rationalism v. constructivism would be the 'fourth debate'.

3 For other recent discussions in this spirit see Fierke and Nicholson, 2001 and Katzenstein, Keohane and Krasner, 1998.

4 See Wendt, 1999: 33–8 for further discussion of these three ways of interpreting the debate.

5 For an exploration of ontological aspects of the debate see Wendt, 1999.

6 For example, Clark, 1998; Finnemore, 1996.

7 For a complementary discussion emphasizing the structure side, see Wendt, 1999: ch. 4.

8 On the latter criticism see Checkel, 1998.

9 For entries into the literature on performativity in IR see Campbell, 1998; Laffey, 2000; Weber, 1998.

10 'Outcomes', such as a balance of power, are understood in this approach as labels for patterns of actions or the result of sets of actions.

11 'Rationally' in this sentence refers to instrumental rationality.

12 Specific features of specific rationalist models/arguments are often taken to task for being unrealistic *per se*, which is probably never a valid criticism.

13 Omitting the case of pure evolutionary models in which 'actions' might be explained as the result of a genetic or totally unreflective cultural 'programming'. This qualification applies at several places in the discussion that follows.

14 Where 'makes sense' is understood in terms of instrumental rationality.

15 Much of this program has been textbookified; see, for a good recent example, Mas-Collel, Green and Whinston 1995. For examples of work closer to the frontier, see Fudenberg and Levine, 1998; Rabin, 1993; Rabin and O'Donoghue, 2001.

16 A major exception being work on the epistemic foundations of decision theory; for an overview see Dekel and Gul, 1997.

17 See, for overviews and main results, Fudenberg and Levine, 1998, Weibull, 1995, or Young, 1998. The political scientist Robert Axelrod has been a pioneer in evolutionary models, and there is now a small community of IR scholars working on evolutionary game-theoretic and related computational models (Cederman, 1997). This intriguing line of work remains curiously disconnected from the much larger set of theoretical results developed by economists and theoretical biologists.

18 On the other hand, one might also ask what is the point of delimiting 'rationalism' in the first place, except to fit out one side for an interparadigmatic battle.

19 In the 1970s, Schelling explored models with agents that were far less than rational by the meanings that evolved in subsequent game-theoretic work. Indeed, his analysis of tipping models illustrates that evolutionary thinking is not a new thing in 'rationalism'.

20 See, for example, Elster in some of his earlier work (e.g., Elster, 1985).

21 For example, Walt, 1998; Wendt, 1992, 1994.

22 See, for example, Kratochwil, 2000; Searle, 1995; Smith, 2000; Wendt, 1999.

23 Meyer and Jepperson, 2000.

24 For discussion and illustration see Emirbayer, 1997; Jackson and Nexon, 1999; Wendt, 1999: ch. 4.

25 There are various ways to render this distinction other than 'causal-constitutive'; this particular language is developed at greater length in Wendt, 1998, 1999.

26 The causal-constitutive distinction is in turn often thought to have implications for the epistemological debate between positivist 'Explanation' and interpretivist 'Understanding' (see Hollis and Smith, 1990); we are not convinced that explanation and understanding require fundamentally different epistemologies, but have chosen to set the issue aside in this chapter.

27 For further discussion see Wendt, 1999: ch. 3.

28 We shall use these terms interchangeably to denote the subjectively perceived wants that actors do have rather than normative or objective wants that they arguably should have, which is how Fearon thinks the concept of interests should be understood (in contrast to non-normative 'preferences'). For an overview of this distinction, with further references, see Wendt, 1999: 231–3.

29 Though Keohane has since made clear that he does not see rationalism as a 'materialist' theory; see Keohane, 2000.

30 See especially Schelling, 1960.

31 See Ashley, 1984; Dessler, 1989; Walker, 1987; Wendt, 1992.

32 For an overview see Wendt, 1999: 113–19.

33 For example, in a Nash equilibrium, players' beliefs about others' actions are such that every player prefers to take the action that confirms the others' beliefs as correct.

34 See Ferejohn, 1991.

35 On the affinities between rationalism and interpretive sociology, see Esser, 1993; Ferejohn, 1991; Norkus, 2000.

36 See March and Olsen, 1998.

37 See Wendt, 2001, for example, for some suggestions along these lines.

38 For a good discussion of this contrast highlighting the role of legitimacy see Hurd, 1999.

39 See, for example, Gary Becker's (1957) explanation of the practice of racial discrimination in terms of 'tastes', or Ellickson (1991) on 'first party enforcement' of social norms (referring to the internalization of a norm as a preference).

40 See Finnemore and Toope, 2001. Note that this is not to say that constructivists have yet generated an adequate theory of obligation; though see Kratochwil, 1989 for a start.

41 For further discussion see Wendt, 1999: 238–43.

42 For a more rationalist discussion of this idea see Sugden (1993) on 'thinking as a team'.

43 But see also the discussion below about whether the constitution of actor identities and preferences can be comprehended in terms of a coordination (or convention) account.

44 Indeed, this is often to be expected. Since the argument that you should do *X* because it is in my self-interest is not likely to persuade unless you happen to care independently about my welfare, there are strong incentives to cast arguments in public-spirited terms even when the underlying motives are selfish. See Elster, 1995 and Fearon, 1998.

45 On this point see Philip Pettit's (1995) interesting discussion of the 'virtual reality' of *homo economicus*, where he argues that instrumental thinking will tend to kick in when it becomes highly advantageous for actors to use it, but otherwise most actors most of the time will do what is socially appropriate.

46 This example raises a deeper question, however, about the idea of following a norm because one believes that it is right to do so. Has one internalized a norm if one's desire to abide by it is in fact conditional on whether others do so?

47 See Wendt, 1998, 1999.

48 For example, Barnett, 1998; Checkel, 1997; Cronin, 1999; Finnemore, 1996; Lynch, 1999; Wendt, 1999.

49 On corporate actorhood see Wendt, 1999: ch. 5.

50 For typologies of identity concepts see Wendt, 1999: 224–30 and Fearon, 2000. Note that for rationalists, at least, the decision about where to locate 'the body' can be a methodological rather than an ontological question. See, for example, Lake and Powell, 1999; Elster, 1986.

51 The terms complex and simple learning are Nye's (1987).

52 Lake and Powell (1999) call it a 'methodological bet'.

53 Though see Stigler and Becker, 1977.

54 On this point see Clark, 1998 and Lake and Powell, 1999.

55 See, for example, Bowles, 1998; Cohen and Axelrod, 1984; Gerber and Jackson, 1993; Raub, 1990.

56 Of course, the account does not explain why the convention is to drive on the right rather than on the left, since driving on the left is equally a social equilibrium. This is an example of the type of 'equilibrium selection problem' that motivated the turn to exploring more evolutionary models in 1990s microeconomic theory; see, for example, Young, 1998.

57 Of course, these physical constraints are constraints only if one has a desire to get from one place to another without injury. Constraints on action are always relative to desires.

58 Indeed, the main criticism of equilibrium explanations in game theory is that they give no causal account of how an equilibrium state of affairs would or does come about. This is the main reason for the great attention to evolutionary models by 'rationalist' economic theorists in the 1990s.

59 Though we could say that the convention does constitute part of the role identity 'good driver'.

60 See Lewis, 1969 for the most philosophically developed effort to understand meaning in terms of coordination in games, and Weingast, 1995 for a rationalist analysis of the institution of sovereignty that highlights a number of constitutive effects on the meaning of state action.

61 Schelling (1960: 92) had hinted that 'roles' in the sociologist's sense might be productively analyzed and understood in terms of a coordination account. For a more developed empirical analysis along these lines see especially Laitin, 1998.

62 For a useful exchange on this and related issues see Fierke and Nicholson, 2001.

63 Although this 'rationalist' approach to the social construction of meanings and identities is hardly unprecedented. Lewis (1969) had suggested it with regard to meanings in language (and he says it is just a development of arguments about convention by David Hume); Sugden (1989) gives an evolutionary account of social norms and normativity in terms of social conventions; and Fearon and Laitin (1996, 2000), Kalyvas (1996) and Laitin (1998) have all pursued arguments along these lines in political science.

64 The potential convergence is also evident in Kratochwil's (1989: 69–94) discussion of the emergence of norms, which draws favorably on 'rationalist' scholars like Schelling and Lewis.

65 This is the central message of Mandelbaum (1955), one of the earliest statements of the holist position in the modern philosophical debate.

66 See Bhargava, 1992; Collin, 1997; Gilbert, 1989; and Lewis, 1969; among others; Wendt (1999: ch. 4) offers an interpretative review of this literature.

Bibliography

- Adler, Emanuel (1997) 'Seizing the Middle Ground: Constructivism in World Politics', *European Journal of International Relations*, 3: 319–63.
- Akerlof, George (1970) 'The Market for "Lemons": Quality Uncertainty and the Market Mechanism', *Quarterly Journal of Economics*, 84: 488–500.
- Ashley, Richard (1984) 'The Poverty of Neorealism', *International Organization*, 38: 225–86.
- Barnett, Michael (1998) *Dialogues in Arab Politics*. New York: Columbia University Press.
- Becker, Gary (1957) *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Bhargava, Rajeev (1992) *Individualism in Social Science*. Oxford: Clarendon Press.
- Bowles, Samuel (1998) 'Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions', *Journal of Economic Literature*, 36: 75–111.
- Campbell, David (1998) *Writing Security*, 2nd edn. Minneapolis: University of Minnesota Press.
- Cederman, Lars-Erik (1997) *Emergent Actors in World Politics*. Princeton: Princeton University Press.
- Checkel, Jeffrey (1997) 'International Norms and Domestic Politics: Bridging the Rationalist–Constructivist Divide', *European Journal of International Relations*, 3: 473–95.
- Checkel, Jeffrey (1998) 'The Constructivist Turn in International Relations Theory', *World Politics*, 50: 324–48.
- Clark, William (1998) 'Agents and Structures: Two Views of Preferences, Two Views of Institutions', *International Studies Quarterly*, 42: 245–70.
- Cohen, Michael and Axelrod, Robert (1984) 'Coping with Complexity: The Adaptive Value of Changing Utility', *American Economic Review*, 74: 30–42.
- Collin, Finn (1997) *Social Reality*. London: Routledge.

- Cronin, Bruce (1999) *Community Under Anarchy: Transnational Identity and the Evolution of Cooperation*. New York: Columbia University Press.
- Dekel, Eddie and Gul, Faruk (1997) 'Rationality and Knowledge in Game Theory', in David M. Kreps and Kenneth F. Wallis (eds), *Advances in Economics and Econometrics: Seventh World Congress*, Volume 1. Cambridge: Cambridge University Press, pp. 87-172.
- Dessler, David (1989) 'What's at Stake in the Agent-Structure Debate?', *International Organization*, 43: 441-73.
- Ellickson, Robert (1991) *Order Without Law: How Neighbors Settle Disputes*. Cambridge, MA: Harvard University Press.
- Elster, Jon (1985) *Making Sense of Marx*. Cambridge: Cambridge University Press.
- Elster, Jon (1986) *The Multiple Self*. Cambridge: Cambridge University Press.
- Elster, Jon (ed.) (1995) 'The Strategic Uses of Argument', in Kenneth Arrow et al. (eds), *Barriers to Conflict Resolution*. New York: W.W. Norton. pp. 236-57.
- Emirbayer, Mustafa (1997) 'Manifesto for a Relational Sociology', *American Journal of Sociology*, 103: 281-317.
- Esser, Hartmut (1993) 'The Rationality of Everyday Behavior', *Rationality and Society*, 5: 7-31.
- Fearon, James (1998) 'Deliberation as Discussion', in Jon Elster (ed.), *Deliberative Democracy*. Cambridge: Cambridge University Press. pp. 44-68.
- Fearon, James (2000) 'What is Identity (As We Now Use the Word)?', mimeo, Stanford University.
- Fearon, James D. and Laitin, David D. (1996) 'Explaining Interethnic Cooperation', *American Political Science Review*, 90: 715-35.
- Fearon, James D. and Laitin, David D. (2000) 'Violence and the Social Construction of Ethnic Identity', *International Organization*, 54: 845-77.
- Ferejohn, John (1991) 'Rationality and Interpretation', in K. Monroe (ed.), *The Economic Approach to Politics*. New York: HarperCollins. pp. 279-305.
- Fierke, K.M. and Nicholson, Michael (2001) 'Divided by a Common Language: Formal and Constructivist Approaches to Games', *Global Society*, 15: 7-25.
- Finnemore, Martha (1996) *National Interests in International Society*. Ithaca: Cornell University Press.
- Finnemore, Martha and Toope, Stephen (2001) 'Alternatives to Legalization: Richer Views of Law and Politics', *International Organization*, 55: 743-58.
- Fudenberg, Drew and Levine, David K. (1998) *The Theory of Learning in Games*. Cambridge, MA: MIT Press.
- Gaeremynck, Elisabeth and Jackson, John (1993) 'Endogenous Preferences and the Study of Institutions', *American Political Science Review*, 87: 639-56.
- Gilbert, Margaret (1989) *On Social Facts*. Princeton: Princeton University Press.
- Goldstein, Judith and Keohane, Robert (eds) (1993) *Ideas and Foreign Policy*. Ithaca: Cornell University Press.
- Hall, Rodney (1999) *National Collective Identity: Social Constructs and International Systems*. New York: Columbia University Press.
- Hollis, Martin and Smith, Steve (1990) *Explaining and Understanding International Relations*. Oxford: Clarendon Press.
- Howe, R. (1994) 'A Social-Cognitive Theory of Desire', *Journal for the Theory of Social Behaviour*, 24: 1-23.
- Hurd, Ian (1999) 'Legitimacy and Authority in International Politics', *International Organization*, 53: 379-408.
- Jackson, Patrick and Nexon, Daniel (1999) 'Relations before States', *European Journal of International Relations*, 5: 291-332.
- Kalyvas, Stathis (1996) *The Rise of Christian Democracy in Europe*. Ithaca: Cornell University Press.
- Katzenstein, Peter, Keohane, Robert and Krasner, Stephen (1998) 'International Organization and the Study of World Politics', *International Organization*, 52: 645-85.
- Keohane, Robert (2000) 'Ideas Part-way Down', *Review of International Studies*, 26: 125-30.
- Kratochwil, Friedrich (1989) *Rules, Norms, and Decisions*. Cambridge: Cambridge University Press.
- Kratochwil, Friedrich (2000) 'Constructing a New Orthodoxy? Wendt's "Social Theory of International Politics" and the Constructivist Challenge', *Millennium*, 29: 73-101.
- Laffey, Mark (2000) 'Locating Identity: Performativity, Foreign Policy and State Action', *Review of International Studies*, 26: 429-44.
- Laitin, David (1998) *Identity in Formation: The Russian-Speaking Populations in the Near Abroad*. Ithaca: Cornell University Press.
- Lake, David and Powell, Robert (1999) 'International Relations: A Strategic Choice Approach', in D. Lake and R. Powell (eds), *Strategic Choice and International Relations*. Princeton: Princeton University Press. pp. 3-38.
- Lapid, Yosef (1989) 'The Third Debate: On the Prospects of International Theory in a Post-Positivist Era', *International Studies Quarterly*, 33: 235-54.
- Legro, Jeffrey (1996) 'Culture and Preferences in the International Cooperation Two-Step', *American Political Science Review*, 90: 118-37.
- Lewis, David (1969) *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.
- Lynch, Marc (1999) *State Interests and Public Spheres: The International Politics of Jordan's Identity*. New York: Columbia University Press.
- Mandelbaum, Maurice (1955) 'Societal Facts', *British Journal of Sociology*, 6: 305-17.
- March, James and Olsen, Johan (1998) 'The Institutional Dynamics of International Political Orders', *International Organization*, 52: 943-69.
- Mas-Colell, Andreu, Green, Jerry R. and Whinston, Michael D. (1995) *Microeconomic Theory*. New York: Oxford University Press.
- Meyer, John and Jepperson, Ronald (2000) 'The "Actors" of Modern Society: The Cultural Construction of Social Agency', *Sociological Theory*, 18: 100-20.

- Moravcsik, Andrew (1997) 'Taking Preferences Seriously: A Liberal Theory of International Politics', *International Organization*, 51: 513-53.
- Norkus, Zenonas (2000) 'Max Weber's Interpretive Sociology and Rational Choice Approach', *Rationality and Society*, 12: 259-82.
- Nye, Joseph (1987) 'Nuclear Learning and US-Soviet Security Regimes', *International Organization*, 41: 371-402.
- Pettit, Philip (1995) 'The Virtual Reality of *homo economicus*', *The Monist*, 78: 308-29.
- Rabin, Matthew (1993) 'Incorporating Fairness into Game Theory and Economics', *American Economic Review*, 83 (5): 1281-302.
- Rabin, Matthew, and O'Donoghue, Ted (2001) 'Choice and Procrastination', *Quarterly Journal of Economics*, 116 (1): 121-60.
- Raub, Werner (1990) 'A General Game-theoretic Model of Preference Adaptations in Problematic Social Situations', *Rationality and Society*, 2: 67-93.
- Reus-Smit, Christian (1999) *The Moral Purpose of the State*. Princeton: Princeton University Press.
- Ruggie, John (1983) 'Continuity and Transformation in the World Polity', *World Politics*, 35: 261-85.
- Ruggie, John (1998) 'What Makes the World Hang Together? Neo-utilitarianism and the Social Constructivist Challenge', *International Organization*, 52: 855-85.
- Schelling, Thomas (1960) *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
- Schelling, Thomas (1978) *Micromotives and Macrobehavior*. New York: W.W. Norton & Co
- Searle, John (1995) *The Construction of Social Reality*. New York: Free Press.
- Smith, Steve (2000) 'Wendt's World', *Review of International Studies*, 26: 151-64.
- Spruyt, Hendrik (1994) *The Sovereign State and its Competitors*. Princeton: Princeton University Press.
- Stigler, George and Becker, Gary (1977) 'De Gustibus Non Est Disputandum', *American Economic Review*, 67: 76-90.
- Sugden, Robert (1989) 'Spontaneous Order', *Journal of Economic Perspectives*, 3 (4): 85-97.
- Sugden, Robert (1993) 'Thinking as a Team: Toward an Explanation for Nonselfish Behavior', *Social Philosophy and Policy*, 10: 69-89.
- Walker, R.B.J. (1987) 'Realism, Change, and International Political Theory', *International Studies Quarterly*, 31: 65-86.
- Walt, Stephen (1998) 'International Relations: One World, Many Theories', *Foreign Policy*, 110: 29-46.
- Waltz, Kenneth (1979) *Theory of International Politics*. Boston, MA: Addison-Wesley.
- Weber, Cynthia (1998) 'Performative States', *Millennium*, 27: 77-95.
- Weibull, Jorgen (1995) *Evolutionary Game Theory*. Cambridge, MA: MIT Press.
- Weingast, Barry R. (1995) 'A Rational Choice Perspective on the Role of Ideas: Shared Belief Systems and State Sovereignty in International Cooperation', *Politics and Society*, 23 (4): 449-64.
- Weldes, Jutta (1999) *Constructing National Interests: The US and Missiles in Cuba*. Minneapolis: University of Minnesota Press.
- Wendt, Alexander (1992) 'Anarchy is What States Make of It: The Social Construction of Power Politics', *International Organization*, 46: 391-425.
- Wendt, Alexander (1994) 'Collective Identity Formation and the International State', *American Political Science Review*, 88: 384-96.
- Wendt, Alexander (1998) 'On Constitution and Causation in International Relations', *Review of International Studies*, 24, Special Issue: 101-17.
- Wendt, Alexander (1999) *Social Theory of International Politics*. Cambridge: Cambridge University Press.
- Wendt, Alexander (2001) 'Driving with the Rearview Mirror: On the Rational Science of Institutional Design', *International Organization*, 55 (4): 1021-51.
- Young, H. Peyton (1998) *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton: Princeton University Press.

Re
the
ha
ab
un
an
ey
lis
ap
alt
of
no
ap

th
ch
sic
so
us
rat
br
su
ch
try
so
wa
bl
wi

co
mu
de