



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Transformational fallibilism and the development of understanding

**Citation for published version:**

Kemp, S 2017, 'Transformational fallibilism and the development of understanding', *Social Epistemology*, vol. 31, no. 2, pp. 192-209. <https://doi.org/10.1080/02691728.2016.1270365>

**Digital Object Identifier (DOI):**

[10.1080/02691728.2016.1270365](https://doi.org/10.1080/02691728.2016.1270365)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Social Epistemology

**Publisher Rights Statement:**

This is an Accepted Manuscript of an article published by Taylor & Francis in Social Epistemology on 23/02/2017, available online: <http://www.tandfonline.com/eprint/GKb8cBtRAnDjlg5gmXvK/full>

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## **Transformational Fallibilism and the Development of Understanding**

Dr Stephen Kemp

Position: Senior Lecturer in Sociology, University of Edinburgh

Email: [s.kemp@ed.ac.uk](mailto:s.kemp@ed.ac.uk)

Phone: 0044 (0)131 650 3895

Address: Sociology, School of Social and Political Science, University of Edinburgh, Chrystal Macmillan Building, 15a George Square, Edinburgh, United Kingdom, EH8 9LD

Author's Biographical Note: Stephen Kemp is a Senior Lecturer in Sociology at the University of Edinburgh, United Kingdom. Correspondence should be addressed to [s.kemp@ed.ac.uk](mailto:s.kemp@ed.ac.uk) or Dr Stephen Kemp, Sociology, School of Social and Political Science, University of Edinburgh, Chrystal Macmillan Building, 15a George Square, Edinburgh, United Kingdom, EH8 9LD

Acknowledgements: I would like to thank Christoforos Bouzanis, Nathan Coombs, and the anonymous referees for their comments on versions of this paper.

## Transformational Fallibilism and the Development of Understanding

**Abstract:** This article argues that inquirers should adopt an active orientation to the limits of their knowledge, an approach referred to as 'Transformational Fallibilism'. Drawing on the Popperian tradition, this approach treats the fallibility of knowledge as more than a philosophical nicety, rather seeing the questioning of claims, including those that have been successful, as a key way to improve the understandings of inquirers. This is illustrated with reference to the example of Newtonian and Einsteinian understandings of gravity and time in the natural sciences, and debates about the role of fathers in child-rearing in the social sciences. I compare this approach to that found within critical realism, arguing that while defenders of the latter acknowledge fallibility to some extent, certain of their arguments also place problematic limits on it. Examining Sayer's work, I argue that his view that practically successful understandings reveal something about the structure of the world involves an over-confidence in these understandings. I suggest that Elder-Vass's use of the idea of approximate truth involves a similar difficulty. By contrast, I argue for the importance of treating all elements of the understanding, including those bound up with practical successes, as potentially in need of reconstruction at any point.

**Keywords:** Fallibilism, Critical Realism, Philosophy of Science, Andrew Sayer, Dave Elder-Vass

### 1. Introduction

I was talking to a fellow attendee at a conference a year or so ago, and the subject of the fallibility of knowledge came up. My colleague was clearly irritated by the concept. 'The fallibility of knowledge!', he said, 'What's so interesting about that? Everyone acknowledges that knowledge is fallible don't they?!'<sup>1</sup> This seemed to me to be insightful, up to a point. At least within the sphere of contemporary sociology and social theory there are few if any thinkers who defend the idea that knowledge is *infallible*. Constructionists, post-structuralists, feminist epistemologists, complexity theorists, and actor-network theorists, to name adherents to a few of the significant positions within the landscape of social theory, are all comfortable with, and often insistent upon, the idea that knowledge is fallible. Even those who retain some enthusiasm for science and its achievements, such as critical realists and neo-pragmatists, nevertheless explicitly espouse the view that knowledge is fallible. Proponents of all of these positions accept that the truth of a knowledge claim cannot be established with certainty, even if they differ on how sceptical and critical one should be about the knowledge claims of scientists and others. As such, this conference conversation pushed me to think harder about whether there is nevertheless something important about the idea of fallibility that is not generally accepted; whether there is something to be learned from thinking carefully about fallibility. My conclusion was that there is a further lesson to be learned, and this article attempts to articulate it.

The approach to knowledge expounded and defended in this article is what I call 'transformational fallibilism'. Whereas some of those who accept fallibilism draw very few consequences from the argument that our knowledge claims are uncertain, transformational fallibilism suggests that there are important lessons for inquiry that should be drawn from this argument. I will be arguing that transformational fallibilism is a valuable approach and elaborating on this by means of a comparison with the approach to fallibilism taken within critical realism, a position which provides a valuable contrast for three reasons. Firstly, it is a key, prominent, contemporary approach to the philosophy of social science. Secondly, critical realism's proponents are explicitly committed to fallibilism and this allows me to explore the difference between what I see as their somewhat limited commitment in this area and the transformational approach being defended here. Thirdly, engaging with critical realism allows me to explore two specific ways in which a commitment to fallibility can be limited, each resulting in what I would see as an over-confidence in (some) beliefs. One is by the argument that practically successful beliefs reveal something about the nature of the subject-matter that knowers are engaging with. The other is by the contention that it is possible for inquirers to identify which beliefs are approximately true. I will be arguing that both of these moves, each of which is defended by an important critical realist thinker, mistakenly reify current beliefs and understandings which are better treated as open to critical appraisal and reconstruction in the service of improving our understandings.

The arguments presented herein resonate with a previous critique of critical realism developed by Justin Cruickshank. Cruickshank (2004) usefully argues that the critical realist approach to ontology is somewhat inconsistent in its attitude to the fallibility of ontological argument itself. At times, critical realists treat their ontology as part of a process of inquiry and as something that could itself be transformed; at other times it is treated as if it is beyond the realm of transitive change. In this article I want to develop this view about the inconsistency within critical realist arguments but do so not by focusing on the status of ontological argument but on the critical realist view of substantive knowledge claims. The intention of this engagement is to show the limitations of the critical realist approach and demonstrate what a committed embrace of fallibilism would involve.

The article will proceed as follows. Firstly I will develop an account of transformational fallibility with reference to other arguments in the philosophy of science and examples from the natural and social sciences. Then I will go on to introduce critical realism with reference to the work of Roy Bhaskar, before exploring its approach to fallibilism in detail by examining one classic work, *Method in Social Science* by Andrew Sayer (1992 [1984]), and one recent book by a prominent realist, *The Reality of Social Construction*, by Dave Elder-Vass (2012). Once I have critically appraised these approaches I will conclude the article by addressing the question of transformational fallibilism's 'realism' or otherwise.<sup>2</sup>

## **2. Transformational Fallibilism and Improving Networks of Understanding**

The argument I want to defend and explore in this section is the claim that inquirers *should* treat all elements of their understanding, including theoretical and factual claims, as

potentially requiring revision in order to take their inquiry forward. Such an approach is fallibilist, but goes beyond a limited adherence to this doctrine insofar as it does not treat the idea that a claim may be untrue as simply a philosophical nicety that an inquirer can acknowledge whilst retaining great confidence in the truth of certain beliefs for the time being. Rather, it argues that in order to develop knowledge, even apparently obvious and hard-to-question beliefs should have their limitations explored. The approach advocated here thus has affinities with Popperian approaches in which fallibility is seen as a fundamental feature of knowledge (see e.g. Popper, 1959). This approach, that I have termed ‘transformational fallibilism’, argues that in pursuing an inquiry the investigator should be willing to subject any element to questioning and ultimately revision or reconstruction in order to improve their network of understandings. This statement immediately raises a whole set of issues about the meaning of its key terms, and explaining these will be part of how I develop and justify my account.

Starting with the notion of inquiry, I am using this in a broad sense to indicate an investigation in which a knower/knowers are trying to better understand some subject matter. This can include, but is not restricted to, attempts to describe it, characterise its properties, explain features of it, relate it to other phenomena, comprehend it, and so on. Next I want to comment on why I use the term ‘understandings’ and what it means to refer to them as a ‘network’. I am referring to ‘understandings’ to indicate a pluralistic orientation to the range of aspects that could be said to be part of inquirers’ attempts to know. This can include beliefs, concepts, theories, factual claims, models, exemplary problem solutions, and so on. In this article I will often be discussing theories and factual claims, but my suggestion is that the same arguments apply to all of the elements of the understanding mentioned above. In relation to the term ‘network’, I use this in order to emphasize the interconnected character of the elements of the understanding. As I argue further below, such elements are not “atoms” with independent identities that happen to be collected together, but are crucially connected with one another in ways which impact upon their meaning and import. In conceptualizing understanding this way I am drawing on the insights of post-positivist philosophers of science such as Quine (1951) and Kuhn (1970 [1962]) who, in different ways, emphasized the importance of the interconnection of understandings.

The last of these definitional points brings me to the deeper question of what it means for an inquirer to improve their understandings. I admit there are many possible ways of analysing such a notion, and also that the very idea of inquirers improving understanding in the first place might be questioned by those of a skeptical bent (cf. Woolgar, 1988). What I am hoping to do here is put forward an idea of improvement that would have at least some elements that would be accepted by many of those who are not committed to skepticism. This proposal is not intended as a complete account of what it means for an inquirer to improve their understanding, but as one part of a wider picture. As such, it would be more precise to refer to it as ‘problem-solving improvement’. Problem-solving improvement, as I am accounting for it here, occurs when an inquirer reasons through anomalies and contradictions within their existing network in a way that increases the coherence of their understandings. The basic idea is that if an inquirer can reasonably address contradictions

and anomalies in their network of understanding in a way that increases the coherence of that network, this can legitimately be seen as an improvement.

The element of this definition that would straightforwardly find sympathy with defenders of a range of approaches is the focus on increasing coherence and the removal of anomalies (inconsistent elements of the understanding).<sup>3</sup> Most obviously this is the case with the approach of Popper and his followers, who argue that theoretical systems are problematic where they have anomalies, and argue for the systematic address of these (Popper, 1959; Lakatos, 1978). However, the value of coherence is recognized in other approaches. For example, defenders of a critical realist approach object to contradictory theories on the basis that the world itself cannot both 'be' and 'not be' in some particular state (cf. Elder-Vass, 2012). Many defenders of a pragmatist view of knowledge would also see contradictory theories as problematic, on the basis that they are disabling in terms of the guidance they supply for our practical actions (cf. Garrison, 1994).

Nevertheless, it is an important part of my argument that increasing the coherence of a network of understanding is not sufficient to generate problem-solving improvement. As writers like Barnes (1982) have noted, it is possible to increase the proportion of one's beliefs that are coherent by the simple expedient of abandoning anomalous beliefs. Given that I have suggested that inquiry involves activities such as describing, characterising and/or explaining the features of some subject matter, it would be an odd account of improvement in this area which would see these as straightforwardly achievable by abandoning beliefs about that subject matter.<sup>4</sup> This is why I have required that in working towards coherence inquirers must be attempting to 'reason through' problems with existing belief. To simply abandon beliefs is not to engage with why those beliefs were held in the first place and what the ramifications of abandoning them will be. Just dropping beliefs also doesn't engage with why the contradiction arose in the first place. Putting the argument more positively, problem-solving improvement results from an attempt to puzzle out why anomalies/contradictions exist in one's understanding and what their significance is, given the subject matter they pertain to. On this view, anomalies are identifying something that the inquirer needs to attend to. What is needed is to puzzle through why the anomalies have arisen and how reconstructing existing understandings to resolve them can allow the inquirer to better understand their subject matter. Let me try to illustrate this idea with two examples, one from everyday life and one from the history of scientific investigation.

Imagine that I leave a biscuit out on the bench one night and discover in the morning that it has gone. I ask my daughter about this and she says that she did not eat it. I cannot think of another explanation for why the biscuit has disappeared and so decide that she must be lying about it. However, my network of understandings includes a belief that my daughter is an honest person and does not tell lies about everyday situations. In such a case I have an anomaly to face, in that my explanation of the disappearance of the biscuit entails that my daughter is lying, yet I believe her to be honest. Now, if increasing the coherence of belief was all that was required to produce a problem-solving improvement, I could achieve this simply by dropping the belief that my daughter ate the biscuit. However, to my mind this would be unsatisfactory, and would not constitute an improvement, because it would not

address the reason the anomaly came up in the first place: that I was engaged in an inquiry oriented to explaining the disappearance of the biscuit. To actually improve my understanding, a more positive response is required, one that explores further the basis of the anomaly. In this case, different avenues could be explored to try and achieve this, including: (1) re-assessing whether I am right to believe that my daughter is honest through e.g. a reconsideration of previous events; and (2) looking for another explanation of the disappearance of the biscuit. Taking the latter tack, I may inspect further and find little droppings on the bench, and take these as evidence that mice have been at work. In such a situation I can attribute the disappearance of the biscuit to the mice and replace my initial attribution of responsibility with a plausible alternative. This removes the incompatibility between my belief in my daughter's honesty, my daughter's statement on the matter, and my account of the disappearance of the biscuit. Furthermore, it does so in a way which uses the stimulus of the anomaly as a motivation to reason through the situation, increase the coherence of my understandings and thus improve them.

Let's take another example now, this time from the history of science. Hanson (1962) notes that it was observed in the 19<sup>th</sup> century that the orbit of Uranus had odd perturbations that were anomalous with other aspects of the network of understandings shared by many 19<sup>th</sup> century scientists. This network included Newtonian gravitational theories, theories of mechanics, and understandings of the position and gravitational influence of the known planets. The difficulty was that observations of Uranus's position did not cohere with other aspects of the understanding. In order to resolve this problem, Leverrier (and others) postulated that there must be a planet beyond Uranus, the gravitational influence of which explained the oddities of Uranus's orbit. Observations were made and it was established that there was a further planet beyond Uranus, which was named Neptune. This improved the coherence of existing knowledge, as the orbit of Uranus was now compatible again with the extant gravitational and mechanical theories, and also incorporated Neptune into the network. Furthermore, it did so by reasoning through the sources of the anomaly rather than, for example, discarding the observations of the orbit of Uranus without giving any reason for doing so. This also exemplifies the more general point that to move towards a problem-solving improvement one needs to be concerned not merely that an anomaly exists but with why it exists, and resolving this issue is part of moving towards coherence.

Having discussed the idea that reasoning through anomalies in order to increase the coherence of a network of understandings is one way for inquirers to improve their knowledge, this now needs to be linked to the question of transformation. The connection is the important idea that in order to improve networks of understanding by positively addressing anomalies, we may need to transform even currently successful understandings.<sup>5</sup> It is true that some revisions may involve altering beliefs that were only tentatively held anyway, and thus are relatively untroublesome to alter. However there are also cases where an anomaly can only be dealt with by reconstructing or rejecting theoretical ideas that have been successful.

What counts as 'success' in this sense is a difficult question which has been extensively discussed by philosophers (for just two of many possible examples see Van Fraassen, 1980;

Rorty, 1991). I am not intending to contribute to these debates so much as give a relatively permissive account which captures a key element for the purposes of this argument. This is that success refers to the belief of inquirers that some theory (or other element of the understanding) is supported by claims that they consider to be in some way a test of that theory – by being relatively ‘empirical’, ‘practical’, or ‘observational’. To return to an earlier example, those investigating astronomical phenomena considered the observations of Uranus to be a success for Newtonian theories, one such success among many. It is this kind of success which often encourages inquirers to hold very strongly to theories that are backed up in this way. Nevertheless, it is central to the account developed here that even successful theories may need to be reconstructed or rejected.

This account of the fallibility of successful theories is most consistently expounded in the Popperian tradition, with Popper (1959), and those inspired by him such as Lakatos (1978) and Larry Laudan (1981), defending the notion that no scientific theory is immune from rejection or transformation as inquiry progresses. Laudan, in particular, is known for exploring the defeasibility of successful theories, with his classic article ‘A Confutation of Convergent Realism’ (1981) identifying a range of theories that were empirically successful but ultimately rejected or fundamentally reconstructed, including the humoral theory of medicine, the phlogiston theory of chemistry, and the caloric theory of heat. One of the important messages of Laudan’s article is that it is crucial not to conflate success with truth.<sup>6</sup> It is worth noting, though that there are other intellectual resources which have been used to come to similar conclusions. For example, John Holmwood has drawn on Parsons’ (1949) [1937] epistemological ideas to argue that in order to deal with ‘residual’ categories, those not integrated into the theoretical system, the system itself may need to be reconstructed (see Chapter 3 of Holmwood, 1996).

A classic example of the fallibility of successful theories in the physical sciences is the reconstruction of the concepts of gravity, space and time necessitated by Einstein’s Theories of Special and General Relativity. Although there are questions about exactly which scientific findings Einstein considered problematic enough for Newtonianism that he was motivated to develop an alternative (see Holton, 1969; Van Dongen, 2009), nevertheless a key aspect of the strength and appeal of Einstein’s work was due to its ability to deal with phenomena such as the advancing perihelion of Mercury (see Hanson, 1962) and the results of the Michelson-Morley experiment, both of which were anomalous to Newtonian theory, but could be coherently accounted for within Einstein’s approach (Hentschel, 1992). In order for this development to occur, Newton’s hugely successful analysis of gravity had to be crucially reconstructed such that gravity was no longer conceived of as a force that produces instantaneous action at a distance from one body of mass to another but as a force emanating from an object with mass that bends space and time through the operation of waves (Will, 1988). Likewise, apparently unproblematic conceptions of space and time as absolute were thrown into question such that, for example, the question of whether two events have occurred simultaneously was no longer seen as having a single, straightforward answer, this answer instead coming to vary depending on the framework which one was inhabiting (Hesse, 1974)



To give an example from social science, let me turn to debates about which forms of family organization are best for the children being raised in them.<sup>7</sup> This is, admittedly, a politically contentious area which ties into wider social and political debates about the appropriate roles of men and women in society. From the range of social scientific research in this field, I want to look at a particular segment of debate in which claims from one participant can be judged to be successful in terms of their consistency with some supporting data but have been shown by later research to be somewhat misleading. This later research is itself part of an on-going investigative dynamic.

The starting point for this example is the claim that fathers contribute something importantly distinctive to parenting such that the level of well-being of a child is typically improved by the involvement of their father. The sociologist David Popenoe has been strongly committed to this viewpoint, and he claims that, when it comes to parenting, 'fathers – men – bring an array of unique and irreplaceable qualities that women do not ordinarily bring' (Popenoe, 1996a: 19). Popenoe discusses various aspects of this, and to make this discussion manageable I want to focus on two specific claims. The first is that fathers are important because they contribute positively to behavioural outcomes by (i) promoting self-control; and (ii) providing a role model to teenage boys who, without a father figure, are 'notoriously prone to trouble' (Popenoe, 1996a: 19-20). The second is that fathers' involvement with children contributes positively to the latter's academic/cognitive skills and achievements (Popenoe, 1996a: 21). For Popenoe, data from a range of studies support these claims because they show that children in two-parent families where the biological father and mother are present typically have more positive outcomes in these areas than children who are brought up solely by their mother (Popenoe, 1996b: 145-8).

Popenoe's argument has some degree of success (as defined above) insofar as the idea that fathers have a distinctive and positive contribution is consistent with some of the studies that he cites.<sup>8</sup> The studies of children that Popenoe discusses do, at least to some extent, point to positive features of the involvement of fathers in child-rearing when it comes to the behaviour and academic/cognitive skills of children. However, as I have been emphasising, theories that are successful may still need to be importantly reconfigured and reconstructed. And this reconstruction has occurred in the analysis of fatherhood where debates and further research in the area have yielded evidence and analyses which provide strong reasons to reconstruct Popenoe's explanation of outcomes. Other writers have argued that it is misleading to associate the kinds of differences in outcome that Popenoe has pointed to with a distinctive value added by the involvement of the father (Biblarz and Stacey, 2010; cf. Anderssen et al, 2002). One key way that this has been questioned is by exploring whether there are differences in outcomes between children in families parented by the mother and father, and those in families parented by two mothers. This seems a more convincing way to analyse the role of fathers because, as Biblarz and Stacey (2010) point out, it avoids conflating number of parents with the input of a father, unlike Popenoe's analysis. That is to say, Popenoe is often comparing outcomes of a two-parent arrangement where the children are raised by the mother and father with outcomes of a one-parent arrangement where the children are raised by the mother. Biblarz and Stacey avoid this problem by comparing outcomes in two-parent families where the difference is

between a mother-father and a mother-mother pair. They do so by conducting a meta-analysis of a range of existing studies that have compared outcomes for children with lesbian and children with heterosexual parents. One important part of their conclusion is that significant differences have rarely been found – for every study that identified a significant difference in outcome, there were at least four others that did not (Biblarz and Stacey, 2010: 8; see also Tasker, 2010). In relation to academic/cognitive skills, those studies that did find a difference identified a higher level of interest, effort and/or success in school from children brought up by two mothers rather than in mother/father parenting arrangements (Biblarz and Stacey, 2010: 8). In relation to behaviour, only one study found a difference, and this, admittedly, found that teachers identified children with lesbian parents as being more likely to have an issue in this area (see Vanfraussen et al, 2002). However, even here, this was only in relation to one aspect of behaviour – attention. In all other aspects of behaviour, including those most relevant to Popenoe’s claim – aggressiveness and delinquency – this study was consistent with others in finding no difference.

In my view, the findings of this meta-analysis present serious anomalies for Popenoe’s argument that fathers bring ‘unique and irreplaceable qualities’ that tame the troublesome behaviour of children and promote academic achievement. The meta-analysis suggests that there is very little difference between children parented by two mothers and by a mother and father in these areas. In the face of these anomalies, I would argue that Popenoe’s argument needed to be reconstructed, despite its initial success in finding support from some studies. Instead of claiming that being parented by a father as well as a mother makes a distinctive contribution to positive outcomes in relation to children’s academic/cognitive achievement and behaviour, it is more plausible to argue that having two parents makes a positive contribution in this respect. Note that to argue this is absolutely not to *blame* single-mothers for the apparent (average) outcomes for their children, as many factors to do with gendered economic inequality, stigma and problematic gender relations more generally surely contribute to these (see Silverstein and Auerbach, 1999).

Of course, it would be inconsistent of me to argue that the success of recent research establishes definitively that two-parent families are best for the wellbeing of children. Ongoing research examines the validity of existing findings, challenges the conceptualization of wellbeing used, and investigates how other configurations, such as those involving more than two parents, might affect the children, all of which will have implications for claims about the role of fathers.

I see these examples, one from natural science and one from social science, as illustrating the importance of reconstructing successful theories in the face of anomalous evidence. Returning from these examples to the general analysis, it is pertinent to explore exactly why it is the case that even very successful understandings may need to be reconstructed in order to deal with anomalies. A key aspect of this is that elements of a network of understandings are connected to one another. That is to say, both the coherent categories and the anomalies are linked together insofar as they are all part of the set of understandings that are being used to analyse the world. Developing on from this, the key point is that the presence of anomalies may be bound up with the coherent categories, as

Holmwood and Stewart have argued (Holmwood and Stewart, 1991). This is because it may well be the character of the coherent categories – the way they divide up and analyse their subject matter – that both contributes to their ability to coherently account for some phenomena and produces the existence of anomalies (that which they cannot coherently account for). Thus, the Newtonian conceptions of gravity, space and time allowed this theory to coherently account for many aspects of planetary behaviour as comprehended within the observations of its era. However, it was the very fact that the Newtonian approach comprehended gravity, space and time in this way which also meant that it could not coherently account for phenomena such as aspects of the positioning of Mercury. This is one example of the approach to fallibility advocated here: the idea that any aspect of a set of understandings, no matter how apparently coherent and successful, may need to be transformed in order to improve our knowledge.

Within the social sciences, the possibility that even successful understandings may need to be reconstructed to improve understanding has not always been recognized by those interested in reflecting on trajectories of inquiry. For example, James Rule (1997) argues that it is a problematic feature of sociology that it has not accumulated reliable insights over time, tending instead to change its viewpoint at regular intervals. Rule's position is that it is accumulation rather than reconstruction which social sciences should be engaging in. From the perspective that I am developing here, a lack of accumulation of substantive insights isn't problematic in itself, as there is an expectation that theories and factual claims will *need* to change at some point to take inquiry forward. What is more important from the standpoint of progress is whether plausible justifications can be given for preferring later positions over earlier ones, and I would argue, without being able to put a case forward here, that these may be given even where substantial transformations have taken place.<sup>9</sup>

Having outlined the approach to fallibilism that I wish to defend, I now want to compare it with that taken by critical realists. I will provide some background on Critical Realism first, and then look in greater depth at the arguments of Sayer and Elder-Vass.

### **3. Critical Realism: Background**

Over the past 40 or so years critical realism has developed into a wide-ranging philosophical approach that analyses many phenomena, from religiosity and ethics to the character of 'race' and gender discrimination (e.g. Porpora, 2005; Carter and Virdee, 2008; New, 2005). Given the focus on inquiry in this article, I will be concerned with critical realism's influential philosophical analyses of natural and social science, and my aim in this section is to briefly introduce three elements: the notion of ontology, the critical realist approach to inquiry, and its treatment of epistemological relativism. I will discuss these with particular reference to the work of Roy Bhaskar, who is usually seen as the founder of critical realism (for a more extended overview of critical realism see Collier, 1994).

A central feature of critical realism is its strong emphasis on ontology, that is, on philosophical reflection about the basic characteristics of entities that are investigated by the social and natural sciences. Critical realists such as Bhaskar argue that ontological

reflection is crucial in order to provide a convincing philosophical analysis of these activities. Bhaskar also contends that it can make an 'under-labouring' contribution to activities of the sciences by removing conceptual confusions and obstacles on the path to successful investigation (Bhaskar, 1997). Arguably, some critical realist approaches go beyond that in order to propose extensive ontological frameworks which are seen as providing the appropriate starting point for empirical research (e.g. Archer, 1995). Whatever the purpose it is recommended for, one of the key ontological ideas for critical realists is that of 'structure'. Rather than seeing the world as constituted by patterns of events, or as intrinsically formless, critical realists argue that the world is made up of structured entities with causal powers, the interplay of which produces the events that occur in the world (Bhaskar, 1997). The two further aspects of critical realism that I want to introduce here are the idea that inquiry is oriented to grasping the real structural properties of what is investigated, and the contention that critical realists should adopt epistemological relativism.

From a critical realist perspective, pure natural science is about the identification of the structures of the natural world and an investigation of their powers (mechanisms). Bhaskar outlines the process of investigation (logic of scientific discovery) in *A Realist Theory of Science*, which was originally published in 1975, though I will refer to the second edition from 1997. For Bhaskar, the logic of scientific discovery has three phases (see Ch 3 of Bhaskar, 1997). The first phase starts from a regularity of events. This may be discovered in nature, but is, for Bhaskar, more likely to be produced within an experimentally controlled situation. The second phase involves 'creative model-building' in which theories about the structures and powers that might generate such a regularity are put forward (Bhaskar, 1997: 145). In the third phase theories are put to the test in order to establish which structure and mechanism actually does account for the regularity in question and is thus 'not imaginary but real' (Bhaskar, 1997: 146). Bhaskar's subsequent work on the philosophy of social science argued that there are relevant differences between social structures and natural structures, as well as contending that human agents play an important role in producing social life. Nevertheless, the orientation to identifying the real properties of these distinctively social entities was retained in Bhaskar's thought, as well as in that of other critical realists (see Bhaskar, 1986; Archer, 1995).

Moving onto the epistemic dimension of critical realism, from the terminology that Bhaskar sometimes uses in *A Realist Theory of Science* it might be assumed that he straightforwardly believes that science does discover the real, and thus that he sees successful theories as true. However, Bhaskar also introduced epistemological arguments that point in a different direction, and have been influential on other critical realists. To see this, we first need to consider his distinction between the transitive and intransitive dimensions of science. For Bhaskar, the transitive is the domain of historically located theories and practices which is very much a social product and, as its name suggests, changes over time. By contrast, the intransitive is the domain of events and objects that are unaltered by changes in scientific understanding of them (Bhaskar, 1997: 21-4). This distinction provides a useful basis for Bhaskar's argument that realists should accept 'epistemological relativism', and reject the 'correspondence theory of truth' (Bhaskar, 1997 : 249). His reason for making these claims

is the familiar post-Kantian one that we cannot compare propositions – transitive elements of science - with states of affairs in the world – the intransitive objects of knowledge. Bhaskar states that, as such, '[t]here is no way in which we can look at the world and then at a sentence and ask whether they fit' (Bhaskar, 1997: 249). Although the intransitive elements of the world do not change as a result of conceptual developments, transitive theories are fallible and can change over time. As Bhaskar has made clear elsewhere, he does not take this to mean that we cannot give good grounds for preferring one theory to another (1986: 72). Thus, although Bhaskar subscribes to epistemological relativism he does not subscribe to 'judgemental relativism'.

From this characterisation we can see that there are arguably two different tendencies in critical realism as developed by Bhaskar, one which sees inquiry as burrowing towards the real, and another which eschews the idea of correspondence and emphasizes the fallibility and changeability of our theories (see for discussion Fay, 1990). What I want to suggest now is that related tendencies can be found in other critical realist work, and I will be arguing that this results in some critical realist arguments being incompatible with the transformational fallibility that I have been presenting as crucial if inquiry is going to improve by means of problem-solving.

#### **4. Practical Success and Preservation: The Work of Andrew Sayer**

Andrew Sayer's *Method in Social Science: A Realist Approach* (1992) [1984] is a highly cited book, and it is not hard to see why. It is clearly written, and is addressed to the social scientific inquirer as much as the philosopher, without this compromising its careful argumentation. There are also elements of Sayer's account that chime in with an orientation to inquiry as a transformative process. Nevertheless, what I want argue in this section is that there are aspects of Sayer's approach which unintentionally retain a conservative orientation to inquiry, and which thus limit his embrace of the more radical consequences of fallibilism.

Sayer's core arguments have close connections to the aspects of realism presented so far. Sayer argues that we must make a clear distinction between 'factual knowledge' and 'facts as things or states of the world' (Sayer, 1992: 47). Putting this another way, Sayer argues that we must separate out 'Thought Objects' from 'Real Objects' (Sayer, 1992: 47). This is in line with Bhaskar's insistence on making a distinction between the transitive dimension of knowledge, which changes over time as theories change, and the intransitive dimension, which does not change when scientific theories do. Another idea that is very much related to Bhaskar's work is Sayer's more general insistence that we need to see the world scientists investigate as 'structured' and 'differentiated'. Sayer argues that this ontological claim about the character of the world is preferable to the conventionalist view that the world has 'a structureless, entirely malleable character' (Sayer, 1992: 70).

There are aspects of Sayer's approach where his orientation to inquiry is close to the approach to fallibility that I am defending. For example, Sayer loosens the hold of 'facts' as something that cannot be misguided and fallible by arguing that it is wrong to make a clear

differentiation between facts and theories such that facts are seen as having an ‘undeniable reality’ and theories are seen as ‘speculative’ (Sayer, 1992: 48). Rather, Sayer argues that we should speak of ‘factual statements’ rather than facts, and he contends that both factual statements and theories are ‘thought objects’ (Sayer, 1992: 52). This means, for example, that the data that we gather through observation or other means are always preconceptualized. Sayer develops this notion in critique of others who might fail to recognize this:

‘Social scientists who treat “data” literally as “given things” (often those who feel most confident about the objectivity of their knowledge and the “hardness” of their facts) therefore unknowingly take on board and reproduce the interpretations implicit in the data: they think *with* these hidden concepts but not *about* them.’ (Sayer, 1992: 52. Emphasis in original)

In this quote Sayer nicely encapsulates a key fallibilist point: that inquirers need to be aware that their data are always interpreted by them. As such, it is misleading to treat ‘facts’ – our factual claims – as having a distinctive hardness that makes them resistant to transformation.

Sayer addresses fallibilism more directly when discussing the notion of absolute truth, stating that:

‘Strictly speaking, then, we can never justifiably claim to have discovered the *absolute* truth about matters of fact, or to have established some absolute foundation for our knowledge (‘foundationalism’). Our knowledge must be admitted to be *fallible*.’ (Sayer, 1992: 67. Emphasis in original)

Although he rejects the notion of absolute truth, Sayer does not reject the idea of assessing the justification of beliefs altogether, and like Bhaskar he criticizes ‘judgemental relativism’ (see Sayer, 2000a: 47-8). Sayer argues that instead of focusing on absolute truth we should assess the ‘practical adequacy’ of beliefs. Sayer states that:

‘To be practically adequate, knowledge must generate expectations about the world and about the results of our actions which are actually realized.’ (Sayer, 1992: 69)

According to Sayer, then, theories can be distinguished by their ability to cope with the world. Sayer argues that this is a criterion implicitly accepted by ‘idealists’ in their everyday actions. Though idealists might explicitly argue that the world is how we construct it, ‘they do not generally try to leave rooms through the ceiling rather than the door’ (Sayer, 1992: 70). That is to say, even idealists pay attention to the likelihood that their actions will or will not be successfully realized; and in doing so, they are deciding to act on practically adequate beliefs (that a door makes a good exit point) rather than impractical beliefs (that one can float through the ceiling).

Having pointed out positive features of Sayer’s approach to fallibility, I now want to argue that there is nevertheless an implicit conservatism in the way that Sayer relates the practical adequacy of a theory to the state of the world to which the theory applies. To see this, let’s

consider Sayer's discussion of what he characterises as two competing 'conventions'. It is worth quoting Sayer at length:

'The reason that the "convention" 1 that we cannot walk on water is preferred to the convention 2 that we can, is because the expectations arising from 1, but not 2, are realized. They are realized because of the nature of the associated material interventions (trying to walk on water) and of their material contexts. In other words, although the nature of objects and processes (including human behaviour) does not uniquely determine the content of human knowledge, it does determine their cognitive and practical possibilities for us. It is not thanks to our knowledge that walking on water doesn't work, but rather that the nature of water makes 1 more practically adequate than 2. The fact that 1 is nevertheless still, in principle, fallible, needn't alter our preference for it over 2.' (Sayer, 1992: 70)

What is problematic from the perspective of transformative fallibilism developed in this article is the way in which the 'nature' of the object being related to is invoked as an explanation for the success of a particular theoretical claim (convention). The success of a theory is being explained with reference to the nature of the world, as in Sayer's statement that 'It is not thanks to our knowledge that walking on water doesn't work, but rather that the nature of water makes 1 more practically adequate than 2'.<sup>10</sup> The problem with this inference is that it rests on a static and present-practice-centred view of the possibilities. Despite having suggested that convention 1 is potentially fallible, Sayer still assumes that its present practical success indicates something about the nature of water rather than about our current state of knowledge. But I would argue that this is to reify present practical success and use it to make a questionable inference about the nature of things.

Let me develop this criticism further. What I want to contend is that it would only be if the claim 'we cannot walk on water' is unshakeable that Sayer could infer something about the nature of water from it. But to argue this is to reify our current state of knowledge and capabilities as if these could not change. An initial way to question Sayer's example would be to point out that we can indeed walk on water – when it is frozen to a certain level of thickness. This might seem a flip response, but let us consider that if we lived at a low altitude in an equatorial climate zone without electrical technologies we might have confidently asserted that water could not be walked on, and been very surprised to discover that it could when it solidified into ice. That would have been an unidentified possibility for us. In response to this, Sayer could refine his convention to argue that we cannot walk on water in its fluid state, but what I want to suggest nevertheless is that it is possible that humans might find some way to walk on fluid water, one which drew on hitherto unidentified practical possibilities.

Let me give a different example from scientific history to try to make this point seem less speculative.<sup>11</sup> An Andrew Sayer from the 1850s might have said the following: "The reason that the "convention" 1 that we cannot generate vast amounts of energy from a few grams of matter is preferred to the convention 2 that we can, is because the expectations arising from 1, but not 2, are realized. It is not thanks to our knowledge that generating vast amounts of energy from a few grams of material doesn't work, but rather that the nature of

matter makes 1 more practically adequate than 2.” 1850s Sayer would have been quite right about the practical success of his beliefs at the time, but quite wrong to infer that this was due to the nature of matter. Later work by Einstein and others meant that it is possible to extract vast amount of energy out of a few grams of matter. Contrary to Sayer, what is practically possible is related to our state of knowledge. To argue otherwise is to imply that our current practical capabilities do not simply represent our current levels of knowledge and practical skill but tell us something definitive about the nature of the world.

The defender of Sayer might want to reply the following: as you admit, Sayer points out that the convention that we cannot walk on water is fallible – so how can he rightly be accused of reifying this claim? My response to this would be in keeping with my general treatment of critical realism: the point is not that critical realists never make arguments which are consistent with the more radical form of fallibilism defended here. Rather, it is that they also make arguments which reify certain claims about the world. If Sayer’s views were completely consonant with transformational fallibilism he could not make the case that it is the nature of water that produces the practical adequacy of the relevant convention. For Sayer to then add that convention 1 is fallible ‘in principle’ is either to use the idea of fallibility in a restricted fashion, treating it as a possibility that is nevertheless essentially inconsequential, or to immediately undermine his prior assertion. That is to say, if Sayer did take the fallibility of convention 1 seriously he could not assert that it is the nature of water that makes it work rather than our state of knowledge, because making this connection between convention and nature undermines the idea that the convention may need to be transformed in the future.

Let me sum up the discussion in this section. I have argued that Sayer does at times acknowledge the fallibility of scientific knowledge in his work. Given that this is the case, it is striking that he nevertheless wishes to infer something about the nature of the world on the basis of successful theories. This displays a tendency to reify the understandings of current successful techno-scientific knowledge and project it out onto the external world with the implication that successful knowledge is underwritten by its adequate relation to the nature of objects. From the perspective of this article, one key problematic outcome of this view is that, when taken seriously, it can undermine any sense that current scientific theories and practices are in need of transformation.<sup>12</sup> The implication of Sayer’s critical realist view is that if there is some practical result that we cannot achieve this must be because of the nature of things, rather than the limitations of our current practice.<sup>13</sup>

## **5. Epistemological Standards and External Perspective: The Work of Elder-Vass**

The other critical realist approach that I want to consider is that of Elder-Vass. Elder-Vass has quickly become an influential figure within critical realism, with his careful work on the notion of emergence and its relevance for understanding structure and agency stirring up debate between critical realists as well as challenging non-realist approaches (see for example Elder-Vass, 2007; King, 2007). More relevant for our concerns here, though, is Elder-Vass’s recent book on *The Reality of Social Construction* (2012). In this work, Elder-



Vass offers a sympathetic critique of social constructionist ideas, arguing that aspects of a constructionist approach are compatible with critical realist analysis whilst also criticising other aspects of constructionist thought. In order to understand the relationship of his theories to the question of fallibility we need some consideration of Elder-Vass's analysis of knowledge, which is most clearly articulated in in Chapter 11 of *The Reality of Social Construction*.

A key aspect of Elder-Vass's orientation is his critical relation to the analytical tradition of epistemology which he sees as attempting to 'justify some specific criterion of true knowledge from first principles' (Elder-Vass, 2012: 209). Elder-Vass argues instead that knowledge should be judged by '*actual* standards that exist independently of philosophers' and these are what we should be concerned with (Elder-Vass, 2012: 209. Emphasis in original). The classic view of knowledge within analytic philosophy is that knowledge is 'justified true belief',<sup>14</sup> and, whilst Elder-Vass accepts that knowledge is about belief, he argues that we should not associate knowledge with truth for the following reason:

'We can never know with certainty that any given knowledge claim is true (though we may be very confident that it is) and hence to identify knowledge with true belief would mean that we could never know whether any given claim was knowledge or not.' (Elder-Vass, 2012: 210)

Underlining this point, Elder-Vass argues that it is 'impossible for us to apply in practice' the standard that knowledge is true belief, and thus rejects the requirement that to be knowledge a belief should be true (Elder-Vass, 2012: 210). However, he does retain the idea that to be knowledge a belief should be considered to be justified. And he gives this justification a social character insofar as he argues that beliefs can be considered justified, and thus as knowledge, when they are formed in line with '*socially authorised practices*' of the appropriate kind (Elder-Vass, 2012: 214).

In order to account for social authorisation, Elder-Vass applies his notion of 'norm circles'. For Elder-Vass 'a norm circle is the group of people who are committed to endorsing and enforcing a particular norm' (Elder-Vass, 2012: 22). In relation to the authorisation of knowledge, Elder-Vass argues that there are two relevant sorts of circles that contribute to a belief becoming knowledge. One of these is 'epistemological circles', the members of which uphold a view about which criteria must be met for a belief to count as knowledge. As an example, Elder-Vass suggests that members of the Azande act as an epistemological circle who enforce the idea that certain oracular practices are a means for obtaining knowledge about who is engaging in witchcraft within the social group (Elder-Vass, 2012: 218). What epistemological circles do is thus endorse a certain way of producing knowledge as valid. Complementing these are 'epistemic circles' which Elder-Vass explores in relation to active, currently-debated scientific knowledge claims. Whereas epistemological circles ratify a certain kind of procedure, epistemic circles authorize specific claims. Elder-Vass suggests that in some scientific fields, there may be a great deal of agreement on which claims to endorse, whereas in others there may be different groups who endorse competing, incompatible claims (Elder-Vass, 2012: 227). Alongside epistemological and epistemic

circles, Elder-Vass argues that there can be other kinds of social influences, such as views about gender, which have shaped scientific knowledge production (Elder-Vass, 2012: 228).

Elder-Vass characterizes his approach as an 'ontological' one, but he is well aware that analyses of knowledge have epistemological consequences, and we can turn to explore these now. In considering these issues Elder-Vass argues that we need to be cautious of broad terms like 'relativism', and he offers a differentiated set of categories to account for his position. Elder-Vass argues that his approach supports 'knowledge relativism' insofar as it acknowledges that what is held to be knowledge may vary across time and place, and also that some beliefs that are held to be knowledge within particular epistemic circles will be false (Elder-Vass, 2012: 230). As part of this position, Elder-Vass acknowledges that one claim may conflict with another but both may nevertheless count as knowledge, and indeed '*be* knowledge' if each is upheld within a different epistemological/epistemic circle (Elder-Vass, 2012: 230-1). Although he upholds knowledge relativism, Elder-Vass wants to reject another form of relativism, relating to truth. Truth relativism involves the idea that incompatible descriptions of the world can all be true despite the fact that they conflict with one another. Elder-Vass rejects such a position, arguing instead that:

'...all true ways of describing the world would be consistent with each other and correspond to a single reality that exists independently of our descriptions of it.'  
(Elder-Vass, 2012: 231)

Thus Elder-Vass insists that true theories will all be compatible with one another.

As well as rejecting 'truth relativism', Elder-Vass rejects the relativist view that all standards are equally capable of warranting convincing knowledge claims. This becomes clear when Elder-Vass states that '*...not all epistemological standards are equal*' (Elder-Vass, 2012: 232. Emphasis in original). What is the difference between them? Elder-Vass states the following:

'Some epistemological standards are better than others in terms of the degree to which they tend to produce what Goldman calls veritistically reliable beliefs (1999)...That our epistemological standards are socially agreed and historically variable, then, does not entail that they are entirely arbitrary. They may *also* be well-suited to the job of generating veritistically reliable, or approximately true, knowledge.' (Elder-Vass, 2012: 232. Emphasis in original)

Furthermore, Elder-Vass argues that approximate truth plays an important role in assessing standards:

'...we can and should examine which epistemological standards are more conducive to generating reliable, approximately true knowledge, and we can often come to rationally justifiable conclusions in such evaluations.' (Elder-Vass, 2012: 233)

It is hard to see how this kind of evaluation can be undertaken without knowing which beliefs are approximately true. The procedure strongly implied in Elder-Vass's discussion is that we use standards to generate beliefs, see which of these are approximately true, and then use this as a means to praise or criticise the standards.<sup>15</sup>

Having outlined some of the key features of Elder-Vass's approach, let me now turn to assess it. Aspects of Elder-Vass's position are consistent with the approach to fallibility defended here. As we have seen, Elder-Vass rejects the association of knowledge with truth on the basis that to require a claim be established as true before it can be admitted to the sphere of knowledge is to place the bar impossibly high. This is consistent with the transformational fallibilist view that identifying beliefs as 'true' illegitimately reifies those beliefs, treating them as 'finalized' rather than 'in play'. Nevertheless, I also want to argue that Elder-Vass's adoption of 'approximate truth' as part of his account is a more problematic move.

I should begin with an admission that the question of 'approximate truth' in Elder-Vass's work is potentially difficult to analyse given that he does not, to my knowledge discuss the notion of 'approximate truth' further than in the quotations already cited. This means that there are unanswered questions, such as why it is that Elder-Vass thinks that the generation of approximately true knowledge is a viable epistemic goal when he did not see the pursuit of truth in the same light. Nevertheless, the issue I want to explore here is as follows. Elder-Vass's analysis relies on the idea that the approximate truth of belief can be established. What I want to argue is that the identification of beliefs as approximately true is problematic if we want, as inquirers, to keep all options on the table in terms of improving our knowledge.

To explore this we need to go further than Elder-Vass and fill out the conception of approximate truth. What I want to suggest here is that identifying a belief as approximately true seems to require a knowledge of the region or range in which the truth can be found. This is apparent, for example, in one of the serious analyses of the notion of approximate truth that can be found in the literature: that put forward by the scientific realist Stathis Psillos in his volume *Scientific Realism: How Science Tracks Truth* (1999)<sup>16</sup>. Psillos' general view is well summarised in this remark:

'...a theory is approximately true if the entities of the general kind postulated to play a central causal role in the theory exist, and if the basic mechanisms and laws postulated by the theory approximate those holding in the world, under specific conditions of approximation.' (Psillos, 1999: 277)

This analysis treats judgements of the 'approximate truth' of a theory as something that is established by comparing the account given by the theory with the truth about the mechanisms and laws holding in the world. To establish that a theory is approximately true, we have to know what is true.<sup>17</sup> Now, it is important to acknowledge that Elder-Vass does not comment on Psillos's notion of approximate truth and I am not suggesting that he is committed to Psillos's approach. Nevertheless, what is relevant about Psillos's approach for our discussion is that it illustrates the point I mentioned above, which is that identifying approximate truth seems to require a knowledge of the region the truth itself is in. Psillos's account is perhaps unusual in the precision required in the reference to 'specific conditions of approximation'. But even more everyday uses of 'approximate' seem to rely on a sense that one knows the range in which the matter being approximated can be found. Thus, if I measure the length of my desk by counting the number of times I can flip a ruler from one

end to the other, I could legitimately describe this as approximating the length of the desk, and I would have a sense of the likely parameters of the length even if I'm not confident I have pinned it down exactly.<sup>18</sup> Thus, it seems plausible to link the notion of 'approximate truth' to the idea that we can identify the parameters which the truth itself is in.

However, I would argue that the idea that we can know the parameters in which the truth resides is deeply questionable. This relies on the idea that the framework of theories and concepts through which we interact with some subject matter are essentially correct, and the only changes that will be required in order to improve our understanding will be small-scale improvements in precision. What this seems to rule out is that more serious revision and reconceptualization may be required in order to deal with existing problems of understanding. Putting this point in terms of usage, would it seem natural to say of a belief that it is 'approximately true but in need of serious revision?' I would argue that this would be a jarring utterance to make about a belief. As such, if we want to defend the idea that all beliefs should be considered open for reconceptualization in order to improve our understanding we will not designate them as approximately true. To be clear, once again, this has required some extrapolation beyond the explicit remarks made by Elder-Vass about approximate truth. Nevertheless, my sense is that these extrapolations do not unreasonably overstate the truth-commitments implied in his use of this term, based on: (i) my reference to an explicit analysis of approximate truth in the scientific realist tradition; (ii) connections with the everyday sense of approximation; (iii) Elder-Vass's own reference to Goldman, whose work shows a strong confidence in human ability to locate truth.

I think it is worth commenting here on why the temptation might arise to define certain aspects of our understanding – say theories - as approximately true. This seems likely to occur when these theories are very successful, and mesh well with factual claims that are widely accepted. One has a sense that the critical realist thinks that there is a degree of sophistry in questioning the approximate truth of a theory in such cases (cf. Porpora, 2004). And, indeed, Elder-Vass's arguments have an intriguing link to Sayer's when the former refers to the possibility of generating 'reliable, approximately true knowledge'. In both cases we can see the temptation to move from the success/reliability of a theory to make some further claim about it: in Sayer's analysis, that it has some connection with the nature of the object, in Elder-Vass's analysis, that it is approximately true.

What I would argue, by contrast, is that a theory can have undoubted and numerous successes, but it is important to avoid being mesmerised by these. The threads that will unravel established theories are often to be found elsewhere. Thus, instead of trying to explain why a theory is very successful by reference to its approximate truth, we should be continuing to explore the anomalies and problems which may ultimately lead to its reconstruction. Simply focusing on the achievements or reliability of Newtonian gravitational theory, for example, would have discouraged attention from the puzzles and anomalies that were challenges to it, some of which ultimately contributed to the need for it to be importantly reconstructed.

Am I overstating the case that a strand of critical realist analysis reifies successful theories alongside its otherwise useful commitment to fallibilism? It might be argued, after all, that

the formulations I have picked up on would be better seen as occasional slips in the way ideas are expressed rather than genuine theoretical commitments. I would contend that this is not the case. My argument is that there is a side of critical realism which, in its orientation to burrowing towards the 'real', and in its view that practical success reveals the character of the real (at least to some extent), leads to a restricted commitment to fallibilism. Let me give one last example of this, returning to the work of Bhaskar. Early in the pages of *A Realist Theory of Science*, Bhaskar states the following:

'We can easily imagine a world similar to ours, containing the same intransitive objects of scientific knowledge, but without any science to produce knowledge of them...In such a world the causal laws that science has now, as a matter of fact, discovered would presumably still prevail, and the kinds of things that science has identified endure. The tides would still turn and metals conduct electricity in the way that they do, without a Newton or a Drude to produce our knowledge of them. The Wiedemann-Franz law would continue to hold although there would be no-one to formulate, experimentally establish or deduce it. Two atoms of hydrogen would continue to combine with one atom of oxygen and in favourable circumstances osmosis would continue to occur.' (Bhaskar, 1997: 22)

Although Bhaskar is making an argument about the persistence of objects and causes independent of science he does so in a way which relies on the reification of present scientific understandings and the identification of them with the real. By referring to 'the same intransitive objects' existing in a world without science, and then proceeding to list what these are, Bhaskar is surely breaking down the barrier that he was attempting to build between the transitive and the intransitive.<sup>19</sup> If the fallibility of the Wiedemann-Franz law was taken seriously, Bhaskar could not say that it would hold whether or not anyone was there to formulate it. If, as I have suggested about knowledge in general, the law is considered to be 'up for grabs', and as potentially in need of reconstruction to improve our understanding, we cannot say that it will apply whether or not anyone grasps it. This is to project the law onto reality itself. Likewise with the statement that hydrogen and oxygen would continue to combine without our knowledge of them. This, again, is to treat our understanding of atoms and their interactions as if it correctly characterizes the world. Of course, one response to this would be to say: "Quite right, too! Who could doubt the atomic constitution of water?!" But two things can be said about such a response. First, those who are inclined to argue that way are tacitly giving up on a serious commitment to fallibilism and throwing their weight behind the view that science produces certainty, at least in relation to some of its claims. Second, one does not have to be a raving skeptic to believe that scientific classifications of atoms and of the constitution of water could change. For example, Hasok Chang (2012) gives an elegant defence of the ways in which the grip of current accounts of water and its atomic constitution could be loosened by conceptualizing the elements and their interactions in a somewhat different manner. The argument I am making is not that current characterisations of water and its atomic constitution should be rejected at the present time. Rather it is that, as with any other knowledge claim, they need to be considered to be potentially fallible so that they can be reconstructed if required in order to make problem-solving improvements in our understanding. Summing up, I see this

quote from Bhaskar's key realist work as providing a further example of the aspect of critical realist thought that conflates successful scientific theorizing and reality, resulting in problematic consequences for thinking about the reconstruction of theories.

## **6. Conclusion**

In this conclusion I would like to briefly address a philosophical question that may strike readers as arising from the discussion so far, and then try to pull together the threads of the arguments made in this article.

The philosophical issue I would like to address is whether the transformational fallibilism that I am defending is best understood as a realist, pragmatist, or constructionist approach to inquiry. I am not entirely comfortable with these categorisations but, if pushed, I would be inclined to characterise my position as a form of 'minimal realism'. My view is realist in the sense that I believe that inquiries can be oriented to investigating the properties of some subject matter rather than being directly concerned with improving the practical success of action (as is arguably the case in pragmatism) or in imposing one's perspective on others (as in some forms of constructionism). In my view, many claims made within investigations are attempts to 'get to grips' with a particular subject matter. However, my realism is 'minimal' because I do not hold that even the most successful scientific (or other kinds of) investigations get in touch with the truth about the real properties, structures, and characteristics of what we engage with through our network of understandings. Our best understandings are based in our own perspective and thoroughly defeasible and subject to reconstruction. My view is that to claim that a belief is true is to misleadingly separate out that belief from its interconnections with other beliefs and to protect it from further reconsideration and reconstruction. Given the apparent richness of the universe, the complex interconnectedness of our understandings, and the limitations of the perspective that we work from, the truth about our subject matter is not even asymptotically approachable or subject to successful approximation. As such, my view that our best understandings are those that currently work best for us in terms of our interactions with the world puts me very close to the border of realism and pragmatism.

The initial intention of this article was to defend the idea that there is still something valuable to be gained from reflecting on the concept of fallibility. To pursue this I have developed a conception termed transformational fallibility. The argument has been that the acknowledgement of possible fallibility should not just be considered a philosophical nicety with no substantial consequences for inquiry but, rather, that all elements of a network of understanding should be actively considered to be fallible and potentially subject to reconstruction so that the network can be improved by productively removing anomalies. Having provided a general analysis of this idea I then sought to show the distinctiveness of transformational fallibilism, and its advantages over other approaches, by contrasting it with the more limited acknowledgement of fallibility within the work of two important critical realists, Andrew Sayer and Dave Elder-Vass. In Sayer's work the limitation on fallibility appears with the argument that successful interactions with a specific subject matter tell us

something about the nature of that subject matter. In Elder-Vass's case, the limit arises out of his defence of the idea that we can identify the approximate truth of claims. Of course, simply identifying a difference in treatment between critical realism and my own approach would not be sufficient to establish the benefits of the latter. As such, I have also been arguing that these moves permit successful forms of inquiry to rest on their laurels, and to transmute their empirical/practical successes into something deeper and more permanent. What I am advocating instead is a strong commitment to impermanence. To continue to improve our understanding, and avoid getting snagged on current successes, we need to see inquiry as an ongoing process where all elements can be subject to criticism, and, over time, all can be expected to change.

## Endnotes

<sup>1</sup> This is my (fallible) recollection of his remarks.

<sup>2</sup> The arguments in this article might have been presented in a different way, such that what I call 'transformational fallibilism' here was argued to be the genuine content of any commitment to fallibilism. The argument would then be that others who claim to be fallibilists simply do not work through the implications of their commitment properly. This would have been a possible way of presenting the argument but given that it is common for adherents to fallibilism to hold a cautious, limited view of its consequences, it seemed clearer to me to identify the approach I am defending as one sub-category of fallibilism.

<sup>3</sup> In making this argument I am not subscribing to a coherence theory of 'truth', rather looking at the relevance of coherence to processes of inquiry.

<sup>4</sup> A similar concern is addressed by Popperians in their considerations about what kinds of developments are legitimate 'scientific' moves to make when a theory is faced with an anomaly and is thus falsified (e.g. Popper, 1959; Lakatos, 1978). One possible Popperian treatment of this is that the replacement theory should have greater 'content' than the theory which was falsified. I cannot go into the complexities of the issue here but suffice to say there has been much debate about this idea including what are, to my mind, telling criticisms of such an approach (see for example Holmwood and Stewart, 1991; Bamford, 1993). I should also note that I am not attempting to set-up an idea of scientificity but rather looking to elucidate one way in which a set of understandings can plausibly be said to have undergone an improvement.

<sup>5</sup> It's important to be clear that in arguing that any element of the understanding may require revision I do not mean that an inquirer should be prepared to change the totality of their understandings in one go, as it is hard to see how a rationale for such a change could be given. One of the limitations of Kuhn's analysis was his tendency to see competing approaches as self-enclosed and mutually exclusive totalities such that converting from one

---

to the other involved something like a complete conversion from one set of theoretical and factual understandings to another (see Demir, 2006).

<sup>6</sup> It is worth noting that some philosophers have worked to 'pick off' cases identified Laudan by questioning whether they were really 'successful' (see for example Leplin, 1997; Psillos, 1999). I cannot properly address their arguments here, but I would comment that the criteria for success that these later philosophers use are often both highly restrictive (excluding very much apparently 'scientific' knowledge down the ages from meeting these criteria, let alone any 'non-scientific' knowledge) and also at a questionable distance from the criteria that seem to be have used by the knowledge-producers at the time.

<sup>7</sup> Thanks to Sharani Osborn for suggesting this example.

<sup>8</sup> Popenoe cites a wide range of studies, and it is possible to question whether all of them support his claims. Nevertheless, in the areas I have identified, it seems to me that at least some of the studies he cites are consistent with his arguments.

<sup>9</sup> It might be wondered how the views defended here relate to Popper's doubts about the idea of justification and his questioning of the giving of 'positive evidence' for knowledge (e.g. Popper, 1959). I would say that the approach I am taking is broadly Popperian in that I agree that knowledge claims are not able to be justified in the sense Popper often focuses on. That is to say, Popper rejects the idea that claims can be shown to be true or highly probable because of positive evidence in their favour, and I follow him on this. A more complex issue is whether I might differ with Popper on the extent to which positive support of any kind can be given to a knowledge claim. Quite what Popper's views on this matter are have been discussed and disagreed upon in debates that address the related question of whether Popper is tacitly committed to induction in some form (for one important contribution see Lakatos, 1978). All I can do here is give the reader a flavour of my own view: that 'positive support' for a claim is not an intrinsically problematic notion, and indeed is part of reasonable argumentation, but that no matter how much 'positive support' a claim has, this does not make it unquestionable.

<sup>10</sup> In case it be thought that this is an anomalous formulation by Sayer, we can also consider his statement that 'the usefulness [of a theory] is not accidental but due to the nature of the objects of knowledge' (Sayer, 1992: 70). See also his criticism of instrumentalism which states that '...it is the structure of the world, rather than our theories about it that make practices possible or impossible.' (Sayer, 1992: 71)

<sup>11</sup> It might be considered illegitimate to draw parallels from Sayer's example, which deals with practical everyday knowledge, in order to extrapolate conclusions about his treatment of scientific claims. However, I would argue that this is not the case. I am not aware of Sayer marking any fundamental gap between practical everyday knowledge and scientific claims. Furthermore, the views that I have cited are raised precisely in a critical discussion of alternative philosophies of science – conventionalism and instrumentalism. My sense is that Sayer chose a somewhat 'everyday' example for the sake of simplicity rather than to distinguish between everyday practical action and scientific analysis in this regard.



---

<sup>12</sup> I am not arguing that the implication of Sayer's arguments about conventions and reality is that science is redundant. Even if my interpretation of this strand of Sayer's argumentation is correct, Sayer could still see science as relevant for investigating other kinds of phenomena where successful conventions have not yet been established, or for generating new levels of precision in relation to known phenomena. My point is rather more specific: it is that the strand of Sayer's approach which is less fallibilist in character implies that once success is established this tells us something about the world and thus encourages us to move science on to other areas or towards refining existing successful understandings rather than seeing these as up for a more fundamental reconceptualization and revision.

<sup>13</sup> See Holmwood (2001) for a related critique of Sayer's ideas which argues that the latter's defence of a distinction between system and lifeworld (Sayer, 2000b) reifies the categories of analysis in a problematic way.

<sup>14</sup> It's worth acknowledging that there are variations on this within analytical epistemology, with some writers arguing for further conditions, and others wishing to dispense with the relevance of justification. For an interesting discussion of some of these positions see Williams (2001).

<sup>15</sup> An analogy would be as follows: if we are going to test out recipes by working out which of them results in a moist cake, we would make the cakes, and then test them for their moistness. In such a case, we have to be able to know which cakes are moist and which are not.

<sup>16</sup> I note here that Psillos's analysis of approximation to truth is intended to avoid the difficulties found in that put forward by Popper (see Chapter 11 of Psillos, 1999). It is perhaps also worth noting that it is questionable whether Popper's own defence of the notion of verisimilitude is really consistent with his fallibilism. But this is too large an issue to address here.

<sup>17</sup> It might seem puzzling that the approximate truth would be of interest if the truth was already available. Psillos is interested in approximate truth because it provides him a means to say that although a general theory may not be strictly true, it can be very close to the truth (the established facts of the matter). For example, Kepler's first law is not strictly true because it does not capture planetary orbits absolutely accurately; nevertheless, for Psillos, it is approximately true because it is close to the truth about those orbits (Psillos, 1999).

<sup>18</sup> The enthusiast for the concept of truth might at this point say 'Ha! Doesn't this example show that you accept an idea of approximate truth in relation to cases such as the truth of the length of the table?!' My response would be that what I believe myself to be identifying about the table using my ruler is based *within* a (simplistic) framework of ideas about length. My sense of what is approximate and exact is within this framework. My understandings are unlikely to be compatible with the deepest considerations about length and measurement currently available in physics and philosophy, and nor do I see the current state of knowledge in those disciplines as the truth, the last word, on the question of length. Thus, I do not think I am capturing the truth of the length about the table by measuring it.

---

<sup>19</sup> Holmwood and Stewart (1991: 18-19) make a similar argument in relation to Bhaskar's analysis of theory change.

**Acknowledgements:** I would like to thank Christoforos Bouzanis, Nathan Coombs, and the anonymous referees for their comments on versions of this paper.

## References

Anderssen, N., Amlie, C. and Ytterøy, E.A. 2002. 'Outcomes for Children with Lesbian or Gay Parents. A Review of Studies from 1978 to 2000.' *Scandinavian Journal of Psychology* 43(4): 335-351.

Archer, M. 1995. *Realist Social Theory: The Morphogenetic Approach*. Cambridge: Cambridge UP.

Bamford, G. 1993. 'Popper's Explications of *Ad Hocness*: Circularity, Empirical Content, and Scientific Practice.' *British Journal of the Philosophy of Science* 44 (2): 335-355.

Barnes, B. 1982. *T.S. Kuhn and Social Science*. London: Macmillan.

Bhaskar, R. 1986. *Scientific Realism and Human Emancipation*. London: Verso.

Bhaskar, R. 1997. *A Realist Theory of Science*. Second Edition. London: Verso.

Biblarz, T.J. and J. Stacey. 2010. 'How does the Gender of Parents Matter?' *Journal of Marriage and Family* 72(1): 3-22.

Carter, B., and S. Virdee. 2008. 'Racism and the Sociological Imagination.' *British Journal of Sociology* 59(4): 661-679

Chang, H. (2012) *Is H<sub>2</sub>O Water? Evidence, Realism and Pluralism*, Dordrecht: Springer

Collier, A. 1994. *Critical Realism: An Introduction to Roy Bhaskar's Philosophy*. Verso: London.

Collier, A. 2003. *In Defence of Objectivity: On Realism, Existentialism and Politics*. London: Routledge.

Cruickshank, J. 2004. 'A Tale of Two Ontologies: An Immanent Critique of Critical Realism.' *The Sociological Review* 52(4), 567-585.

Demir, I. 2006. 'Thomas Kuhn's Construction of Scientific Communities.' In *Returning (to) Communities: Theory, Culture and Political Practice of the Communal*, edited by S. Herbrechter and M. Higgins, 89-109. New York: Rodopi.

Elder-Vass, D. 2007. 'For Emergence: Refining Archer's Account of Social Structure.' *Journal for the Theory of Social Behaviour* 37(1): 25-44.

Elder-Vass, D. 2012. *The Reality of Social Construction*. Cambridge: Cambridge UP.

Fay, B. 1990. 'Critical Realism?' *Journal for the Theory of Social Behaviour* 20(1): 33-41

Garrison, J. 1994. 'Realism, Deweyan pragmatism, and educational research.' *Educational Researcher* 23(1): 5-14.

Hanson, N. 1962. 'Leverrier: The Zenith and Nadir of Newtonian Mechanics.' *Isis* 53(3): 359-378.

Hentschel, K. 1992. 'Einstein's attitude towards experiments: Testing relativity theory 1907-1927.' *Studies in History and Philosophy of Science Part A* 23(4): 593-624

Hesse, M. 1974. *The Structure of Scientific Inference*. London: Macmillan.

Holmwood, J. 1996. *Founding Sociology? Talcott Parsons and the Idea of General Theory*. London: Longman.

Holmwood, J. 2001. 'Gender and Critical Realism: A Critique of Sayer.' *Sociology* 35(04): 947-965.

Holmwood, J., and A. Stewart. 1991. *Explanation and Social Theory*. London: Macmillan.

Holton, G. 1969. 'Einstein, Michelson, and the "crucial" experiment.' *Isis* 60(2): 133-197.

King, A. 2007. 'Why I am not an Individualist.' *Journal for the Theory of Social Behaviour* 37(2): 211-219.

Kuhn, T.S. 1970. *The Structure of Scientific Revolutions*. Second edition. New York: New American Library.

Lakatos, I. 1978. *The Methodology of Scientific Research Programmes: Philosophical Papers Volume 1*. Edited by J. Worrall and G. Currie. Cambridge: Cambridge UP.

Laudan, L. 1977. *Progress and its Problems: Towards a Theory of Scientific Growth*. Routledge and Kegan Paul: London.

Laudan, L. 1981. 'A Confutation of Convergent Realism.' *Philosophy of Science* 48(1): 19-49.

Leplin, J. 1997. *A Novel Defense of Scientific Realism*. Oxford: Oxford UP.

New, C. 2005. 'Sex and Gender: A Critical Realist Approach', *New Formations* no. 56: 54-70.

- Parsons, T. 1949. *The Structure of Social Action*. Second Edition. New York: The Free Press.
- Popenoe, D. 1996a. 'A World Without Fathers.' *The Wilson Quarterly* 20(2): 12-29.
- Popenoe, D. 1996b. *Life Without Father: Compelling New Evidence that Fatherhood and Marriage are Indispensable for the good of Children and Society*. London: Martin Kessler.
- Popper, K. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.
- Porpora, D. 2004. 'Objectivity and Phallogocentrism.' In *Defending Objectivity: Essays in Honour of Andrew Collier*, edited by M. Archer and W. Outhwaite. Routledge: London.
- Porpora, D. 2005. 'The Spiritual Turn in Critical Realism.' *New Formations* no. 56: 147-161.
- Psillos, S. 1999. *Scientific Realism: How Science Tracks Truth*. London: Routledge.
- Quine, W. 1951. 'Two Dogmas of Empiricism.' *The Philosophical Review* 60(1): 20-43.
- Rorty, R. 1991. *Objectivity, Relativism and Truth*. Cambridge: Cambridge UP.
- Rule, J. 1997. *Theory and Progress in Social Science*. Cambridge: Cambridge UP.
- Sayer, A. 1992. *Method in Social Science: A Realist Approach*. Second Edition. London: Routledge.
- Sayer, A. 2000a. *Realism and Social Science*. London: Sage.
- Sayer, A. 2000b. 'System, Lifeworld and Gender: Associational versus Counterfactual Thinking.' *Sociology* 34(4): 707-725.
- Silverstein, L.B. and C.F. Auerbach. 1999. 'Deconstructing the Essential Father.' *American Psychologist* 54(6): 397-407.
- Tasker, F. 2010. 'Same-sex parenting and child development: Reviewing the contribution of parental gender.' *Journal of Marriage and Family* 72(1): 35-40.
- Van Dongen, J. 2009. 'On the role of the Michelson–Morley experiment: Einstein in Chicago.' *Archive for History of Exact Sciences* 63(6): 655-663.
- Van Fraassen, B. 1980. *The Scientific Image*. Oxford: Oxford University Press.
- Vanfraussen, K., I. Ponjaert-Kristoffersen and A. Brewaeys. 2002. 'What does it Mean for Youngsters to Grow up in a Lesbian Family Created by Means of Donor Insemination?' *Journal of Reproductive and Infant Psychology* 20(4): 237-252.
- Will, C. 1988. *Was Einstein Right?* Oxford: Oxford UP.

Williams, M. 2001. *Problems of Knowledge: A Critical Introduction to Epistemology*. Oxford: Oxford University Press.

Woolgar, S. 1988. *Science: The Very Idea*. Tavistock: London.