

Pluralism about rationality in economics

C. Tyler DesRoches and Thomas Wells

[I]t would be a mistake to adopt a single interpretation of rational choice theory: how the theory is best interpreted depends on the questions it is addressing, and the circumstances in which action is being viewed.

Satz and Ferejohn (1994, p.86)

We begin by noting three observations. Firstly, the word 'rationality' connotes a number of overlapping but distinguishable ideas and connections, calling to mind, among others, intelligence, reasonableness, correctness, truth, usefulness, efficiency, and choice. 'Rationality' is a term that seems in better company with complex and contested terms like 'freedom' than the more straightforward 'inflation'. As with 'freedom' it is difficult to pinpoint exactly what the concept of rationality denotes because no single definition seems able to accommodate all that we may want to mean when using it. However, this need not be a problem, since it is quite possible to develop a particular theoretical concept of rationality which explicitly draws on and fixes a manageable number of its connotations which are applicable to a specific context. This seems to be how theoretical concepts of rationality are actually developed. But our second observation is that theoretical concepts of rationality developed in this way have a tendency to expand beyond their initial constrained domain, and to become understood by both practitioners and critics as *the* definition of rationality *per se*. Finally, our third observation is that this expansionist move is linked to a realist rhetoric which justifies a particular concept in terms of its truth and the project of scientific unification.

We argue for pluralism about rationality, in which concepts of rationality are explicitly understood in terms of their particular theoretical construction. Within the social sciences, and even within the different sub-disciplines of economics, different questions are asked which call for the development and use of different theoretical concepts, such as of rationality. Since these tools are designed to answer certain questions, their success should be understood in terms of their effectiveness in answering those questions rather than their general 'truth'. We should not judge a paintbrush a failure because it isn't very good at hammering nails. Indeed an excessive focus on the tools social scientists use in their research distracts from the correct level of evaluation, namely, are their *questions* the best ones? A better understanding of what concepts such as rationality are designed to do will help us to understand their capacities and limitations.

Admittedly, the argument that the origin of concepts has implications for their use may seem obvious or even trivial to many social scientists. But our three observations on the practice of much social science suggest that the point, though perhaps commonplace, is not properly taken up. To address this we go further than the commonplace assertion that there may be some implications to develop a theoretical structure that allows and encourages the implications to be explicitly identified and systematically addressed. We term the combination of research questions with the methodology and interests of the relevant researchers the 'original problem context'. Each original problem context implies a unique set of 'explanatory commitments' for conducting research i.e. the shape of the theoretical tools that will be developed and employed.

We formalise this by means of an idealising abstraction that allows us to map different theoretical concepts (in our case, ‘rationality’) along a few key dimensions of explanatory commitment with each theoretical concept occupying a unique point.

Our thesis is that a particular theoretical concept of rationality can only be successfully applied where its explanatory commitments are specified, and our framework supports this by systematically distinguishing each conception of rationality in terms of its unique set of explanatory commitments, thus making explicit the particular domain of each concept and the limits to its extension. For the purposes of illustrating our argument about rationality we focus on only three of the most striking explanatory commitments. These are internalism-externalism, individual-social, and explanation-prediction.

Although we believe that our argument applies equally to all theoretical concepts of rationality, a main conception in the debate is of course *the* Rational Choice Theory (RCT) of economics, and no discussion of rationality in the social sciences would be convincing without addressing this reigning paradigm. Indeed, RCT provides a good test and illustration of our argument and its implications are worth covering in depth. Much of the extensive criticism aimed at RCT has been empirical, along the lines of, ‘since people in reality do not make choices as described in RCT it should be rejected in favour of ‘theory x ’. Our approach is quite different and more constructive because it suggests that RCT may be good for answering certain questions, but not others. More than this, it disentangles RCT from inappropriate use and criticism by systematically reconstructing it in terms of a coherent and consistent set of explanatory commitments, thus allowing a clearer picture of where its appropriate domain lies and why. Empirical challenges to RCT only demonstrate that it is unable to answer certain kinds of questions, i.e. that its effective domain is narrower than commonly thought, but they don’t have any binding implications for the ‘truth-value’ attached to RCT since it is still the best concept for answering certain questions and it is a different matter whether those questions are the ones we social scientists should be interested in. Our paper redirects attention to *the questions* we researchers choose to ask, since it is the problem context that determines the kinds of conceptions (of rationality) one constructs and employs. Alternative theoretical concepts of rationality, such as the bounded rationality of behavioural economics, is better understood as asserting the importance of acknowledging certain questions (how do *humans* make decisions?) rather than a rivalry over which rationality is the ‘true’ one.

The paper proceeds as follows: Section (2) discusses the promotion of various theoretical concepts of rationality and shows that they are actually arguments for the priority of particular problem contexts; Section (3) develops our theoretical framework in more detail; Section (4) focuses on the case of RCT and shows how our framework complements and extends Satz & Frejohn’s (1994) account; Section (5) considers some objections; Section (6) concludes with a general discussion of the implications of our argument.

2. Concepts of rationality in economics

[E]veryone more or less agrees that rational behavior simply implies consistent maximisation of a well-ordered function, such as a utility or profit function. [Becker 1962, 1]

[T]he concept of program-based behaviorprovides a unifying paradigmatic framework, at the same level of generality as rational choice theory, within which such other approaches can be readily integrated and be related to each other in a coherent manner. [Vanberg 2002, 9]

Rationality is interpreted here, broadly, as the discipline of subjecting one's choices - of actions as well as of objectives, values and priorities - to reasoned scrutiny [Sen 2002, 4]

Claims such as these by Gary Becker and Viktor Vanberg are frequent in the literature on rationality. We are presented with the appearance of a competition between various claims about what rationality really is, where the prize at stake is of determining the new universal paradigm underlying social science. In such a winner-takes-all competition the claims often tend to the metaphysical – that rationality just has to be like ‘this’ because of the immutable laws of logic (Becker) or the overwhelming evidence of evolutionary psychology (Vanberg), although in their calmer moments most proponents would modestly claim that they are merely proposing different perspectives. Indeed, it is the calmer moments that should prevail since a deeper look at the arguments supporting their claims reveal a traditional debate for primacy between sub-disciplines, arguing that their particular *questions* are the most important ones to ask. This is worth illustrating using the examples of Becker and Vanberg.

Becker's 1962 article is directly concerned with justifying a version of Rational Choice Theory (one consequence of our argument is to suggest that there is no monolithic 'Rational Choice Theory' but that it fragments under scrutiny). He argues that the economist's concept of rationality concerns the rationality of the *market*, and is quite compatible with the irrational behaviour of individual market *actors*. But this has been misunderstood by both critics and proponents, who have understood the relationship of the market and individual actors in the wrong way, so that "Confidence in market rationality led some into stout defences of rationality at all levels, while confidence in household irrationality misled others into equally stout attacks on all rationality [Becker 1962, 8]." Becker's argument can be understood as an attempt to distinguish and clarify the particular theoretical understanding of rationality of interest to the economist from the many other possible connotations of rationality that might be interesting in other circumstances. Economists had become confused by the re-directing of comment and analysis away from the market and towards the individual, but it was questions about market reactions to changing variables that was the economist's main interests [Becker 1962, 2]. Connecting the dots, we would add that the economist's understanding of rationality was itself developed as a tool for that ‘original problem context’ and this understanding became confused precisely when its application was extended to quite different questions such as about the decision-making of individuals. It is all about the questions, and this is further illustrated by Becker's extension of his clarified concept of rationality to firms as well as households, because firms should be understood as belonging to the same problem context for which the same tools are useful [Becker 1962, 12].

Vanberg, on the other hand, isn't defending a paradigm - he's out to topple one. His paper therefore takes a different course than Becker's, beginning by discussing the right questions and then moving on to the correct concept of rationality that the right questions require. On the first page, Vanberg admits the troubling fact that despite a wealth of criticism from heterodox quarters, Rational Choice Theorists remain unshaken in believing that "at least for their *explanatory purposes*, the rational choice model is the best analytical tool available [Vanberg

2002, 7-8, my italics]." The battle is therefore fundamentally over the right explanatory purposes. Vanberg pulls the reluctant Rational Choice Theorist towards his perspective by i) presuming agreement on the explanatory commitments, ii) disparaging RCT's theoretical resources, and iii) talking up the exciting new vistas that adopting his concepts would allow. Vanberg begins by characterising RCT as an attempt to provide "an explanatory account of purposeful human action [Vanberg 2002, 10]," though of course many Rational Choice Theorists (presumably including Becker) would decline to go long with this. Vanberg then immediately identifies a problem, which he tugs away at throughout the paper: RCT takes rationality itself for granted and makes no effort to explain it! [Vanberg 2002, 10].

Vanberg then suggests that a shift of focus is needed from examining how actors behave rationally to examining how human rationality itself came about. In this interpretation, the Rational Choice Theorist is characterised as attempting to understand the proximate causes of rational behaviour and failing because he lacks the essential context, the 'big picture', of the ultimate causes of rationality [Vanberg 2002, 18]. RCT is 'isolated' from empirical and theoretical progress in modern behavioural research [Vanberg 2002, 9] so it simply lacks the resources to answer these important questions. Repeating the point with a twist, Vanberg suggests that RCT is concerned with actors' pursuit of ends, when it should be concerned with human behaviour as exercises in problem solving, which again points towards a need to explain the origins of the human problem solving apparatus and knowledge. Vanberg's own concept of rationality is of 'rules' or 'programmes' evolving through a process of trial and error (at the genetic, cultural, and individual level). This is a backward-looking forward-facing approach in which rationality means adaptedness and the focus becomes not the rationality of individual behaviour, nor Becker's rationality of the market, but the rationality of these programmes. Taking up Vanberg's explanatory commitments and concepts is supposed to promise more progress. For example, irrational behaviour that RCT can only see as anomalies can be explained as a result of evolved problem-solving programmes maladapted in specific and systematic ways to the current problem.

Amartya Sen's approach contrasts with both Becker and Vanberg. Sen is far too aware of the nuances of 'rationality' to ever fall into the trap of claiming his interpretation is the final, true one. He recognizes the richness of 'rationality' in ordinary language and uses this as his starting point. Sen's argument is instructive because it works differently from the Vanberg and Becker cases and is substantially in line with our own approach. Sen effectively argues that the tool of RCT has taken over how social scientists think about rationality in general, and now dictates the kind of questions that we ask, leaving us blind to important aspects of choice. Sen criticises RCT not as false, but as insufficient for the role of governing paradigm. Thus, RCT in its various forms is neither a necessary nor a sufficient definition of rationality because it fails to capture all the important connotations of 'rationality', and that the connotations it does pin down are not the most important ones. The result is an unnecessarily narrow and distorted definition of rationality that excludes individual agency from the analysis, including the phenomenology of decision making, freedom, the moral contexts of decisions, and consideration of goals and values as well as means. For Sen, at core of rationality is a sophisticated concept of agency closely related to Kantian moral autonomy. No axiomatic approach to rationality is capable of capturing real agency, and when such approaches are tried they actually reduce our understanding of agency and choice. Indeed Sen's own interpretation of rationality as 'reasoned scrutiny' is agent centred and deliberately sets minimal boundaries on what can count, beyond a kind of moralised common-sense. This definition is not rigorous since it doesn't require adherence to logical rules

of coherence and consistency. However it is normative in asserting the role of reasons and the responsibility of the reasoner. The focus shifts to how the morally autonomous individual should consider his choices.

The discussion so far has sought to show that arguments about what 'rationality' means should be understood differently from usual. Underneath the surface claims that 'this is what rationality really is' lies a debate about which questions social scientists should be asking. Concepts of rationality only come in later, as a way to answer the question. Vanberg's case demonstrated the significance of 'explanatory purposes' for social scientists choice of concepts. Both Becker and Sen emphasised that concepts of rationality can be over-extended beyond what they are good for. While Becker sought to bring economists back to the right questions (about market behaviour), Sen seemed to be more concerned with making space for important questions (about agency). If correct, this analysis suggests that it is the purposes of the social scientists, formalised as their 'explanatory commitments', which come first and do the work in theory choice and construction. It is the explanatory commitments that shape and constrain the theoretical concepts used, rather than the other way around. The next step is to embed this idea in a deeper theoretical structure that will define explanatory commitments more precisely and show systematically how they limit the extension of particular concepts of rationality.

3. Mapping explanatory commitments

All theories make idealisations in order to focus on particular causal mechanisms which are supposed to be significant [Maki 1992, 2001]. All theories are therefore unrealistic in the strict sense that some of their statements are clearly untrue. Clearly then, truth *per se* cannot be an legitimate requirement for a good theory, nor can establishing strict falseness be sufficient to rule out a theory. But truth and falsity should matter somehow in theory choice if we are to avoid arbitrary choices and account for scientists' own claims. Successful idealisation requires that a theory be significantly true and not harmfully false [Maki 2001, 384]. That is, the causal mechanisms focussed on (endogenous to the theory) are the ones which are actually significant, while the details left out of the theory (exogenous) are either negligible or irrelevant. Debates about significance and harmfulness bring us to consider the various constraints on theorising characteristic of different disciplines.

Social scientists frame research problems in ways influenced by their experience, particularly influence by their discipline's traditional concepts, concerns, subject area and methodology. These create constraints on theory choice and development in three principle ways characterised by Uskali Maki [2001] as empirical (what kind of evidence counts and how), social (including the social norms for persuading and interesting one's colleagues), and ontological (beliefs about the 'way the world works'). These constraints do not determine specific theories but they do rule out many approaches and, especially in combination, may shape theory choice and construction in particular directions [Maki 2001, 385]. The combination of these is a researcher's 'original problem context'. Straightforwardly, differences in the original problem context are reflected in different disciplines' characteristic concerns with different domains of phenomena, so that the concern of Edgar the economist may be to accurately predict consumer behaviour in the Dutch blue cheese market while Mildred the sociologist seeks to explain the implications of the constraint of institution 'x' on society 'z'. But of course, different disciplines will also have different questions about the same apparent domain. Corruption may be

investigated by an economist in terms of incentive structures, or, by a New-Institutionalist, in terms of agency theory as a problem of managing fiduciary trust, but an economic sociologist would be wary of abstracting from the social context and be immediately interested in the social aspects of how the incentive structure was set up and how it comes to have particular meanings and values for the actors [Granovetter 2007, 152]. Quite different mechanisms are claimed to be significant, and quite different abstractions to be harmful so the particular concerns, methodology and questions of researchers can be seen to have a strong relation to the choice of explanandum and explanation.

Such an original problem context can of course be given a rich anthropological description but our purposes require a schema in which different problem contexts can be *systematically* distinguished and compared. This seems to require some abstraction from the deep description of actual situations in social science, and the choice of some suitable variables. Among other properties that can be attributed to the original problem context are their 'explanatory commitments', which we define as the researchers' commitments to how they understand the explanandum. Explanatory commitments structure and constrain the kinds of questions that can be asked and the conceptual tools that can be developed to answer them. It seems possible to distinguish each original problem context in terms of their differing sets of explanatory commitments and, if we have chosen the 'right ones', these explanatory commitments should in turn explain the nature and proper extension of the theoretical concepts employed. The theoretical concept of interest here is, of course, 'rationality', and the three kinds of explanatory purposes which seem most relevant are internalism-externalism, individual-social, and explanation-prediction. Of course, it may be that other dimensions turn out equally or more important, but the simplicity of these makes this initial development of our framework more intuitively appealing.

3.1 Internalism-Externalism

This section draws on the distinction made by Satz and Ferejohn (1994) between the *internalist* and *externalist* accounts of rational choice theory. Each embodies a rather distinct kind of research questions. Internalist accounts of rationality are committed to explaining rationality in terms of what goes on within the actor during rational decision making. They therefore refer to psychological mechanisms presumed to be active in the actor (which may be a firm as well as the standard human individual), such as beliefs, intentions, desires, and so on, and take these psychological mechanisms seriously. This means that internalist accounts work out the implications of rational behaviour in terms of these psychological mechanisms and must provide a plausible account of what these psychological states are and how they work. Internalist accounts can be normative in a way that externalist accounts cannot. Rationality here can provide a standard against which actual choices, and the psychological states that give rise to them, can be evaluated. If one wants x, then one *ought* to do y. Furthermore, one's preferences and beliefs can also be criticised as more or less rational, thus allowing various non-instrumental accounts of rationality. Obviously the reach of such accounts will be limited when the psychological mechanisms cannot be reasonably assumed to operate.

In contrast, externalist accounts seek to provide an account of rational behaviour rather than an account of rational decision making, so like most forms of RCT, they are not in the business of explicating a rational psychology or isolating psychological mechanisms. Indeed, externalist accounts need not take psychological mechanisms seriously, even when they posit

them in an *as if* way [Satz & Ferejohn 1994, 76]. Behaviour is understood *as if* it were maximising a goal under a set of constraints and rationality requires merely formal consistency with this function. Radically externalist accounts entirely deny the causal role of intentionality and are equally applicable to any patterned behaviour, such as the behaviour of gas particles when heated or consumer behaviour [Satz & Ferejohn, 1994, 75]. Other possible positions include the moderate externalism promoted by Satz and Ferejohn, which does accept the causal role of human intentionality for social science questions, and so adds the condition that motivation must be lined up somehow with the supposed goals for the system to work. Externalist accounts have the advantage of being applicable to a wider range of behaviour types, but they are purely instrumentalist. They are limited to predicting what would happen if the attributed goals, constraints, and *ceteris paribus* clause holds, and this means that in cases where the posited instrumental relation doesn't hold, such as the so-called voting paradox, they will be difficult to apply. Externalist accounts, are also unable to make normative claims about how actors *ought* to make decisions or what they *should* decide because they can only suggest what would occur if a certain instrumental relation holds. Of course, this last point is more often than not taken to be a virtue by the those who employ externalist accounts of rationality.

3.2 Individual-social

Consciously or not, researchers also have foundational commitments to explain phenomena at a certain level of reduction, i.e. the level of complexity at which their explanans is situated, most obvious in the division between explaining at the level of the individual or at the level of society. The question of reduction is where the causal chain ends for the purpose of the analysis. Consider for example, the debate between methodological individualists and methodological holists about what level is fundamental, i.e. should society activity be considered as only the aggregate outcome of the actions of individuals, or does society have emergent aspects, such as institutions, which are independent of their current members and may also cause and shape individuals and their actions. The most radical accounts of these can be found in the easily caricatured *homo economicus* and *homo sociologicus*. Here the former is supposed to be a completely detached, self-seeking agent while the latter is supposed to be entirely a product of society and operates according to internalised socially shared rules and norms [Elster 1989, 97]). But the individual-social dualism does not cover all levels of reduction of interest to the social scientist. It is also possible for social scientists to seek explanations at the sub-individual level, for example as deriving from neurological properties, or posited multiple selves or mental programmes interacting within individuals. For social scientists there is a spectrum of possible explanatory commitments between the sub-individual and social level. Under the sub-individual level is the neurological (explanations are in terms of brain states) and then interaction of sub-individual elements (multiple selves, multiple programmes); under the individual level is individual behaviour and behaviour as an agent (able to critically reflect on his own actions and values); finally, under the social level is aggregate behaviour and then social institutions (including social values, rules, customs, etc.). While not definitive this list covers the major positions that social science theories take.

3.3 Prediction and explanation in rational choice theory

Some social scientists attach great importance to a theory's ability to make accurate predictions while some concentrate on explaining the phenomenon in question. We take

‘explanation’ as directed to making a phenomenon comprehensible by describing the relevant structure or operation or circumstances, while ‘prediction’ is directed at forecasting presently unknown activity of the phenomenon into the future. Explanation and prediction are associated with different theoretical and conceptual virtues. Social scientists concerned primarily with explanation will seek to elucidate and understand the causal structures and mechanisms of the phenomenon in detail i.e. how it really *works*. At the extreme this explanatory focus is associated with a concern for the ‘veristic realisticness’ or representing the referents of a theory in a manner that is true, likely to be true, or close to true [Mäki 1996]. Social scientists more concerned with successful prediction may take more ‘short-cuts’ to bypass the task of elucidating and checking the causal structure and move on to testing what a theory pays, its predictive power. At the extreme such a concern is often associated with the instrumentalism, where theories are designed and assessed according to their ability to generate successful predictions or conclusions and their truth is considered irrelevant to this task [Boland 1979, 508-509]. For example, Milton Friedman’s famous suggestion that “[T]he relevant question to ask about the ‘assumptions’ of a theory is not whether they are descriptively ‘realistic’, for they never are, but whether they are sufficiently good approximations for the purpose at hand [Friedman 1953, p.15].”

3.4 A brief application

Applying these explanatory commitments to our three examples of rationality in the social sciences we see that each can be roughly characterised in a unique point in the space mapped by these three dimensions. Becker’s account of RCT is committed to strong externalism, strong prediction, and explanation at the level of aggregate behaviour,; Vanberg’s approach is committed to moderate internalism, moderate explanation, and explanation at the level of sub-individual elements (programmes); Sen’s rationality seems committed to strong internalism, explanation and explanation at the level of the individual agent. As we might expect, these three kinds of explanatory commitments may be related. This appears in the clustering of concerns with externalism, prediction and the aggregate level on the one hand and on the other, of internalism, explanation, and the individual. Some judgement is involved in assessing each concept with the framework and this may be challenged. In particular, some original problem contexts may be ambivalent between levels on the criteria, as behavioural economics sometimes seems concerned with explaining the behaviour of individuals in terms of psychological mechanisms, and at other times uses individual behaviour to deduce the existence of psychological mechanisms. Since our framework is built on the assumption that explanatory commitments are consistent, it automatically exposes such inconsistencies and implicitly judges them as potentially problematic both for the concept’s success as a tool and for the coherence of the underlying original problem context.

4. Illustrating the case further: Rational Choice Theory

RCT has various conflicting versions, a fact that has led to much confusion for both its critics and defenders. What most Rational Choice Theories have in common seems to be a instrumentalist conception of rationality, but this can be interpreted in different ways. As well as the briefly discussed account of Becker, various other accounts can be distinguished, including Savage’s internalist (and normative) account [see for example, Robert Sugden 1991] and Samuelson’s externalist Revealed Preference account, among others. With RCT, as elsewhere in the use of ‘rationality’, there is a tendency for different theoretical approaches to become

entangled in practice. Indeed, as discussed above, it is a common feature of articles concerning RCT to begin by claiming that it has been incorrectly understood, and that the true RCT should be understood quite differently. More philosophical analyses of RCT often analyses the link between concepts of rationality and questions asked which we have argued is central [for example, Lovett 2006, Satz & Ferejohn 1994], although such analyses are often ‘partisan’ (in the sense of analysing only one kind of question-concept relationship).

Our approach is consistent with this up to the point where the ‘true’ version of RCT is revealed. Our framework is a general one that allows us to roughly distinguish between different concepts of RCT in terms of consistent and distinct sets of explanatory commitments. Although this is linked to the kind of questions each concept can answer, our framework does not make final judgements about either the effectiveness of different concepts for their questions or the importance of those questions themselves. Rather, our framework provides the broad analytic resources that are necessary to answer such questions by providing a systematic and functional taxonomy of the different concepts in a neutral way. Furthermore, our framework keeps this taxonomy explicit and prevents different versions of RCT from becoming blurred or entangled in practice while also maintaining certain clarity about what each can and cannot do.

Partisan accounts do have the virtue of exploring in depth the relationship, both forwards and backwards, between explanatory commitments and their associated theoretical concepts, and research concerns. This is a necessary complement to the analytic overview we seek to provide because it fleshes out the links which we merely sketch. Satz & Ferejohn [1994] for example provides an excellent analysis of the relationship between the commitment to moderate externalism and certain kinds of questions that we now turn to in order to illustrate the point. Satz & Ferejohn reject the common understanding of RCT as a straightforward account of individual human psychology and instead focus on what it can achieve as a moderately externalist account explaining how the structural constraints in particular situations determine ‘choices’.

The conception of human rational agency in terms of maximising over a complete and consistent set of preference orderings is not psychologically realistic....Many of the most important uses of [rational choice theory] need not rely directly on any theory of human psychology. The theory’s use in economics and sociology, for example, is largely aimed at illuminating structures of social interaction in markets, governments, and other institutions. In many social-scientific explanations, we are not interested in explaining a particular agent’s behaviour, but in the general regularities that govern the behaviour of all agents. [Satz & Ferejohn 1994, 74]

Satz & Ferejohn accept the formal definition of rationality as ‘behaviour consistent with goal-seeking’ and detach it from any particular psychological account, although they don’t deny that it must be consistent with at least some plausible psychological mechanism. Their point is that commitment to a particular psychological mechanism isn’t required by the questions they consider interesting. They then add back in the goals generated by plausible external theories of interests. The proviso they add is that because humans are not automatons but chock-full of various interests and beliefs, the external theory of interests must be very powerful for the explanation, and hence theoretical predictions, to work. This occurs with structuralist explanations where the interests of the players are so constrained by their place in the structure that the players themselves (and their slippery personal psychologies) don’t matter anymore.

Satz & Ferejohn argue that whether individualist or structural-level accounts of social phenomena are appropriate depends on the purpose of the explanation. For some purposes, the appropriate focus is on individual agency and choice. For many social-science questions, however, the appropriate focus is on how social structures and features of the agent's environment exert constraints on her action. This is not an a priori feature of social-science explanation. Rather, it is a feature of the types of questions in which social scientists are typically interested [Satz and Ferejohn 1994]. This supports our argument for that it is the original problem context which sets the explanatory commitments that in turn shape the precise theoretical concepts deployed. We agree with Satz & Ferejohn's conclusion that social scientists' should look back along this chain when considering how to approach a problem. They should step back and consider whether this problem context fits with the explanatory commitments that their conceptual tool kit was designed to meet. Satz & Ferejohn also demonstrate the value of systematically disentangling different concepts of rationality and use the single criterion of internalism-externalism effectively to distinguish three approaches. Like ours, their approach implicitly demands theoretical consistency and allows for the implications of such explanatory commitments to be pursued, such as that externalism accounts cannot have normative implications.

5. Some possible objections to our framework

We can imagine a number of possible objections to our framework. To some, our framework may suggest a laissez-faire instrumentalism because we don't say which concept of rationality is the 'right' one, the 'best' one, or even how to build a better one. The dimensions we chose for our framework may seem arbitrary or otherwise unsatisfactory. The examples we used to illustrate our account were unusually reflective, and so perhaps did not reflect the ordinary use of concepts of rationality in practise.

It is not the aim of our argument to settle disputes about which kind of rationality we should use. We are just claiming that there are different conceptions of rationality for different purposes. Indeed one of our aims is to downplay the significance of the theoretical tools so that debates about which questions are important can come to the fore. We see our framework as providing a neutral way of mapping different concepts so that they can be clearly distinguished from each other and explicitly related to their theoretical functioning. Our framework directly supports evaluation in terms of the functional assessment of different concepts of rationality by identifying the correct space for evaluation i.e. which questions they were designed to answer. This also allows cases of unwarranted extension of rationality concepts to be identified, and then our argument suggests that either the questions should be adapted to fit what the concept can do, or different concepts should be selected or created to fit the new explanatory requirements.

The selection of the dimensions of level of reduction, explanation-prediction, and internalism-externalism was based on two criteria. Firstly, the principle of explaining more with less suggests that we choose the smallest number of key variables that would allow the theories to be adequately distinguished. Secondly, we were concerned with choosing the dimensions which seemed most relevant to the case of rationality. On this, we accept that a better list, or a better description may be possible and desirable. Nevertheless, this kind of bare-boned Model T Ford framework makes the point that we want to make.

It is true that we have only considered reflective accounts in our examples, either by thoughtful practitioners or specialised philosophers of the social sciences. Using such accounts allowed us to build on and complement the analysis of ‘rationality’ that is already going on in the social sciences and which captures broad swathes of the relevant literature, without having to start from scratch. Nevertheless it may be a weakness of our approach that it is yet to be tested on an operationalisation of theory since the gap between theoretical and practical understanding of rationality does seem quite wide.

Finally we come to the chicken and egg problem of tools and purposes (questions). We have suggested that purposes are prior because they shape the selection and creation of tools. Nevertheless it is obvious that in practical situations the available tools also limit the questions that can be asked and how. This explains why most economists use marginal analysis – because it is in the tool-box of economists. We believe that purposes are prior in two ways. Firstly, they are logically prior in the sense that it is questions that motivate investigation rather than tools. Secondly, while there may seem a two-way relationship of questions and tools in the short term (i.e. where practical considerations are important), in the long term it is questions which drive innovation in tool design.

6. Some implications

It is a tremendously liberating move to insist that no single account of rationality can account for everything we may want to analyse as rational. Social scientists should not feel constrained to use only one account (such as RCT) using brute force or ingenious and contrived reasoning to somehow apply it to an original problem context that really doesn’t fit. They have more freedom to consider other more suitable conceptual tools.

But our framework also disciplines the development and use of such tools in two ways. Firstly we insist on consistency between the explanatory commitments implied by the original problem context and the concept of rationality taken up. This prevents the entanglement of quite different accounts of rationality under the same title, such as accounts of RCT which claim to be about both individual behaviour and aggregate behaviour. Such blurring makes application and criticism problematic (as in the recent history of RCT). Secondly our framework makes this relationship explicit and limits the extension of concepts of rationality to other problem contexts where these explanatory commitments don’t make sense.

Our taxonomic framework focusses attention away from discussions of the truth of the theoretical tools of analysis and towards the understanding and evaluation of researchers’ concerns and purposes. Tools are limited by their construction in both their scope of application and the meaningfulness of their answers. It is the questions and not the tools which should take the lead in theory choice and development. This suggests that good criticism of a concept of rationality will be about its relevance to the current problem context: Is this an aggregate level phenomena? Are we trying to explain or predict? Concepts of rationality are seen to be limited by the kind of questions they can ask rather than their traditional subject areas. As Gary Becker’s ‘economics of the family’ and other expansionist neo-classical economists have demonstrated, it is possible to deploy RCT in non-traditional domains, and this amounts to finding ingenious ways to ask neo-classical questions outside of their original problem context. Expanding the scope of a tool therefore requires convincing others that this new problem context significantly resembles

the original. Of course this can be contested, but the debate should be between different visions of how the world works rather than the truth of the tool deployed. The right questions would be: Does this extension include the most significant causal mechanisms? Does it harmfully exclude significant causal mechanisms?

References

Becker, Gary S. (1962) 'Irrational Behavior and Economic Theory', in *The Journal of Political Economy* Vol. LXX (1), pp 1-13.

Elster, Jon. 1990. *Nuts and Bolts of the Social Sciences*. Cambridge: Cambridge University Press.

Elster, Jon. 1989. *The cement of society*. Cambridge: Cambridge University Press.

Friedman, Milton (1953) 'The methodology of positive economics', in *Essays in Positive Economics*. University of Chicago Press, pp 3-43.

Granovetter, Mark (2007) "The Social Construction of Corruption". Pp. 152-172 in Victor Nee and Richard Swedberg, editors, *On Capitalism*, Stanford University Press.

Hausman, D. and McPherson, M. 2006. *Economic Analysis, Moral Philosophy, and Public Policy*. Cambridge: Cambridge University Press.

Lovett, Frank (2006) 'Rational Choice Theory and Explanations', in *Rationality and Society* Vol. 18 (2), pp 237-272.

Mäki, Uskali (1992) 'On the Method of Isolation in Economics', in *Idealization IV: Intelligibility in Science* (ed. Craig Dilworth), special issue of *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 26, pp. 319-54.

Mäki, Uskali. 1996. Scientific realism and some peculiarities of economics, in *Realism and Anti-Realism in the Philosophy of Science*, R.S. Cohen et al. (eds.), Boston Studies in the Philosophy of Science, Vol. 169. Dordrecht: Kluwer: 425-445

Mäki, Uskali (2001) 'The Way the World Works (www): towards an ontology of theory choice', in *The Economic World View* (ed. Uskali Mäki), pp 369-389.

Rutherford, Malcolm. 1996. *Institutions in Economics: The old and the new institutionalism*. Cambridge: Cambridge University Press.

Satz, D. and Ferejohn, J. 1994. "Rational Choice and Social Theory". *The Journal of Philosophy*, pp.71-87.

Simon, H.A. 1965. 'Mathematical Constructions in Social Science', in *Philosophical Problems of the Social Sciences*, ed. D. Braybrooke, pp.83-98. New York: Macmillan.

Sen, Amartya (2002) *Rationality and Freedom*, Cambridge and London, Harvard University Press

Vanberg, Victor J. 2002. "Rational Choice vs. program-based behaviour", *Rationality and Society*, 14(1), pp.7-54.