

# Critical Realism from a Kantian Point of View

Alex Viskovatoff

April 15, 2000

## Contents

1	<b>Introduction</b>	1
2	<b>Critical realism and economic methodology</b>	2
3	<b>Difficulties of critical realism</b>	4
3.1	<i>Critical realism is compatible with J. S. Mill's methodology</i>	4
3.2	<i>Transcendental realism is less true to science than Kant</i>	7
4	<b>Kant's critical philosophy</b>	9
4.1	<i>Two types of transcendental arguments: constitutive and regulative</i>	10
4.2	<i>The three components of science</i>	11
4.3	<i>Bhaskar's "transcendental argument"</i>	14
4.4	<i>Going beyond Kant's transcendental idealism</i>	17
5	<b>Grounding social theory</b>	17
5.1	<i>Transcendental deduction of rule-following</i>	17
5.2	<i>'Ruling out' decision theory</i>	21
5.3	<i>'Validating' social system theory</i>	23
6	<b>Conclusion</b>	24

## 1 Introduction

According to Roy Bhaskar, there are three main traditions in the philosophy of science: empiricism, transcendental realism, and Kant's transcendental idealism (Bhaskar 1975: 24–25). While Bhaskar provides a sustained critique of empiricism, his arguments against transcendental idealism are somewhat cursory. This is odd, in view of the facts that empiricism had been widely discredited by the time he wrote his (1975) (by Wittgenstein, Quine, John Austin, and Wilfrid Sellars, among others), that analytic philosophy was then moving into what Richard Rorty has called "its Kantian stage" (Sellars et al. 1997: 3); and that the philosophy of science is currently experiencing a Kantian renaissance. While Bhaskar's critique of empiricism is compelling, his transcendental realism involves substantial ontological commitments. Kant's critique of empiricism has been around much longer than Bhaskar's, has withstood the test of time (for example, Sellars built upon it), and involves no such commitments. Therefore, if critiquing empiricism is the aim, turning directly to Kant instead of Bhaskar's

naturalist appropriation of transcendental argumentation should at least be considered. On the other hand, the lessened ontological commitments that Kant allows would seem to come at a cost—Kant’s transcendental idealism, that is, his anti-realism.

Kant called his philosophy the “critical philosophy”. He inaugurated a new, critical phase in philosophy, by being able to successfully accuse his predecessors of what he called “dogmatism”—for putting forward positions they could not properly justify. By adopting the position of transcendental realism (which was named by Kant, and which Kant subjected to critique), Bhaskar in effect adopts a precritical position, giving the task of justification a diminished importance, under the guise of avoiding what he calls the “epistemic fallacy”. In responding to the inadequacies of Kant’s system, I shall argue in this paper that instead of falling back into a precritical position, it is preferable to accept the force of Kant’s arguments against transcendental realism (and against empiricism and rationalism) while trying to avoid his anti-realism.

As I shall also argue, there are two other reasons why critical realists should consider moving more closely to a Kantian position than they have until now. First, Kant’s transcendental deduction essentially yields the result of the necessity of rule-following for conscious beings. One thus obtains this result by a much more direct route than this is currently done in critical realism—by first arguing for the three-domain picture of reality, and then, building upon that, arguing that rules must be present in the deepest, real domain. Second, critical realists argue that choice must be seen as ‘real’ (Lawson 1997: 30), but do not indicate what is the source of the ‘freedom’ of human actors. Kant’s distinction between the realm of laws and the realm of reasons, with the necessity of the former being causal but that of the latter being *normative*, provides a worked-out philosophical account of this freedom.

## 2 Critical realism and economic methodology

Starting with his 1975, Roy Bhaskar has sought to work out a philosophy of science capable of ‘grounding’ a ‘committed’ social theory. In so doing, to combat the empiricism that is all-too-prevalent in post-World-War-II social science, he has—very appropriately—turned to Kant. However, apparently to remain true to Marx’s “critique” of German Idealism, Bhaskar has done this by attempting to ‘naturalize’ and ‘socialize’ Kant’s method of transcendental argumentation. In his transcendental deduction of the categories, Kant offered an argument against Hume’s skepticism by arguing that if one were not already employing categories such as substance and causality, the legitimacy of which Hume doubted, experience (the existence of which, unsurprisingly, Hume did not doubt) would not be possible. Here, the experience that is being considered is that of a solitary, “transcendental” subject (‘transcendental’ in the sense that it is only the ‘form’ of subjectivity that one is considering, and not the qualities of any particular, concrete experience): Kant “employ[s] the transcendental procedure . . . in an individualist and idealist mode” (Bhaskar and Lawson 1998: 4). According to Bhaskar and Lawson however, “Bhaskar demonstrates that *there is little need to be so restrictive*” (p. 4; emphasis added). And accordingly, instead of taking the

possibility of experience for a transcendental subject as his starting premiss, Bhaskar takes as his starting premiss the possibility of natural science.

What Bhaskar deduces from this is the three-layered ontology which is central to his system. Reality consists of three overlapping domains: the empirical, the actual, and the real. Experiences make up the empirical and are subjective; the actual consists of objective, subject-independent events which produce these experiences. (This ontology is thus closer to Hume's than to Kant's, since Bhaskar's fundamental category for what is objective is events, whereas for Kant it is the object. According to Kant, one would not be able to discriminate events unless one already had a sufficiently rich "sensory manifold" to be able to situate objects in it.) In general, phenomena at the level of the actual will not exhibit regularity: usually when I step on the accelerator of my car, it will speed up, but this will not always be the case (such as when I have run out of petrol, for example). Since, because of our ability to conduct controlled experiments in the laboratory, we know that the world *can* exhibit regularity, according to Bhaskar, we must postulate a level of "the real" 'below' that of the actual, producing actual events and governed by lawful regularities: for why should we think that the phenomena observed outside of the laboratory, which exhibit little regularity, are produced by different "mechanisms" than those observed inside the laboratory, which do exhibit regularity? Without this regularity at the level of the real, science would not be possible: this is Bhaskar's "transcendental argument".

Since empirical realism looks for lawfulness at the level of "constant conjunctions" of experiences, under it (for which Bhaskar sees Hume as canonical) the "three domains of reality are collapsed into one" (Bhaskar 1975: 57). This makes empirical realism unable to account for the possibility of science, since, as Hume demonstrated, it cannot derive the necessity of laws from the only thing it has to work with—the constant conjunction of events. On the other hand, transcendental idealism is unable to recognize that the natural world does not depend upon the experiences of human beings. And thus, transcendental realism provides "the best account [of science] currently available—in so far as it is at present uniquely consistent with the historical emergence, practical presuppositions and substantive contents of the sciences . . ." (Bhaskar 1989: 185).

Once one has this philosophy of science in place, one can apply it to social science. Tony Lawson does this in two main ways, one negative and one positive. First, the Bhaskarian ontology makes clear why mainstream economics is not succeeding as a science, and why it cannot succeed. This is because what lies at the bottom of mainstream economics is *deductivism*: the practice to search for laws as constant conjunctions of events at the level of the actual. In the natural sciences, this kind of approach can work, because of the possibility there to by means of controlled experiments to construct "closed systems", that is, systems in which the various mechanisms operating at the level of the real are separated from each other, so that the lawfulness of the real domain becomes exhibited in the domain of the actual. Because of the nature of the social however, social phenomena are always inherently produced by open systems, so that experiments are not possible. Therefore, no economic science following deductivist principles is possible. Second, Bhaskar's ontology tells us what economists

must do in order to practice a successful science: they should look for regularity in the domain of the real, not in the domain of the actual (as mainstream economists do), and try to decide what kinds of entities make up this domain in the case of society. The basic kind of entity that Lawson settles upon is rules: rules constitute the mechanisms which produce social phenomena, but since social systems are open, they will not do so in a way that exhibits uniformity at the level of the actual.

### 3 Difficulties of critical realism

Despite its demonstration of difficulties with empirical realism, transcendental realism is not without difficulties of its own: in fact, as we shall see, it is not substantially different from versions of realism less uncompromising than Hume's. Furthermore, as indicated in section 1, Bhaskar's critique of transcendental idealism is much less developed than his critique of empiricism, thus making one dubious of his claim to provide "the best account" of science "currently available".

#### 3.1 Critical realism is compatible with J. S. Mill's methodology

As we have seen, central to critical realism is the distinction between the domains of the real and the actual, the domains in which 'underlying' causal mechanisms operate and in which these mechanisms jointly manifest themselves in concrete situations outside of the laboratory. Because these mechanisms (which do exhibit lawful behavior) combine in complex ways, and in any given situation some will be present and some not, uniformity will be present in the domain of the real, but not of the actual. Since it "collapses" these two domains, empirical realism commits the fundamental error of looking for lawful behavior on the level of the actual. This accusation may well hold true for Hume's 'fundamentalist' version of empiricism<sup>1</sup> However, another empiricist profoundly influential on mainstream economics—John Stewart Mill—was sensitive to the possibility of a lack of regularity at the 'phenomenal' (actual) level in the social sciences, explicitly drawing attention to it. Consider the following passage, with analogues to Bhaskar's framework pointed out in the text with insertions in square brackets:

Suppose that all which passes in the mind of man is determined by a few simple laws [real domain]: still, if those laws be such that there is not one of the facts surrounding a human being, or of the events which happen to him, that does not influence in some mode or degree his subsequent mental history, and if the circumstances of different human beings are extremely different, it will be no wonder if very few propositions can be made respecting the details of their conduct or feelings [actual domain], which will be true of all mankind. (Mill 1965: 863; cited by Skorupski 1989: 265)

As is well known, the way Mill tried to cope with this problem was by means of the concept of a *tendency law*. A tendency law states that one phenomenon is causally

---

1. However, even on this score, one may have doubts: see Parsons (1999: 156–164).

dependent upon others, but only if certain conditions are met, which are usually not specified precisely, but merely referred to in the form of a *ceteris paribus* law. Mill indicates how this work in his discussion of ethology, which, states Mill

may be called the Exact Science of Human Nature; for its truths are not, like the empirical laws which depend upon them, approximate generalizations, but real laws. It is, however, (as in all cases of complex phenomena) necessary to the exactness of the propositions, that they should be hypothetical only, and affirm tendencies, not facts. *They must not assert that something will always, or certainly, happen*; but only that such and such will be the effect of a given cause, so far as it operates uncounteracted. (Mill 1965: 870; emphasis added)

Lawson defines deductivism as “a mode of explanation which involves deducing the explanandum from a set of initial conditions plus regularities that take the *form* ‘whenever this event or state of affairs then that event or state of affairs’” (Lawson 1999: 224). It would be hard to find a clearer rejection of deductivism than the passage from Mill just quoted. I think the two quotations just given indicate the strong affinities between Mill’s philosophy of science and critical realism.<sup>2</sup> This is all the more the case when one considers that the concept of ‘tendency’ plays a prominent role in Lawson (1997), and Lawson distinguishes his understanding of the concept from “many of its interpretations in the economics literature” in the following way: “It is not about events that would occur if things were different but about *a power that is being exercised whatever events ensue*” (Lawson 1997: 23; emphasis in original). In the passage just cited, Mill also thinks of the “given cause” as operating, whether or not the expected effect actually occurs.

Now, although the claim that mainstream economics is deductivist is central to Lawson’s critique of the latter, this is one of Lawson’s claims that has come under the most criticism: the view among most economic methodologists (among those anyway who think it is worth asking such questions) is that the best characterization of mainstream economics’ methodology is that it follows *Mill’s* philosophy of science.<sup>3</sup> Hands (1999) has raised precisely this point against Lawson. In his response, Lawson states that “If a capacity of sorts is posited, then, if deductivist (typically formalistic) modelling is to proceed, it must be supposed that this capacity in the given situation is always exercised and its activity invariably realised or *actualised*” (Lawson 1999: 224; emphasis in original). I completely fail to see this. Mill could endorse Ricardo’s abstract chains of reasoning, without needing to hold that the conditions of the premises would always be met, and he said as much in the passage I have quoted.

---

2. A difference between the two is that Mill tends to speak of laws in cases where Bhaskar and other critical realists tend to speak of causal mechanisms, but this is immaterial, since transcendental realists do hold that causal mechanisms are governed by laws.

3. Of course, to say that mainstream economics follows a Millian methodology by itself is neither to commend or critique it: it all depends on what one thinks of Mill’s philosophy of science. Hausman (1992) is most associated with the thesis that economics follows a Millian methodology, and in overall sees this in a positive light. In my (1999; forthcoming) I adopt this thesis, but argue that it provides a cogent explanation for the failure of mainstream economics.

The obvious danger here is that if transcendental realism is similar to Mill's philosophy of science, while the latter serves as the basis for the methodology of mainstream economics, then transcendental realism may not be able to deliver on one of its main objectives, namely, to provide a philosophical critique of mainstream economics. One way of 'testing' this suspicion is to consider whether there is anything in transcendental realism that leads specifically to the kind of social ontology which Lawson puts forward, involving "rules, relationships, positions and the like" (Lawson 1997: 57), or whether an advocate of 'rational choice theory' could use critical realism to justify his own ontology, too. In a discussion of how to develop an ontology for social science, Lawson states that

The *ex posteriori* fact of human intentionality and choice indicates that there are real material causes or structures which facilitate intentional action. But it does not yet follow that there are structures which can be said to be clearly *social*. Now if the term social is to designate anything specific here, it must be a dependency on human intentional agency. . . . [I]n determining the real possibility of social science we must acknowledge that science employs not only a perceptual, but also a causal, criterion for the ascription of reality to a posited object. . . . Entities which cannot be observed directly can be known to exist through the perception of their consequences at the level of actual events and states of affairs. . . .

Once we accept the property of depending upon human agency as criterial for the social, and acknowledge the causal criterion for ascribing reality, it is easy enough to see that identifiable social structures do exist. Items such as (societal) rules, relations and positions clearly depend on human agency as well as condition our every day (physical) activities. (Lawson 1997: 31)

Now, I see no reason why a rational choice theorist could not agree with everything in this passage up to the last sentence, and amend the latter as follows, so that he can agree with it too: "Well-ordered preferences and expectations clearly depend on human agency as well as condition our every day (physical) activities." That is to say, I am not able to find anything in transcendental realism which would lead one to choose a holistic, institutionalist brand of social theory over an individualistic, rational-choice variety.<sup>4</sup>

I share Lawson's intuitions about what society consists of. The point I am making here is that these intuitions are not supported by transcendental realism—neither by any Kantian considerations, nor by realism itself—but by Lawson's intuitions about what is real. Rules and positions are real; utility functions and demand curves are not. One can agree, but the point is that transcendental realism doesn't provide us with any argument to that effect.

---

4. To be sure, just before the passage cited, Lawson observes that mainstream "economists are seen to be unable to reconcile real choice with their project of economic 'modelling' ", and this can be taken as a reason to reject rational-choice theory. But as Parsons (1999: 166) points out, if all human actions are governed by mechanisms as transcendental realism asserts, then how can human choice be real according to transcendental realism?

Another instance when Lawson has very good intuitions, but receives no help from the official doctrine of transcendental realism, is those passages in *Economics and Reality* which do explicitly take up a position directly opposed to Mill's methodology of social science. This is in the sections "The Method of Isolation" (pp. 131-133) and "Abstraction and Economic Modelling" (pp. 234-237). The basic point is that, in his belief that economics could be a "separate" science, Mill was assuming that in society one finds a "mechanical combination" rather than a "chemical composition" of causes (Mill 1965: 371). Lawson finds that "contra both Mill and recently Cartwright, but as Keynes clearly recognized, social material does not usually conform to [the requirement that causes be mechanically decomposable] at all" (Lawson 1997: 235). Significantly, it is Keynes—someone who had spent much effort in understanding the economy as a working scientist—not Bhaskar whom he sights in support.

### 3.2 Transcendental realism is less true to science than Kant

Bhaskar claims that transcendental realism "is at present uniquely consistent with the historical emergence, practical presuppositions and substantive contents of the sciences" (Bhaskar 1989: 185). When one relates Bhaskar's three-tiered ontology to both the history of science and particular contemporary sciences, it is hard to take this claim seriously. Consider for example molecular biology. In what sense are the "generative mechanisms" at the level of the "real" in a different "ontological domain" than the effects they produce?? The effects on a patient, such as coughing, skin lesions, or whatever of a viral infection are easier to observe than the viruses themselves, but what is the point of turning this rather technology-dependent distinction into an absolute ontological difference? (The distinction between the empirical and the actual does not at all exhibit this kind of fuzziness: my sense impressions do not suddenly get up and walk away if I concentrate on them to make them as vivid as possible.)

Bhaskar's doctrine that one must always suppose that there are "generative mechanisms" underlying any observed regularity goes against an important episode in the history of science: what to make of the "action at a distance" which seemed to be implied by Newton's universal law of gravitation. According to the physical ontology current at the time Newton introduced his theory, the world consists of particles in motion. So how could two bodies attract each other, if they were not in direct contact? As Newton's theory became increasingly well-confirmed, gradually natural philosophers became willing to think of gravitation as *nothing more than* a law, with no "mechanism" associated with it. Later, Kant revised this somewhat ad hoc solution to the conceptual problem, by respecifying the concept of matter so that gravitational attraction is a property of matter in the same way that extension and mass were already considered to be, thus enabling the development of field theories. Here we have a case where an *empirical* development led to the abandonment of the search for "generative mechanisms", and a continued search for them would have indeed impeded the development of the discipline.

It is not difficult to come up with an example of a regularity with no associated 'underlying generative mechanism' in the social domain: inflation. The 'causal mechanism' involved in inflation is straightforward (even if producing a mathematical model

that can mimic time series for a given economy is not). If I am a businessman, and I see that my costs are rising, I think to myself that I had better raise my prices or I shall soon be making losses, so I do raise my prices. There are two points to be noted here. First, all the causality occurs on the level of the businessman (together with the communication to him of raised prices for his input goods and his communicating to others his raised prices for the goods and/or services he produces), and it is hard to see in what sense the causality here needs to be discovered: the businessmen can follow his own reasoning, and observers successfully guessing his intentions can follow it as well. It would make no more sense to look for a special mechanism in the case of the businessman than it would for me to look for a special mechanism to explain why I turned on the light when I walked in to a room: the room was dark, so I turned on the light. The difference between the two cases is not philosophically interesting. Second, while there is an aggregate, higher-level regularity ‘produced’ by the ‘combined’ actions of the participants in an economy which we call inflation, there is no mechanism, hidden or otherwise, associated with this regularity, in the way that there is a mechanism—an electric motor—associated with the correlation between the voltage of the current entering the motor, and the rotational speed of the shaft connected to it. There are just the intentional actions of participants in the economy which, the ‘causality’ of which, since the actions are intentional, is in principle accessible to us, and the higher-level ‘epiphenomenon’ of inflation. (Of course, once people start noticing the inflation, it is no longer just an epiphenomenon, but starts affecting people’s actions; nevertheless, all of the causality is still occurring in people’s intentional action and communication.)

To conclude this section, I will make a brief remark about Kant’s idealism—which seems to be the main reason why Bhaskar sees Kant’s philosophy as so misguided as not to be worth subjecting to a sustained critique. It must be kept in mind that whatever the weaknesses of the idealism of Kant’s philosophy, these weaknesses pertain only to it taken as a philosophical doctrine, and not as a scientific methodology. Contrary to the way Bhaskar puts it, Kant never held that the objective, natural world depends on us: he merely held that although things do have an existence independently of us, we can know nothing about these “things in themselves”. On the other hand, the empirical objects that *are* constituted by our senses are the very same objects that empirical science studies. This doctrine may seem counterintuitive to most, but that has no implication one way or another as to the ‘correctness’ of the associated scientific methodology. (Kant’s main motivation for the doctrine was to produce a refutation of Cartesian skepticism yielding certainty, something Bhaskar nowhere mentions: and no one has been able to produce as compelling a refutation by a less idealist route.) In any case, idealism should not be an issue for the social sciences, since (“dialectical materialism” notwithstanding) few people today would want to argue that social phenomena are determined by ineluctable “material forces”, as opposed to the thoughts and actions of human beings, which exist in the realm of ideas, not matter.



#### 4 Kant's critical philosophy

The standard way of situating Kant in the history of philosophy, and one encouraged by himself, is to see him as confronting with each other two great traditions in philosophy—empiricism and rationalism, and specifically Hume and Leibniz—and combining them in such a way as to provide for the first time, Kant claimed, a secure foundation for knowledge. By turning reason on itself, Kant was able to uncover the unfounded presuppositions of empiricism and rationalism, and by demonstrating the limits of knowledge, to make what fell within these limits that much more well-grounded. But to fully understand Kant's project, and the sense in which his philosophy was the *critical* philosophy, it is necessary to situate his project in the larger context of the Enlightenment. Robert Pippin has forcefully expressed how to do so:

As Kant himself explains his own position, he makes it quite clear that he considers himself to be moving beyond the (as he saw it) half-hearted attempts at providing a foundation for modern science in the likes of Descartes and Leibniz and Spinoza and Hume, attempts still tied to unwarranted metaphysical and theological commitments, and/or unjustified, inconsistent views about human psychology. Moreover, he also insists, contrary to many views of his project then and now, that he is not doing so by re-working standard (and ultimately dead-end) modern positions, by arguing that objects are “really” mental constructions, or that his own enterprise is either a metaphysical account of the mind's inner nature or a psychological inventory of the faculties at work in experience. The issue he virtually invented, the “conditions of the possibility of experience,” is, he insists, a radically new formulation of philosophy's task, and so provides for the first time the appropriate way of thinking about the deepest philosophic issues of the modern revolution. The self-grounding required for modernity to be modernity can now be accomplished without empiricist foundationalism or metaphysical fancies. The modern subject will determine for itself, completely and unconditionally, what to accept as evidence about the nature of things and, ultimately, what to regard as an appropriate evaluation of action. It will be completely self-determining, not bound to the “given” as foundation, not bound to the dogmatic belief that “the order of thought and the order of things are one,” and insistent that only if “pure reason can be practical,” can indeed be the exclusive object of the will, can human beings be said to be free. (Pippin 1999: 46).

Kant was thus intent on bringing the project of modernity to fruition, by demonstrating how human beings can be rationally self-determining. This entailed finding a means to stop the “endless controversies” raging in the “battleground . . . called metaphysics” (Kant 1781/1997: A viii). This means amounted to “institut[ing] a court of justice, by which reason may secure its rightful claims while dismissing all its groundless pretensions” (A xii). Evidently, if this court was to serve its function, it must be able to deliver findings the certainty of which is binding. It was this apodictic certainty which the transcendental arguments of the *Critique of Pure Reason* were intended to fulfill.

#### 4.1 Two types of transcendental arguments: constitutive and regulative

The arguments for necessary conditions for any possible experience are found in two sections of the *Critique of Pure Reason*, the Transcendental Aesthetic and the Transcendental Analytic. In the first, Kant argued that space and time are the subjective form of human sensibility, rather than independently existing entities, as Newton held them to be, or simply relations between objects, as Leibniz held them to be. While Kant's position is certainly counter-intuitive, Kant argued that it has the virtue of making comprehensible how the science of geometry is possible, something which is otherwise mysterious, since we seem to have 'intuitions' about geometric truths which do depend neither on experience, but also are something more than simply the consequences of definitions of geometric objects, since they do not depend on our doing a deductive proof. While this part of Kant's doctrine is still of interest to philosophers of physics, it has no bearing on social science, so need not concern us further. In the Transcendental Analytic, Kant tried to demonstrate by means of his "transcendental deduction" that certain categories, such as causality and substance, apply universally to any possible experience, since they are required for any 'self-aware' experience of an object.<sup>5</sup> Since both these arguments deal with allegedly necessary conditions for any possible experience, they may be valid or invalid (and few people today would find either of them valid as they stand), but they cannot be affected by any empirical finding. Kant believed this type of argumentation to be one of his greatest innovations, since, according to him, it is the only source of synthetic a priori knowledge, that is, knowledge that does not derive from experience, but is also more than just analytic knowledge (i.e., knowledge of a logical truth).

As already noted, Kant had a strong interest in the science of his time, so it should come as no surprise that the Transcendental Aesthetic and Analytic were of relevance to it. For example, from the category substance, when applied to nature (considered as the sum total of the objects of possible experience), he obtained the principles that the substance in nature remains constant throughout all change and that every alteration has a cause. However, Kant did not believe the doctrines developed in the *Critique of Pure Reason* to be sufficient to 'ground' science. This is because even if, contra Hume, one has shown the legitimacy of the principle of causality, it does not follow that the actual causal processes observed empirically will be such as to lead to any scientific understanding: nature might be so disordered as to be simply incomprehensible to humans. Therefore, according to Kant, another transcendental deduction is required, this time to derive the "orderliness" of nature. He gave this in the section "On the regulative use of the ideas of pure reason" of the *Critique of Pure Reason* and at greater length in his *Critique of Judgement* (1790/1951) (only calling it a transcendental deduction in the latter). This argument can be stated quite simply: in studying nature, one must proceed as if nature is ordered (or, as he put it in his (1781/1997), as if it is the product of a rational designer), for if nature is not so ordered, the project of science would not make any sense. Kant puts the point this way:

---

5. I will present a related argument in section 5.1 below, to the effect that thinking subjects necessarily follow rules.

[I]t cannot even be seen how there could be a logical principle of rational unity among rules [i.e., the laws of nature discovered by science] unless a transcendental principle is presupposed, through which such a systematic unity, as pertaining to the object itself, is assumed *a priori* as necessary. For by what warrant can reason in its logical use claim to treat the manifoldness of the powers which nature gives to our cognition as merely a concealed unity, and to derive them as far as it is able from some fundamental power, when reason is free to admit that it is just as possible that all powers are different in kind, and that its derivation of them from a systematic unity is not in conformity with nature? For then reason would proceed directly contrary to its vocation, since it would set as its goal an idea that entirely contradicts the arrangement of nature. (Kant 1781/1997: A650–A651/B678–B679).

We thus see that Kant put forward two different transcendental deductions: one based on the premise that there is experience, and another in effect on the premise that science *must be* possible. There is an important difference between them. In the first, in which the categories are deduced from the possibility of experience, the categories are *constitutive* of experience: without the categories, there would be no experience. In the second, in which the orderliness of nature is deduced from the goal that we set ourselves to study nature, the principle “proceed as if nature is ordered” is *regulative*: this principle merely directs us as to how we should go about doing our research (look for connections between apparently different kinds of phenomena, etc.)—nature need not be ordered for it to exist, so the principle is not constitutive of nature. Only the first kind of argument leads to certainty, and is an instance of what Kant calls “the ‘apodictic’ use of reason. In the second, whether the conclusion is in fact correct or not “is still a problem”; Kant calls this “the ‘hypothetical’ use of reason” (Kant 1781/1997: A646–A647/B674–B675).

## 4.2 The three components of science

With the exception of the *Metaphysical Foundations of Natural Science*, Kant did not in his post-critical writings devote a work specifically to science, but touched upon topics that would now be considered to belong to the philosophy of science in his more general works, specifically, the first and third *Critiques* and in his *Lectures on Logic*. It has therefore been difficult for philosophers to get a sense of what the overall implications of Kant’s philosophy would be for scientific method. This task has been made significantly easier by Gerd Buchdahl’s (1992) recent Kant interpretation.

Buchdahl’s entry point into the Kantian system is Kant’s observation in the introduction to his *Logic* that aside from confirmation, there are two other criteria that determine the acceptability of a hypothesis: its degree of scientific systematic unity or integration, and its ‘real’ (not merely ‘logical’) possibility (Kant 1992: 586; see also Kant 1781/1997: A770/B798). Systematic unity is a desideratum that has received attention both in logical empiricist (Oppenheim and Putnam 1958) and later philosophy of science literature (Friedman 1974; Kitcher 1981) and in the philosophy of economics literature (Viskovatoff 1994). It has even been recognized by economists themselves: Adam Smith made it the basis of his account of science, while the main-

stream game theorist Robert Aumann for example emphasizes it in his ‘justification’ of game theory (Aumann 1985). Probably the criterion of possibility has less currency in contemporary philosophy of science. Kant attempts to clarify it by means of the following example:

If, for example, to explain earthquakes and volcanoes we assume a subterranean fire, then such a fire must be possible, if not as a flaming body, yet as a hot one. For the sake of certain other appearances, however, to make the earth out to be an animal, in which the circulation of the inner fluids produces warmth, is to put forth a mere invention and not a hypothesis. For realities may be made up, but not possibilities; these must be certain. (Kant 1992: 586)

(Evidently, Milton Friedman (1953) is not an advocate of Kantian philosophy of science!) So, when Kant speaks of possibility, he means that a hypothesis must be judged as possible, given background scientific and general knowledge.

Buchdahl (1992) proposes that Kant’s scientific methodology be understood as implying that science consists of three ‘components’, each corresponding to one of the criteria probability (i.e., empirical confirmation), possibility, and unity just considered. Buchdahl calls these respectively the *probative*, the *explicative*, and the *systemic* components; they are shown diagrammatically in figure 1.<sup>6</sup> The component corresponding to the criterion of possibility is called “explicative” because one demonstrates the possibility of a given hypothesis by explaining how what it describes could come about: Newton’s theory of gravitation, with its action at a distance, provoked controversy precisely because, if physical processes were produced merely by particles and empty space, it could not be explained how action at a distance could come about.

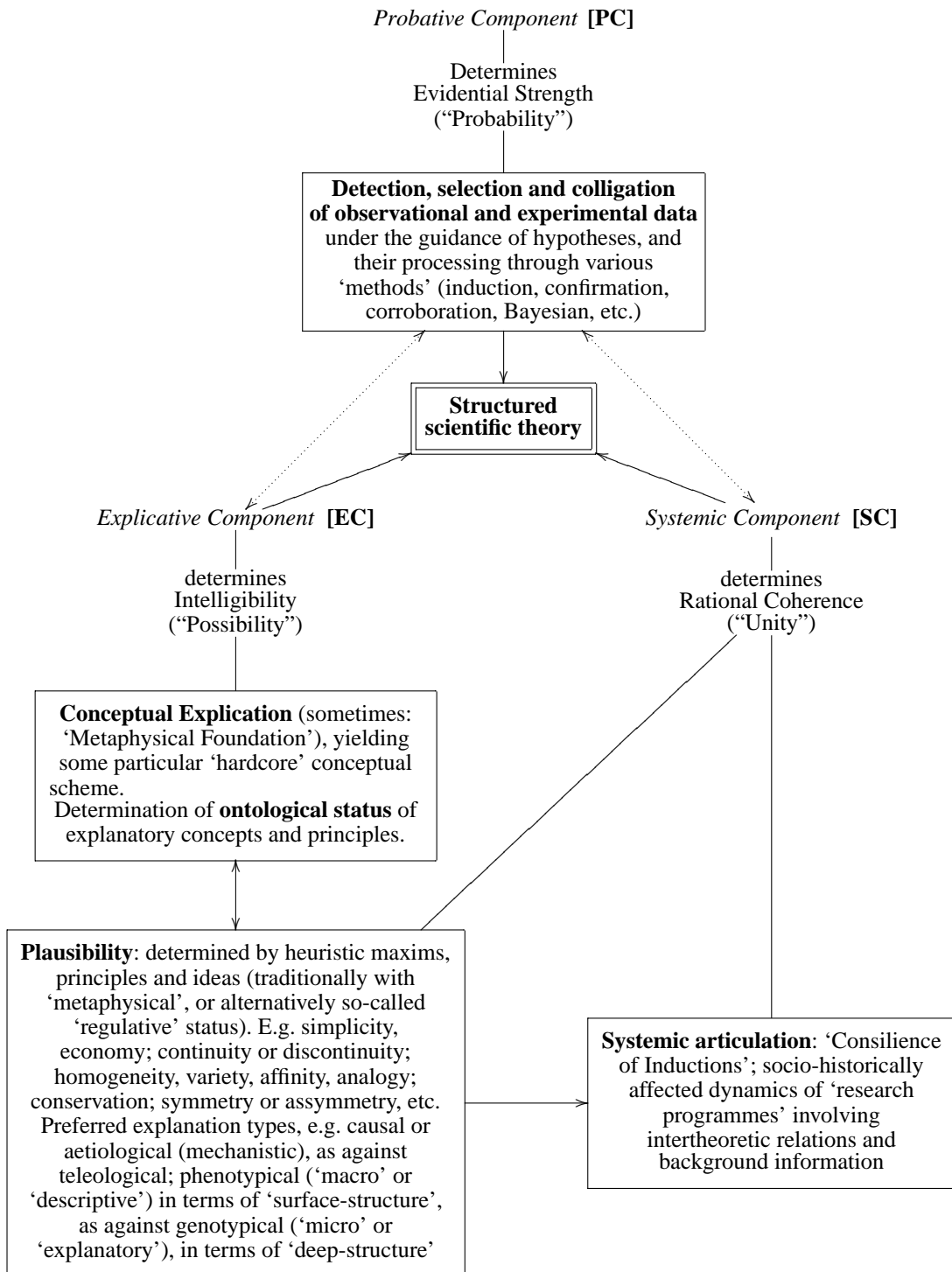
Science is produced and progresses by the interaction of all three components, which can affect one another. Thus, in the case of the reception of Newton’s theory of gravitation, due to the well-confirmed status of that theory, the probative component influenced the explicative component, by inducing a conceptual change in the theoretical framework, as we have noted. However, the influence can run in the other direction as well, as when the adherence of a theory to conservation or symmetry principles (explicative component) will make physicists more confident in the truth of the theory (probative component). And of course, the systemic component influences what kinds of theories or models will be thought worth being subjected to empirical test, as exemplified by the pursuit by contemporary theoretical physicists of a ‘grand unified theory’; similarly, one’s confidence in the correctness of Darwin’s theory of evolution derives largely from its universality—from its ability to explain the origins and development of all forms of life.

It is to be expected that Kant would apply his method of transcendental argument to the philosophy of science, and indeed his critical system does impinge on science, in three ways. First, there are the arguments of the *Critique of Pure Reason*, which describe the conditions of the possibility of experience. At the transcendental level, the causal principle is in place with apodictic certainty, since without it, experience would not be possible. Thus, the transcendental aesthetic describes experience or na-

---

6. The diagram is taken from Buchdahl (1992: 20).

Figure 1: Methodological components of scientific theory



ture “in general”, certainly of relevance to science, albeit not often brought into the center of the discussion. However, science does not normally function at the transcendental level, but at the empirical level. Therefore, the presumption that we make in science that nature is ordered is not something that we know with certainty, but rather what Kant calls a “projection”: the principle here is regulative (of our investigative activity) not constitutive (of experience and nature). This regulative use of the principle of explanatory unification is the second sort of transcendental argument that is in play: if nature were not ordered, we could not learn anything about it, so scientific activity would have no point; so if we are to study nature, we must assume that it is ordered. The third kind of a priori argument that comes into play in science according to the Kantian account is the kind of argumentation that Kant engaged in in the *Metaphysical Foundations*. (It is probably not correct to call it transcendental, since while the possibility of particular objects is being demonstrated (even the possibility of experience of those objects), the possibility of experience *as such* is not.) This kind of argumentation is considerably weaker than the other two kinds, since all that it does is essentially engage in a kind of conceptual exploration, relating concepts implied by (mathematically obtained) physical theories to other theoretical concepts and the categories of Kant’s critical philosophy.

Kant’s intentions in the *Metaphysical Foundations of Natural Science* have often been misunderstood by commentators: specifically, he is often believed to hold that for a regularity that has been observed to hold universally to be a true law, it must be deducible a priori. However, Kant is quite clear that Newton’s laws for example are contingent, empirical laws (despite for instance his ‘derivation’ of the inverse square law by his consideration that, if force diffuses itself as it moves away from a central point equally on a spherical surface, the ‘density’ of the force will decrease as the inverse square of the distance from that point): as he notes,

in natural science there are an infinity of conjectures in regard to which certainty can never be expected, because natural appearances are objects that are given to us independently of our concepts, to which, therefore, the key lies not in us and in our pure thinking, but outside us, and for this reason in many cases it is not found; hence no certain account of these matters can be expected. (Kant 1781/1997: A480/B508)

Hence, we come to the somewhat ironic finding that the method Kant followed in the *Metaphysical Foundations of Natural Science* may be more applicable to the social sciences—for social processes certainly do lie inside us, in our thinking.

### **4.3 Bhaskar’s “transcendental argument”**

We are now in the position to evaluate Bhaskarian philosophy of science from a Kantian perspective. Compared to Kant’s philosophy of science, Bhaskar’s suffers from the following drawbacks.

*Bhaskar does not recognize the explicative and systemic components.* We saw in section 4.2 that Kant’s methodology of science is considerably richer than an em-

iricist methodology of science: the latter (virtually by definition) only recognizes the probative component, neglecting the explicative and systemic components. This should surprise no one. What is perhaps surprising though is that despite his intention to critique empiricist philosophy of science and to provide a superior alternative to it—and explicitly drawing upon Kant at that!—Bhaskar also does not draw attention to the explicative and systemic components, and is perhaps not even aware of them.<sup>7</sup> This explains why transcendental realist methodology has strong affinities with Mill's, as I have argued in section 3.1. As one knows from the experience of mainstream economics essentially following Mill's methodology, overlooking the explicative and systemic components is especially harmful in the case of economics, since in that discipline the probative component alone is not sufficient to confirm or reject a theory, leaving the discipline with no adequate criterion of theory selection (Viskovatoff 1999).

*Bhaskar misunderstands the kind of transcendental argument that is required to demonstrate that nature is ordered,* taking the constitutive variety as his model instead of the regulative one. Bhaskar says that his transcendental argument is “revisable”, but we do not have to leave it at that. The status of his argument hinges around how one interprets the “generative mechanisms” of the real domain. Either the fundamental laws of physics—which (on the model of the universal law of gravitation) are nothing but mathematically expressible laws, with there being no mechanism in the usual sense of the word to describe in addition to the regularity itself—are mechanisms in Bhaskar's sense of the word, or they are not. If they are, then Bhaskar's claim that “there must be generative mechanisms” reduces to the Kantian “we must assume that nature is ordered”. If they are not—so that a “generative mechanism” can never be simply a regularity like the one described by the universal law of gravitation—then Bhaskar's claim (and hence his “trancendental deduction”) is simply wrong, because of examples from the history of science like the one of the universal law of gravitation.

If one chooses the first interpretation, then Kant's demonstration that nature is ordered is still superior to Bhaskar's for two reasons. First, it gets by without introducing and heavily relying upon a new ontological distinction, that between the real and the actual, which is not at all plausible or stable. (I turn off a light. The light goes off. Why must I place the event of the light going off in a different ontological domain from the mechanism which produced the event—the light switch?) Second, Kant's demonstration is categorical (by describing what “must be”, not a state of affairs), whereas Bhaskar's demonstration must portray itself as corrigible. What sense does it make to

---

7. One can infer from his realism that one should look for real mechanisms, and that can be taken as a reference to the probative component. This is however a purely formal stipulation: how is one to know what is real? The candidates to consider are those entities that are (really) possible, given one's overall theoretical framework. Upon a charitable interpretation of what ‘look for what is real’ could mean in practical terms to the working scientist, this precept is still found to state implicitly what Kantian methodology makes explicit. In any case, I have nowhere seen Bhaskar discuss the importance of explanatory unification—the systemic component—for science. Unfortunately, (Bhaskar 1975) does not have an index. However I looked for terms such as unification, explanatory unification, unity of science, system, and systematicity in the index of (Archer et al. 1998), and did not find any.

say that “I know that nature is ordered (i.e., there is a real domain consisting of generative mechanisms beyond the domain I exist in, that of phenomenal events), but I could be wrong?”

*Bhaskar’s adoption of a naturalist position precludes his using an “idealist” transcendental argument to directly demonstrate the reality of rule-following.* By not shying away from idealism however, we are able to demonstrate (as shown in section 5.1) that rule-following is constitutive of the rationality of any thinking subject, a very useful result for a critique of mainstream economics.<sup>8</sup> As I have argued, there is nothing in Bhaskar’s transcendental realism itself which would lead one to reject the rationality concept of mainstream economics.

*Bhaskar’s claim that there is an epistemic fallacy gives him the license to abandon the standards of philosophical argument set by Kant’s Critical philosophy and to fall into a pre-Critical mode of philosophizing.* Bhaskar holds that there are scientific ontologies and there are philosophical ontologies, and that there is nothing wrong with creating an ontology without a scientific basis, just as long as the ontology specifies merely the form of entities (they exhibit lawfulness) and not the specific laws they follow. But this is to fall into a pre-Critical (i.e., pre-Kantian) metaphysics, analogous to the Scholastic metaphysics in which one argued whether matter was fundamentally continuous or consisted of particles. Bhaskar defines the “epistemic fallacy” as “the view that statements about being can be reduced to or analysed in terms of statements about knowledge” (Bhaskar 1975: 36).<sup>9</sup> He also mentions a possible objection, and a rejoinder to it:

The view that statements about being can be reduced to or analysed in terms of statements about knowledge might be defended in the following way: ontology is dependent upon epistemology since what we can know to exist is merely a part of what we can know. But this trades upon a tacit conflation of philosophical and scientific ontologies. For if ‘what we can know to exist’ refers to a possible content of a scientific theory than that it is merely a part of what we can know is an uninteresting truism. But a philosophical ontology is developed by reflection upon what must be the case for science to be possible; and this is independent of any actual scientific knowledge. (Bhaskar 1975)

Since Bhaskar elsewhere admits that his “transcendental deduction” is corrigible, he has not actually based his own “philosophical ontology” on what *must* be the case, but on ‘what is probably the case’ or ‘what seems at the moment to be the case’. Therefore he has not met his own standards for creating a philosophical ontology. Furthermore, Kant’s argument to the effect that nature *must* be ordered does satisfy this standard, and Kant does not set up any ontology in addition to that of science (except for an ontology

---

8. Since I base myself on Pettit (1993), the argument I rehearse is expressed in naturalist terms. I would agree however with Bhaskar that transcendental arguments that deal with the individual are inevitably idealist; see (Pippin 1988).

9. The supreme commiter of the “epistemic fallacy” was surely Hegel, which makes it puzzling that Bhaskar has lately developed a strong interest in that particular philosopher.



of what is necessary for *experience* to be possible, of course, the motivation for all of which is epistemological, not that of ‘ontology left to its own devices’). Therefore, Bhaskar’s rejoinder to the objection does not work, which means that there is no such thing as the epistemic fallacy as he defines it.

The cost that critical realism pays for not making ontological considerations subject to epistemology are high. As I argue in this paper, stepping back from an idealist, Critical position has the effect of preventing critical realism from mounting a conclusive attack on mainstream economics: lowering the critical standards for oneself also means that one has less critical machinery with which to critique others.

#### **4.4 Going beyond Kant’s transcendental idealism**

Certainly, Kant’s anti-realism and his claim that we can not know things in themselves is unsatisfactory. But the way to overcome it is not by falling into a pre-Critical position, by declaring that we can know the nature of things that are beyond the bounds of our experience, by adopting a “God’s-eye view” of the world to construct a realist ontology. After Kant, the only philosophically adequate way of adopting a realist position is to follow Hegel by accepting Kant’s critical insights and in particular his insight that all that we know we know by means of our concepts, but to argue against Kant that 1) our concepts are simply our way of grasping the world as it really is; and 2) Kant’s idea of some kind of direct intellectual perception of objects without conceptual mediation is incoherent.

## **5 Grounding social theory**

### **5.1 Transcendental deduction of rule-following<sup>10</sup>**

Let us keep to the general practice of starting from the concept of intentionality, and begin by asking what makes up an intentional system. What we mean by an intentional agent is an agent that has beliefs and desires and that the agent is able to interact with the world so as to make representations of it, to change these representations as a consequence of its interactions with the world, and then to affect the world in a way that is appropriate given its beliefs and desires. We suppose that there will be regularities in the beliefs and desires that are held by the agent. These regularities can be classified into two main categories. One class of regularities determines what it is to be *evidentially rational* in the attitudes the agent holds: the agent must be responsive to whatever evidence might come up that has a potential impact on the beliefs that it is rational for the agent to hold. If the agent does not have sufficient information to form beliefs about a given set of contingencies that would allow it to choose a course of action that would serve to satisfy the agents’ desires, for example, the agent should

---

10. I have not yet written up this section and the next so that they match my current ‘transcendental’ orientation. I am therefore including an extract from an earlier paper, based on (Pettit 1993), who takes a more naturalist approach than I am now inclined to follow. I hope that, since I will in any case have to move my deduction of rule following in a Wittgensteinian direction, while Pettit uses Wittgensteinian arguments, the incongruity between this section and the rest of the paper is not too great.

go about trying to collect more information. The second set of regularities determines what it is to be *responsively rational* in the responses the agent makes. The agent must respond to environmental stimuli in such a way as to preserve itself and further its own goals, given the beliefs that the agent has. A system will count as an intentional agent to the extent that its interactions with the environment, or at least some of its interactions, are governed by such regularities.

From what we have said so far, all kinds of things can be intentional agents. For example, the Pathfinder probe that landed on Mars was an intentional agent: it moved towards rocks and took samples because this behavior is appropriate in order to determine the composition of the Martian soil, and determining this composition is one of the purposes for which it was designed. Similarly, animals are intentional agents, as one can see from watching any wildlife documentary, of a lion stalking an antelope for instance. We surmise that the lion's behavioral repertoire has been 'designed' in such a way by the process of natural selection so as to make it behave most effectively in hunting its prey, given its sensory and motor abilities.

So much for what makes up an intentional system. It is there that most of the philosophy of mind literature has left it, as has the economic methodology literature reporting on the former I have already cited. As we saw, a robot or cat can be an intentional system.<sup>11</sup> Does the notion of intentionality fully exhaust what we mean when we say that human beings are rational beings? According to Pettit, it does not, and to explicate the difference between human beings and other kinds of intentional systems, he introduces the concept of *thinking systems*. All the actions performed by an intentional system will be intentional (in the second sense of the word meaning that they will be guided by the system's beliefs and desires). How the thinking system differs from the non-thinking intentional system is that the range of things the former is capable of doing intentionally is greater. Specifically, the thinking system can deliberately act in such a way to make it more likely that the requirements of rationality will be met—so that it will be more likely that its beliefs are true, or that it will be more likely that its actions will serve to further its desires:

Perhaps the best way of capturing this contrast is to see that while the thinking system must have a desire to be rational—for example, a desire to have beliefs that are more likely to be true than false—the non-thinking subject will have no such desire in its make-up. It may be designed generally to adjust, say in the light of new evidence, so as to have true beliefs; it may in that sense be a truth-seeking system. But it will not have a desire to have true beliefs that will lead to action on the basis of this or that channeling of belief: say, the belief that by taking time, by investigating further, or by listening to others, it can increase the expectation of having true beliefs. It will not have a desire for truth, or more generally a desire for rationality, in the common-or-garden sense of desire. It counts as a believer but, by contrast with the thinking subject, it is a

---

11. John McCarthy, the inventor of LISP—the original hacker's language—as well as of the term 'artificial intelligence', has indeed observed that "Machines as simple as thermostats can be said to have beliefs, and having beliefs seems to be a characteristic of most machines capable of problem solving performance" (McCarthy 1979: 3).

blind believer, not a conscious one. (Pettit 1993: 6)<sup>12</sup>

It is desirable to specify what requirements must be satisfied by a system for it to be capable of thought so construed. Pettit argues that there are two such requirements, which can be called *intentional ascent* and *rule-following*. By intentional ascent we mean that a subject must be capable of having not only this or that belief, but also of having beliefs and other intentional attitudes about that belief. That is, the subject must not only be able to have the belief that *p*, but must also have beliefs about the proposition that *p*. For instance, it must be able to have the belief that it must be the case that *p*, given *q*, that it is unlikely for both *p* and *q*, and so on. If the subject did not have such beliefs, it could not engage in the project of making it more likely that its beliefs are true.<sup>13</sup>

To turn now to the second requirement, that of following rules: a subject must be capable not only of having beliefs about propositions, but also be able to treat propositions and the components out of which they are constructed as rules of thought. That is, a subject must be able to treat a given proposition as a rule, identifying the constraint that it represents for the practice of belief-formation and setting itself intentionally to form beliefs in accordance with that constraint. For example, if the proposition is ‘The cat is on the mat’, then the subject must be able to grasp the condition that makes it appropriate at the relevant time and place to believe that the cat is on the mat and must be able, in addition, to abide by that condition. If the subject were not able to follow rules in this way, it would not be able to reflect on and ‘disassemble’ its reasoning processes in such a way as to try to improve them.

The topic of rule-following has been at the center of a philosophical debate since Wittgenstein (1953), and Pettit develops a solution to the problem of how rule-following is possible. First, we follow the rule associated with a proposition by following more basic rules associated with elements of the proposition. Thus, in order to follow the rule given by the proposition ‘The cat is on the mat’, the subject must know among other things the rules which pick out cats and mats from other things in the world and the rule which tells it when one thing is on top of the other. Now, a difficulty that arises with the notion of rule-following, raised by Kripke (1982), is the following. Consider the proposition ‘That is a cat’. I follow the rule for belief dictated by this proposition by following the rule associated with the property of being a cat; this more basic rule dictates that I group some things together as cats, but exclude others. How might the rule associated with the property be presented to me, as something that dictates responses over an indefinite range of cases and as something that I can identify and try to conform to in those cases?

---

12. John Haugeland has made a similar point, introducing the concept of “ersatz” intentionality to be distinguished from Searle’s intrinsic rationality, to indicate the difference between what Pettit has called non-thinking intentional systems and thinking systems: “[A] computer has no understanding whatever about business, accounting, or even arithmetic. *Understanding* a domain and its entities is understanding the principles according to which that domain and those entities are constituted; and such understanding can be nothing other than a *commitment* to those principles” (Haugeland 1998: 301).

13. It is likely that language plays an important role in allowing human beings to have such beliefs, but we can omit that detail here.

Since the rule must dictate responses in an indefinite range of cases, it cannot be presented to me in a finite set of cases. If that were so, it would always be possible to treat this set as a subset of a larger, arbitrarily constructed set. Thus, there is no *logical* reason why a set of cat-exemplars could not be treated as a paradigm for objects that are either cats or bushes, or cats until a certain point in time and bushes thereafter, and so on; there will be no necessary link between the exemplars and any particular class. But on the other hand, for the rule to be accessible to a finite mind like mine, it is hard to see how else it could be revealed to me other than by some finite set of exemplars of some kind. The problem then is how to reconcile these two opposing demands: how to make the rule have an objective validity on the one hand, so that it can be treated as normatively binding, while on the other hand allowing the rule to be accessible to finite subjects, so that it can be something that anyone can intentionally try to follow.

Pettit's solution is the following. Given a finite set of exemplars, I have an *inclination* (presumably derived from nature) to interpret the set as representing a particular definite rule. Thus, given the objects presented to me as examples of cats, I make the 'natural' kind of generalization of what kind of objects the term 'cat' refers to, and not one of the other formally admissible ones of the kind that Kripke suggests. Because this generalization derives from my inclination however, there is the danger that it is not objectively valid. But does not the account just given suggest that what it means to be a cat is whatever I take it to mean, with that being the end of the story, and there being no possibility of my being mistaken?

Pettit avoids this danger by adding to the extrapolative inclination another 'mental habit'. Suppose that, either by nature or by social convention, if I find that there is a divergence between my inclination and either the determination I had made under similar circumstances at a different time, or the determination made under similar circumstances by someone else, then I assume that

something is amiss. Suppose I am disposed in such cases to reserve commitment until a perturbing condition has been identified, a condition that means that the inclination that it affects should be discounted. In that case, and it surely has a certain plausibility, the rule and kind that I should be said to identify on the basis of the initial exemplars is not whatever kind is fixed by my actual extrapolative inclination, but whatever kind is fixed by that inclination as it operates under circumstances that do not occasion interpersonal or intertemporal inconsistency: by that inclination, as it operates under those favourable conditions that survive negotiation with myself across time or with other individuals. (Pettit 1993: 8)

Pettit calls the kind of story just told as 'ethocentric'. It employs two kinds of materials: habits of response (extrapolative inclinations) and practices of negotiation. Guided by my extrapolative inclination, I obtain a rule for judging whether something is a cat from the exemplars I initially confront. This rule then serves to yield judgments in my future encounters with different objects. But I do not act as if I cannot go wrong in my specification of the rule and application of it; this is because I give way to my inclination only so long as I can do so under conditions such that a convergence between the inclination and my inclinations at other times or the inclinations

of other persons proves to be attainable. Thus I do not see myself as an authoritative employer of the rule, “for the rule to which I address my efforts of construal is authoritatively presented only in a voice that is beyond my control: the concerted voice of intertemporally or interpersonally consistent responses” (Pettit 1993: 9).

If Pettit’s account is more or less correct, then the upshot is that if we were not by our nature disposed to follow certain habits and inclinations and not others, then thought would not be possible at all. Since without thought, there would be no experience (of the kind human adults have, anyway), we have the makings of a transcendental argument that human rationality is constituted by rule following. Because the argument has no empirical component, it yields apodictic certainty.

## 5.2 ‘Ruling out’ decision theory

Let us turn now to the implications of what Pettit calls the “intentional psychology” that has been developed in the previous section for the explanation of human action. In so doing, let us keep in mind in particular the Kant’s criterion of ‘real possibility’ introduced in section 4.2.

Usually, one thinks that the causally relevant properties in the case of intentional behavior are the intentional system’s beliefs and desires. But according to the picture just developed, the rules that are followed are just as important. The causally relevant factors thus include propositionally characterized perceptions and propositionally characterized beliefs, together with the habits of inference involved in moving from perceptions to beliefs and from beliefs to other beliefs, together with the desires involved in moving from beliefs to actions. Thus what is involved in explaining an agent’s action is in effect to produce an argument that we can take the agent to have endorsed:

This style of explanation involves reconstructing the agent’s pattern of inference or deliberation, the pattern of explicit or implicit reasoning that she is taken to have followed. It need not represent the agent as justified in what she did, since it may indicate that the premises in the light of which she responded were not true or were not the whole truth, or it may point us to a slip that the agent made in her reasoning. Still, the explanation does introduce us to the viewpoint of the agent herself. Even if it supports a critical stance on her responses, it will have to invoke an intelligible emotion, a common oversight, a standard fallacy, or something of that kind, to make sense of her failure; otherwise it will leave us in the dark about how she, an intentional and thinking system, came to be moved as she was. (Pettit 1993: 234)

Pettit calls approach followed in this kind of explanation *inference theory*, observing that it amounts to what is now commonly described as interpretation, and has been earlier referred to as *Einfühlen* (Herder) and *Verstehen* (Dilthey, Weber).

Inference theory has the name it does because it sticks rather closely to patterns of inference that agents actually use to develop their beliefs and choose their actions; it does not depart substantially from what philosophers call ‘folk psychology’, but merely elaborates it. Here it stands in marked contrast to the theory used by economists for the explanation of human behavior, known as decision theory. Decision theory

postulates that agents have stable and well-defined preferences, but takes these preferences as given; that (in contrast to inference theory) preferences do not affect beliefs, and vice-versa; and that when agents face uncertainty, they are able to assign precise values to what they believe the probabilities of the various possible outcomes are. Decision theory thus departs significantly from folk psychology. Economists do not see this as a problem however, but actually as a scientific necessity, and a mainstream economist will not feel that a genuinely 'scientific' explanation of a given action has been given unless it is 'explained' by means of decision theory. On the basis of what has come before, what can we say about the relative merits of inference theory and decision theory?

Pettit believes that there is an argument for the former and against the latter, and I believe that he is correct. The argument is based on the presumption that human subjects are not just intentional systems, but thinking systems. The argument is that in order to be a system capable of thought, the human being has to form new beliefs and choices in the manner depicted by inference theory:

To be a thinking subject is to be an intentional system that acts intentionally with a view, among other things, to having rational intentional states. Such a system will be subject to certain causally relevant pressures in the formation of beliefs and intentions and other relevant states. If an agent is generally to be rational, she will be subject to pressures that generally make for rationality. And if she is to be capable of thought, capable of intentionally achieving such rationality, then she will be able to identify the sorts of pressures at work, to assess them for their tendency to promote rationality in any given case, and to reinforce or inhibit them as the assessment requires. (Pettit 1993: 244)

Under the inference-theoretic picture, there is no difficulty about seeing how the thinking subject can meet these conditions. A person can see that to make a decision wisely, she must assess all relevant matters, pay careful attention to evidence, reassess her beliefs if new evidence comes along, and so on. So there is no difficulty about seeing how a person can act intentionally to improve the rationality of her intentional states, that is, how she can have the ability to think.

Consider now how things stand with decision theory. Say a new piece of evidence comes along, for example. Applying decision theory, it will usually be supposed that Bayes' rule is used in updating subjective probabilities. The person would not only have to be able to give a numerical value to the strength of her prior belief (and this implies having an unusual capacity for introspection), but also be able to apply Bayes' rule, which is not computationally trivial. So it is very hard to see how an intentional system could use decision theory in practice to improve the rationality of her beliefs and actions. Thus, given that people are thinking systems, we cannot plausibly see them as intentionally responding just to the sorts of factors identified in decision theory as constitutive of rationality.

Pettit concludes from this the following:

Decision theory . . . does not give an account of the factors that thinking systems like you and me consider as we find our way to revisions of our beliefs and to decisions about what to do. For such an account we do better

to rely on some version of the inference-theoretic picture presented earlier. It develops an image of the springs of beliefs and choices that allows us to make ready sense of how an intentional subject can have the capacity for thought: how she can have the ability to act intentionally with a view to promoting her own intentional states. (Pettit 1993: 245)

Thus, the neoclassical formalism for describing human behavior is not able to show us how we are capable of our highest form of rationality, the form of rationality that makes us different from animals or machines. Furthermore, given what we know both from our ‘transcendental deduction’ of rule-following for thinking systems as well as from our background knowledge about what kinds of factors individuals are able to take into account in their decision making, it is clear that the decision-theoretic, utility-maximizing model of human decision making does not satisfy the Kantian criterion of real possibility. From the standpoint of Kantian philosophy of science therefore, utility maximization is not a legitimate scientific hypothesis.

### 5.3 ‘Validating’ social system theory

The next thing to do is to justify our concept of social system. The way to do that is to follow the form of argumentation that Kant follows in his *Metaphysical Foundations of Natural Science*: we demonstrate how social systems are possible, given our concept of the rule-following actor/subject. Actors are able to successfully communicate with one another, which means that they must share some rules. (Wittgenstein’s (1953: §§243–315) private language argument even suggests that being a member of a community with collectively binding rules is constitutive of thought, and Pettit picks up this idea.) All that one needs at this point to get to the concept of a (Luhmannian) social system is to show how the system achieves ‘autonomy’ with respect to actors. This can be done by means of an argument from complexity: once the set of shared rules becomes sufficiently complex, in those contexts which fall within the perspective which ‘pertains’ to a given social system, actors will in general not be able to reason their way through to developing their own perspective and to subsequently negotiate their way through to actualizing it, so that communications and actions produced will be attributable to the set of rules and hence the social system itself, not to individual actors.

Finally, when it comes to ‘deriving’ ‘higher-level’ concepts of social theory, such as that of symbolically generalized communications media such as money, power, or truth, the type of a priori considerations involved will be the regulative ideas of simplicity, unity, and generality. Upon casually surveying modern societies, one finds that they seem to have developed various ‘spheres’, each operating according to its own particular ‘logic’. Bringing all of these various logics—the use of money to make claims on goods, the use of empirical research and careful argumentation to determine truth, the use of law to regulate cases of conflicting interests—under one general concept is justified by the ideals of unification and generalization.

## 6 Conclusion

Critical realists are to be congratulated for wanting to develop a critical methodology relevant to practicing economists, for wanting to do this via the philosophy of science, and for their intuition that the philosophy to turn to specifically is that of Kant. But Bhaskar's watered-down 'transcendental philosophy', with its naturalism inspired by "dialectical materialism", will not do. To go beyond empiricism, it is necessary to study Kant much more closely, and to adopt a realism that does not neglect Kant's critique. Fortunately, there now exists a literature which makes it significantly easier to reappropriate Kant's critique of empiricism.

## References

- Archer, Margaret, Roy Bhaskar, Andrew Collier, Tony Lawson, and Alan Norrie, eds. 1998. *Critical Realism: Essential Readings*. Critical Realism: Interventions. London: Routledge. 15, 24
- Aumann, Robert J. 1985. What is game theory trying to accomplish? In *Frontiers of Economics*, edited by Kenneth J. Arrow and Seppo Honkapohja, 28–76. Oxford: Basil Blackwell. 12
- Bhaskar, Roy and Tony Lawson 1998. Introduction: Basic texts and developments. In Archer et al. (1998), 3–15. 2, 2
- Bhaskar, Roy 1975. *A Realist Theory of Science*. Leeds: Leeds Books. 1, 1, 2, 3, 15, 16, 16
- Bhaskar, Roy 1989. *Reclaiming Reality: A Critical Introduction to Contemporary Philosophy*. London: Verso. 3, 7
- Buchdahl, Gerd 1992. *Kant and the Dynamics of Reason*. Oxford: Blackwell. 11, 12, 12
- Fleetwood, Steve, ed. 1999. *Critical Realism in Economics: Development and Debate*. London: Routledge. 24, 25, 25
- Friedman, Michael 1974. Explanation and scientific understanding. *Journal of Philosophy*, 71:5–19. 11
- Friedman, Milton 1953. The methodology of positive economics. In *Essays in Positive Economics*, 3–43. Chicago: University of Chicago Press. 12
- Hands, D. Wade 1999. Emirical realism as meta-method: Tony Lawson on neoclassical economics. In Fleetwood (1999), 169–185. 5
- Haugeland, John 1998. *Having Thought: Essays in the Metaphysics of Mind*. Cambridge, Mass.: Harvard University Press. 19
- Hausman, Daniel M. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press. 5
- Kant, Immanuel 1781/1997. *Critique of Pure Reason*. Cambridge: Cambridge University Press. 9, 10, 11, 11, 11, 14
- Kant, Immanuel 1790/1951. *Critique of Judgment*. New York: Hafner. 10



- Kant, Immanuel 1992. *Lectures on Logic*. Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press. Trans. and ed. by J. Michael Young. 11, 12
- Kitcher, Philip 1981. Explanatory unification. *Philosophy of Science*, 48:507–31. 11
- Kripke, Saul 1982. *Wittgenstein on Rules and Private Language*. Oxford: Basil Blackwell, rev. edition. 19
- Lawson, Tony 1997. *Economics and Reality*. London: Routledge. 2, 5, 5, 6, 6, 7
- Lawson, Tony 1999. Critical issues in *Economics as Realist Social Theory*. In Fleetwood (1999), 209–257. 5, 5
- McCarthy, John 1979. Ascribing mental qualities to machines. Stanford University Computer Science Department. <<http://www-formal.stanford.edu/jmc/ascribing.ps>> (4 March 1999). 18
- Mill, John Stuart 1960–1991. *Collected Works of John Stuart Mill*. Toronto: University of Toronto Press; London: Routledge and Kegan Paul. 25
- Mill, John Stuart 1965. *A System of Logic Ratiocinative and Inductive Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. Mill (1960–1991), vol. 7–8. 4, 5, 7
- Oppenheim, Paul and Hilary Putnam 1958. Unity of science as a working hypothesis. In *Concepts, Theories and the Mind-Body Problem*, edited by Herbert Feigl, Michael Scriven, and Grover Maxwell, volume 2 of *Minnesota Studies in the Philosophy of Science*, 3–36. Minneapolis: University of Minnesota Press. 11
- Parsons, Stephen D. 1999. Why the ‘transcendental’ in transcendental realism? In Fleetwood (1999), 151–168. 4, 6
- Pettit, Philip 1993. *The Common Mind: An Essay on Psychology, Society, and Politics*. New York: Oxford University Press. 16, 17, 19, 20, 21, 21, 22, 23
- Pippin, Robert B. 1988. The idealism of transcendental arguments. *Idealistic Studies*, 43:97–106. 16
- Pippin, Robert B. 1999. *Modernism as a Philosophical Problem: On the Dissatisfactions of European High Culture*. Malden, MA: Blackwell, second edition. 9
- Sellars, Wilfrid, Richard Rorty, and Robert Brandom 1997. *Empiricism and the Philosophy of Mind*. Cambridge, Mass.: Harvard University Press. Taken from *Minnesota studies in the philosophy of science*, v. 1, *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. 1
- Skorupski, John 1989. *John Stuart Mill*. London: Routledge. 4
- Viskovatoff, Alex 1994. Unity of theory: The role for theoretical maps. In *Perspectives on the History of Economic Thought*, volume 10, 3–18. Aldershot, Hants, England: Edward Elgar. 11
- Viskovatoff, Alex 1999. Neo-transcendental foundations for economics. Paper presented at the European Association of Evolutionary Political Economy annual meeting, Prague, November 1999. 5, 15
- Viskovatoff, Alex forthcoming. Review of Deborah Redman, *The Rise of Political Economy as a Science*. *Research in the History of Economic Thought and Methodology*. 5

Wittgenstein, Ludwig 1953. *Philosophical Investigations*. Oxford: Basil Blackwell.  
19, 23