

# SYMPOSIUM ON EXPLANATIONS AND SOCIAL ONTOLOGY 1: RATIONAL CHOICE THEORY AND SOCIAL EXPLANATION

JOHN FERREJOHN  
*Stanford University*

---

## 1. INTRODUCTION

In the *Common Mind*, Pettit argues that rational choice theory cannot provide genuine causal accounts of action. A genuine causal explanation of intentional action must track how people actually deliberate to arrive at action. And, deliberation is necessarily enculturated or situated "... we take human agents to reason their way to action, using the concepts that are available to them in the currency of their culture" (p. 220). When deciding how to act, "... people find their way to action in response to properties that they register in the options before them, properties that are valued in common with others and that can be invoked to provide at least some justification of their choices" (p. 272). That people seek to make justified decisions implies that, at times, their own goals or objectives will be modified in deliberation. Something that rational choice theory cannot allow.

Rational choice explanations cannot explain in this way because they do not represent people as fully situated deliberating agents. The theory instead represents agents "... as means–end automatons: as black boxes designed, we know not how, for the instrumental pursuit of certain goals" (p. 268). Ordinarily, these goals are taken as fixed and given by the economic situation of the agent. Rational choice explanations "... will not tell us how the agents themselves are supposed to work out their decision. And it will not allow that they might be as concerned about the ends they should be pursuing, as they allegedly are about the means that are best suited to the given ends" (pp. 268–9). But, "... no culture encourages people to reason just terms of the economic and

social gains that different options represent, at least, outside of certain narrow contexts like that of the market" (p. 220). So, on Pettit's view, rational choice theories point to the wrong sorts of considerations as guiding action; and it ignores the fact that preferences are neither fixed nor firmly anchored in self interest. Nevertheless, Pettit does believe that rational choice theory can provide other kinds of explanation of action: standby or resiliency explanations.

A standby explanation of action does not explain it by pointing to its causal history, but provides instead an account of forces that would work to keep the action or action pattern stable or robust under counterfactual perturbations. Pettit offers this example to illustrate resiliency explanation:

A ball rolls along a plane on a straight line, say in accordance with Newton's laws. Suppose the path of the ball is marked out with little pinball posts on either side and that these posts are equipped to serve a double role: they will deflect any force that would push the ball off its path; and they will tend to return the ball to the straight path in the event that a perturbing force is effective. In such a case not only does the ball actually roll on a straight path; it rolls on such a path resiliently . . . Suppose now that we want to explain the resilience of the ball's rolling in a straight path. In this case we will need to mention the pinball posts, although those posts play no role in actually producing the straight-path behaviour; they are merely standby causes. (pp. 276–7)

When it comes to social theory, Pettit argues that rational choice theory can play the same role as the pinball posts. Suppose there is in place an established norm, such as the convention of driving on the left side of the road, and that this convention is regarded as giving reasons for action. Normally, drivers will treat the existence of this convention or norm as the reason to drive on the left hand side of the road and they will tend to drive on the left. Or, perhaps, in the presence of such a settled convention or norm, they form habits of conforming. A causal account of agent action will have to reflect these mechanisms. Pettit thinks it implausible that the rational choice account of conformity – that failing to conform could be extremely costly to the agent – could be a causal explanation of her action. Still, he thinks that the rational choice account might explain the resiliency of the convention. He quotes David Lewis to this effect: "If that habit ever ceased to serve the agent's desires, it would at once be overridden and corrected by conscious reasoning." (p. 281). The idea is that the self interests of the agent would be employed to override habit if, counterfactually, the habit were to be self defeating. So, while the agents do not choose actions as rational choice theory dictates, the resilience of their actions is due to the fact that they are in the self interest of the agents.

There is a lot to be said in favor of Pettit's characterization of

standby or resiliency explanations, in that it helps give sense to common forms of social theory that appear to be problematic from other viewpoints. He argues, for example, that functional explanations – explaining actions in terms of their consequences – can be understood better as resiliency rather than causal accounts. The trouble for functionalism is that the effects of an action follow the action and cannot play a role in a genuine causal account. But a resiliency explanation would not provide a causal story of how the actions were chosen but would say only that if there was, counterfactually, a departure from the action pattern that jeopardized its beneficial consequences, those who suffered the consequences could and would intervene to correct the deviation. There are, on this picture, causal processes that would come into play if the beneficial effects were jeopardized.

Pettit is led to adopt the resiliency account because of his requirement that social explanation account for the actions of thinking creatures and not merely intentional systems. He thinks that the agents of a rational choice theory are driven by narrowly defined self interest, and regards that self interest as arising non-deliberatively. And if the agents of rational choice theories do not deliberate about their preferences, but merely calculate about which acts to take given fixed preferences (and perhaps mechanically automatic beliefs too), they are merely intentional systems and not thinking creatures at all. So, if we want to find an explanatory role for rational choice theory for thinking creatures like us, that role cannot be to provide actual causal explanations. The aim must be, in a sense, lower.

As attractive as resiliency explanations can be, the restriction of rational choice theory to this kind of explanation is unnecessarily restrictive and I shall argue against it. I do not think that rational choice theory is always, or even usually, restricted to resiliency accounts. Indeed, within ordinary market contexts, Pettit himself thinks that it is plausible and legitimate that thinking agents will take deliberative account of their rational self interest.<sup>1</sup> But even outside markets, it seems to me that resiliency explanations depend on rational choice considerations playing an actual and not merely a standby role. This is because resiliency explanations depend for their adequacy on the existence of causal mechanisms that would come into play in counterfactual situations. Unlike the physical case of the pinball posts, those mechanisms are activated by rational choosers who would act to preserve threatened benefits. It is a fact that agents in the actual world anticipate, correctly, how they would rationally respond if the benefits were

<sup>1</sup> Some kinds of markets do raise problems for rational choice explanations as Pettit understands them. Labor markets, for example, are domains in which non-self interested factors ordinarily play an important role in deliberation and choice.

threatened, that they have reason to keep conforming to the norm or convention.

More fundamentally, I think that while Pettit does succeed in showing that certain specific kinds of rational choice explanations – those driven by narrowly defined self interest – may not figure prominently as causal accounts outside the market setting, this does not imply that rational choice theory, when separated from strictly the pursuit of narrow self interest, cannot supply causal explanations outside market contexts. Pettit's argument that rational choice cannot provide causal explanation is grounded in his claim that rational choice theory is tied to the pursuit of narrowly defined and fixed self interest. But if the theory is shorn of this aspect – if it is taken as a kind of coherence theory in which agents actions, beliefs and desires fit together as the axioms require – there is no reason that culturally defined value standards cannot be represented within it. Rational choice explanations in political science, for example, commonly assume that agents are ideologically motivated. And if a political ideology is taken to be a set of beliefs and policies that are aimed at making the world better for some or all people, ideology can serve to justify actions. And, plainly, ideologically motivated agents can criticize and change subsidiary goals and objectives. So, as far as I can see, Pettit's complaint is not really directed at rational theory itself, but against the particular version of rational choice theory in which agents pursue narrow and fixed self interest.

So, in the succeeding sections I will pursue two arguments. The first is that even if rational choice agents are understood in the way that Pettit understands them – as aiming only to pursue their narrow self interest – the idea of a resiliency explanation is not sharply separable from causal explanation. Then I will argue that, if rational choice theory is understood more broadly, there is no reason why it cannot provide causal explanations for the actions of thinking creatures (or human beings).

## 2. EXPLAINING INSTITUTIONS AND NORMS: RESILIENCY EXPLANATION

In the last few years an important project of rational choice theory has been to explain the emergence of norms or institutions among rational agents who are motivated only by narrowly defined self interest. This neo-Hobbesian project can be understood as an attempt to come up with a kind of reductionist account of norms and institutions.<sup>2</sup> Recently,

<sup>2</sup> There are reasons to be skeptical about such a project, at least as a general explanation for altruistic behavior or morality. Some of these reasons are internal to the explanatory project itself. Do these explanations actually work according to the internal criteria of

theorists have begun to make use of noncooperative game theory for the purpose of constructing accounts of emerging institutions. This technology has the advantage of forcing analysts to represent explicitly both the strategies and the information available to the agents. Noncooperative representations appear to model causal processes or, at least, causal processes that would occur following specific sequences of play. They are supposed to supply causal mechanisms by which self-seeking rational agents would invent and adopt moral norms, or institutions for “solving” various collective action problems and be guided by them, out of considerations of self interest.

In some ways Pettit’s strategy of resiliency explanation resembles noncooperative game explanations of norm emergence. Both accounts rest on notions of what agents would do at events that could occur counterfactually or “off the equilibrium path” of play. However, in a game theoretic account, these responses by the actors are common knowledge conjectures as to how the game would continue following an unanticipated action. Indeed, these beliefs are what leads the agents to keep playing equilibrium strategies. In resiliency explanations, however, the status of these counterfactual causal processes is less clear. Pettit argues only that if the deviation were sufficiently threatening to the interests of an agent, then she would, in fact, respond by choosing an

rational choice theory? There are two ways that existing efforts have been inadequate. First, is it the case that the behavior in question can be supported as equilibrium behavior among rational self-interested agents? The classical example of this problem is “anonymous” tipping. How can we account for the existence of an effective norm that directs strangers to leave tips in restaurants that they will never visit again? While it may be easy to show that such a norm might be desirable, it is much harder to see why it would be complied with in conditions of anonymity without positing an *ad hoc* motivation for compliance. Second, even if the behavior in question is supportable as an equilibrium, is that equilibrium unique? In general, whenever altruistic behavior can be supported as an equilibrium among self-interested actors, so can all kinds of other (non-altruistic) behavior.

Other reasons for skepticism are methodological: in what sense are self seeking agents more basic or simple building blocks of social theory than agents with other kinds of motivation? Why not prefer Aristotle’s starting point that takes man to be a social animal and suggests that his constitution as a self-seeking individual is constructed rather than primary? Pettit’s objection seems to fit within this latter line of criticism but he offers a new way of putting the objection. Real human beings are thinking agents and, as such, do not make decisions in the mechanical way that simple self-seeking (merely intentional) systems would. As a result, to assume that humans are self-regarding in the simple way that some rational choice theories do, is to ignore the possibilities that real people would not. Specifically, real people might well take their preferences, as well as their actions, as subject to deliberation rather than regarding them as fixed or parametric. To rule out this possibility by assumption is to place artificial limits on our capacity to understand human behavior. But, while the reductionist project might not be promising a general explanation of the emergence norms, it might provide valid explanations for particular norms and institutions.

action that defends that self interest. But, is it known in advance by the agent, or by everyone, that she would take her material interests as reasons for action? If such responses were expected by everyone, then it seems that the agents are actually taking into deliberative account their self interest (and that of others) in deciding on what to do or what to refrain from doing. Pettit may wish to avoid that conclusion by assuming that the agents are unaware of how they would respond to perturbations. But, if these responses are not known by everyone, but only by the social theorist, then it is hard to have much confidence in the existence of such standby mechanisms without actual evidence of their use.

Milgrom, North and Weingast (MNW) attempted to explain the emergence of cross-border trading in medieval Europe by means of a game of repeated interaction. The problem with cross-border trade is that no over-arching state or legal system exists to enforce contracts, and so contractual promises will tend to lack credibility. Nevertheless, such trade did emerge in northern Europe and, at times, showed as much vitality as within-state trading. Why? MNW argue that a system of private courts emerged – they call this the law merchant – where traders could, for a fee, bring suit against contractual non-compliers. While the courts lacked enforcement powers, they did publish their findings, and non-compliers were shunned as trading partners. In other words, the traders themselves punished the non-compliers.

It is easy to see how such an arrangement would be stable – how it could be supported as an equilibrium – and therefore why, once it was in place, people would tend to play according to its rules. But, there is nothing in the MNW story that suggests that the courts were set up with this purpose in mind. In that sense the MNW story is a functional account – their argument explains the emergence of the institution based on its beneficial effects for those whose compliance is required – and as such is a good candidate for a resiliency explanation. Indeed, the authors do not try to provide a specific genetic or causal account of the emergence of the courts or of the practices of the merchants who made use of them.

However, if no causal explanation is available, a resiliency explanation of this kind seems to be little more than a “just so” story. After all, resiliency explanations depend on the existence of causal mechanisms that would work in if the action pattern was perturbed. And, in the case of explaining norms and institutions, where we are looking at an action pattern or institution that remains in place over long time periods, there is little doubt that such perturbations would actually occur from time to time, and that the specified standby causal processes would have had to come into play. The adequacy of the explanation depends, in some degree, on evidence coming to the fore that substantiates such responses. So, at least in explaining complex and long-lasting action

patterns – institutions, conventions and norms – resiliency explanations cannot plausibly rely on causal processes that always stay in the background.

### 3. THINKING CREATURES

In the *Common Mind*, Pettit develops a view of social explanation – holistic individualism – that has two components. Holism is the idea that social explanation is aimed at explaining the behavior of thinking subjects – subjects who have the capacity to be normatively guided – rather than the behavior of merely intentional systems. Individualism requires that acceptable social explanations of any kind and at any level must be instantiated through causal processes that operate at the level of individuals. These are both attractive, if elusive, ideas that are worth exploration; and they have implications for how rational choice explanations should be understood and evaluated. As I find the notion of individualism much less controversial than Pettit's idea of holism, I shall spend less time on it. It is the holistic requirement of social explanation that makes rational choice explanations problematic and so it is here that I shall concentrate my attention.

Pettit offers a specific view of rational choice theory. He sees rational choice theory as psychological in the sense it explains actions by reference to psychological states or intentional attitudes. But, he also thinks that what is distinctive about rational choice explanation, what separates it from what he calls ordinary intentional explanation, is that it posits causal antecedents to its intentional attitudes. True, he permits the content of preferences to depart somewhat from the wealth maximization idea common in some models, but preference is still driven by self interest in a fairly restrictive way. Rational choice explanations, then, are taken to be committed to explaining actions in terms of the exogenously given self interest of the actors.

There is no doubt that much actual rational choice theory is well described by this characterization. Elsewhere, I called this kind of rational choice model an example of thick rationality, meaning that it entailed making extra assumptions about preferences over and above what rationality itself requires. Thick rational theories often have strong and empirically testable implications. This added empirical leverage is typically why analysts are tempted to add the extra assumptions. But, such theories, because they have so many restrictive assumptions, are only applicable (approximately true) in a narrow range of circumstances (in the sense that they can be rejected in other circumstances). Practitioners of the Chicago school of economics, for example, usually assume that agents are motivated in one or both of the ways that Pettit supposes. Where that assumption is warranted it can be a fertile source



of explanation, but sometimes theorists export these thick assumptions to settings where there is little evidence in favor of them. One ought to be suspicious of such explanations even where they seem to be superficially successful.

I doubt that either of Pettit's claims – that rational choice theory is psychological and that its agents are motivated by self interest – are necessary for rational choice explanation. While rational choice theorists often describe their theories using psychological or mental terms (beliefs, preferences, etc.), I think that this is better seen as a convenient shorthand than a statement of anything fundamental to rational choice theory. The agents of rational choice theory must act in certain constrained ways and one account of the constraints involved is psychological and self-interested. But other accounts can be given too. Elsewhere, Debra Satz and I offered structural explanations for how these constraints may come to be satisfied in some circumstances, and others are available. What is necessary for agents to be rational is that their action patterns have to satisfy some coherence constraints. How they come to do that seems an open issue.

I also do not think rational choice theory is committed to any particular substantive conception of self interest. Nor do I believe that rational choice accounts are tied to the idea that preferences must come first, prior to action. It is true, that at some point, agents must come to have goals or objectives and that to be rational, these goals and objectives must be seen as “fitting” with agents' actions and beliefs. But the sense of fit seems to me to be open, at least, in principle. So, I see much less of a distinction than Pettit does between rational choice theory and ordinary intentional explanation. Rational choice theory is simply one variety of intentional explanation that has sometimes been formalized, and restricted, for various explanatory purposes.

Pettit also thinks that rational choice theory may be committed to a certain kind of more or less formal decision theory – what he calls Bayesian decision theory – and that its accounts of action depend on this formalism in some important way. In particular, Pettit thinks that Bayesian decision theory puts “aside the question of how agents are supposed to think and deliberate as they find their way to action ... Second, the preferred models tend to represent people as instrumentally rational ... taking ends for granted and pursuing the best means of realising them. In a phrase, the models used in rational choice depict human agents as means–ends automatons ...” (p. 268).

It is not clear that rational choice theory is actually committed to the Bayesian picture, however. Indeed, Pettit argues that holism requires an interpretation of decision theory – inference theory – that departs from the standard decision theoretic understanding, but does not actually contradict decision theory itself. In the inference theoretic account,



agents have beliefs about the world and evaluative beliefs about desirability. They deliberate in order to form true beliefs and to prefer desirable things. The set of true propositions and valuable objects is, in this account, naturally taken to be fixed and antecedent to choice.<sup>3</sup> Thus, inference theory takes a subject's desires (and beliefs) as aimed at actual desirability (and actual likelihood) and, so understood, we would expect desires and beliefs to be responsive to reasons for thinking that an object is more desirable or a state of affairs more likely. Preferences, on the inference theoretic account, are not seen as fixed, but are the subject of deliberation, in the same way that actions are. This way of looking at preferences and beliefs permits us to see them as corrigible and, in this sense, endogenous to the circumstances of decision making. Preferences can be expected to change in the course of decision making as agents consider what is entailed in pursuing this or that goal and as they deliberate to arrive at a decision. This is an attractive interpretation of rationality that illuminates the role of deliberation in decision making.

Pettit thinks that the kind of formalism represented in standard interpretations of decision theory – the formalism that represents preferences as fixed and beliefs as subject only to mechanical revision – implies that rational choice explanation falls short of satisfying the holism criterion. The creatures of rational choice theory fail to deliberate about what to want, believe and do in the way that thinking agents would. For this reason, Pettit doubts that rational choice theory can provide a causal explanation of the actions of thinking agents. What people do, they do as thinking and deliberating agents and the best explanation for their actions must always take fully into account the fact that people are guided by norms and reasons. That it is inconvenient and costly to have a child just now cannot explain why Mary chose to have the abortion. The explanation for her action must take account of norms that she accepts, perhaps considerations of identity, and other possible

<sup>3</sup> Years ago Gary Becker (1965) and others put forward a similar model – household production theory – as an interpretation of standard decision theory. The idea is this: there are things that are desirable, but these are not generally found in the market. Actual commodities that can be purchased in the market can be used, along with other inputs, such as the time of the consumer, to produce the basic or desired goods. Persons may be somewhat uncertain as to these production relations and need, therefore, to form conjectures or beliefs about how much desirable stuff can be produced from a specific bundle of goods purchased in the market. Desires over marketed products, therefore, are dependent on beliefs about how these products can be converted into desired things. Like Pettit's theory, household production theory supports an idea of deliberative rationality. When deciding what to purchase, a person should be trying to figure out how to get the optimal amounts of the basic desirable stuff, subject to whatever resource constraints she faces. The point is, from the standpoint of actual rational choice theories, household production theory is one interpretation. Moreover, as far as I can tell, whether or not that interpretation is adopted has no consequences for the formalism of the models.

sources of reasons for action. The economic costs, such as the cost to her career of having a child or the costs of having the abortion, would need to be brought deliberatively into play as constituting parts of reasons for or against action. And Pettit doubts that thinking creatures would bring such considerations into play except in extraordinary circumstances. If keeping the child would lead to the ruin of the family and to the certainty that the child itself would have a horrible life, those considerations may come into balance with other deliberative concerns. But, normally, life decisions of this kind would not be decided by taking self interest directly into account.<sup>4</sup>

In any case, as Pettit goes on to point out, the distinctive contribution of rational choice theory to social explanation is anyway not really to get a deeper understanding of the deliberations and decision making of individuals. "Most interesting rational choice explanations . . . involve an appeal to aggregative mechanisms . . . as well as to motivational ones." (p. 269) As examples, he points to Adam Smith's invisible hand, prisoner's dilemmas, free-rider problems, and Schelling's explanation of residential segregation. Each of these phenomena arises from aggregating the behavior of self-seeking, instrumentally rational agents. Thus, while rational choice theory is psychological in one sense, its aim is not really to provide a causal account of the agents' actions by invoking their mental states, but to understand how they interact with each other in institutional settings to produce aggregated results. It may not matter, therefore, that rational choice theory does not track the deliberative process very well if its value is elsewhere. But, if this is so, it is not so clear that the models producing the aggregate results that Pettit values, are models of thinking creatures rather than models depicting how simpler, "merely" intentional systems would interact. Indeed, it seems that Pettit sees each thinking creature as embedded in a merely intentional system that provides or would provide occasional reminders to us not to risk material ruin when we deliberate to arrive at a decision. Our guardian "angel" would not usually offer such advice and we would not usually take it, except in those exceptional circumstances in which our deliberations tempted us to such a disastrous course of action.

Pettit's holism – the idea that to be a thinking, as opposed to a merely intentional, subject, an agent must be interacting with others in certain non-causal ways – is also attractive but it is less clear to me which of its features are crucially relevant to disputes in social science. As I understand him, Pettit argues that thinking subjects are capable of being

<sup>4</sup> I use an abortion example to show that I think that even here Pettit's argument is not really fully convincing. After all, the self-interest considerations are not narrowly tied to Mary's well being, but to others in her family, including the child's, if it were to be carried to term.

guided by norms and that it is this capacity that allows such agents to be guided by considerations of deliberative and practical rationality. Holism seems to require that agents are motivated to act rationally, for example, and not simply to pursue whatever seems the best course of action at each moment. A holistic agent, as far as I can see, wants to act consistently with norms she holds or endorses, even where doing so may involve some sacrifice of value to her. Perhaps, too, holism entails that the agents take, or could take, a critical perspective on norms or rules.

That agents can be so guided allows their fellows (and observers generally) to have a kind of interpretive access to their thinking. This is so for two reasons: the first is ontological – there is *something* that is being thought: for thinking subjects “there has to be a certain determinacy of content” (p. 236). Thus, when we aim to understand what a thinking subject was doing by her own lights, there is a target we are aiming at. The second is epistemological: there is something about thinking subjects that makes their thought “commonable” or graspable to an observer.

I can see how Pettit’s holism, if true, makes intentional interpretation a coherent enterprise and, no doubt, a necessary part of understanding social phenomena. But, I must say, in spite of his attempts, there are some steps in the arguments that I do not really understand. Take, for example, determinacy of content. Why must thinking subjects have determinate thought contents? This seems to follow from the fact that such subjects have intentional control over such contents: “it [a thinking subject] also tries to be faithful to certain contents in the intentional states it forms; it tries to form such states as circumstances make it rational to hold” (p. 236). This seems to be saying that contents are determinate because the agent is engaged in some kind of optimization program and that determinacy follows because such a program always has a unique solution. But this seems false as a general statement. I can follow a course of action without settling on my objectives – such a course would amount to keeping my options open or not deciding things until I need to. Indeed, I think it is common that people sometimes shift and sometimes clarify or specify their objectives in the course of acting. Besides, even if we ignore this point, why should we think that thinking subjects will always succeed in exerting control over the contents of their intentional states?

I also have some doubts about the epistemological claim as well, and these are problems that Pettit recognizes on pages 238 to 239. I grant that thinking subjects will sometimes want their thoughts to be commonable to their fellows – the people they are in interaction with, and to their future selves as well. But how accessible does that actually make their thoughts to others in their community, to people outside their community, or to outside observers? Pettit thinks, and I agree, that this

problem (of opacity) is most difficult when trying to interpret what other species are up to. Here he offers a pragmatic justification for the effort, but offers no guarantee that we will ever really grasp what the cat or the monkey is really up to, from the inside. But, unlike Quine, he is more sanguine about other cases. He grants that historical interpretation will remain somewhat indeterminate because we are not in actual contact with the subjects, and are unable to negotiate with them in the appropriate way about how rules should be applied in new cases. In contemporary cases, he thinks the prospect for interpretive understanding is much better. But what if we take account, as some anthropologists and others have urged, of the possibility of genuine pluralism as between communities and subgroups? I will say more about this below in the context of a concrete historical/cultural example.

Pettit's holism thesis might also have quite controversial implications for social science practice; implications that might point away from the ecumenicism he embraces in his discussion of individualism. Let us grant that, subject to the caveats raised above, it is possible to get sufficient access to thinking subjects to make it possible to grasp the rules they are following (as a common possession) and to understand causal histories at the individual level. Other approaches to understanding individual choice, such as rational choice theory, seem to make only limited attempts to interpret in this strong sense. I have thought that the difference between rational choice theory and deeper hermeneutic approaches in this regard is partly a matter of degree and partly a matter of focus. But, perhaps something more profound is involved. I will go into this later, but suppose that there is a sharp difference between the kind of knowledge about thinking subjects rational choice theorists can obtain and the kind available to interpreters. Is that difference the difference between grasping a rule (internally) and the pragmatic (external) understanding we can have of cats or chimpanzees (presumably, not thinking subjects at all)? Put another way, does rational choice theory, as a methodology, permit explanatory access only to merely intentional systems, or only to that part of the behavior of thinking systems that could be understood without taking account of their being thinking systems? If so, the warm ecumenicism that Pettit endorses in the context of considering historical, structural and (what he considers) rational choice explanations, may not be warranted.

#### 4. RATIONALITY AND RATIONAL CHOICE THEORY

The notion of rationality can be characterized in various ways. Rationality might be taken merely as a consistency condition on choice behavior. If an agent makes a certain choice from one set of options, rationality requires that he choose the same option when it is part of a

smaller subset of the same options. Understood in this way (there are many alternative consistency conditions that have been proposed), there is no prior commitment as to what particular mechanism or mechanisms brings about the consistency. There must be something that does that job, of course, but the rationality hypothesis that does the work in rational choice theory is not committed to any particular mechanism for attaining consistency. The rationality assumption can, of course, be given a particular folk psychological (FP) or intentional interpretation. Rational agents are those whose choices maximize the satisfaction of their desires, given their beliefs; rational agents are consistent, on this account, because they aim (successfully) to satisfy their preferences in light of their beliefs. The FP version of rationality seems to correspond to Pettit's idea of merely intentional systems. It also seems to support a causal view of the relationship between beliefs, desires and actions. But the FP interpretation of rationality is not a requirement of rational choice theory; it is an optional interpretation of it.

Thinking of rationality as a mere consistency condition does not entail unrealism about how agents make choices or decisions. There has to be some causal explanation for how choices get made and that explanation may involve mental states in causal roles. But, whether they get made in the particular way that FP interpretation goes, does not seem to matter to rational choice theory as such. Indeed, empirical psychology seems to suggest that the specific mechanism in the FP interpretation is a bit shaky, at least, as general grounding for the rationality hypothesis. This worry about FP should not be read as a doubt as to whether people are more or less well described in its terms. Indeed, Pettit and I both think that, under normal conditions, people's behavior is very well described in FP terms and that, for that reason, people successfully understand themselves and others in the terms of ordinary folk psychology.

However, as I said at the beginning of this essay, I do not understand rational choice explanations in the way that Pettit does. For Pettit, rational choice explanations are psychological – in that they attempt to explain action in terms of mental states – and that they are committed to positing some substantively defined (and more or less fixed) desires or preferences as having a causal role in bringing about actions. Moreover, agents' preferences are taken to be based on their self interest in a fairly narrow sense of (usually) being concerned to pursue wealth or status (social acceptance). But, even if these were not the goals attributed to agents, I believe that Pettit would say that some such substantive assumption about desires must be made in any rational choice theory. He writes that:

rational choice explanation is a strategy of psychological explanation ... distinguished among psychological approaches by two assumptions: that

people are predominantly concerned with self regarding ends like economic gain and social acceptance; and that in pursuing those concerns they conform to formal models of rationality of a kind illustrated by decision theory. (p. 269)

I believe that whether or not rational choice explanations are psychological explanations is a matter of interpretation. If the rationality hypothesis is taken merely as a kind of consistency condition on patterns of choice – and this is the way rationality works within formal models – and if the mechanisms by which this consistency condition is maintained are left open, the psychological interpretation is no more than one interpretation. Secondly, whether or not desires or preferences are to be taken as given, or antecedent to choice, is also a matter of interpretation. As far as I can see, there is no requirement within the formal theory that choices adapt to beliefs and desires, rather than that desires and beliefs adjust to choices. It may be harder to make (inner) psychological sense of such adjustments, but, from an external or explanatory viewpoint, a rational choice account is only necessarily committed to consistency.<sup>5</sup> In any case, examples of such adjustment are familiar. Pascal tried to change his beliefs by adopting a course of action that would lead to his holding those beliefs. And Aesop's fox changed his beliefs as to the tastiness of grapes which are out of reach to justify giving up the futile effort to get them. As for desire adjustment, Stoics and Buddhists both think that tastes for material things are habit forming and can be resisted through abstemiousness.

In any case, looking at actual explanatory practice, it seems misleading to take rational choice theory as committed to any particular motivational assumptions. Indeed, practitioners, when considering any new explanatory endeavor, always start by worrying about what it is that motivates the agents in the target domain.<sup>6</sup> Sometimes these worries get very fine: when studying political competition, there is a great deal of (theoretically consequential) disagreement about whether politicians are vote maximizers (simple-minded seekers of re-election as the parody has it), or probability-of-election maximizers, or idealistic policy seekers, or whether they are maximizers at all. One response to the discovery that different assumptions about agent preferences produce different empirical implications, is to try to develop a theory that is robust to

<sup>5</sup> For example, in simple coordination games with multiple equilibria, we can describe each equilibrium in terms of the system of beliefs needed to support it. If, in one equilibrium everyone drives on the left-hand side of the road, that practice is supported by the belief that others will be driving on the same side (based on their own beliefs). The most that a game theorist can say about such games, if she can say even this much, is that one of the equilibria, with its associated system of beliefs, will be observed.

<sup>6</sup> This initial worrying seems to correspond to the hermeneutic aspect of rational choice theory.

variations in the plausible range of desires. Another response, more empirically motivated, is to take advantage of the consequence that different agent objectives produce different behavior to get more empirical leverage on the phenomenon. In either case, both these explanatory practices suggest that rational choice theorists do not generally adopt any fixed or narrow self-interest assumption, but leave the question of the content of agent preferences open for investigation. At least, that is what I would call “best practice”.

In this sense, rational choice theory is not a specific theory any more (and probably less) than physical theory is a specific theory. It is an approach, a methodology, or a schema for understanding some aspects of social interaction and, as such, embraces an indefinite number of specific theories. These theories have in common only a general commitment to the formalism of intentional psychology and, technically, probably not even that much. There are certainly thinkers within rational choice theory who challenge the notion of maximization, for example, or who doubt that preferences are well ordered in the way that is standardly assumed in many models. There are lots of such theories and explanations within the general enterprise of rational choice theorizing and the Chicago school approach represents a small but influential brand.

This point can be illustrated by an example from neoclassical economics. Neoclassical economics recognizes two kinds of agents, consumers and firms. Firms (in competitive markets) are supposed to have substantively specified objectives – maximizing profits at fixed prices and given technical constraints. Consumers, on the other hand, are taken merely to have preferences and (in extensions, beliefs) and to make their choices to maximize preference satisfaction. The reason for this thinness in consumer rationality is that the economist is in no position to place substantive restraints on consumer preferences, and contents herself with imposing only weak assumptions of convexity, smoothness, non-satiation, and so on. Obviously, given the lack of constraints on consumer preferences, there are few theorems in the theory of the consumer. But both consumer theory and firm theory are rational choice theories as I use the term. Using Pettit’s terminology, it seems more useful to think of rational choice theorists as committed only to a weak version of intentional explanation: concerned with explaining the individual and collective behavior of intentional systems. Obviously, where there are good reasons for it (as I think there are in the neoclassical theory of the firm) theorists may add substantive assumptions and derive empirically fertile implications.

Secondly, while it is true that rational choice theorists do, as part of their enterprise, try to bring the phenomena they are studying under a rational choice description – showing how what the agents did could be



understood as rational under the circumstances – this is seldom the main point of the enterprise. The aim is to explain some suitably characterized phenomenon and not merely to show that it can be understood as the product of the choices of rational agents. And the explanation sought typically relates more or less measurable or observable phenomena to each other. The theorems sought for explanatory purposes are ones that posit regular connections between, for example, prices and quantities, or between the rules of an electoral system and the number of parties that can be observed.

This might be illustrated with an example. Suppose an economic historian – whom we take to be a card carrying rational choice theorist – wished to explain historical variation in fertility behavior (birth rates, or male–female ratios). She would begin by *assuming* that the agents in question were rational (they had some objectives they were trying to maximize under constraints) and possibly that they were specifically motivated to maximize wealth. Then she would try to represent the constraints on the agent’s maximization problem and establish the relationships between these constraints and choice behavior. These “comparative static” relations could then be evaluated statistically. The point of the enterprise is not to show that the agents are rational, or that they are wealth seekers – those are maintained hypotheses – but to understand how parametric shifts in the constraints would affect rational behavior. I suppose that if it were found that much of the historical or comparative variability in fertility could be statistically accounted for by such a model, it would give a reason (but a weak one at best) to think that rational choice theory, or one version of it, is useful in explaining these phenomena. But, since the theory is not compared to “non-rational” alternatives, it is hard to see vindication of the rational choice approach as the economic historian’s concern in any stronger sense than this.<sup>7</sup>

I think that this is pretty much the standard way that rational choice theory is applied to explaining social phenomena. The purpose of the enterprise is to develop what practitioners call “positive” theory: a set of testable propositions that follow from the rationality assumptions combined with the constraints faced by the agents. There are some additional principles or maxims that most practitioners use, too, as guides in developing theory. One is implicitly illustrated above: the constraints rather than preferences or beliefs should do most of the explaining. This is a view advanced by Gary Becker, who believes, as a

<sup>7</sup> To take an alternative favored by Pettit, satisficing agents would be sensitive to shifting constraints on choice in more or less the same manner that maximizing agents would be. Qualitatively, both accounts would probably generate the same comparative statics predictions in this case. So, finding that the comparative statics predictions of both models are confirmed in fertility data, cannot separate the two models.

practical matter, that not much can be known (by observers) about preferences and, consequently, that good rational choice explanations ought to be driven by variations in constraints rather than variations in preferences. Recently, with the advent of informational issues, there is some methodological reluctance to build very much on the fact, or possibility, of disagreement in beliefs. But these maxims are both controversial and not necessary features of rational choice theory.<sup>8</sup>

### 5. INTERPRETATIVE EXPLANATION

As an approach to explanation, rational choice theory is often contrasted with the hermeneutic or interpretive approaches. In these latter approaches, the observer tries to take the point of view of the agent to see her actions as intentionally chosen in light of her own situation (preferences, beliefs, choice contexts). As Pettit notes, this approach is common to anthropology, history and some kinds of sociology (of the kind associated with Irving Goffman, for example). The point of the enterprise is to display choices as intelligible in the view of the agents. This typically entails discovering what the subjects want and believe and, indeed, this is sometimes the whole point of the enterprise. Rational choice theory, as I described it above, obviously has an embedded interpretive perspective. As part of getting to the theorems, the analyst considers how things look from the perspective of the agent. The agent in question, though, is typically a fairly generic character and not usually a concrete (situated) person with a particular history and embedded in a specific community or group.

Pettit's focus on rule or norm following, and his endorsement of the inference-theoretic approach, suggests that it is here that the thinking part of the thinking subject comes crucially into play. Interpretive social science is, presumably, aiming at getting hold of the commonable rules that the agents are following or (in historical applications) did follow. It seeks access to the deliberative and evaluative and practical criteria employed by thinking agents. Its practitioners are able to do this because, according to Pettit's account, thinking agents must be following rules that are public, in a way that makes their actions and mental states accessible to observers (even those at distant removes of time and culture, in addition to members of their own community.)<sup>9</sup>

<sup>8</sup> Indeed, experimental economists have developed various techniques to "induce" preferences in experimental subjects (to be sure, preference induction in human subjects typically relies on a background assumption about subject preferences outside the experiment: they are assumed to prefer more money to less).

<sup>9</sup> I have some questions about this account. The scope for an anthropologist to negotiate with subjects over new cases seems pretty limited. Is it really possible for such an observer to grasp or share a rule with her subjects, or is her knowledge of rules always external?

Perhaps Pettit would say that economists, when they set up the agents in their models, are not aiming at commonable rules at all but are merely using the interpretive perspective as a device or stance, serving to set up the model and derive testable propositions. Model agents are not, after all, real agents at all and so the economists' "interpretation" of their model is not directed to a proper target. But, prior to setting up an economic model, presumably economists try to get some sense of how the real agents, whose behavior they are trying to explain, are actually motivated. Moreover, this exploration of agent motivation generally continues within the scientific enterprise. What the agents want is, again, only a maintained hypothesis, not a dogma. The effort to determine agent motivation seems to involve genuinely interpretive activity, even if economists are typically less than explicit about how this is done. Typically, rational choice theorists appeal to intuition or to notorious "stylized facts" when writing down an analytic model. They exert too little effort to learn how the target agents see things.<sup>10</sup> There is much common ground between the intuitions of rational choice theorists and the interpretive enterprise of humanistic social scientists. That common ground merits exploration.<sup>11</sup>

So, let us assume that hermeneutics is, in some sense, an intrinsic part of individual level social explanation, as much for rational choice theory as for anthropology. Still, there are reasons to think that the interpretive part of the enterprise, alone, will often yield indeterminate or conflicting results. Consider the interpretation of a specific social practice – let us say the conduct of elections in seventeenth-century England.<sup>12</sup> Historians of elections in that period fall roughly in two camps based on what they believe elections were about. The first group sees elections for public office in that period as more or less like elections at any time or in any place. That is, elections are aimed at choosing among candidates for public office – the people we want to lead us in making law and policy. The other group sees elections in that period

<sup>10</sup> In some domains of social science such efforts have been undertaken in a serious way. For example, rational choice theorizing about legislative behavior has profited greatly from the detailed and sensitive ethnographies of the American Congress that Richard Fenno (1965, 1973 and numerous other books) has developed over the course of his career. His work has influenced many of the formal models of legislatures. This is not to say that any of those models have taken up Fenno's characterizations in full. But there has been a continuing dialogue in this area between ethnographic and formal accounts, which has benefited both traditions.

<sup>11</sup> My claim that economists engage in two kinds of interpretive activity – one directed at real agents, the other at model agents – suggests that these two activities are connected. Indeed, I think they are. Through trial and error in model development, the interpretation of model agents helps the rational choice theorist gain some intuitive grasp over what real agents might actually be doing.

<sup>12</sup> This is drawn from John Ferejohn (1991).

(early seventeenth-century England, or perhaps in early modern Europe generally) as occasions for endorsing the order of things by, for example, accepting as political leaders the natural leaders of social life (aristocrats and gentry with inherited claims on the office). On this latter account, elections should normally be uncontested. In this context, voting against a legitimate official is not a free expression of preference. Rather, it is an expression of profound alienation that would not normally be felt by enough people to generate a contested election – at least, not if things in the community are proceeding within socially acceptable bounds.

As a matter of fact, however, elections in the early seventeenth-century became steadily more competitive (there was an increasing frequency of contests with two or more candidates, more voting against “legitimate” office holders, more disputes about the results, more voter mobilization) and this competition peaked at the start of the Civil War. As it happens, two prominent camps of historians differ profoundly on how to understand this sequence of events (increasing electoral competition, widening franchise, more evidence of active contestation). The first approach takes the increase in competitiveness as a more or less natural working out of the logic of electoral competition in a period in which parliamentary office was becoming more valuable, because of the general growth in wealth and the more extensive role of government in distributing the spoils. Insofar as the growth in industry and wealth is part of modernity, the development of competitive election processes is seen as a kind of natural concomitant of modernizing England.

The second group of historians regards the same general facts as evidence of profound social rupture and breakdown. As the old and accepted medieval order failed to contain new social dislocations – arising from religious disputes, rising commercial and industrial activity, and increased mobility – ancient electoral practices broke down. Elections became new battlegrounds; and contested elections, specifically, were seen as socially fracturing outbreaks of war and hostility among families and social groups. On this account, the rise of competitive elections is not a forward-looking indicator of modernization, but is a signal of the breakdown of an older system of social order.

My view is that both of these schools are seeing something that is true in the transformation of election practices. Both breakdown and modernization were occurring within the same practice; and there is evidence that the participants themselves were divided as to how to understand what was going on. One way to think about the situation is in terms of a proto-typical example. Suppose the local aristocratic family has controlled the parliamentary seat for a long time and, according to shared understandings in the community, deserves to be selected again. But larger issues, of religious division, foreign policy, trade, and so on, were increasingly having an impact on the local people and one of these

locals (from a family with no “entitlement” to the seat) decides to stand for the office in order to push for policies of common concern. By standing for office, the candidate expresses either of two positions: first, that he is not any longer following the old rule (support for and deference to the local leaders) as before, but is following, or trying to establish, a new one (contesting for power under emerging normative rules governing elections); or, second, that he is following the old rule in new circumstances (by challenging the normative claim of the incumbent to the seat on grounds the rule does, or should, recognize). This disagreement – are we at a stage in a smooth historical process or at a point of rupture? – is roughly what divides the two historical interpretations.

The point is that, given the available record, it does not seem very likely that this dispute can be conclusively settled. Indeed, there has been an oscillation between the two schools for generations, each time with new evidence and new sources. This disagreement may be due to the incompleteness of historical records. There may be a right answer here, but we simply do not have enough data to pin it down. Or it may be that there was not a unique commonable rule being followed at all. There may have been two such rules (or more) and participants might have disagreed as to which was normatively compelling. One would not be too surprised to learn that those with inherited claims on office might have seen elections quite differently from those who did not. More interestingly, participants and observers might well have been internally conflicted over the appropriate norms. I would expect disputes about norms to be common in periods of transition and dislocation.

I am not sure how Pettit would respond to this example, but I do not think this kind of issue is uncommon in disputes among historians. It seems that he could respond at either the ontological or epistemological level. If there are two norms offering conflicting accounts and appraisals of a social practice, each with some appeal to participants, the situation is one of genuine conflict. In this case, the conflict seems quite profound and intractable: there is no unique way to go on in applying the norms to new circumstances. Or, perhaps it is not yet clear to the participants (or to later historians) whether there is one rule or two. That is, the issue is being disputed (this might be a case of ontological indeterminacy). Or, perhaps it is a case of epistemological indeterminacy – arising merely from the inability of historians to negotiate with their subjects over how the common rule is to be applied to new situations.

## 6. HISTORICIST AND STRUCTURALIST EXPLANATIONS

Pettit’s individualism provides a way of reconciling different levels of causation of social facts; by taking advantage of the supervenience of

social facts on individual facts, it gives a kind of priority to individual level causal stories. It is an instantiation of causal fundamentalism applied to social phenomena; and the notion that higher levels of causation get their efficacy through “programming for” lower level causes seems very plausible. I am not sure how controversial such a view is among philosophers of science but, at the abstract level, I cannot see very much to be said against it. However, when Pettit applies the doctrine to particular problems – such as reconciling historical and structural with individual level explanations – I have some objections.

Pettit argues, against Jon Elster’s preference for fine-grained explanations, that more individual level information will not necessarily improve an historicist explanation. An historicist explanation is one that asserts a causal relation between events separated in time. The example he gives is this: that the widespread belief that the Soviets would not use force to back up the puppet regimes in Eastern Europe led to the collapse of those regimes. He admits that higher level (historical) causal statements can be true only in virtue of there being some individual level events that actually bring about the collapse of the regimes. But, given the general belief that the Soviets would not act, it was more or less inevitable that if some particular people had failed to rise up, others would have and the result would have been the same. I gather that the kind of study that Pettit would criticize would be one that goes into great detail in tracing individual level events in Germany, and insists that this information somehow adds to our understanding of the collapse of the communist regimes. Such information would, for Pettit, be irrelevant and distracting. If this is Pettit’s claim then I agree in part. The additional individual level information does not really help answer the question: “why did the communist regimes fall?” By suggesting that the actions of particular agents produced the fall (which was true) the individual level account attributes too much causal efficacy to those particular agents.

But suppose no such individual account was true or was consistent with individual level rationality. Suppose, for example, that the communist regimes actually were known to possess the resources and will to put down any foreseeable uprising (based on available information and the hypothesis that people were generally rational) even without Soviet help. In other words, suppose that the decline of Soviet willingness to intervene did not change the calculus of potential insurgents in any way. In that case it would not be true that the belief that the Soviets would not intervene was sufficient to bring about the collapse. Assume, further, that the particular people who rose up did it out of misunderstanding or extreme bravery or whatever. If by acting in this way, these agents provoked a widespread uprising that was not foreseeable (on available information), then it would not be misleading to emphasize their particular roles (and their mistaken beliefs or strange preferences). In

that case, it seems, those particular people did play a necessary causal role. Without exploration at the individual level, we do not have any hope of knowing whether particular agents brought about the fall in some such way or not.

I would put the matter this way. The fall of the communist regimes was a small set of events and the widespread belief that the Soviets would not intervene was a belief about a more or less singular phenomenon. The assertion of a high level causal connection between such things is bound to be pretty unconvincing taken by itself. So, it seems likely that the production and investigation of individual level information, while possibly not being part of the best explanation itself (if it turns out that the individual regimes were unable to put down an uprising, so that Pettit is right that if it had not been these people it would have been some others), can give confidence or erode confidence that the historicist account is a good one. Evidence that the people taking the first rebellious steps believed that their local regime would not retaliate – or that soldiers would not act on orders to do so – would increase confidence that Pettit's explanation is adequate. In this sense the availability of individual level information can increase the confidence in an historical account without actually becoming part of the account itself.

This same point can be made with respect to structural explanations. A structural explanation tells a causal story that connects aggregate regularities. Pettit's favorite example is the claim that unemployment increases crime. Pettit supposes that this could be a true causal account only if an increase in unemployment "programs for" individual level activities that do the work of producing an increase in crime. It seems that many particular mechanisms could do the job. An increasing crime rate needs only a few individuals, located anywhere in society, to change their behavior, so would not depend on any particular individuals' actions. A typical account might go this way: when someone becomes unemployed, either the cost of criminal activity decreases or its benefit increases. Someone who just barely preferred the straight life while employed, would be pushed over the edge to preferring a criminal career after the loss of a job. Sociologists may, in fact, already have in hand an individual level account of how it is that unemployment increases crime. The availability of that account (or some account) is what makes the structural story plausible. And research on individual level mechanisms can work to increase or decrease confidence in the structural account.<sup>13</sup> The accounts of the behavior of particular indivi-

<sup>13</sup> Pettit shows that such an account depends on some background assumptions – perhaps in a sunny beachside community the increase in unemployment did not make criminal activity more attractive relative to sunbathing. In that case (in economesse once again) the opportunity cost of crime would not have declined.



duals will not figure in the best explanation of the increase in crime (which is structural) but those accounts are relevant to the question of whether that explanation is adequate.

Elsewhere, Satz and I have argued that good explanations of social phenomena – including structural and historical ones – must satisfy (only) a weak “rationality constraint” which requires that it is possible that the involved agents could find it rational to act as they are required to in order to make the explanation true.<sup>14</sup> Roughly, this means that higher level explanations must pick out outcomes that can be supportable as strategic equilibria among the agents populating the system. This is a weak constraint in that it only asks that it is possible that agents could rationally have acted in the ways the explanation requires. It does not require that rational agents involved in the situation would necessarily pick out that action pattern (because there could be other equilibria), but only that no one of them would rationally have had a reason to change her actions.

A chronic problem with individual level rational choice explanations is that, typically, they are incomplete in the sense that strategic interaction among rational agents can usually support many different outcomes as equilibria. More can be said or explained about the actual outcome than can be gotten from the rationality of the agents. Thus, the rationality constraint does not require that all explanations work only at the individual level, or that better explanations do, and certainly does not support any kind of eliminativism. Structural, historical and cultural information ought not to be thrown away and can often help pin down good explanatory accounts.

## 7. CONCLUSION

Pettit’s conception of social theory in *The Common Mind*, is an extremely important contribution to the methodology of the social sciences. His insistence that social theory aim at explaining the behavior of thinking systems is particularly important for those working in rational choice theory. Recognizing this point might lead many of them to abandon the idea that preferences must be understood as given or fixed. There is nothing in the formalism of rational choice theory that prevents theorists from taking this point on board. At the formal level, the theory requires only that actions, beliefs and preferences “fit” together in the right way. The theory is not committed to the further view that people can only adjust their actions to bring this fit about.

I think, too, that the notion of resiliency, or standby explanation, is a valuable addition to the repertoire of explanations for stable action

<sup>14</sup> *Op. cit.*

patterns or institutions. I have argued that such explanations, like game theory explanations, must take into account causal explanatory structures, and so are not as sharply different from causal accounts as Pettit argues. Obviously, the sense in which resiliency accounts depend on causal mechanisms is somewhat different from Pettit's own notion of a causal explanation. For him, a causal explanation reports on actual causal history, and excludes what would occur in counterfactual situations. In the case of stable patterns of action – norms and institutions – I doubt that such a sharp distinction is possible.

I think, finally, that Pettit's conception of rationality is far too narrow. Indeed, his objections to rational choice theory seem to be aimed at materialist motivation more than at the notion of rationality. Perhaps I am simply drawing the line between rational choice theory and hermeneutics in a different place. At least, when it comes to explaining behavior in non-market contexts, attempts at rational choice explanation normally ask what it is that the agents are trying to do. Sometimes those agents are pursuing wealth or status, sometimes they are pursuing other objectives. Once the objectives are established, the machinery of the rationality assumption comes into play. But, as I have suggested, while this two-stage procedure is normal explanatory practice, there is no particular reason to insist that the stages are actually separated, either for the analyst or for the agents themselves. Analysts can, and often do, go back and forth between trying to fix agent objectives and calculating their best actions in light of those preferences. I believe that agents do so as well. One makes a tentative judgement as to what to pursue and then sees what actions would be rationally recommended. If those actions are sufficiently unattractive, agents may very well adjust their goals or objectives. Agents acting in this way look to be deliberative or thinking agents as well as rational ones.

#### REFERENCES

- Becker, Gary. 1965. A theory of the allocation of time. *The Economic Journal*, 75:493–517
- Fenno, Richard. 1966. *The Power of the Purse*. Little Brown
- Fenno, Richard. 1973. *Home Style*. Little Brown
- Ferrejohn, John. 1991. Rationality and interpretation: parliamentary elections in early Stuart England. In *The Economic Approach to Politics*. Kristin Monroe (ed.), HarperCollins