volume II ♦ issue 4

2011

Logos & Episteme

an international journal
of epistemology

Romanian Academy
Iasi Branch
◆◆◆
"Gheorghe Zane" Institute
for Economic and Social
Research

INSTITUTUL EUROPEAN

Contact address:

Institutul de Cercetări
Economice şi Sociale „Gh.Zane”

Iaşi, str.T.Codrescu, nr.2, cod 700481
Tel/Fax: 004 0332 408922
Email: logosandepisteme@yahoo.com

www.logos-and-episteme.proiectsbc.ro

Logos & Episteme

ISSN-L: 2069-0533
ISSN: 2069-0533

This journal was edited within The Knowledge Based Society Project supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU ID 56815

TABLE OF CONTENTS

ARTICLES

FOR A POST-HISTORICIST PHILOSOPHY OF HISTORY. BEYOND HERMENEUTICS[1]

Adrian COSTACHE

ABSTRACT: With the publication of *Being and Time* and *Truth and Method* philosophical hermeneutics seems to have become the official philosophy of history, with exclusive rights on the questions arising from the fact-of-having-a-past. From now on the epistemological approach of the German historical school, reaching a peak in Dilthey's thought, is unanimously recognized as definitively overcome, *aufheben*, by the ontological interrogation of hermeneutics. But, with the same unanimity, it is also recognized that the reasons behind this overcoming and their validity are not readily apparent. For, as it has been shown in the literature, Heidegger's critique of Dilthey proves to be partial and lacunar, whereas Gadamer's is straightforwardly ambiguous. Our paper assumes as its first task a re-evaluation of these critiques and of the hypotheses proposed in the literature with regard to what could be the problem with Dilthey's epistemology. In this sense the paper argues that the problem resides in that the fundamental concepts on which it is based are bound to miss the peculiarity of history by idealizing it and masking the power relations inhabiting it. As a second task, our paper proposes an investigation of whether philosophical hermeneutics itself manages to rise to the expectations through which Dilthey's thought is evaluated. As it will become manifest, the answer to this question is in the negative. That is why, in the end, we will defend the necessity of a post-historicist and post-hermeneutic philosophy of history.

KEYWORDS: epistemology of history, ontology, historical experience, historical meaning, historical understanding

1. Introduction

Sometimes it is worth doing philosophy with the hammer even if you do not have Nietzsche's genius and are not planning the transvaluation of all values. A wrong question, a brutal inquisitive gesture can open paths imagination alone would not have discovered.

[1] This paper was made within The Knowledge Based Society Project supported by the Sectorial Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU ID 56815.

ARTICLES

FOR A POST-HISTORICIST PHILOSOPHY OF HISTORY. BEYOND HERMENEUTICS[1]

Adrian COSTACHE

ABSTRACT: With the publication of *Being and Time* and *Truth and Method* philosophical hermeneutics seems to have become the official philosophy of history, with exclusive rights on the questions arising from the fact-of-having-a-past. From now on the epistemological approach of the German historical school, reaching a peak in Dilthey's thought, is unanimously recognized as definitively overcome, *aufheben*, by the ontological interrogation of hermeneutics. But, with the same unanimity, it is also recognized that the reasons behind this overcoming and their validity are not readily apparent. For, as it has been shown in the literature, Heidegger's critique of Dilthey proves to be partial and lacunar, whereas Gadamer's is straightforwardly ambiguous. Our paper assumes as its first task a re-evaluation of these critiques and of the hypotheses proposed in the literature with regard to what could be the problem with Dilthey's epistemology. In this sense the paper argues that the problem resides in that the fundamental concepts on which it is based are bound to miss the peculiarity of history by idealizing it and masking the power relations inhabiting it. As a second task, our paper proposes an investigation of whether philosophical hermeneutics itself manages to rise to the expectations through which Dilthey's thought is evaluated. As it will become manifest, the answer to this question is in the negative. That is why, in the end, we will defend the necessity of a post-historicist and post-hermeneutic philosophy of history.

KEYWORDS: epistemology of history, ontology, historical experience, historical meaning, historical understanding

1. Introduction

Sometimes it is worth doing philosophy with the hammer even if you do not have Nietzsche's genius and are not planning the transvaluation of all values. A wrong question, a brutal inquisitive gesture can open paths imagination alone would not have discovered.

[1] This paper was made within The Knowledge Based Society Project supported by the Sectorial Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU ID 56815.

Take for example: how many books dealing with the philosophy of history have been written in the 20th century?

How should one approach such question? How could one answer it? It is practically impossible to determine the number of books written in *one* language during a *single* year on any topic whatsoever. Books can take years to be published. Either because they are deemed not interesting by the criteria of their time only to be recognized as important afterwards; or because they get lost and are rediscovered only years later.

But if one disregards the practical difficulties involved in such an endeavor and, with a completely unreflective gesture, heads straight to the catalogs room of the nearest library or a database, while counting titles, one will be struck by the following fact: in the 20th century history was a problem mainly for the hermeneutic and phenomenological philosophy. Hence it was taken as an ontological problem and it received an ontological treatment. Judging by the number of books carrying titles with such keywords as 'narrative,' 'historical understanding,' 'phenomenology,' 'life-history,' etc. that come up during a basic search, all the other approaches to the problem are secondary.

Now, this is certainly a remarkable fact if we bear in mind that no later than the end of the 19th century the problem of history was considered to be an epistemological problem *par excellence*. As it is well known, for the German historical school, the first philosophical movement to study history in a systematic manner, the central problem was the possibility of an objective knowledge of history and the scientific character of the historical sciences.[2]

Thus, what happened so that the epistemology of history faded out of the philosophical scene and the ontological approach gained a hegemonic position?

The present paper proposes itself a threefold task: first of all, to retrace the origins of this shift in the history of the philosophy of history; second of all, to examine the theoretical justification (or lack thereof) of the abandonment of the epistemological approach to history; and, third of all, to examine whether the ontological approach proposed by the hermeneutic project is truly more suitable for the study of history and if it really does justice to the peculiarity of this field of investigation.

[2] See in this sense Jean Grondin, *Introduction to Philosophical Hermeneutics*, trans. Joel Weinsheimer (New Heaven & London: Yale University Press, 1994), 76-90; Jeffrey Andrew Barash, *Martin Heidegger and the Problem of Historical Meaning* (New York: Fordham University Press, 2003), 1-64; Paul Hamilton, *Historicism* (London & New York: Rutledge, 1996).

2. Brief history of the philosophy of history in the 20th century

If one were to retrace the history of the philosophy of history back to its turning point at the beginning of the 20th century, one would find the first signs of the shift from epistemology to ontology in the works of Wilhelm Dilthey and the process completed through the works of Martin Heidegger and Hans-Georg Gadamer.

Dilthey commences his life's work on the premises of the historical school whose methodologist[3] he considered himself to be. His central question is the historical school's question: is an objective knowledge of history possible? Do the historical sciences really deserve the name 'science'? In contrast to the other members of the historical school though, Dilthey approached these questions by way of a "critique of historical reason" – paralleling Kant's "critique of pure reason" – centered on the constitution of knowledge in the historical sciences and their specificity in comparison with the natural sciences. It is in the elaboration of this critique that the shift towards ontology appears for the first time.

A failed attempt to offer for human sciences a methodological foundation through the development of a descriptive psychology determines Dilthey to approach the intertwining of knowledge and life head on. Once this reflection on life appears, it seems that nothing else matters any more and the fate of the epistemological approach to history is sealed. In fact, everything happens as if in both Dilthey's work and the history of philosophy such an epistemological inquiry has always been a means to an end, but never an end in itself. The way in which Heidegger describes Dilthey's thought in "Wilhelm Dilthey's Research and the Struggle for a Historical Worldview" provides us with a telling example.

> Dilthey's formulation of the question of history *in terms of the history of the sciences* began with his essays on Schleiermacher's hermeneutics and those dealing with the study of the history of the sciences, etc. These essays are not simply historical research on the history of the sciences but rather attempts to understand how in earlier times human life was interpreted. Their ultimate theme is a question about the concept of life.
> Dilthey's *epistemological* formulation of the question of history had the same motive. Here too we need to emphasize his question about the concept of life. […] [H]is interest was not a doctrine of method and system; he was not

[3] Grondin, *Introduction to Philosophical Hermeneutics*, 84.

> concerned with the question of how to classify particular sciences within different domains. Such matters later became the interest of Rickert.[4]

For Heidegger, the elaboration of the concept of life is Dilthey's greatest merit and fundamental contribution to the philosophy of history. This is the reason why, of all the philosophers of history before and after him, Heidegger takes Dilthey as partner of dialogue in the development of his thought. On the other hand though, Dilthey's greatest failing is that he did not go far enough in the line of questioning he started. Heidegger notes:

> Dilthey penetrated into that reality, namely, human Dasein which, in the authentic sense, is in the sense of historical being. He succeeded in bringing this reality to givenness, defining it as living, free, and historical. But he did not pose the question of historicity itself, the question of the sense of being, i.e., concerning the being of beings.[5]

Heidegger's contention is that precisely because he does not pose the question of the meaning of being Dilthey fails to complete the project of the critique of historical reason. That is why both his *Ideas Concerning Descriptive and Analytic Psychology* – the work containing the first formulation of the critique of historical reason – and in his *Introduction to the Human Sciences* – containing the second formulation – are bound to remain unfinished. Thus Heidegger feels he has to take it upon himself and carry further the task remaning incomplete after Dilthey and the historical school. As he writes:

> We need to repeat his questioning and to do this on the basis of a type of research – namely, *phenomenology* – that provides us with the suitable resources for advancing further than Dilthey's own position.[6]

His insights into historicity and reflections upon the being of history are meant to be just that: a repetition of Dilthey's questions but this time asked 'correctly,' i.e., so that they can be answered. These brief reflections upon historicity from "Wilhelm Dilthey's Research…" will receive an in-depth elaboration in the fifth chapter of *Being and Time*. But nowhere does Heidegger say either in what sense is phenomenology more suitable for the investigation of history or why the ontological examination of history constitutes a "further advancement" by comparison with Dilthey's epistemological inquiry.

---

[4] Martin Heidegger, "Wilhelm Dilthey's Research and the Struggle for a Historical Worldview," in Martin Heidegger, *Supplements. From the Earliest Essays to* Being and Time *and Beyond*, ed. John van Buren (Albany: State University of New York, 2002), 155.

[5] Heidegger, "Wilhelm Dilthey's Research," 159.

[6] Heidegger, "Wilhelm Dilthey's Research," 159.

It might seem that it is precisely in order to overcome this lack that, in *Truth and Method*, Gadamer engages in an extended discussion of the historical school, taking up the first half of the Part II. This work set out explicitly to "inquire into the consequences for the hermeneutics of the human sciences of the fact that Heidegger derives the circular structure of understanding from the temporality of Dasein"; i.e. to carry further, at the ontic level, Heidegger's insights into the questions of history. Obviously, this task, whereby the shift away from epistemology in the study of history becomes complete, could have done without that "Historical preparation" culminating in "Overcoming the hermeneutical problem through phenomenological research" in Dilthey's thought.

The only problem is that neither *Truth and Method* makes clear in what sense Dilthey's epistemology of history is overcome through phenomenology's ontological research. As it has been remarked in the literature, although

> The general outlines of Gadamer's estimation of Dilthey and the latter's contribution to hermeneutic philosophy are fairly well known"[7] […], "[t]he exact details of Gadamer's interpretation and critique of Dilthey and their justification (or lack of justification) are less clear – perhaps not in the least because Gadamer's own remarks are for the most part rather general and oftentimes ambiguous as well.[8]

In fact, the only clear thing when it comes to what constitutes for Gadamer the problematic character of Dilthey's thought is that announced in the titles of the sections dedicated to him: that there is a conflict between science and life and that this somehow gets Dilthey entangled in the "aporias of historicism." That is why the interpreters of Gadamer's work have adopted all the logical possibilities on the basis of what is said in the text.

For some, Gadamer accuses Dilthey for the psychologism and subjectivism of his epistemology.[9] For others, the problem would be the objectivism of his views on historical understanding.[10] Or the fact that he proves himself to be an

---

[7] Thomas Nenon, "Hermeneutical Truth and the Structure of Human Experience: Gadamer's Critique of Dilthey," in *Dilthey-jahrbuch fur Philosophie und Geschichte der Geisteswissenschaften* 8 (1992-1993): 75.

[8] Nenon, "Hermeneutical Truth," 75.

[9] See Anthony Giddens, "Hermeneutics and Social Theory," in *Hermeneutics. Questions and Prospects*, eds. Gary Schapiro and Alan Sica (Amherst: The University of Massachusetts Press, 1984), 225 and David Couzens Hoy, *The Critical Circle. Literature, History and Philosophical Hermeneutics* (Berkeley, Los Angeles, London: University of California Press, 1978), 11.

[10] See Joel Weinsheimer, *Gadamer's Hermeneutics: A Reading of* Truth and Method (New Heaven & London: Yale University Press, 1985), 148-155.

objectivist and a subjectivist at the same time.[11] Or that his thought is purely and simply ambigous.[12]

Given the symmetry of these positions it is obvious that each institutes itself as a counter-argument for its opposite and thus that they eventually annul each other.

This puts our brief history of the philosophy of history in a difficult position. By the very movement through which this history becomes manifest, by the same movement it also becomes opaque and incomprehensible. We have now a clear image of the steps that lead to the hegemony of the ontological approach to history but we have no clue as to why we have actually ended up here. In fact, given that no one seems to know in what it consists, the step beyond Dilthey and the historical school proposed by Heidegger and Gadamer in their respective hermeneutic projects ought to have taken us nowhere. And yet it does…

3. Why abandon the epistemological approach to history

In order to overcome this difficulty we would like to venture the following interpretive hypothesis: *the problem with Dilthey and the reason why his epistemology of history is to be left behind becomes manifest if one approaches the ambiguous text of* Truth and Method *from the point of view of Heidegger's reproach from "Wilhelm Dilthey's Struggle…" for not having asked the question of the being of history.*

The first indication in favor of this interpretive hypothesis appears already in that, as we have seen, Gadamer presents his work as a continuation of his master's. A second favorable sign for our hypothesis is to be found in the very terms in which Gadamer expresses the question that will guide *Truth and Method* after he learned the lessons from Dilthey's failure. These are precisely the general terms in which Heidegger discussed Dilthey's thought in "Wilhelm Dilthey's Struggle…," terms which appear only once in the whole Gadamerian corpus. Gadamer asks:

> What is the relation between power and significance, between forces and ideas, between the facticity and the ideality of life? This question must decide how knowledge of history is possible.[13]

[11] Thomas Nenon, "Hermeneutical Truth," 77.

[12] James Risser, *Hermeneutics and the Voice of the Other. Re-reading Gadamer's Philosophical Hermeneutics* (New York: State University of New York Press, 1997), 64-65.

[13] Hans-Georg Gadamer, *Truth and Method*, trans. Joel Weinsheimer and Donald G. Marshall (London & New York: Continuum, 2004), 221.

From the point of view of this question, problematic in Dilthey's thought appear to be the basic concepts, the fundamental epistemic cuts on the basis of which he constructs his philosophy of history. For Gadamer, the problem of Dilthey's epistemology is that it proves itself incapable of showing how lived experience (be it individual or collective) becomes historical? How can it leave marks upon the future and future generations? Due to this, Gadamer shows, Dilthey is also prone to misunderstanding the concepts of historical meaning and historical understanding. For he ceaselessly falls prey to the temptation of overstating the ideality of existence and minimizing the play of forces immanent to the flux of historical life and the power relations constituting it.

3.1.

For Gadamer, in the first formulation of the critique of historical reason Dilthey misses completely the problem posed by historical experience. The task he assumed at this point was to institute the facts of consciousness as foundation for historical knowledge. The presupposition underlying this endeavor is announced explicitly by Dilthey:

> [t]he first condition of the science of history is that I myself am a historical being, that the person studying history is the person making history.[14]

For him, just as the life of the individual is constituted through the continuous structuring and unification of the meaning of his/her personal experiences history is constituted by the meaning of the different particular events structured around certain unifying centers. And, just as the whole of someone's life can be understood on the basis of the particular experiences he/she had, history as such can be understood starting from any particular epoch.

As one can see, here the homogeneity between the subject and the object of historical knowledge, between the historian and the past to be known is postulated purely and simply. And with this it is postulated also the possibly historical character of human experience. Hence Dilthey's interest in the biographies of exceptional people supposed to open for us the royal path for understanding history. For history is nothing more, nothing less than a reflection of their genius.[15]

As Gadamer observes,

[14] Gadamer, *Truth and Method*, 217.

[15] See in this sense Wilhelm Dilthey, "Poetry and Experience," in *Wilhelm Dilthey: Selected Works. Volume V*, eds. Rudolf A. Makkreel and Frithjof Rodi (Princeton: Princeton University Press, 1985).

> This, however, is no solution to the epistemological problem that Dilthey posed. Rather, posing homogeneity as its condition conceals the real epistemological problem of history.[16]

This is clearly visible in that it never crosses Dilthey's mind to ask the reverse question: why would not the geniality of exceptional people be the reflection of the historical movements in which they are caught?

3.2.

In the second formulation of the project of historical reason the problem of historical experience is finally recognized as a problem. By approaching the task he assumed for himself in terms of lived experience, Dilthey is finally in possession of the means to show why individual experience and historical experience are homogenous.

In contrast to the facts of consciousness, the lived experience invoked now is preconscious, prior to the subject-object dichotomy and defines itself as the smallest unity of meaning that can be taken as basis of the nexus of psychic life. And, still in contrast, it is teleologically oriented towards its exteriorization in expression (which can take different forms, above all linguistic), just as this exteriorization (the expression), through the very way in which it is constituted, paves the way for its understanding.

Through the correlation lived experience – expression – comprehension Dilthey sees though another one – *that between life and comprehension*. Behind this triad we discover the fact that life bears within itself from the very beginning a dimension of knowledge. Of course, this is a prescientific knowledge, but by taking the form of legend, work of art, economic order, law, etc., not less objective. Here we can find the ground of the homogeneity between the knower and the known in history, between the historian and the past under scrutiny. Gadamer is in perfect agreement with Dilthey on this point:

> In language, moral values, and juridical forms the individual – the isolated being – is always already beyond his particularity.[17]

But, as *Truth and Method* shows, when it comes to the consequences to be drawn from all these, it would seem that we should keep away from Dilthey's thought.

Out of the correlation lived experience – expression and out of the latter's possibility to 'solidify' itself as objective spirit Dilthey jumps straightforwardly to the conclusion that

---

[16] Gadamer, *Truth and Method*, 217.

[17] Hans-Georg Gadamer, *Le probleme de la conscience historique* (Louvain: Publications universitaires de Louvain, 1963), 28.

> Each single manifestation of life re-presents something common and shared in the realm of objective spirit.[18]

Is this not another way of affirming the identity of individual and historical experience? This time though Dilthey goes even further by showing in fact that individual experience is not just potentially but *de facto* historical. For if every gesture of the individual is a manifestation of the objective spirit and only because of this it can be understood, then the individual participates to history by every single gesture he/she makes. Here is the precise way in which the second formulation of the critique of historical reason misses historical experience.

By missing the peculiarity of historical experience and the particular way in which it differs from individual experience though, as we were saying, Dilthey fails to arrive at a suitable understanding of the concepts of historical meaning and historical understanding as well.

In the case of historical meaning everything happens as a direct consequence. Since there are just historical experiences, the meanings in which these are expressed and objectified, in their turn, cannot be but historical, every meaning being thus forced to reflect, like a Leibnizian monad, the whole history of humanity and to pave the way towards a complete understanding of the past, present and future. On this point, the only difference between the two formulations of Dilthey's critique of historical reason is that in the first all meaning it is just potentially, whereas in the second it is actually historical.

If one connects the dots between the problematical character of historical meaning and historical experience, the problem with historical understanding comes to the fore, too. For, irrespective of the way in which Dilthey approaches the question of historical knowledge, the understanding it presupposes will be forced to take as normative ideal the phantasm of complete transparency. Since any individual experience comes to express itself in an objective meaning and since this is always historical, understanding becomes a simple 'deciphering' of the past able to render it completely transparent. The only condition here is to pursue it with sufficient determination.

Thus, no wonder that when it comes to explaining how understanding takes place, Dilthey takes Schleiermacher's textual hermeneutics as a model. Gadamer remarks:

---

[18] Wilhelm Dilthey, "The Formation of the Historical World in the Human Sciences," in *Wilhelm Dilthey: Selected Works. Volume III*, eds. Rudolf A. Makkreel and Frithjof Rodi (Princeton: Princeton University Press, 2002), 168.

> We can thus understand why Dilthey starts from romantic hermeneutics. […] Romantic hermeneutics came to the assistance since, as we saw, it took no account whatsoever of the historical nature of experience. It assumed that the object of understanding is the text to be deciphered and its meaning understood. Thus for romantic hermeneutics every encounter with a text is an encounter of the spirit with itself. Every text is strange enough to be fundamentally intelligible even when we know nothing about it except that it is text, writing, an expression of mind.[19]

So, irrespective of the way in which he poses the problem, Dilthey levels out the peculiarity of history, erasing its factical dimension and transforming it into a history of the spirit or an intellectual history.

Through this failure, Dilthey's thought functions like an index to the entire historical school and gives us the reasons for abandoning the epistemological approach to history as such. For, if we take a step back and regard the causes of Dilthey's failure from afar we will see that these are actually constitutive of the research presupposition of the classical epistemological project. The very definition of knowledge as a process carried by a subject upon an object forces us to look for its foundation in the direction of the subject and leaves us no other choice but to take either the consciousness of the individual or his or her life as the ultimate ground. This is the only possible choice because no other human trait is generic enough to characterize the knowing subject.

4. Beyond hermeneutics?

That Dilthey and, in general, the historical school's approach to history deserves to be abandoned on account that it does not do history justice is one thing, whether it should, is another. Even in the hard sciences a theory is not given up simply because it is wrong; at least not until there is another to replace it.

As we recall though, Heidegger and Gadamer's claim was that their ontological approach is meant explicitly as a replacement of the historical school's epistemology. At the beginning of our paper we noted that Heidegger envisaged (his) phenomenology as an "advancement" of Dilthey's position. Just as Gadamer spoke of the "overcoming of the epistemological problem through phenomenological research." But nothing said so far tells us whether this is actually so. And nothing says why it would be. Thus, in order to establish the superiority of the ontological approach of Heidegger and Gadamer's hermeneutic projects, at the very least, we have to ask them the same question they asked Dilthey and the historical school:

[19] Gadamer, *Truth and Method*, 233.

"What is the relation between power and significance, between forces and ideas, between the facticity and the ideality of life?"

Insomuch as that is *their* question, insomuch as it is a question anticipated by Heidegger and asked explicitly by Gadamer, at first sight, it might seem that this is a futile endeavor. Of course both of them are going to answer this question, probably in an exemplary manner. At first sight, this is really the case.

From the very first steps of the formulation of his ontological project Heidegger takes the necessary measures to avoid Dilthey's trap of reducing history to a history of the spirit. Already in his 1921 lecture course on Aristotle, Heidegger straightforwardly defines human existence in terms of facticity, as immersed in a horizon of opacity in which and beyond which the light of our reason cannot go. As Heidegger notes:

> The noun 'life,' has a rich and autonomous meaning, which we can briefly articulate into three senses:
>
> 1. Life in the sense of the unity of succession and maturation…
> 2. Life, grasped as such a delimited unity of succession: now in the sense of something that specifically bears possibilities, ones matured partially in life itself and for it. Life of which we say that it can bring all things, that is incalculable…
> 3. Life understood in the sense in which 1. and 2. intertwine: the unity of extension in possibility and as possibility – lapsed possibilities, laden with possibilities and laden with itself, forming possibilities – and this whole taken as reality, indeed as reality in its specific opacity as power, *fate*.[20]

And in *Being and Time* we are shown that precisely Dasein's finitude is that which confronts it with its own 'destiny,' its authentic historicity on which any history whatsoever is grounded.

> Once one has grasped the finitude of one's existence, it snatches one back from the endless multiplicity of possibilities which offer themselves as closest to one – those of comfortableness, shirking, and taking things lightly – and brings Dasein into the simplicity of its *fate* [*Schicksals*]. This is how we designate Dasein's primordial historizing, which lies in authentic resoluteness and in which Dasein *hands* itself *down* to itself, free for death, in a possibility which it has inherited and yet has chosen.[21]

---

[20] Martin Heidegger, *Phenomenological Interpretations of Aristotle: Initiation into Phenomenological Research*, trans. Richard Rojcewicz (Bloomington and Indianapolis: Indiana University Press, 2001), 64.

[21] Martin Heidegger, *Being and Time*, trans. John Macquarrie and Edward Robinson (San Francisco: Harper & Row, 1962), 435/[384].

By contrast, Gadamer will adopt a more elaborate strategy. We have seen, he agrees with Dilthey on the way in which individual experience becomes historical. For him too, history is the spirit objectified in legends and literatures, in laws and customs, in general, in tradition. But, in order to avoid Dilthey's false conclusions, just like Heidegger, Gadamer will try to show that *tradition has a factical character and is the ontic sign of our finitude.* Or, otherwise put, that

> … the prejudices of the individual, far more than his judgments, constitute the historical reality of his being.[22]

> Long before we understand ourselves through the process of self-examination, we understand ourselves in a self-evident way in the family, society, and state in which we live. The focus of subjectivity is a distorting mirror. The self-awareness of the individual is only a flickering in the closed circuits of historical life.[23]

To this end *Truth and Method* undertakes a demonstration developed in several steps that begins with the well-known rehabilitation of tradition and prejudices. In this first step Gadamer will seek to show that tradition is something else than a collection of curiosities of the past capable of raising only the interests of antiquarians and that, through the knowledge it carries within itself, it offers each and every one of us a solid point for anchoring our lives.

Against a widespread opinion since the Enlightenment the originary meaning of the concept of prejudice makes it quite clear that these cannot be reduced to the status of unfounded judgments, deprived of any value whatsoever. Even though they are not methodically validated, as the ideal of objectivity of modern rationalism requires, prejudices have an experiential foundation confirmed over and over again in the passing of time. That is why they offer the individual a guiding light in all the situations in which his or her reason is confronted by its limits.

In a second step Gadamer shows that this horizon of knowledge brought along by prejudices never exists in and for itself. It is always immersed in another horizon, a larger one, that safekeeps (*bewahrt*) our entire past. That is why tradition presents itself to us as a dual nature. It is at the same time perfectly transparent and completely opaque; it is at once familiar and strange, the discovery of its opacity and strangeness behind its familiarity being experienced as a "pulling up short."[24]

---

[22] Gadamer, *Truth and Method*, 278.

[23] Gadamer, *Truth and Method*, 278.

[24] In German Gadamer says "Erfahrung des Anstoßes," an expression which makes utmost clear the shocking and violent nature of the encounter with tradition. For 'Stoß' actually means

The prejudices regulating social interactions are the privileged example. Through them almost every one of us is able to recognize what counts as the 'right' behavior in a given situation but, most of the time, no one can indicate either why that behavior is considered as such or whether it is truly so. And as soon as we realize our impotence in this regard, the great majority of us will experience frustration.

Thus, we have to recognize that

> That which has been sanctioned by tradition and custom has an authority that is nameless, and our finite historical being is marked by the fact that the authority of what has been handed down to us – and not what is clearly grounded – always has power over our attitudes and behavior.[25]

The opacity of tradition reveals its factical character and indicates quite clearly our finitude. Precisely in this nameless authority we can find the dimension of force of history along with or behind its spirituality. And thus it would seem that the steps of the demonstration taken so far would suffice to prove the superiority of Gadamer's ontological approach to history by comparison to Dilthey's epistemological questioning. All the more in these first two steps taken Gadamer finds already a basis solid enough for rethinking the concepts of historical meaning and historical understanding.

Due to the dual nature of tradition, the hermeneutic 'object'[26] – the bearer of traditionary meaning in the passing of time – is itself not a unity in itself but the unity of a duality. As Gadamer shows us, the hermeneutic 'object' is in fact constituted in and through the tension between the initial meaning intended by the agent/author of the event, work of art or the text making history and the meanings attributed to it through the different interpretations it received throughout time. It is true, as David Couzens Hoy notes:

> We do not see Plato as Descartes or as Kant saw him, but we certainly see Plato differently because of Descartes and Kant.[27]

'shock,' 'kick,' 'impact,' etc. See his *Gesammelte Werke, Band 1, Hermeneutik 1. Wahrheit und Methode* (Tübingen: J. C. B. Mohr (Paul Siebeck), 1990), 272.

[25] Gadamer, *Truth and Method*, 281.

[26] We have to use the scare quotes here for the hermeneutic object is not an object *per se*. Insomuch as, through the prejudices orienting it, tradition is constitutive of the individual, the traditionary meaning is not simply in front of us and we cannot relate to it as to an object. The relation between the individual and tradition is not simply a relationship between subject and object and thus it cannot take the form of a rapport of knowledge.

[27] David Couzens Hoy, *The Critical Circle. Literature, History and Philosophical Hermeneutics* (Berkeley, Los Angeles, London: University of California Press, 1982), 41.

Through this it becomes manifest, on the one hand, that historical meaning cannot be just an expression of the spirit and that it presents itself as a "fluid multiplicity of [ontological] possibilities" anchored in the moment in which they occur. And, on the other hand, that historical understanding is inexorably tied to the tradition and prejudices from which Dilthey, just like the entire historical school, was trying to keep away. Otherwise put, because history is not a text, its understanding cannot be an act of deciphering, but an act of life reflecting clearly its finitude.

Despite the fact that they finally give us the possibility of recognizing the reality of history in historical meaning and historical understanding though, the steps took so far in his demonstration are only seemingly sufficient for probing the superiority of hermeneutics' ontological approach to history over Dilthey and the historical school's epistemological one. Even though they attest the factical character of tradition and give us reasons for seeing in its opacity the reality of history, in the end, they do not demonstrate anything. For one can always reply – as Jürgen Habermas actually does in his debate with Gadamer in the 60s[28] – that this opacity is strictly circumstantial, that it is a sign of a momentarily weakness of reason and that, given the appropriate scientific means, it can easily be overcome.

Gadamer was very much aware of all these. Precisely this is why in the third part of *Truth and Method* he undertakes an "ontological shift of hermeneutics guided by language" whose focus is to show that language, as the essence of tradition, is an ontological determination for us. For it is responsible for the humanity of man and its singularity in comparison to all the other creatures on the face of the earth.

As Gadamer explains, if, for man, there is a world as for no other creature in the world, if in contrast to the other creatures which live in an environment and are so caught up within it that they cannot escape it even if they venture as far away as possible from their natural habitat, man *has* a world, this is because man can have a different "posture" towards what he/she encounters from the surrounding world (*Umwelt*). To have a world means to be free from the environment by having an "orientation" (*Verhalten*) towards it, and thus, by being able to situate it in front of your eyes and to present it to yourself as is. Man has a world because he has risen above the surrounding world and he is able to put the latter into a perspective.

[28] See in this sense Jürgen Habermas, "A Review of Gadamer's *Truth and Method*" in *The Hermeneutic Tradition. From Ast to Ricoeur*, eds. Gayle Ormiston and Alan D. Schrift (New York: State University of New York Press, 1990), 213-245.

But, as Gadamer argues, in this sense, man has a world only due to the *variability of language*. In the same language one and the same thing or state of facts can be described in different ways. One description centers on one particular aspect of the thing or the state of facts described while other descriptions center on other aspects. That is why with every such new description a new facet of the thing or the state of facts comes to appearance and, with this, our perspective upon it is enlarged, the distance separating us from the thing or the state of facts becoming wider and wider.

With this Gadamer crosses once again paths with his master, the thesis of the co-originarity of language and world being one that traverses Heidegger's thought through and through, irrespective of the *Kehre*. If anything, in this particular regard the *Kehre* brings about a radicalization. For if in *Being and Time* the meanings of language are taken as expressions of the significance constituting the worldhood of the world, in Heidegger's later work they themselves come to constitute the world. The "Origin of the Work of Art" says:

> Where there is no language, as in the being of the stone, plant, and animal, there is no openness of what is, and consequently no openness of what is not and of the empty. Language, by naming beings for the first time, first brings beings to word and to appearance. Only this naming nominates beings *to* their being *from out of* their being.[29]

And some ten years later Heidegger notes:

> It is because language is the house of Being that we reach what is by constantly going through this house. *When we go to the well, when we go through the woods, we are always already going through the word 'well,' through the word 'woods,' even if we do not speak the words and do not think of anything relating to language.* […] All beings – objects of consciousness and things of the heart, men who impose themselves and men who are more daring – all beings, each in its own way, are qua beings in the precinct of language.[30] (italics are mine)

But, if everything in the world and the world as such has a lingual constitution; if all beings – those having the being of the historical included – are what they are in and through language, then how can the hermeneutic project, be it Gadamer's or Heidegger's, avoid the reduction of history to the history of spirit that set it in motion? Isn't hermeneutics affirming in a different way Dilthey's old conclusion: "Everything in history is intelligible, for everything is text. 'Life and

[29] In Martin Heidegger, *Poetry, Language, Thought*, trans. Albert Hofstadter (New York & London: Harper Colophon Books, 1975), 73.

[30] Martin Heidegger, "What Are Poets For?" in *Poetry, Language, Thought*, 132.

history make sense like the letters of a word.'"[31]? But was this not the conclusion that gave birth to Heidegger and Gadamer's critique of Dilthey? Wasn't for Heidegger and Gadamer precisely this the sign that the epistemological approach to history must be "overcome"?

In view of all these, it becomes apparent that the moment of reality just discovered is in fact the moment of a spiritual reality. The force of the nameless authority that imposes itself upon the individual and proves to be stronger than the justifications of his or her reason is a spiritual force. For the language in and through which everything 'is' (in an ontological sense) is, obviously, spirit. With this, any precaution taken by defining existence in terms of facticity and instituting finitude as a condition of possibility of historicity vanishes. If we do not know individually where we come from or where we are going, collectively, as people,[32] insomuch as we ourselves and everything that surrounds us 'is' language, we can gain a quite clear idea of where we have been.

But this detour through the question of historical being and the reality of history asked by the hermeneutics of Heidegger and Gadamer does not prove itself to be just unnecessary. It is not just that through it the hermeneutic project fails to defend its superiority over the epistemological approach of Dilthey and the historical school; due to it the hermeneutic approach proves itself to be even more problematical. If for Dilthey's epistemology the Gordian knot were the concepts of historical experience, meaning and understanding, for the hermeneutic project to this list one has to add the concept of historical change as well.

How is the new possible? How do historical changes come about and how are they recognized as such? How does the moment of rupture occur in the continuum of our lives and how is such a moment perceived? Grounding his philosophy on a metaphysics of genius, Dilthey never had any trouble in answering these questions. Genius is, by definition, an agent of change, originality being its fundamental trait. When historical being becomes lingual though, what will be will be conditioned in its being by what has been, by an already constituted horizon of meaning. Through this, what comes to be, the evental future, what Derrida was designating through *avenir* is reduced to *future*, the expected future, in fact, the present that is not yet present, but when it will be, it will be precisely as expected. Gadamer notes it explicitly in *Truth and Method*:

[31] Gadamer, *Truth and Method*, 234.

[32] As it is known, for Heidegger the people are the true agents of history. See in this sense *Being and Time*, 435/[384].

> Even where life changes violently, as in ages of revolution, far more of the old is preserved in the supposed transformation of everything than anyone knows, and it combines with the new to create a new value.[33]

Given all these a last question imposes itself upon us: if the hermeneutic project does not manage to go further in the study of history than the historical school it was trying to overcome, then should we not endeavor to overcome the hermeneutic project itself? Should we not abandon it as it teaches us to abandon the epistemological approach to history of Dilthey and the historical school before it?

Unfortunately these questions are only rhetorical.

[33] Gadamer, *Truth and Method*, 282-283.

WHO IS AN EPISTEMIC PEER?

Axel GELFERT

ABSTRACT: Contemporary epistemology of peer disagreement has largely focused on our immediate normative response to *prima facie* instances of disagreement. Whereas some philosophers demand that we should withhold judgment (or moderate our credences) in such cases, others argue that, unless new evidence becomes available, disagreement at best gives us reason to *demote* our interlocutor from his peer status. But what makes someone an epistemic peer in the first place? This question has not received the attention it deserves. I begin by surveying different notions of 'epistemic peer' that have been peddled in the contemporary literature, arguing that they tend to build normative assumptions about the correct response to disagreement into the notion of peerhood. Instead, I argue, epistemic peerhood needs to be taken seriously in its own right. Importantly, for epistemic agents to count as peers, they should exhibit a comparable degree of reflective awareness of the character and limitations of their own knowledge.

KEYWORDS: epistemic peers, peerhood, disagreement, ignorance

1. Introduction

Recent years have seen an explosion of the epistemological literature on peer disagreement. How should we, as individuals, react when we encounter someone who disagrees with us on a point of fact, yet who we have every reason to believe is our epistemic equal? Disagreement is as much part of our everyday epistemic lives as reliance on others for knowledge, and thus merits attention from epistemologists – perhaps especially so at a time when traditional sources of institutional testimony (experts, media, higher education, science, etc.) are widely perceived as becoming more diverse and polarized.

In trying to understand this phenomenon, epistemologists of disagreement standardly operate with an idealized situation, in which two equally well-informed 'peers,' on the basis of the same evidence, come to different conclusions regarding a specific subject matter. At issue is the question of whether *mere* acknowledgment of disagreement is sufficient to require us to abandon belief, or at least to revise our corresponding degree of belief downwards. One prominent view, the *equal weight* (EW) view, holds that two epistemic peers, upon learning that they assign different credences to a given proposition *p*, should revise their

credences to be 'roughly equal'[1]; in other words, two epistemic peers who find themselves disagreeing on whether *p* should 'split the difference' and update their respective credences so as to reflect the average of the two initial credences. The equal weight view has been subject to severe criticism, with some critics arguing that, in 'splitting the difference,' we are in effect discarding whatever initial evidence we had – 'pre-disagreement,' as it were – for our original belief. As one critic puts it, this "does not seem to be a virtue in an epistemological theory."[2] On what has been called the *right-reasons* (RR) view, we owe it to our initial reasoning (which led us to conclude that *p*) that we regard our epistemic peer, who claims that not-*p*, as "having gotten things wrong" on this occasion.[3] Rather than revising our own credence, we treat the fact of disagreement essentially as a reason for *demoting* our disputant from his position as epistemic peer – after all, we already know that one of us must have incorrectly assessed the force of the evidence, and for all we know it is our disputant who is mistaken.

Such, in general outline, is the shape of the debates that have placed peer disagreement at the heart of recent epistemology. Curiously, however, whereas there has been no shortage of discussions of the phenomenon of *disagreement*, comparatively little attention has been paid to a systematic analysis of what makes someone an *epistemic peer* in the first place. The present paper attempts to fill that gap. In Section 2, I survey the notions of 'epistemic peer' that have been peddled, often implicitly, by contributors to the epistemology of disagreement. In Section 3, I challenge two key assumptions concerning the character, and alleged frequency, of peer disagreement. Section 4 argues that determining epistemic peerhood is a more important task than adjudicating between different responses to the mere fact of disagreement; importantly, epistemic peerhood comes in degrees. In Section 5, I argue that, beyond being equally knowledgeable, epistemic peers should also display a similar degree of reflective awareness of the limitations of their own knowledge. The paper concludes by suggesting a possible refocusing of the philosophical debate on disagreement.

---

[1] David Christensen, "Epistemology of Disagreement: The Good News," *The Philosophical Review* 116 (2007): 193.

[2] David Enoch, "Not Just a Truthometer: Taking Oneself Seriously (but not Too Seriously) in Cases of Peer Disagreement," *Mind* 119 (2010): 969.

[3] Adam Elga, "Reflection and Disagreement," *Noûs* 41 (2007): 486. Note that Elga is here *characterizing*, not *endorsing*, the right-reasons view.

2. Epistemic peerhood: the received view

Lack of unanimity is a common experience in social interactions, and in everyday language we readily help ourselves to the term '*disagreement*' and its cognates. By contrast, the expression 'epistemic peer' is a technical term, invented by philosophers in order to bring into sharper focus a set of theoretical questions. Interestingly, the term makes its first appearance in the philosophy of religion: If there was only isolated dissent about the (alleged) proper basicality of such claims as "God exists," we might rationally dismiss disagreement about a proposition, but, as Gary Gutting argues, "the disagreement of substantial numbers of those who, as far as I can tell, are my *epistemic peers* (i.e. my equals in intelligence, perspicacity, honesty, thoroughness, and other relevant epistemic virtues) is surely another matter."[4] Jonathan Kvanvig, by contrast, insists that "if disagreement is important to justification, it cannot be disagreement among epistemic peers" since one of the peers might, for purely contingent reasons, be "in a better epistemic situation."[5]

Early uses of 'epistemic peer' rely on an understanding of peerhood that stresses parity with respect to general epistemic virtues, as enumerated by Gutting. This contrasts with the dominant view in the debate, which adds the requirement that, for two epistemic agents to count as peers in a factual dispute, they must be "equals with respect to their familiarity with the evidence and arguments which bear on that question."[6] In other words, peers must have "been exposed to the *same* evidence and have worked on it comparably long, carefully, etc."[7] It seems reasonable to assume that, barring special conditions, such sweeping similarity in epistemic outlook will lead to epistemic peers being similarly reliable as sources of information. Indeed, this is sometimes seen as the defining feature of epistemic peerhood, as when David Enoch defines an epistemic peer as "someone who is, somewhat roughly, antecedently as likely as you are to get things right (on matters of the relevant kind)"[8]; others have argued for the same view precisely "because it generates the puzzles about disagreement" without entailing any substantive

[4] Gary Gutting, *Religious Belief and Religious Skepticism* (Notre Dame: University of Notre Dame Press, 1982), 83.

[5] Jonathan Kvanvig, "The Evidentialist Objection," *American Philosophical Quarterly* 20 (1983): 54.

[6] Thomas Kelly, "The Epistemic Significance of Disagreement," *Oxford Studies in Epistemology* 1 (2005): 174.

[7] Bryan Frances, "The Reflective Epistemic Renegade," *Philosophy and Phenomenological Research* 81 (2010): 424. In the contemporary debate, even 'slight differences' in evidence are seen as undermining peerhood; see, for example, Nathan King, "Disagreement: What's the Problem? *or* A Good Peer is Hard to Find," *Philosophy and Phenomenological Research* (2011, Early View: doi: 10.1111/j.1933-1592.2010.00441.x): 13.

[8] Enoch, "Not Just a Truthometer," 956.

similarities between the agents' evidence or aptitude.[9] Thus, what started off as an attempt to identify agents with broadly comparable epistemic virtues (though possibly different exposure to the evidence), gradually became assimilated to a probabilistic point about the relative reliability of epistemic agents as sources of information.

At the beginning of this section, I noted that 'epistemic peer' is an epistemological term of art. As such, any definition is bound to have a stipulative element. Nowhere is this stipulative character more evident than in Adam Elga's – at the time, by his own admission, "nonstandard"[10] – definition of (perceived) peerhood in terms of our (pre-disagreement) assessment of how likely each agent is to be mistaken. On this view, "you count your friend as an epistemic peer with respect to an about-to-be-judged claim if and only if you think that, conditional the two of you disagreeing about the claim, the two of you are equally likely to be mistaken."[11] While this obviously sits well with the equal weight view – of which Elga is a leading proponent – others have questioned why determinations of peerhood should be concluded *before* a claim is about to be judged: Why "exclude from one's conditionalization process the disagreement itself as reason for demoting" our interlocutor from his position as one's peer?[12] In the remainder of this paper, I shall argue that rather than building normative assumptions about the correct response to disagreement *into* our definition of 'epistemic peer,' we need to disaggregate both phenomena and develop a richer notion of epistemic peerhood.

3. Acknowledgment and the paucity of peer disagreement

Why worry about disagreement in the first place? The intuitive reason is clear: Instances of disagreement are, at least *prima facie*, good occasions to reflect on the fallibility of the methods by which we acquire beliefs (including the belief at issue). It would seem, then, that upon encountering a disagreement with a peer, we should engage in some epistemic soul-searching with the aim of identifying weak links in our reasoning processes and eliminating possible sources of error. But note that this is not how contemporary epistemologists of disagreement typically approach the problem: Their concern is with the *pro tanto* epistemic reason given to us by the brute fact of disagreement alone, irrespective of any

[9] Alex Bundy, "In Defense of Epistemic Abstemiousness," *Logos & Episteme* 2 (2011): 288.
[10] Elga, "Reflection and Disagreement," 499.
[11] Elga, "Reflection and Disagreement," 499.
[12] Enoch, "Not Just a Truthometer," 977.

changes that we would make to our belief systems, were we to independently reassess our individual processes of belief formation.[13]

3.1. Acknowledgment and the fact of disagreement

For the fact of disagreement to take on a compelling degree of urgency, it must be acknowledged by at least one party to the dispute.[14] If both peers were blissfully unaware of their doxastic differences, the mere counterfactual observation that they *would* disagree with one another *if* quizzed on the point in question, can hardly be expected to have an effect on their beliefs and credences. (Indeed, it is doubtful whether we could legitimately speak of a 'disagreement' at all in this case.) Exactly what follows from the acknowledgment of the fact of disagreement, is of course the bone of contention among contemporary epistemologists of disagreement. For the defender of the EW view, acknowledgment of the fact of disagreement – combined with the recognition (taken for granted by the EW view) that our interlocutor is our epistemic peer – immediately requires us to adjust our credence downwards and 'split the difference.' By contrast, the RR view considers recognition of the fact of disagreement sufficient for demoting our interlocutor from his position as peer; it takes more than *mere* disagreement to sway what was, after all, our initial considered judgment.

3.2. The paucity of (acknowledged) peer disagreement

Both EW and RR theorists agree on the crucial role of acknowledgment and the purely auxiliary function of judgments of epistemic peerhood – where the latter are seen either as *preceding* the disagreement (as in the EW view) or as *being superseded* by it (as insisted by the RR view). At the same time, parity – both in terms of general epistemic virtues and concerning equal exposure to (and consideration of) the evidence – does remain an important background assumption: After all, disagreements with obvious epistemic inferiors would hardly inspire the sense of urgency that fuels the debate about disagreement among peers. Yet it is worth questioning just how often we do, in fact, find ourselves in genuine situations of *acknowledged* peer disagreement.[15] Parity in terms of the reliability of getting it

[13] On this point, see also Enoch, "Not Just a Truthometer," 957.

[14] Other authors prefer to speak of 'revealed' disagreement, e.g. the editors in their Introduction to the volume *Disagreement*, ed. Richard Feldman and Ted A. Warfield (Oxford: Oxford University Press, 2010), 4.

[15] This is not to say that acknowledged peer disagreement, even if found to be rare, does not pose a genuine theoretical challenge, but only that the contemporary debate derives much of its force from the alleged *pervasiveness* of peer disagreement.

right on a specific point – say, the question of whether p – requires considerable fine-tuning of the specific epistemic histories of both agents (in terms of their overall experience, relevant background beliefs, etc.), which may often outweigh similarities in general epistemic virtues. As Nathan King notes, "the path toward equal reliability is not straightforward,"[16] and even subtle differences in the dispositional response to evidence undermine the fine-grained notion of peerhood as 'equal familiarity with the evidence' and with relevant arguments.[17] Moreover, many longstanding disputes – for example in science – are driven by (often axiological) disagreements about which findings should count as evidence in the first place[18]; in such cases, "equal familiarity with relevant evidence and arguments" as a criterion for determining peerhood is itself under dispute, and parties to such disputes – no matter how objectively well-acquainted with the facts – will rarely acknowledge each other as epistemic peers.

4. The primacy of peerhood

What the preceding discussion suggests is that determining whether two disputants are indeed epistemic peers – and, if they are not, identifying in what ways their relationship falls short of peerhood – has primacy over the question to what extent the mere fact of disagreement offers a *pro tanto* reason for each party to adjust their credences. In this section, I wish to take a first step towards developing a richer notion of epistemic peerhood, before arguing, in the next section, for a plausible connection between peerhood and questions of epistemic value, which so far appear to have been ignored in the philosophical debate about disagreement. For now, I wish to argue that our understanding of peerhood can be enriched by recognizing that peerhood comes in degrees.

Early definitions of 'epistemic peerhood' lacked specificity, insofar as they emphasized the overall similarity in epistemic character, while discounting the important influence of contingent differences (e.g. in exposure to relevant evidence). By contrast, recent attempts to reduce peerhood to *mere equal likelihood to be right on a particular occasion* are overly narrow, insofar as they offer no guidance as to how best to judge whether someone is an epistemic peer or not. Instead of oscillating between these two extremes, what is needed is a 'middle ground' that can account for two sorts of scenarios: First, the possibility that, due to contingent features in their epistemic histories, two equally virtuous agents may fail to be

[16] King, "Disagreement," 13.

[17] See Section 2.

[18] See Larry Laudan, *Science and Values. The Aims of Science and Their Role in Scientific Debate.* (Berkeley: University of California Press, 1984).

'epistemic peers' on a specific question; and second, the possibility that two epistemic agents who happen to be equally likely to get things right *on this occasion*, may nonetheless fall short of peerhood, due to more fundamental differences in epistemic outlook.

Acknowledgment of peerhood is at least as important an element in true cases of peer disagreement as acknowledgment of the disagreement itself. Reflecting for a moment on how peer status – whether objectively warranted or not – is actually accorded by one group (or individual) to another, it seems plausible that judgments of epistemic peerhood, like those of trustworthiness, will often be bound up with social markers of similarity (e.g., indicators of social background, professional affiliation, or academic credentials). Benjamin Wald,[19] following a suggestion by Mark Vorobej, develops a useful distinction between *close* peers (who not only assess the evidence pertaining to a particular topic in similar ways, but also have good reason to believe that they both have similarly good track records in forming true beliefs on the basis of evidence), *distant* peers (who fail to meet one of these conditions), and *remote* peers (who fail both of these conditions). Importantly, distance – in the sense discussed here – can be the result of lack of familiarity with what would constitute a good track record for the other party, and such lack of familiarity can in turn be a side effect of social distance. This is especially pertinent in cases of disagreement among experts from different disciplines (say, disagreement between nuclear engineers and radiation ecologists on the safety of nuclear power plants). As Wald notes, "members of different epistemic communities can count as remote peers to one another", and persistent disagreement between such communities "need not be due to any failure of rationality."[20] The limits of rational disagreement thus need not coincide with the limits of (acknowledged) peerhood.

5. Peerhood and Socratic ignorance

The example of disagreement among experts, on matters of public concern, supports the observation that judgments of peerhood depend, at least in part, on what is at stake. When the stakes are high (and the choices among possible courses of action are stark), it is reasonable to put time and effort into determining the relative epistemic status of disagreeing parties. By contrast, when the claims in question are inconsequential, reserving judgment in the face of disagreement may well be the most prudent thing to do.

---

[19] See Benjamin Wald, "Dealing With Disagreement: Distinguishing Two Types of Epistemic Peers," *Spontaneous Generations* 3 (2009): 113-122.

[20] Wald, "Dealing With Disagreement" 121.

By extension, we demand of epistemic peers (and, even more so, of experts whom we entrust with policy advice) not only that they be as reliable and well-informed as us, but also that they share, by and large, our commitments as to *what it is important to know*. Epistemic peers should not only get their facts right, but should also agree on which facts it is important to get right.[21] Or, if this seems too strong, two epistemic peers – beyond being equally knowledgeable – should at the very least be equally aware of the limitations of their own knowledge. As Philip Kitcher puts it, *reflective* ignorance – that is, being ignorant about the truth value of a first-order proposition, but believing, correctly, that one does not know – constitutes an improvement on *mere* ignorance, since the former "can be the start of something better – of an inquiry that can lead to valuable knowledge."[22] A further improvement would be a state of *Socratic* ignorance, when an epistemic agent is reflectively ignorant with respect to a given claim p and also holds a correct belief as to the relative importance of knowing the correct answer to the question of whether p.[23] In order for epistemic agents to count as peers, they should exhibit a similar degree of reflective awareness of both the character and limitations of their own knowledge, as well as of the extent to which it speaks to live issues of concern.

The significance of being aware of one's own epistemic predicament – not least with respect to one's larger epistemic environment – is not adequately reflected by traditional definitions of epistemic peerhood in terms of either (individual) epistemic virtues or mere reliability on a given occasion. In most cases of persistent disagreement, the relative epistemic standing of the disagreeing parties is far from self-evident. Rather than taking epistemic peerhood for granted and battling over the correct normative response to *prima facie* instances of disagreement, epistemologists would be well-advised to pay greater attention to the causes of disagreement and its persistence, and to the many ways in which peerhood can be undermined by tacit commitments or failure of reflective awareness of one's own epistemic predicament.

[21] Even when the fact in question is itself one concerning the relative importance of, say, different empirical findings.

[22] Philip Kitcher, "How Ignorant? Let Me Count The Ways," (Manuscript of a talk delivered at ZiF, University of Bielefeld, 31 May 2011), 5.

[23] Kitcher, "How Ignorant?" 5.

WHAT IS INTERESTING?

Stephen GRIMM

ABSTRACT: In this paper I consider what it is that makes certain topics or questions epistemically interesting. Getting clear about this issue, I argue, is not only interesting in its own right, but also helps to shed light on increasingly important and perplexing questions in the epistemological literature: e.g., questions concerning how to think about 'the epistemic point of view,' as well as questions concerning what is most worthy of our intellectual attention and why.

KEYWORDS: epistemology, value, interest, curiosity

Finding out the truth with respect to certain topics clearly seems to matter more than finding out the truth with respect to other topics – at least, when we consider things from 'a practical point of view.' Thus finding out the truth with respect to whether North Korea has nuclear weapons clearly seems to matter more, from a practical point of view, than finding out the truth with respect to how many grains of sand are in some random patch of the Sahara. Or again, finding out whether human beings are contributing to global warming clearly seems more significant, from a practical point of view, than finding out how many times the word 'the' appears throughout this paper.

Does the same hold, however, when we look at things not from a practical point of view but rather from a 'purely epistemic point of view?'[1] That is, does it again seem that finding out the truth with respect to certain subjects matters more than others? It would appear so. Take the question about the grains of sand again, and compare it to the sorts of 'pure' scientific questions that are being pursued by the researchers at CERN in Switzerland: for instance, the question of whether corresponding particles of dark matter exist for every known particle of matter. By most lights, finding out whether this thesis about dark matter is true would seem to be of vastly greater epistemic or intellectual importance than finding out the truth about the sand, even if no practical benefit were to emerge from either true

[1] At this point in the paper I am assuming for the sake of argument that a clear distinction between the 'epistemic point of view' and the 'practical point of view' can be maintained, or at least that there is some important difference between the 'points of view' appealed to here. One point that will emerge later on is that this distinction is not as viable as many have supposed.

belief.[2] Indeed, it is tempting to think that finding out the truth about the grains of sand is not only vastly less important but that it matters not at all – again, at least if we are considering things from an epistemic point of view.

But what is it that accounts for the greater epistemic interest or importance of some topics over others – of the topics addressed at CERN, for instance, as opposed to topics concerning the grains of sand? What is it, in other words, that makes the one sort of topic more epistemically interesting or important than the other?

These questions seem worth asking for a number of reasons. For one thing, and self-reflexively, they seem worthwhile in their own right: figuring out what is interesting or important, from an epistemic point of view, itself seems interesting or important, from an epistemic point of view. For another, the questions seem relevant to the disputed issue of what it even means to evaluate beliefs 'from an epistemic point of view' (an issue I will return to in the following section).

But the questions also seem worth asking because of their relevance to wider issues concerning what is worthy of our intellectual attention and why.[3] Notice, for instance, that when as professors we try to stimulate a love of our subjects in our students (a love of philosophy, or of history, or of physics), it looks like what we are doing is trying to get our students to appreciate what it is that is interesting or even fascinating about these subjects. The idea is therefore not that these subjects are worth studying only because of their relationship to other goods – of acquiring a job, perhaps, or of impressing people at cocktail parties. Rather, the idea is that these subjects are interesting or important 'in their own right' – jobs and cocktail parties aside.

I take it that it would therefore be good to be able to articulate just what it is about certain subjects that makes them interesting or important in this way, and I will try to take a first step in that direction in this paper. Although I will try to

---

[2] For some advocates of this claim see Thomas Hurka, *Perfectionism* (New York: Oxford University Press, 1993), 100; Susan Haack, *Evidence and Inquiry* (Oxford: Blackwell, 1993), 199-203; Michael DePaul, "Value Monism in Epistemology," in *Knowledge, Truth, and Duty*, ed. Matthias Steup (New York: Oxford University Press, 2001), 173; Philip Kitcher, "Veritistic Value and the Project of Social Epistemology," *Philosophy and Phenomenological Research* 64 (2002): 191; Michael Bishop and J. D. Trout, *Epistemology and the Psychology of Human Judgment* (New York: Oxford, 2005) 93; and Robert Roberts and W. Jay Wood, *Intellectual Virtues: an Essay in Regulative Epistemology* (New York: Oxford University Press, 2007), chap. 2.

[3] As Dennis Whitcomb, "Curiosity was Framed," *Philosophy and Phenomenological Research* (forthcoming), and "The Problem of Epistemic Significance" (Manuscript); and Roy Sorenson, "Interestingly Dull Numbers," *Philosophy and Phenomenological Research* (forthcoming), note, these issues are especially significant for universities and governments trying to decide how to allocate scarce resources.

make the case that 'interestingness' is an objective property of certain topics or questions, towards the end of the paper I will also note some of the ways in which interestingness seems to be subjective or person-relative.

I. Alston and Goldman

I suggested a moment ago that this set of issues was relevant to the question of what it means to evaluate beliefs from an epistemic point of view. Exploring this connection further should help to highlight its relevance to contemporary controversies in epistemology, as well as to distinguish our questions from a few different ones in the neighborhood.

What does it mean, then, to evaluate a belief from an epistemic point of view? One natural response here, and the one influentially endorsed by both William Alston and Alvin Goldman, is that to evaluate a belief from an epistemic point of view (or, more briefly, just to evaluate it 'epistemically') is to evaluate the belief from the point of view of our epistemic goals and concerns.[4] Consider for instance the belief of the hospital patient who manages to convince himself, in the teeth of the evidence, that he will recover from his illness. Although this belief might well earn high marks from a practical or prudential point of view (the belief might, for example, help to keep his spirits up in his final days), since it goes against his evidence it will presumably not score well from an epistemic point of view, where it is only our epistemic goals and concerns that matter.

But while this line of thinking is obviously appealing, it naturally invites us to ask just what our epistemic goals and concerns are, exactly, and this is a question that is harder to answer than it might first appear.[5] Thus while it might seem that the primary thing that matters to us from an epistemic point of view is simply to acquire the truth – that is, to acquire the truth with respect to any and every topic – Alston and Goldman are both quick to argue that this cannot be quite right, because there are countless topics, what we might think of as 'trivial' topics, that do not seem worth caring about at all, from a purely epistemic point of view. Thus as Alston notes, if acquiring the truth were all that mattered, then it looks like it would be at least prima facie worthwhile, from an epistemic point of

[4] See, for example, William Alston, "Concepts of Epistemic Justification," in his *Epistemic Justification* (Ithaca: Cornell University Press, 1989), 83-84; and Alvin Goldman, *Knowledge in a Social World* (New York: Oxford University Press, 1999), chap. 3.

[5] Again, later in the paper I will raise further doubts about whether a distinction between the 'practical' points of view, as opposed to the 'epistemic' point of view, can really be sustained.

view, to spend our time memorizing phone books – which he takes to be absurd.[6] And Goldman too offers a long list of topics that seem to be altogether lacking in epistemic significance.[7] Given that our interest in acquiring the truth does not seem unrestricted, both therefore agree that (in Goldman's words), "We can no longer suggest that higher degrees of truth possession are all that count in matters of inquiry."[8] Instead, both claim that what matters from an epistemic point of view is not just finding out the truth with respect to just any topic but rather finding the truth with respect to "topics of interest or importance to us."[9] Once our epistemic goals and concerns are relativized in this way – to topics of interest or importance to us – it might then seem that we have a principled way of ruling out the phone book truths (and the like) from the list of our epistemic goals and concerns, and ruling in the things that really are of interest or importance to us.[10]

But something about this refinement is not quite right; more exactly, it looks like the concern about trivial truths that Alston and Goldman were trying to sidestep has not really been sidestepped at all. After all, acquiring the truth with respect to even the most trivial of topics might well be *of* interest or *of* importance to me. Thus finding out the truth with respect to the 323rd number in the Wichita, KS phone directory[11] might well be *of* interest or *of* importance to me – most likely, if I wanted to phone that person up, or perhaps if I were a fact checker for the phone book. Closer to home, virtually all of the topics that occupy my mind throughout the day (sans pure daydreaming, perhaps) are presumably *of* interest or *of* importance to me, even though it would be bizarre to call finding out the truth with respect to most of these topics something I care about from a 'purely epistemic point of view.' Thus finding out the truth with respect to topics like

---

[6] William Alston, *Beyond Justification: Dimensions of Epistemic Evaluation* (Ithaca: Cornell University Press, 2005), 32.

[7] Goldman, *Knowledge in a Social World*, 88; see also Alvin Goldman, "The Unity of the Epistemic Virtues," in his *Pathways to Knowledge: Private and Public* (New York: Oxford University Press, 2002), 61.

[8] Goldman, *Pathways to Knowledge*, 61; cf. Goldman, *Knowledge in a Social World*, 89.

[9] Alston's words; Goldman just says "matters of interest." It is also worth noting that Alston's way of putting things here (*Beyond Justification*) represents something of a change from his earlier essay "Concepts of Epistemic Justification." There he characterizes the epistemic goal in terms of "maximizing truth and minimizing falsity in a large body of beliefs," 83-84, and does not relativize this goal to topics of interest or importance.

[10] For more on restricted vs. unrestricted truth goals, see Stephen Grimm, "Epistemic Goals and Epistemic Values," *Philosophy and Phenomenological Research* 77 (2008): 725-744; and Stephen Grimm, "Epistemic Normativity," in *Epistemic Value*, eds. Adrian Haddock, Allan Millar, and Duncan Pritchard (New York: Oxford University Press, 2009).

[11] An example from Goldman, *Knowledge in a Social World*, 88.

whether my train to work is delayed or on time, or where my dentist's office is located, or what is on sale at the super market, etc., are all (at a given time) clearly of interest or importance to me, but the interest or importance of these topics would seem to be wholly practical. It should be clear, indeed, that finding out the truth with respect to virtually any topic might come to be of interest or of importance to me, provided I had the right incentive.[12]

It seems that what Alston and Goldman need, then, in their characterization of our purely epistemic goals and concerns, is a notion along the following lines: perhaps, of topics that are not just of interest or importance to us in any old way, but rather that are of interest or importance to us, from a purely epistemic point of view; alternatively, what they need is a notion of topics that are simply interesting, given that when we call a topic interesting we already seem to have an epistemic focus in mind.[13] As far as I can tell, however, Alston and Goldman nowhere offer an account of what such a notion might look like, so as things stand their analysis of the 'epistemic point of view' is significantly incomplete.

A few further distinctions are worth bearing in mind, as we try to zero in on our topic. First, it seems entirely possible to acknowledge that some topic is interesting, and yet not be particularly interested in it. For one thing, there might be so many draws on our time that we simply can't focus our interest – our attention or mindpower – on the interesting thing. More significantly, there might also be occasions where we find a topic interesting and yet 'lose interest' in the topic, not because we come to think that we are mistaken about its 'interestingness,' but rather because we come to think that we don't have the ability to grasp or comprehend the truth about the topic. The fact that Stephen Hawking's *A Brief History of Time* was on *The New York Times* best sellers list for several years, for instance, suggests that most people found the main topic of the book – roughly,

[12] Even memorizing a phone book might be of interest or of importance to me, if some maniac were threatening my life unless I got the numbers right.

[13] Michael Brady argues, relatedly, that what we need is a notion of "interesting truths" (rather than interesting questions) to make sense of the notion of our "epistemic aims or goals" (Michael Brady, "Curiosity and the Value of Truth," in *Epistemic Value*, eds. Adrian Haddock, Allan Millar, and Duncan Pritchard (New York: Oxford University Press, 2009) 281-82). And Marian David claims that our interest in the truth is most plausibly restricted to "important and interesting propositions" (Marian David, "Truth as the Primary Epistemic Goal: a Working Hypothesis," in *Contemporary Debates in Epistemology*, eds. Matthias Steup and Ernest Sosa (Malden: Blackwell, 2005), 299). For further appeals to the notion of interesting truths, see Roberts and Wood, *Intellectual Virtues*,157-59; and Jason Baehr, "Credit Theories and the Value of Knowledge," *Philosophical Quarterly* (forthcoming).

cosmology – extraordinarily interesting.[14] But from anecdotal evidence at least it seems clear that after reading a few pages most people put the book down in disappointment. Why? Not because, I submit, they came to think that the topic was uninteresting, but rather because they came to think that they did not have the cognitive wherewithal to tackle it.

A second point is that even though several philosophers have recently attempted to explain the notion of the epistemically interesting or important in terms of the notion of curiosity, this seems like a mistake.[15] For while it is true that we sometimes gesture towards the intrinsic epistemic interest of a topic by saying that we are 'just curious' about it, there are many topics that we are 'just curious' about which nonetheless do not seem, on the face of it, to be interesting. Suppose you hear a familiar song, but you can't for the life of you remember who sang it. You might then be intensely curious about this topic – you might then, indeed, drop everything to try to figure it out – but it would seem like a mistake to call the question of who sang a certain song an interesting one. Oftentimes, when we are 'just curious' about a certain topic, this is simply because we are sensitive to a gap in our information about the topic.[16] But not all of the topics that fall within those gaps are interesting ones.

As we begin, then, we need to keep in mind not just that a topic (the location of my dentist's office, say) might be *of* interest to us and yet we might not

[14] This nice example is from Michael Brady, "Curiosity and Intellectual Virtue," (manuscript).

[15] See especially Whitcomb, "Curiosity was Framed," and "The Problem of Epistemic Significance." For other attempts to explain the notion of epistemic importance by appeal to curiosity, see Goldman, *Knowledge in a Social World*; Gilbert Harman, *Change in View* (Cambridge: MIT Press, 1986); Philip Kitcher, *Science, Truth, and Democracy* (New York: Oxford University Press, 2001); and Nenad Miscevic, "Virtue-Based Epistemology and the Centrality of Truth: Towards a Strong Virtue Epistemology," *Acta Analytica* 22 (2007). For insightful criticism of this move, see Brady, "Curiosity and the Value of Truth," and "Curiosity and Intellectual Virtue." Thanks to Dennis Whitcomb for drawing some of these sources to my attention.

[16] For an early advocate of the 'gap' model of curiosity, see William James, *Principles of Psychology* (New York: Oxford University Press, 2001), 429, and for a more recent, detailed defense George Loewenstein, "The Psychology of Curiosity: a Review and Reinterpretation," *Psychological Bulletin* 116 (1994). In response, some psychologists – e.g. J. A. Litman, "Curiosity and the Pleasures of Learning: Wanting and Liking New Information," *Cognition and Emotion* 19 (2005), and J. A. Litman and Paul Silvia, "The Latent Structure of Trait Curiosity: Evidence for Interest and Deprivation Curiosity Dimensions," *Journal of Personality Assessment* 86 (2006) – have suggested that there are really two distinct kinds of curiosity: curiosity as a feeling of interest, and curiosity as a feeling of deprivation. If this is right, then it would be curiosity as a feeling of interest (or as a response to the interesting) that we are trying to track here.

take it to be interesting, but also that we might take a topic to be interesting, and yet (in some sense) not be particularly interested in it, perhaps because we think we lack the ability to grasp the truth about the topic. In both of these cases, however, there seems to be some underlying sense of topics that 'are interesting in their own right,' or perhaps, of topics that are worthy of our interest, from an epistemic point of view. That is what I would like to focus on here.

II. The strategy

So where should we begin? Since there is very little literature on this topic, the answer to this question is not entirely clear, but I think any account of the interesting should at least try to respect our judgments about the following cases:

Uninteresting	**Interesting**
location of my dentist's office	fundamental scientific laws
number of grains of sand in a random patch of the Sahara	the existence of God
323rd number in Wichita, KS phone directory	[?]
number of redheads in Beiseker, Alberta[17]	

Topics on the left-hand side seem easy to multiply; I have just taken a few of the examples we have discussed so far and added a new one (from Thomas Hurka) to fill things out. For the right-hand side, it seems best to err on the conservative side, at least to begin with, and suppose that if any topics count as interesting surely topics having to do with the basic laws of science, or with whether God exists, should do so.

With these judgments in mind, one way to try to illuminate our notion of the interesting, and the strategy that I shall adopt in what follows, will be to try to identify what look like certain basic topics (or, perhaps better, questions) of epistemic interest, relative to which the items on the right could then be seen as

[17] From Thomas Hurka: "Plainly some beliefs are more worth having than others. It is better to know a fundamental law of the universe than the number of redheads in Beiseker, Alberta, or the workings of a friend's personality than the exact length of his forearm. On the most attractive view, the value of someone's knowledge depends on two factors: how many truths she knows, and their quality or importance" Hurka, *Perfectionism*, 100.

instances – that is, basic topics that seem to underlie our interest in the topics that plausibly belong on the 'interesting' side of our ledger.

But which topics are plausibly of 'basic' epistemic interest? In the following sections I will make the case that three topics (or questions) underlie a great many – perhaps even all – of the topics that we take to be of basic epistemic interest. Namely:

What is there?

How does it work?

How did it get to be that way?[18]

After considering these questions in more detail in Sections 3-5, one issue will be whether this exhausts the list of basic topics of epistemic interest. Another will be whether these questions are all equally basic,[19] or whether instead one most basic question explains the interest of the others. Although this last result would perhaps be particularly satisfying, at the outset we should be open to the possibility that there are simply a plurality of basic questions that are interesting in their own right.

3. Question #1: What is there?

At first glance, the basic epistemic interest of the "What is there?" question seems uncontroversial. What are the far reaches of the universe like, exactly? And what are its fundamental parts? Even if we could not clearly identify a practical benefit that came from exploring the universe's limits, or from discovering its basic building blocks, it would still seem to be worth doing; it would still be interesting to find out what kind of world we lived in. This is an interest, moreover, that seems to be reflected in Aristotle's famous claim at the beginning of the Metaphysics that we take a natural pleasure in looking around us – in seeing how things are, or in finding out what our world is like.

But even though the "What is there?" question might seem to be as good of a candidate for a basic topic of epistemic interest as one might find, it immediately invites the worry that it gets our list from a moment ago all wrong. After all, whether there are so-and-so many grains of sand in a certain patch of the Sahara is

[18] Thanks to Karen Snetselaar for pointing me in the direction of these questions. She claims to have found them on the website of the Anthropology Department at Berkeley, as part of a description of the scientific method. I have not been able to find mention of them on the department's current site, however.

[19] Or, alternatively, whether they are in fact basic, if one is not happy with the idea of things being more or less basic.

presumably a question about what there is in the universe (or this patch of it). And similarly for questions about the number of redheads in Beiseker, Alberta, and so on. Should we suppose, then, that questions about the sand and the redheads really are interesting after all?[20] Or should we instead suppose that the "What is there?" question is not, in fact, a question of basic epistemic interest?

I think that the right answer lies between these two extremes, and that the way to see this is to notice that the "What is there?" question that plausibly drives people like the researchers at CERN is in fact more specific than its surface form might suggest. More exactly, I think that the interest fueling the "What is there?" question is best thought of not as an interest in cataloging every detail of the world but rather in figuring out what kinds of things there are in the world, or the sort of stuff that makes it up. If it is appropriate to think of the world as a beast, then what we have an intrinsic interest in doing, it seems, is identifying its joints – in accurately distinguishing, say, the various elementary particles from one another, or the various chemical elements, or the various biological species.[21]

How does this way of understanding the basic epistemic interest of the "What is there?" question help? By my lights, because it suggests that the reason why we think that finding out the truth about topics like the grains of sand in some random patch of the Sahara is not interesting is because we think there is no joint of nature there – there is no kind of thing 'this patch of the Sahara,' so we are not interested in the properties of this non-thing (such as how many grains comprise it). And the same thing, I think, holds for the redheads in Beiseker and

---

[20] Though perhaps just of such modest epistemic interest or importance that their interest typically gets swamped in favor of topics of greater interest and importance. This seems to be the line taken by Jonathan Kvanvig, *The Value of Knowledge and the Pursuit of Understanding* (New York: Cambridge University Press, 2003), and "Pointless Truth," *Midwest Studies in Philosophy* 32. Also Michael Lynch, *True to Life: Why Truth Matters* (Cambridge, MA: MIT Press, 2004), chap. 4, and "Values of Truth and Truth of Values," in *Epistemic Value*, 227.

[21] Indeed, understood in this way I think we get at a question that drives not just scientific inquiry but metaphysical inquiry as well. What both sorts of inquirers are interested in finding out, in their different ways, is what kinds of things constitute the world, or what it is made up of. Compare Ted Sider: "The goal of inquiry is having one's belief state accurately reflect the world, which in addition to lack of error and lack of triviality requires one to think of the world in its terms. The ideal inquirer must therefore carve the world at its joints, otherwise her beliefs do not adequately conform the world," (Ted Sider, *Writing the Book of the World*, manuscript, 50). Sider's claim here assumes realism about structure rather than argues for it (52). Although I think this is a legitimate assumption, it is clear that anti-realists will want to get off the boat at this point (or perhaps, with respect to at least some of these joints, such as the biological joints, rather than others).

even more obviously for things like the 323rd number in the Wichita phone book. Since there is no joint of nature there, these topics do not inherent any interest from our interest in the "What is there?" question, given that this question is more accurately understood as an interest in the "What kinds of things are there?"

Further support for this way of thinking about the "What is there?" question can also be found by reflecting on some of the things that human beings as a matter of fact find particularly interesting, even particularly fascinating, in the world around them. Think, for example, of things like fireflies, or water bugs, or polar bears. One of the things that makes these creatures so fascinating, I suggest, is the way that they confound or blur the joints that we take the world to have. Thus fireflies for instance have a property – crudely, the ability to light up[22] – that at least for many people animate things not only do not have but (in some sense) cannot or perhaps even should not have.[23] Similarly, water bugs do things that commonsensically look to be impossible for animate things – namely, glide across the surface of the water – and polar bears do something we think should be impossible for huge, ravenous, warm-blooded mammals – namely, survive in an environment where virtually all other animals would quickly die from exposure and lack of food.[24] On the view I am proposing here, the reason why these topics (or phenomena) stand out is because they seem to blur the joints that we commonsensically take nature to have; they seem to behave in a way that seems impossible for things that we are inclined to categorize alongside them. By hypothesis, then, since we are naturally interested in learning about the joints of nature – in the kinds of things there are in the world – we are even *more* interested in things which seem to violate these joints, or to blur them. Perhaps even: violations of these joints rivet our attention because of a nagging doubt that perhaps we have nature wrong.[25]

---

[22] Less crudely, bioluminescence.

[23] If however bioluminescence were as common as non-bioluminescence actually is, then fireflies would arguably lose their special interest, because they would not be seen as joint-blurrers. I return to this idea in Section 10.

[24] This also, I think, accounts for why people since have perennially found 'monsters' or 'freakish' things so interesting: the five legged horse, the elephant man, the bearded lady, etc. On the theory on offer here, the interest or fascination with these topics derives from the interest we have in carving up nature in the right way – in getting some grasp of what things can and cannot do.

[25] I would argue that this is also the deeper explanation for why we find novelty so interesting, as psychologists beginning with Daniel Berlyne, *Conflict, Arousal, and Curiosity* (New York: McGraw-Hill, 1960), have noted.

One final benefit of thinking of the "What is there?" question in this way is that it helps to shed light on the passage from Aristotle's *Metaphysics* mentioned earlier. Although many have cited this passage – and in particular the claim that "All men by nature desire to know" – to support the idea that we have a natural interest in knowing anything and everything about the world (including grains of sand truths, redhead truths, and so on), reading a bit further suggests that Aristotle had something more specific in mind:

> All men by nature desire to know. An indication of this is the delight we take in our senses; for even apart from their usefulness they are loved for themselves; and above all others the sense of sight. For not only with a view to action, but even when we are not going to do anything, we prefer seeing (one might say) to everything else. *The reason is that this, most of all the senses, makes us know and brings to light many differences between things.*[26]

The passage in italics is significant, because what it suggests is that on Aristotle's view when we are looking at the world or taking it in, we are not just doing this in any random way: rather, what we are trying to do is to locate "the differences between things." But what does it mean to locate the differences between things? Presumably, it means to mark out how one sort of thing differs from another – how they differ, for instance, in their properties, or in their causal powers. That is to say, it seems to be a search for joints (or perhaps, as I will suggest later, for structure).

In the following two sections I will now turn, though at less length, to the "How does it work?" and "How did it get to be that way?" questions, and begin to ask how (if at all) these three questions might be related.

IV. Question #2: How does it work?

Evidence for the basic interest in the "How does it work?" question seems easy to find. Young children, for instance, seem to have an instinctive interest in figuring out how things like zippers, buttons, or latches work – that is, in figuring out how the various elements of these things interact with one another. Zippers, buttons, and the like are thus working systems that naturally attract our interest, but our interest in figuring out how things work also seems to apply to much larger systems. Scientists for instance often speak of having a natural interest in figuring out how either the universe as a whole, or some particular part of it, works.[27]

---

[26] Aristotle, *Metaphysics*, Book I.

[27] See, for example, Richard Feynman, "The Pleasure of Finding Things Out," in *The Pleasure of Finding Things Out*, ed. Jeffrey Robins (New York: Basic Books), 1999, and Sander Bais, *In Praise of Science: Curiosity, Understanding, and Progress* (Cambridge, MA: MIT Press, 2010),

In all of these cases it would seem that when we are interested in how a system (large or small) works, the driving desire seems to consist in figuring out how the various elements of the system relate to, and depend upon, one another. One benefit of thinking in these terms, moreover, is that it helps to shed light on why the laws of science should be so universally regarded as interesting, for what these laws seem to tell us is precisely how different elements in the world are related. Thus a law such as Newton's f=ma tells us how the world 'works' in the sense that it encodes information about how properties such as force, mass, and acceleration relate to, and depend upon, one another. That is, what a law of this sort encodes is information about how changes in the value of one of these parameters will lead (or fail to lead) to changes in the value of one of the other parameters, ceteris paribus. The basic interest of the "How does it work?" question therefore helps to explain why it is that we are so interested in identifying the fundamental laws of nature.

What's more, just as we suggested that the special interest of 'joint-blurrers' can be explained in terms of our more basic interest in getting the joints right, so too does the present understanding of the "How does it work?" question help to explain the special interest we have in phenomena that that seem to violate our sense of how the world works. Why? Here again it is plausible to think that the violations attract our interest because of our concern that perhaps we do not have the relationships right; on this view, the 'interestingness' of the violations derives from our prior interest in identifying the relationships themselves.[28]

---

chap. 1. Our natural interest in figuring out how things work, moreover, seems not just to apply to physical systems in the world (or in how the world as a whole works) but also to more abstract systems. Consider, for instance, the natural interest we often have in figuring out how a certain series of numbers "works" (e.g. 3, 8, 12, 42…). In this case, what we are interested in is identifying the rule that explains the series, or upon which the series depends. For more on this see Roy Sorenson, "Interestingly Dull Numbers," *Philosophy and Phenomenological Research* (forthcoming).

[28] This view also helps to explain why certain topics that are apparently not interesting 'in themselves' might well become interesting when considered against the backdrop of our basic interest in identifying how the world works. Suppose, for instance, you were to tell me that the number of redheads in Beiseker is equal to the number of brunettes and blondes combined. Given that this violates the rule or pattern I have come to expect, I will naturally be interested in figuring out what it is that accounts for the difference. The important thing to note, however, is that the primary object of interest here is presumably the pattern (or its violation) rather the fact about the number of redheads in Beiseker considered in isolation. The fact that all sorts of things (topics, facts) might 'become interesting' when considered in light of a particular pattern should therefore not be confused with the thought that the things are interesting or worth learning about in themselves.

Although brief, these remarks about the "How does it work?" question also reveal ways in which it overlaps significantly with our earlier "What is there?" question. For instance, if the "What is there?" question is at bottom inspired by a desire to identify the joints of nature, then it might seem that among the joints of nature we will find precisely the sorts of dependency relations that codified in laws or law-like relationships. 'Joints,' on this broad understanding of the term, would not just be found at the boundary of one physical particle and another, or one biological species and another; rather, joints would also be found wherever we find relationships of nomological dependency.[29]

Second, if we suppose that differences among kinds of things are due to differences in causal powers, then one compelling thought is that the reason why we are interested in carving nature at the joints is because of our even more basic interest in tracking the different causal powers that things have. If joints mark differences in causal powers (among other things), then our interest in joints could then be explained in terms of our interest in figuring out how the various parts of the world relate to one another – or perhaps better, in how they would relate to one another, were they to interact.

If this is right, then it would seem not just that there is a significant overlap between the "What is there?" and the "How does it work?" questions, but that the "How does it work?" question would have a fair claim to explanatory primacy. That is, a case can be made that the reason why we are interested in keeping track of joints is because we are interested in keeping track of differences in causal powers, and the reason we are interested in keeping track of differences in causal powers is because of our interest in grasping how the various parts of the world relate to one another (or would relate to one another, were they to interact). By contrast, taking the "What is there?" question as basic makes it a bit mysterious why we should be particularly interested in tracking joints in the first place, as opposed to counting grains of sand or redheads or what have you.

V. Question #3: "How did it get to be that way?"

We can turn now to our last candidate for a basic topic or question, the "How did it get to be that way?" question. The thought here is that with respect to almost any X, finding out 'the story of X,' or the causal history of X, is interesting to at least some extent: the story of the cotton gin, the story of the finches on Galapagos, the story of Rutherford B. Hayes, and so on. Again, even when nothing

[29] Looked at in this way, the search for joints could also be profitably taken to be a search for structure, where 'structure' perhaps better picks out the factors that mark the various parts of the world off from one another, but that also links them. For more on this, see the next footnote.

of practical importance seems to hang on finding out how things came to be a certain way, the topic seems to naturally spark our interest.

Of course one salient question here is whether this formula really holds for just any X, or even for most Xs (is the story of this saucer of mud really interesting? or the story of the unknown person next to you on the subway?), but rather than try to flesh out this question in more detail in this section I will mainly focus – even more briefly than before, in order to round out this part of the discussion – on asking how if at all the "How did it come to be that way?" question relates to our two previous questions.

Here again, I take there to be significant overlaps. For instance, when I am interested in how a person came to be a certain way, what I am presumably interested in is identifying the factors from the person's past that helped to shape her: the overbearing father, the inspirational book, and so on. That is to say, what I am looking for are the factors in virtue of which she is the way she is, or the factors upon which her current character depends. But if that is right, then it would seem that in asking the "How did it get to be that way?" question I am once again interested in identifying dependencies – not dependencies that hold at the present moment, but rather dependencies that hold between the current character of a thing and some aspects of the past. If we think of 'joints' in the broad sense mentioned above, so that there are joints in nature wherever there are relationships of dependency, then one might think that our interest in joints that fuels the "What is there?" question applies to these cross-temporal dependencies as well.

Perhaps a more natural overlap can be found, however, between the "How does it work?" question and the "How did it get to be that way?" question. After all, in asking how something came to be a certain way, I seem again to be interested in identifying factors from the past that helped to shape its present state. Suppose, for example, it makes sense to think of a person such as Rutherford B. Hayes as a system stretching through time, a system with elements that depend upon one another across slices of time. In figuring out how the Hayes system 'works,' I would therefore be figuring out how the later slices of the system depend upon the earlier slices – that is to say, I would be figuring out how the later slices of the system "got to be that way."

I realize that these remarks – here and in the previous two sections – about the three candidates for basic questions I originally proposed are quite broad-brush,[30] I think our discussion so far nonetheless points to a few significant

[30] I have, for example, claimed that in a broad sense 'joints' can be found not just between (say) different elementary particles but also wherever there are real dependency relations in the

conclusions. First, that what ties together the three basic questions we have considered so far is a common interest in dependence – in particular, an interest in figuring how the various elements of a system relate to, or depend upon, one another. Second, and perhaps more controversially, that this interest in dependencies seems to be best expressed, at its most basic level, in the "How does it work?" question. In other words, that we are not interested in identifying joints or structure in nature 'for their own sakes,' but rather because we are interested in grasping or figuring out how the variously elements of the world are structurally connected—that is, in grasping how they relate to, or depend upon, one another.

In the remainder of the paper, in any case, I will adopt these claims as working hypotheses and now turn to ask: Are there other good candidates for questions of basic epistemic interest that we have so far left out?

VI. Is That It?

In this section I will suggest that any list of questions of basic interest that stopped with our previous three questions would be essentially incomplete, for two reasons.

First, arguably the most interesting, the most fascinating, question has not yet been considered. This is the "How should I live?" question – the question that virtually every human being confronts at some point in their lives, and that Socrates for one took to be of greater interest than any other. Any account of the interesting that left this off the list would have very little to say for it, it seems to me.

Second, there are certain topics that are so universally regarded as interesting that we should try to make sense of them in some way. Here I have in mind topics involving the lives of others, and the rich and famous in particular. As George Loewenstein notes, it is a conspicuous fact about human beings that,

> After one consumes a large restaurant meal, any additional food seems unappealing; however, even after a dinner companion has regaled us with the

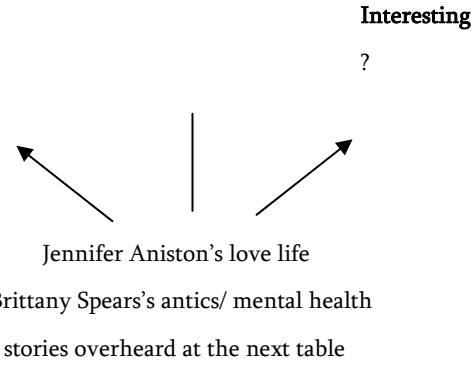
world, both synchronically and diachronically. But it is not obvious that even a broad understanding of 'joint' is broad enough to cover all these roles. Indeed, perhaps it would be better here to speak not of 'joints' but of 'structure.' Sider, for one, seems to use these terms more or less interchangeably: "Discerning 'structure' means discerning patterns. It means figuring out the right categories for describing the world. It means 'carving reality at its joints,' to paraphrase Plato. It means inquiring into how the world fundamentally is, as opposed to how we ordinarily speak of or conceive of it as being" (Sider, *Writing the Book of the World*, 9).

> latest gossip about mutual friends, the muffled conversation at the next table retains its distractive potency.[31]

In other words, the muffled conversation at the next table retains its distractive interest, even when nothing of practical importance seems to hang on it. When it comes to information about the latest doings of the rich and famous, moreover, this natural interest is if anything amplified, e.g., with respect to topics like whether Jennifer Aniston will ever manage to find true love, or in the latest antics of Brittany Spears. Here again, these topics are so obviously interesting to a huge number of people that any account of 'interesting' questions should try to accommodate them in some way.

I will return to the "How should I live?" question in a moment, but one initial reaction that some might have to our interest in others is worth noting: namely, that while the lives of others, and especially the rich and famous, are undeniably *of* interest to huge numbers of people, this interest is driven more by something like a base love of gossip than by anything like the actual 'interestingness' of these topics. Indeed, according to Jason Baehr,[32] beliefs about celebrities amount to a kind of "epistemic rubbish," the having of which actually "undermines or contaminates" someone's mind by distracting him from other, more genuinely "worthy" or "interesting" topics.[33] Although Baehr does not say just which topics qualify as genuinely worthy or interesting, it is clear that on his view the travails of Jennifer Aniston, Brittany Spears, and the like do not.

So where should we put topics like the following on our chart:



[31] Loewenstein, "The Psychology of Curiosity," 81.
[32] Baehr, "Credit Theories."
[33] Baehr, "Credit Theories," 11-12.

If topics of this sort belong on the right, it would be good if we could explain why they belong on the right – what it is that makes these topics legitimately interesting. If they belong on the left, we should be able to explain why people are so systematically misguided about this: that is, why they take these topics to be so interesting when in fact they are not.

VII. The right

At the risk of glorifying People magazine, in this section I want to make a case that these topics do in fact belong on the right: that topics like the trials and triumphs of other people, and perhaps of the rich and famous in particular, deserve to be counted as interesting, and worthy of our attention.[34] Why? I want to say: because of how importantly they bear on the question of "How should I live?" which I mentioned a moment ago, and which seems to be an undeniably interesting topic. Let me explain.

Although presumably everyone wants to live well (or to flourish, or to be happy), it is not at all obvious how to live well, or what constitutes living well. Many people think, for example, that living well involves acquiring a great deal of money, or being famous, or enjoying sustained sensual pleasure. Others think that living well involves being virtuous, or exercising one's highest abilities. And so on. Moreover, even after one comes to think that a particular sort of life is the best, it seems clear that the other lives still hold their appeal. Thus even the person who comes to think that (say) the moderate, virtuous life is the best will still (or so it seems) often be tempted by the lives of fame or power or sensual pleasure – and vice versa. Alternative lives often retain their attractiveness even when we think we know better.

With that in mind, I want to suggest that one reason why the rich and famous are so (legitimately) interesting is because they provide actual examples of people who have managed to achieve certain goods that virtually everyone finds attractive. And what we are plausibly interested in, when we consider their lives, is how the possession of these goods bears on their happiness. Does the money, for instance, lead to more misery than contentment? Does the fame interfere with their ability to form relationships, or to have a stable marriage, or well-adjusted children? The rich and famous are interesting, then, because they are something like experiments in living. Suppose one actually achieved these goods – what else would one have to give up? And is it worth it? These are questions that cannot be

[34] Notice that I am not saying that an obsession with celebrities is healthy, just an interest.

answered from the armchair, but which we can make some progress on learning about the lives of others.[35]

Once one appreciates this point, moreover, it is clear that the rich and famous are not the only people who can shed light on what it takes to live well. Since presumably everyone is trying to live a good life, it is not surprising that we take a natural interest in how their lives are faring – in whether they seem happy or satisfied – because we ourselves have the same goal. Moreover, even once we have come to think that a certain life is best, since it is often not clear what that life requires of us from day to day we look to other people for examples: not just examples of successes, but examples of failures, of people who had the same goal that we did and yet were unable to realize it for one reason or another. By hypothesis, at least part of the reason why we find these stories of success and failure of such interest – at least part of the reason why we find them so interesting – is because they help to guide our own project of living well.[36]

So (no surprise) other people are interesting, and especially their stories of success and failure. On the theory here, moreover, the reason why they are interesting is because they help to shed light on the basic question of how to live well.[37] We are now in a position, finally, to try to connect up these various thoughts about the nature of the interesting.

---

[35] Indeed, one reason why I think the divorces and rehab stints of celebrities is of more interest to us than their occasional long-lasting celebrity marriage is because they show in a dramatic way not only that these goods are not sufficient for happiness, but even that they interfere with it.

[36] Again, I emphasize the 'at least part.' For of course, our interest in the trials of the rich and famous might have other sources as well: e.g., it might be motivated by the perverse pleasure we feel in seeing others fail, and especially the powerful. In the usual case, I suspect our motives are mixed.

[37] As C. A. J. Coady notes: "Most people are curious to know the truth not just about the physical world or mathematics but also about the deeds and misdeeds of other people. Sometimes this curiosity has a functional point in orienting us towards the people with whom we are going to interact: this is the idea that social psychologists are getting at with their somewhat simplified talk of norm reinforcement and the like. But sometimes the satisfaction of the curiosity is simply fascinating in itself, even if it can also be useful. We are interactive social beings who spend a great deal of our lives in conversation, much of it about other people. News about their journey through life with its pitfalls and triumphs is intrinsically interesting to most people, and it is often even more interesting when we know or believe that we are not going to hear it from them." (C. A. J. Coady, "Pathologies of Testimony," in *The Epistemology of Testimony*, eds. Jennifer Lackey and Ernest Sosa (New York: Oxford University Press, 2006), 259.)

VIII. The bottom

Added together, we have now considered a few different kinds of topics, or questions, that are prima facie of basic interest. There were our original three:

> What is there?
>
> How does it work?
>
> How did it come to be that way?

And to that list we added one more:

> How should I live?

The first three topics or questions, I have suggested, are broadly motivated by an interest in how the world hangs together or 'works': in the sorts of causal powers things have, or in the various ways in which things depend upon, or relate to, one another. And the last question is something like an ethical or existential one, the timeless question of how we should live our lives.

It might be thought, moreover, that the final question bears no deep relationship to the other three; that maybe they are both 'interesting' in there own way, and that's all that can be said. But there is a different story that one could tell about how these various topics are related, and one that seems to have a lot going for it: namely, a story on which our interest in these dependence relations is not intrinsic – not something we are interested in 'for its own sake' – but rather derives from our more basic practical interest in prediction and control. James Woodward puts this idea in the following way:

> I suggest… that the distinguishing feature of causal explanations… is that they are explanations that furnish information that is potentially relevant to manipulation and control: they tell us how, if we were to change the value of one or more variables, we could change the value of the other variables…. [This] has the advantage of exhibiting an intuitively appealing underlying rationale or goal for explanation and the discovery of causal relationships: if these are relationships that are potentially exploitable for purposes of manipulation, there is no mystery about why we should care about them.[38]

Woodward speaks here of "causal explanations," but his point would seem to apply to explanations that appeal to dependence relations more broadly. After all, what causal explanation afford, if Woodward is right, is information about how the various parameters in question relate to, or depend upon, one another. But then this general thought scheme would seem to apply to relations of

[38] James Woodward, *Making Things Happen: a Theory of Causal Explanation* (New York: Oxford University Press, 2003), 6-7.

dependence that go beyond what we normally think of as causal relations: e.g., mereological explanations, relations between the past and the present, and so on. In any case, if Woodward is right, then our interest in figuring out how the world hangs together or works in this way is again not intrinsic but arguably instrumental. That is, it is an interest that derives from our still more basic interest in predicting and possibly controlling how the around us world will unfold. And it is no mystery why we would want that: if we can predict and possibly control how the world will unfold then we can try to steer it in a way that promotes our well-being.

So how do these thoughts bear on the 'interestingness' of the question of how we should live? My proposal is as follows: what we are primarily interested in is figuring out how to live well – in figuring out what it takes to flourish, or what constitutes happiness. This is why we find the lives of others so interesting, because they help us to work out both what sort of life is the best or most fulfilling as well as what it takes to live out that life on a daily basis. But the crucial point to appreciate, and the point that helps to draw together all of the questions that we have considered so far, is that coming to a conclusion about what life is best is really only half the battle, for we still have to try to bring that life about. That is, we still have to take steps to make ourselves famous, or rich, or powerful (or what have you).[39] Yet notice that we can't do any of this unless we know how the world works; that is, unless we know how to bring these conditions about. Put another way, we cannot take steps towards living well unless we know the sorts of effects our choices will have, or how the world will respond to our choices. The reason why topics having to do with how the world works are interesting, I submit, is because of their deep and essential connection to our interest in living well.

In short, to the question of why we find the affairs of other people so interesting, my answer is: because of the light they shed on living well. And to the question of why we find figuring out how the world works so interesting, I want to say: because coming to some conclusion about the best life to live is only half the battle. We still need to bring that life into existence, and that requires worldly know how. Finally, to the question of why we find "How should I live?" so interesting, my view is: we just do. The "How should I live?" question is the most

[39] It might be thought that one can take steps to make oneself more virtuous, say, without knowing much about the world, if one thinks that whether one is virtuous is primarily an internal affair (a disposition of the will, say). But even if that picture of virtue is accurate, one still needs to provide for the external conditions necessarily to remain alive, and that would require knowing how the world works.

interesting question on offer, and the one that on this account explains the interestingness of all the rest.

IX. Further issues

That is my attempt to illuminate the nature of the interesting. Essentially, topics are interesting because they bear some sort of special, or perhaps necessary, relationship to our well-being. In this section I will consider a few different objections to the account.

The first is that if this way of thinking about the interesting is on the right track, then it threatens to erase one of the most popular distinctions in the epistemological literature, and one which I myself appealed to right at the beginning of this article: namely, the distinction between what matters to us 'from a purely epistemic point of view' as against what matters to us 'from a practical point of view.' For if things matter to us from a practical point of view just in case they impact our well-being, and if it turns out the underlying reason why topics are of epistemic significance or interest to us – that is, are interesting to us – is because of their special relationship to our well-being, then it begins to look like our 'purely epistemic interests' will in fact be practical at heart.

Rather than think of this as an objection to the account, however, I would suggest instead that we take it as a kind of discovery or clarification. After all, and as we saw with Alston and Goldman, although appeals are often made to something like 'our purely epistemic goals and concerns,' the way this idea is typically developed leaves us with no good sense of what counts as one of our purely epistemic goals or concerns. Worse, given the way the notion of the 'purely epistemic' has been relativized to topics 'of interest' to us, it looks like things like finding out where my dentist's office is located might count as one of my 'purely epistemic concerns' – again, not an appealing result.

That said, I do not want to claim that there is no way to preserve the distinction between things that matter from a purely epistemic point of view as opposed to from some other point of view (practical, moral, etc.) – indeed, in a moment I will explore one such suggestion. If the discussion so far has been accurate, however, then one conclusion I do want to draw is that such a distinction is more complicated than has often been supposed.

A second objection is that in many cases the connection between the 'interestingness' of a topic and our well-being seems implausibly remote; put another way, the objection is that a connection to our well-being is not a necessary condition on interestingness. Suppose for instance that my son comes to learn about solar eclipses in school, and wants to know everything about them; by

his lights, they simply could not be more interesting. But since it is implausible to suppose that his interest in eclipses derives from a desire to predict or control his environment in a way that favors his well-being, then it seems more straightforward to suppose his interest here is in some way basic or intrinsic, or in any case, not just of instrumental significance.

There are two ways one might respond to this concern. The first would be to say, again with Woodward, that even though the 'interestingness' of figuring out how the world works might originally have been fueled by a desire to predict and possibly control one's immediate environment, it should come as no surprise that over time interestingness would attach to systems beyond our immediate environment, or outside of our control. As he writes:

> On this view, our interest in causal explanation represents a sort of generalization or extension of our interest in manipulation and control from cases in which manipulation is possible to cases in which it is not, but in which we nonetheless retain a concern with what would or might happen to the outcome being explained if various possible changes were to occur in the factors cited in the explanans. If we had been unable to manipulate nature – if we had been, in Michael Dummett's example, intelligent trees capable only of passive observation – then it is a reasonable conjecture that we would never have developed the notions of causation and explanation and the practices associated with them that we presently possess.[40]

On this view, we should not suppose that the interest of things outside of our control is in some sense intrinsic; instead, it is best thought of as a kind of offshoot or hangover of our original instrumental interest, now applied outside its original domain. Along the same lines, Peter Lipton suggests that even though our interest in figuring out how the world works "has gone far beyond our practical concerns, this overshooting is not particularly surprising.... [for] we know that activities and traits originally caused by practical considerations may run way beyond the reasons for which they were originally selected, rather as an inclination to save potentially useful objects may lead to philately."[41]

A second way to respond to this objection does not deny that these topics are genuinely worth pursuing for their own sake – it does try to explain away, say, my son's implicit sense that finding out how eclipses work is intrinsically worthwhile – but it claims that this view only makes sense in light of a particular conception of how we should live, or of what constitutes the best sort of life. For notice that even though we earlier focused on accounts of well-being that focused

[40] Woodward, *Making Things Happen*, 11.

[41] Peter Lipton, "What Good is an Explanation?" in *Explanations: Styles of Explanation in Science*, ed. John Cornwell (New York: Oxford University Press, 2004), 15.

on attaining fame or wealth or pleasure, we should not lose sight of the fact that among the traditional list of contenders for the best sort of life has always been the life of contemplation – or, perhaps, more broadly, a life of intellectual endeavor. On this view, the best life is precisely the life that exercises the highest or noblest thing in us – our minds or our reason or our intellect – strictly for its own sake, or because it is good for creatures like us to exercise our highest powers.

But if the life of contemplation or intellectual endeavor is indeed the best sort of life, then grasping how the world works would then be worthwhile not just instrumentally but intrinsically as well – or, better, because figuring out how the world works would partly constitute living well, or achieving the best sort of life. The picture would then be as follows: although at one level interestingness attaches to topics because of their deep though nonetheless instrumental connection to our well-being, at another level certain topics are indeed worth pursuing for their own sake, or because doing so constitutes living well. Moreover, finding out the truth with respect to certain topics – topics in fundamental physics, for example – might be worthwhile for more than one reason: on the one hand because of our instrumental need to figure out how the world works (if we are to bring about a life that we take to be good), and on the other hand because living well (on this view) is constituted by exercising one's intellectual talents in grasping the deep structure of the world.

In this section I have focused on just two objections (again, the objection that it wipes away the distinction between the epistemic point of view and other points of view, and the objection that it cannot make sense of the intrinsic, non-instrumental interest or importance that many topics seem to possess), but obviously there are more issues to be considered. For instance, I do not have a good answer to the question of why certain topics that are essentially related to our well-being nonetheless do not seem interesting. To wit, topics like where I can find food or drink or shelter seem essentially related to my well-being, and are obviously of real interest to me – but intuitively, a topic like where my next meal is coming from is not an interesting one. It is of great interest, no doubt; but as we noted at the outset, topics can be of interest, even great interest, without (apparently) being interesting.

Another question that needs to be explored is why certain topics might shift from being interesting to uninteresting over time. Suppose People magazine, for instance, were to chronicle some celebrity's eighth stint at a rehab clinic. Rather than judge this topic to be interesting, we might well think of it as tiresome. This is a problem for the theory proposed above, because it seems like the travails of others should be abidingly interesting to us, but perhaps (by way of response) there

simply comes a point where we 'learn our lesson' from someone's behavior – we have the relevant data about what this sort of life is like – and we do not continually have to relearn it. Then again, perhaps there are other reasons why a topic might shift from interesting to tiresome for one person, reasons that have to do with the relativity of the interesting. I will consider that way of looking at things in the following, penultimate section.

X. Subjective or objective?

We can begin to draw our discussion to a close by addressing a question that I have touched on in passing a few times but have yet to consider adequately: namely, the question of whether interestingness is an objective property of a thing, a subjective property, or some sort of mix. For instance, although we spoke earlier as if interestingness was an objective property that some things possessed (especially, certain questions or topics), this has seemed to some psychologists to be patently mistaken:

> The central flaw of this model is that there is no evidence that some things are inter- esting to nearly everyone – variability is clearly the norm. In our research on interest and curiosity, we see huge variability in the extent to which people find pictures, poems, text, random images, classical paintings, and social encounters to be interesting…. [N]ovelty and complexity are subjectively appraised, not objectively discerned. Nothing is necessarily novel or complex to everyone, so it is unrealistic to assume that some things are interesting to everyone.[42]

Something about this claim clearly seems right, I think: some event that I take to be extremely interesting you might find quite dull. Moreover, it seems right that whether we take the thing at issue to be interesting will often depend not just on (a) whether the thing is unexpected or unusual (as Silvia and Kashdan note), but also on (b) whether the truth with respect to whatever topic is at issue is in some sense 'open' or undecided. It is worth exploring for a moment just how these variables make our judgments about the interesting person-relative, before we return to the question of how these subjective elements of the interesting might be squared with our earlier claims about its objectivity.

To get a better sense of the subjectivity of the 'unexpected' consider again our example of the fireflies, but now imagine them in a world where virtually all creatures are bioluminescent (an Avatar-like world, perhaps), and where only a

[42] Paul Silvia and Todd Kashdan, "Interesting Things and Curious People: Exploration and Engagement as Transient States and Enduring Strengths," *Social and Personality Psychology Compass* 3/5 (2009): 787.

few rare things – our common houseflies perhaps – are not In that world, it seems fair to say that the interesting things would be the houseflies, and that things like fireflies would be taken for granted. Or again, consider phenomena that seem to violate commonly accepted laws or principles – perhaps, a phenomenon such as Brownian motion at the turn of the 20th Century. Although this phenomenon will strike those who have a misguided sense of what the laws allow as extremely interesting (as, indeed, it struck the physics community at the time as extremely interesting), it is not hard to imagine a community that had the laws right from the outset, which would fail to find motion interesting. For them, this is just what one would expect, given how the world works.

The interestingness of a particular question or topic can also depend, as just suggested, on whether we take the question to be suitably 'open.' Take for instance one of our earlier paradigm examples of an interesting question, the question of whether God exists. If our earlier list of questions of interest (or, of interesting questions) was even roughly correct, it is not hard to see why this question should be so universally regarded as interesting. If God exists, after all, then any account of the kinds of things there are would be massively incomplete without such a being.[43] Again, if God exists then he would presumably be intimately responsible not just for the universe 'working' one way rather than another but also for it 'getting to be' one way rather than another. And of course, if God exists then this would have an enormous impact on our views about how we should live, or about what sort of life is best.

That said, suppose you die and come face to face with God, making it abundantly clear that he exists. It would then seem that the question of whether God exists would no longer be an interesting one, for you, and the reason seems to be that the question is no longer suitably 'open.' You now have a definitive answer to your question. Or suppose you are an atheist who is fully convinced that God does not exist. In that case, the question of whether God exists would then presumably be no more interesting to you than the question of whether leprechauns exist. Even though you realize that many people are extremely interested in this question, you will think they are misguided. Judgments about interestingness therefore seem to depend, in certain cases, on one's current stock of beliefs.

Apart from these influences on our judgments of interestingness, it is noteworthy, finally, that sometimes our judgments of interestingness result from a subtle interplay of subjective and objective elements. Suppose you are learning

[43] Even though, traditionally, God is no kind of thing – no species. He is, rather, the being that brings species into existence and sustains them.

about special relativity as a teenager and you are told that clocks rocketing through space move more slowly than clocks here on earth. That clocks behave in this way, we can imagine, will strike you as interesting primarily because it seems so (subjectively) unexpected; broadly speaking, that this happens violates your commonsense picture of what the world is like, or of how it works. But now suppose you grow up to become a leading expert in special relativity. It does not seem a stretch to suppose that this clock behavior will still strike you as quite interesting, even perhaps as still fascinating. If that was indeed the case, the reason why would seem to be that despite the fact that the behavior was entirely expected, based on your knowledge of how the world works, the commonsense view of the world would still have enough of a grip on you to make it seem continue to seem unexpected, still a bit startling, at another level.

Suppose that the commonsense grip went so deep, moreover, that virtually everyone naturally regarded the clock behavior as interesting, even fascinating – even after they came to see that it was to be expected given relativity. Would it make sense to think of the behavior as objectively interesting, given that it was wedded to a particular (optional) picture of the world? Here I think it is not clear what to say, but it nonetheless helps to bring out how contingent factors can bear on our judgments of interestingness.

How, then, can we square these two aspects of the interesting? On the one hand, that we take certain topics or questions to be objectively interesting – so that people who fail to be moved by, say, the "What is there?" question, and thus claim not to care what the far reaches of the universe are like, are objectively failing to track an interesting question. And, on the other hand, that whether we take certain phenomena to be interesting – such as fireflies, or Brownian motion, or what have you – might legitimately vary from person to person, in accordance with the person's background beliefs?

Now, one tempting way to try to resolve this conflict is to notice the subtle difference in the way I just presented the alternatives in the last paragraph: on the one hand, as a claim about interesting topics or questions and on the one hand a claim about interesting phenomena.[44] Taking a cue from this difference, one natural thought is that the two claims can be reconciled by supposing that objective interestingness is a property only of questions or topics, rather than phenomena (taking the notion of phenomena broadly, so that it encompasses concrete things as well as events). When it comes to phenomena, by contrast,

[44] Notice again, for example, that when Silvia and Kashdan point to the variability of the interesting they point to things such as "pictures, poems, text, random images, classical paintings, and social encounters" – in other words, to concrete things or phenomena, rather than to particular questions.

whether one finds a given thing or event legitimately interesting will depend almost entirely on one's expectations and background beliefs. Thus someone who failed to find the question of how he or she should live interesting would thereby be deficient – that is, be failing to appreciate a question or topic that really is interesting – whereas someone who failed to find fireflies interesting would not necessarily be deficient at all.

Although I think this way of squaring the two ideas is very attractive, and probably in broad outline correct (it is where I would put my money, in any case), it brings to light some significant difficulties with the categories we have been using so far. For example, suppose someone claimed not to find the Big Bang interesting. It would then seem that, just as with someone who failed to find the "What is there?" question interesting, the failure would lie more with the person than with the Big Bang (as it were). But how exactly should we think about the Big Bang? If it is an event, then it looks like events (as concrete phenomena), and not just questions or topics, can indeed be objectively interesting. But it also does not seem like too much of a stretch to think of the Big Bang as not just an event but also a topic – moreover, a topic that is intimately connected to what looks like the intrinsically interesting question, "How did we get to be this way?" So is the Big Bang only interesting because of how importantly it bears on the question of how we got to be this way? Or is it interesting in its own right (as an interesting event, e.g.)? Rather than attempt to answer these questions (I am not sure how to), I will simply flag them, and note that they present a problem for the proposal above, that only questions or topics might be the bearers of objective interestingness.

In any case, and despite the problems for this way of marking the difference between objective and subjective interestingness, unlike Silvia and Kashdan I do think that there is a difference here to be marked, and that the most promising route appeals to this distinction between questions or topics, on the one hand, and concrete phenomena, on the other.

XI. Conclusion

Summing up, I hope to have shed light on questions of increasing significance and perplexity in the epistemological literature – questions, for example, concerning how to think about 'the epistemic point of view,' as well as how to think about what makes a topic epistemically interesting (or important or significant). If am right, then at least one influential way of approaching these questions (e.g., of the sort we see in Alston and Goldman) is fraught with problems. Moreover, if I am right then in order to make sense of the claim that certain topics or questions are

intrinsically (rather than just instrumentally) worthy of inquiry, then we need to make appeal to a particular substantive, and controversial, conception of the good life – one on which contemplation, or intellectual endeavor, helps to constitute the best sort of life. Finally, and if nothing else, I hope to have opened up ways of thinking about the interesting, and revealed outstanding questions, that will help to guide further research in this important area.[45]

[45] Thanks to Jason Baehr, Andrew Bailey, Michael Brady, Jordan Glass, Sandy Goldberg, John Greco, Hud Hudson, Ralph Kennedy, Michael Pace, Sarah Wright, and especially Allan Hazlett for helpful comments and discussion on earlier versions of this essay, as well as to audiences at the University of Edinburgh, Loyola Marymount University, and Wake Forest University.

UNKNOWABLE TRUTHS

Paul HUMPHREYS

ABSTRACT: This paper addresses a solution due to Michael Fara to the Church/Fitch paradox of knowability. Fara's solution has significant interest but the paradox can be resurrected within his approach by considering a slightly more complex sentence. The issue of what counts as an epistemological capability for enhanced agents is then discussed with some emphasis on the developmental heritage of agents and their ability to transcend conceptual frameworks.

KEYWORDS: paradox of knowability, epistemic limits, Michael Fara

1. Introduction

The question of where the limits of knowledge lie has usually been discussed with the tacit understanding that the knowledge concerned is knowledge accessible to humans. That is natural, but for philosophers of science, and especially those who hold that science has access to knowledge that lies beyond the epistemic grasp of humans, a broader perspective on the limits of knowledge is required. By 'lying beyond the epistemic grasp of humans,' I do not mean the kind of knowledge that is usually the focus of limit science, the knowledge that will be available once the scientific method has completed its task. For that limit is construed as the limit of science constrained by observations, where observations are observations made by, or accessible to, humans.[1] What I mean here by 'lying beyond the epistemic grasp of humans' is for the potential knowledge to be about some aspect of reality that is inaccessible to human perceptual and cognitive capacities, but may be accessible by enhanced agents or agents with different epistemic capacities that practice an identifiable kind of scientific method.

[1] The infamous 'model-theoretic argument.' Hilary Putnam, *Reason, Truth and History* (Cambridge: Cambridge University Press, 1981) is an exception to this because the 'observational' sentences used there are simply a set of constraints that are formulated in some special sub-vocabulary of the language.

2. The Church/Fitch Argument

What has come to be known as the Church/Fitch knowability result has a direct bearing on this discussion of the limits of scientific knowledge, construed broadly. Here is the special form of the argument, special in the sense that the argument is specifically about knowledge in the standard sense, restricted to human knowledge. There are three premises in the argument. Let the sentential operator K represent 'some epistemic agent at some time knows that' and □the standard modal operator representing metaphysical possibility, that is, truth at some metaphysically possible world. Then the first premise is:

1K. $\forall p(p \rightarrow \Diamond Kp)$

which asserts that for any truth p, it is possible for at least one epistemic agent to come to know that p. Premise 1K is often taken to state a verificationist principle, one that denies the realist position that there are parts of reality, representable in some conceptually accessible language, that it is impossible for any agent ever to know. Note that 1K is not stated as a necessary truth. In the original argument, this could be strengthened because verificationist positions generally link truth conditions and verification conditions so tightly that the concept of an unknowable truth is incoherent.[2] In contrast, the realist position is best construed as contingent. It is certainly metaphysically possible for there to be a world in which every contingent truth is not only knowable, but known; it simply happens that the evidence we have strongly suggests that our world is not like that. A solipsistic and idealistic world inhabited by an inquisitive simpleton is an example of the former, because the only contingent truths that need to be known consist in the few thoughts that pass through the sparsely populated mind of the agent. Once mathematical truths are allowed, the existence of a gap between what is known and what is knowable becomes more plausible.

The second premise is:

2K. $\forall p \, \Box(Kp \rightarrow p)$

I shall take this as the unproblematical assumption that knowledge requires the truth of the sentence that is known.

The third premise is also unproblematical, simply asserting that knowledge of a conjunction entails knowledge of the conjuncts:

3K. $\forall p \Box(K(p\&q) \rightarrow (K(p) \,\&\, K(q)))$

[2] Traditional forms of intuitionism in mathematics take this form.

The Church/Fitch argument is then a straightforward consequence of 1K, 2K and 3K. Take as an instance of 1K the sentence p & ¬Kp. This asserts that there is a true sentence that is never (actually) known by any agent. It is taken as an empirical fact that there are unknown truths in that sense.[3] Typical examples are historical facts, such as the exact number of bacteria in the body of Joan of Arc when she was burnt at the stake in 1431. Assuming clear countability criteria for bacteria and the impossibility of resurrecting the pre-cremated Joan of Arc, this is a good candidate for a truth that no agent has known or ever will know. Although an extremely strict verificationist might reject this sentence as a candidate for knowledge, it is metaphysically possible that the truth of that sentence is known and so satisfies 1K. If the appeal to unobservables such as bacteria troubles you, take instead the number of hairs on Joan of Arc's head at the appropriate time. Using this instance in 1K we have (p&¬Kp) → ◊K(p&¬Kp) and given the truth of p&¬Kp we have ◊K(p&¬Kp). Using 3K, we have that ◊(Kp & K ¬Kp). Since K¬Kp entails ¬Kp, using 2K, we conclude with ◊(Kp & ¬Kp), which is inconsistent.

Once again assuming that 2K and 3K are uncontroversial (and also that the inferences made are unproblematical) we have three choices for what to do.[4] The first is to retain the verificationist principle 1K, and to maintain that sentences of the form p & ¬Kp do not exist, despite the considerations given above. The verificationist would then be committed to the position that:

CV ∀p(p → Kp)

that is, that all true sentences are not just knowable but are known by some agent at some time. That is not an inconsistent position, and it perhaps captures what a strict verificationist is committed to, but at the least it drives the verificationist to deny the existence of truths that correspond to states of affairs that occurred before the evolution of cognitively capable agents, as well as many other kinds of contingently inaccessible facts. This is a sufficiently unattractive position that I shall not pursue it.

The second option is to reject assumption 1K on the basis that the verificationist position is false. A third option is to claim that 1K does not represent what verificationism intends and that an alternative premise should be used. Before discussing this third option, let us generalize the argument. Let O be a sentential operator satisfying the following three conditions:

---

[3] For arguments to this effect see Nicholas Rescher, *Unknowability* (Lanham, MD: Lexington Books 2009). Further reasons are given in section 2 below.

[4] An alternative approach is to reject one of the inference rules used in the proof on the grounds that one or more of them is unacceptable to a verificationist.

1K*. $\forall p(p \rightarrow \Diamond Op)$ (Possibility Principle)

2K*.$\forall p \, \Box(Op \rightarrow p)$ (Veracity Principle)

3K*. $\forall p \Box(O(p\&q) \rightarrow (O(p) \,\&\, O(q)))$ (Conjunction Principle)

From these three premises, if some claim of the form $p \,\&\, \neg Op$ is true, then it is provable that

CK*. $\forall p(p \rightarrow Op)$

Here is a proof that can be found in Rescher's *Unknowability*.

1. $\Box(O(p\&\neg Op) \rightarrow (Op\&O\neg Op))$	by substitution in premise 3
2. $\Box(O\neg Op \rightarrow \neg Op)$	by substitution in premise 2
3. $\Box\neg O(p\&\neg Op)$ modus tollens	From 1, sentential equivalent to 2,
4. $\neg\Diamond O(p\&\neg Op)$	From 3 by definition of modal operators
5. $(p\&\neg Op) \rightarrow \Diamond O(p\&\neg Op)$	By substitution in premise 1
6. $\neg(p\&\neg Op)$	From 4,5
7. $p \rightarrow Op$	From 6.
8. $\forall p(p \rightarrow Op)$	By generalization

3. Fara's Argument

The first part of my discussion will be based on a recent analysis by Michael Fara of the argument.[5] Fara demonstrates that the general form of the Church/Fitch argument can be applied to a number of non-epistemic situations, such as the actions of an omnipotent agent. I shall concentrate here on an application (not discussed by Fara) to scientific agents that have access to knowledge lying outside the domain actually accessible by humans, but constrained by the norms of science so that they are not omniscient. Three features of 1K then require clarification.[6] What counts as the scope of quantification over p is important. I shall take p as representing an arbitrary sentence in some interpreted language, rather than a proposition, so that the concepts occurring in p must be specifiable. p describes some actual and not merely possible fact (state of affairs); the facts can be about individuals or can be generalizations, including laws. The modal operator

[5] Michael Fara, "Knowability and the capacity to know," *Synthese* 173 (2010): 53-74.

[6] I shall assume here that quantifying over the set of all sentences is unproblematical, setting aside here concerns raised by Patrick Grim, "There is No Set of All Truths," *Analysis* 44 (1984): 206-208.

◊ in 1K is usually taken to represent metaphysical possibility. Because the knowledge operator K occurs within the scope of the modal operator, the agents, past, present, or future, must be located in the appropriate possible, usually non-actual, world. 1K is then a curious mixture of epistemic and metaphysical content. The premise says that for every actual truth, in some metaphysically possible world an agent knows that truth. Fara explicitly represents p as being preceded by an actuality operator A, so that the premise 1K becomes

$$\forall p\,(Ap \rightarrow \Diamond KAp) \qquad (A)$$

According to Dorothy Edgington what is needed to capture the anti-realist position is to interpret (A) as[7]:

(p is true at the actual world w) → ∃w'(It is known by some agent at some time at w' that p is true at w) (B)

As Fara then points out, when p is not known in the actual world, the consequent of (B) claims that an agent located outside the actual world knows at least one factual truth about the actual world and "this kind of counterfactual knowledge of actuality seems starkly impossible."[8] We can have knowledge of non-actual situations, because it is commonly held, for example, that we can stipulate for a given world that it have the same laws as our world. But because we are concerned with scientific claims having empirical content, a priori modes of investigation or stipulative claims about what the actual world contains are inappropriate and this extra-worldly epistemic access is impossible. This means that 1K does not correctly represent what a verificationist needs.

Modifying Fara's own solution a little, his interpretation of premise 1K is this:

actually p → actually ∃x(x has the capability to know that actually p) (C)

where 'having the capability to know that' means (a) that the agent possesses the appropriate epistemic equipment, whether exercised or not, that would allow the agent in appropriate circumstances to know that p and b) 'to know that' is an operator that satisfies both 2K and 3K.[9]

We then have by substituting

'p&∀y¬(y knows that p)' (D)

[7] Dorothy Edgington, "The paradox of knowability," *Mind* 93 (1985): 557-568.

[8] Fara "Knowability," 64.

[9] Fara uses the term 'capacity' rather than 'capability' but the former has nuances in the philosophy of science that are best avoided here.

for p in (C):

> actually(p&∀y¬(y knows that p)) → actually ∃x(x has the capability to know that actually [p and that no cognitive agent knows that p])

and, as Fara notes, this does not does not lead to a contradiction because the agent x has the capability to know that (actually) p and also has the capability to know that (actually) no cognitive agent knows that p. In the Joan of Arc example, we can see that, on a reasonable interpretation of 'has the capability to know that,' this is true for the sentence about the number of hairs on Joan of Arc's head. A key aspect of this response to the Church/Fitch argument is that we do not run afoul of the analog of 2K. It does not follow from the fact that an agent has the capability of knowing that p that p is in fact true.

4. Another Problem

Fara's solution to the paradox is appealing, not the least because it has a robustly epistemological focus. But the paradox can be revived in what is perhaps a more worrying form. We can begin by asking whether (C) correctly captures what a verificationist needs. According to Joe Salerno the kind of anti-realism involved in the argument asserts that:

> ...barring vagueness and ambiguity, we could in principle know the truth value of any fully understood proposition, given only finite improvements to our epistemic capacities, resources, or environment.[10]

Key placeholders in this claim are 'We,' 'in principle,' 'fully understood,' and 'finite improvements.' As with most epistemological discussions, the agents are taken without question to be we humans. Concordantly, the 'in principle' refers to idealizations of our existing cognitive capacities, as does the 'finite improvements.' Let us see what happens if we relax this restriction. Because Fara's (C) has no modal force beyond the capability feature, the truth of (C) depends on the contingent development of cognitive agents and science in our world. So we have to make some decisions about where to draw the line between what lies within the domain of applicability of an acceptable capability and what does not. Traditional empiricist-based verificationism about factual truths ties verification procedures tightly to the perceptual capabilities that evolution has provided to humans. Even if not universally agreed upon, it is plausible that evolution could have taken a different course and endowed humans with different kinds of perceptual abilities. In addition, the advent of technological supplements to our

[10] Joe Salerno, "Introduction to knowability and beyond," *Synthese* 173 (2010):1.

perceptual and intellectual capabilities are developments that should be welcomed by a scientifically sympathetic verificationist. On the other side, epistemic agents endowed with supernatural powers would be unacceptable to most verificationists. So I shall take as an acceptable cognitive agent any agent that could have developed during the lifetime of our universe, including agents possessing artificial intelligence or hybrids that combine human cognition with the different capabilities of instruments and computational science, the existence of which is compatible with the actual laws of nature, and including those that for contingent reasons did not come into existence.

We now need to ask whether it is reasonable to argue, under the more liberal account of epistemic agents, that (C) is true. The answer, I think, is that the weakening of 1K from a metaphysical possibility to the actual existence of an agent capable of knowing that p, as (C) asserts, renders (C) false. For if we need sophisticated technology or superhuman computational abilities to know that p, as we do in many scientific cases, then there is evidence that no *actual* cognitive agent, whether one that is human and supplemented by artificial devices, or one that is entirely artificial, is capable of knowing that p for a sentence p the content of which involves an extreme degree of unobservability or uncomputability. Here, the capability can even be taken as capability in principle as long as the idealizations involved do not take us out of the domain of what is actually possible given the laws of nature and the finite cognitive resources available.

To deny the truth of (C) is, of course, to deny what a particular kind of verificationist asserts. I am primarily interested in whether Fara's modification of the argument allows the verfication ist to completely escape the problems stemming from the original argument. As I shall now show, the modification used in (C) is insufficient to completely escape from trouble.The substitution instance (D) appropriately follows the original line of argument in taking a true but unknown sentence as the target case. But consider a different substitution instance of (C):

(D*): 'p&$\forall$y$\neg$(y has the capability of knowing that p)'

where the quantification is over actual agents. Does it follow from the fact that y has the capability to know that p, that p is true? One might dispute this because on some interpretations of capability, it is not true that for every sentence, necessarily if some agent has the capability to know that p, then p is true. Capability is a dispositional property, available for use on a wide variety of applications, if the conditions are right. I have the capability to know that my front door is painted purple were it to be in fact purple, but as a matter of fact it is not. However, Fara restricts himself to actual truths, and any substitution instance of C satisfying the

antecedent will therefore involve actual truths. Now consider what the relevant instance of (C) says under substitution of (D*):

> actually (p & ¬∃y(y has the capability of knowing that p)) → actually ∃x(x has the capability of knowing that actually [p and ¬∃y(y has the capability of knowing that p)]).

Consider the antecedent. p is an actual truth and in the entire history of the actual world, no agent has the capability of knowing that p. Now take the sentence schema

> 'there are exactly N objects of mass 10,000 kg in a closed orbit around star S' (E)

where S is the name of a star, N is a non-negative integer, and S is sufficiently distant from us that the measurement of N would require technological innovations that are never actually carried out, let us say on grounds of the expense needed to make the required instruments. For some N, possibly zero, the sentence schema is true.[11] Call that instance p. Because the necessary instruments were never constructed, no agent possessed the capability of knowing that p. Remember that some cognitive agent in the history of the universe must actually have that capability, not just be in a position to possibly have it. So we can detach the consequent and distribute the capability over the conjunction to obtain:

> actually ∃x(x has the capability of knowing that actually p and x has the capability of knowing that actually ¬∃y(y has the capability of knowing that p).

Simplifying the conjunction, we have that ∃x(x has the capability of knowing that actually p). But we also have from (D*) that ¬∃y(y has the capability of knowing that p). And this is a contradiction since p is an actual truth.

One aspect of this argument needs attention.[12] There is a sense of 'capability' for which it is true that if an agent has, for example, the technological capability to construct an instrument that would provide evidence for a true sentence p, and the existence of that instrument would endow the agent with the epistemic capability to know that p, then the agent has the capability to know that p. There is another sense of 'capability' in which collapsing the successive capability claims is illegitimate. There are two things to mention about this dual

[11] In assuming that some instance of the sentence schema is true, I am assuming that classical logic holds, at least in the form that the truth values of all instances of the schema exist. This assumption is not stronger, I believe, than the factual assumption that there is a true but unknown sentence, an assumption that is used in the original argument.

[12] The need to explicitly address this point was suggested by Majid Razvi and Ken Akiba.

sense of 'capability.' First, Fara makes a distinction between capacities and abilities in this way:

> ... one has the capacity to perform some feat provided one's internal constitution does not rule out the performance of that feat; to have the ability to perform that feat one must, in addition, be disposed to succeed in performing that feat if one tries.[13]

I believe that Fara's account of capacities (called here 'capabilities') does not permit us to say that an agent lacking the physical equipment to carry out a cognitive task has the capacity to do so. Just as a blind person does not have the capacity to perceive visual shapes, so an agent lacking the physical equipment to determine the number of objects orbiting a given star, however contingent that lack is, does not have the capability to know that number. Secondly, since both senses of 'capability' are legitimate, there is one interpretation of (C) under which the inconsistency exists for that stricter kind of verificationist. Moreover, the more liberal version of verificationism, the one that allows the compression of iterated capabilities, will need to provide some kind of criterion about where the boundary between in principle capabilities and in practice capabilities lies.[14] Too far towards the in principle wing and one loses much of the original motivation for being a verificationist. Too far towards the in practice wing and the risk of permitting inconsistencies returns.

To summarize: Fara's modified representation of verificationism, as contained in (C), itself leads to a contradiction taking plausible auxiliary assumptions as true. So if (D*) is true, (C) must be false and there are actual truths that lie beyond the epistemic capabilities of any actual agents. The truth of (D*) depends upon contingent facts about what extensions of scientific capabilities are actually made and what kinds of capabilities are needed to access all factual truths. The first of those facts will influence the strength of what verification position is reasonable, whereas the second will determine the epistemic gasp between a realist position and that verificationist position. Once one reaches the extreme of an agent that is omniscient about factual truths (rather than possible truths), not only will (D*) be false but the modal collapse of knowability into knowledge will not be objectionable in that situation.[15]

---

[13] Fara, "Knowability," 68.

[14] For a discussion of this distinction with respect to traditional empiricism and contemporary computational science, see Paul Humphreys, *Extending Ourselves* (New York: Oxford University Press, 2004).

[15] Thanks to Gene Mills for raising the issue of where on the spectrum between current humans and omniscient knowers we should draw the appropriate epistemic boundaries.

5. Further Issues

We can now explore some more general issues involved with unknowable truths. In our discussion above, we left aside Salerno's requirement that the propositions involved in the argument should be fully understood. It is here that some worries might enter. For (D*) to be false, it is enough for a single agent to be capable of knowing that p, but such a situation would violate standard norms for the objectivity of scientific knowledge. Even for cases of non-scientific knowledge, situations in which the justification for p could not be conveyed to others would be grounds for suspicion but we can address that problem at the price of a small amount of imprecision by requiring that a necessary and sufficient condition for (D*) to be false is that a small class of agents is capable of knowing p.

This requires a specification of the conceptual resources that agents have and that they can understand. For cognitive agents with extended abilities, this specification is not easy to give because the quantification in (D*) is over all agents. Conceptually weaker agents do not have the ability to understand what it is that cognitively more powerful agents can know and a given agent cannot understand, let alone fully understand, a proposition formulated using a set of concepts he or she does not have and cannot comprehend. It is often claimed that scientists of bygone eras could not have known true propositions about contemporary theories such as quantum field theory because they lacked the appropriate concepts. This is no doubt correct for historically situated agents, but the more general issue is whether there are concepts that no agent could understand. Because we are dealing with sentences and not propositions, the linguistic relativity of conceptual resources does seem to preclude languages that are incommensurable with all known languages from being understood even by the cognitively most powerful agents. One kind of reason for maintaining a realist position would thus be to maintain that there are sentences that it is impossible for any cognitive agent to ever know because the concepts involved lie beyond understanding. I am not sure that it is possible to provide any clear content to that claim, if only because it requires the idea of an interpreted language that no agent could ever understand.

When discussing science as it might be conducted by aliens, Rescher says that the science of different civilizations is

> inevitably closely tied to the particular pattern of their interaction with nature as funneled through the particular course of their evolutionary adjustment to their specific environment.[16]

[16] Rescher, *Unknowability*, 26.

and that

> ... in science, as in other areas of human endeavor, we are prisoners of the thought-world that our biological and social and intellectual heritage affords us.[17]

Although various degrees of constraint are mentioned, this conceptual framework is said to "determine what is seen as an appropriate question and what is judged as an admissible solution."[18] Let us begin with the social and intellectual heritage.

At least within our own species, one of the most striking features of science has been its ability to transcend in large part, if not completely, the limitations of the social and intellectual environment, at least in the natural sciences. Saying "It's a science thing, you wouldn't understand" is an exhortation to get into the lecture room and learn something, not a claim that because you are from Nepal or Rwanda, the concepts of quantum field theory are blocked by your culture. The extensive debate about the existence of cultural universals such as kin groups, sexual attraction, and concepts of fairness has revealed the existence of at least several hundred solid candidates for transcultural concepts.[19] So claims about the effects of cultural environments must be about a deep kind of environment that perhaps rests on the biological underpinnings, in ways analogous to the ways in which surface grammars reflect cultural accidents and deep grammars reflect a common linguistic apparatus, because the differences in terrestrial environments that allow a common kind of scientific understanding to exist show that surface environments cannot preclude understanding. There is a plausible case that can be made for a certain degree of first-person perspective knowledge that can only be acquired via cultural embedding (*Verstehen*) but at best it will rule out some features of psychological and social knowledge, even if we allow such knowledge to count as scientific.

So it must be the biological heritage that constitutes the barrier to knowledge at least in the sense that the biological equivalent of embedding yourself in a different cultural environment is possible only to a limited extent. There exist virtual reality simulations that allow one to experience, at least to a certain extent, the experiences undergone by an individual suffering from paranoid schizophrenia. Current evidence suggests that schizophrenia is a heritable trait, so the simulation does allow us to enter a world that is in some way biologically different from, for most of us, our own. It is also worth noting that humans who have had their visual abilities made operative by cataract surgery

[17] Rescher, *Unknowability*, 32.
[18] Rescher, *Unknowability*, 26.
[19] Donald Brown, *Human Universals* (Philadelphia: Temple University Press, 1991).

have been able to acquire concepts they previously lacked.[20] We should also keep in mind that 'naive physics' or 'intuitive physics' is not only very wrong about the physics of our world, but has incorrect conceptualizations about what counts as a natural state and other matters. To overcome this, we use forms of representation that are increasingly remote from our instinctive conceptualizations. So, fractal dimensions, although analogous to spatial dimensions, are not the kinds of things that our ordinary ways of interacting with the world give us, even though fractals are approximately realized in objects such as crystals. This suggests, although it is not presently possible to test it, that a species equipped with different kinds of sensory inputs could arrive at scientific results that we could understand. To take a simple but slightly fanciful example, dogs, which operate in a massively olfactory world that we will probably never properly understand, had they brains of the appropriate level of representational and computational power, could have developed molecular chemistry. This is not necessarily a matter of our results from science converging on results from other ways of doing science; it is simply reflects my scepticism about the degree to which accidental contextual factors can determine what we can and cannot know.

One of the things we have learned from the intensive debates about neural networks is that they possess an enormous degree of conceptual plasticity. Unlike Kantian frameworks in which the fundamental concepts are fixed, inputs leading to a redistribution of the connection weights in the network result in changes in the sub-conceptual representations used by that network. Maybe we could not understand what advanced aliens with connectionist based intelligence know, but they have the capability to understand what we know. Moreover, despite Kuhn's denial, based on a behaviorist approach to concept formation, we can understand what previous scientists did, so this indicates some kind of progress.

I want to finish with a couple of remarks about the notion of limit science. As Nick Jardine noted thirty years ago, the very notion of a limit for scientific activity is not well defined and we have no evidence either that such a limit exists or that science will not perpetually shift from one representational framework to another.[21] These are reasons for not taking seriously discussions of scientific realism and anti-realism, physicalism, and especially what can and cannot be known that are couched in terms of limit science. Although such discussions often take place in the context of the temporal development of science, the issues are equally relevant to the development of reductive programs. Kuhn focused on the successive replacement of

[20] John Z. Young, *Doubt and Certainty in Science* (New York: Oxford University Press, 1960).

[21] Nick Jardine, "The possibility of absolutism," in *Science, Belief, and Behaviour: Essays in Honour of R.B. Braithwaite*, ed. D.H. Mellor (Cambridge: Cambridge University Press, 1980), 23-42.

theories about particular domains over time. Yet if incommensurability arguments based on theory-relative semantic content are valid, they should apply to synchronic reduction as well, presenting serious difficulties for any reductive procedure similar to Nagelian inhomogeneous reduction. For example, 'gold' has a different meaning at the nano-scale than it does at the human scale and similarly with 'water' and 'H_2O,' as Mark Johnston has argued.[22] This synchronic incommensurability affects the program of fundamental physicalism because the evidence in favor of fundamental physicalism is supposed to come in part from progress in reducing other sciences to physics. It also renders suspect the familiar argument by Putnam that because terms such as 'red' can apply to unobservable entities as well as observable entities, the division between the observable and the unobservable cannot consistently be maintained. This kind of argument is fallacious when an inference from the syntactic identity of terms to the sameness of semantic content is made.[23]

[22] Mark Johnston, "Manifest Kinds," *Journal of Philosophy* 94 (1997): 564-583.

[23] Earlier versions of this paper were read at the *Limits of the Knowledge Society* conference, Romanian Academy, Iasi, Romania, the University of Pittsburgh workshop on Scientific Progress, and Virginia Commonwealth University. Thanks are due to each of those audiences for comments.

EVIDENCE AND TRANSMISSION FAILURE

Mark McBRIDE

ABSTRACT: Some philosophers (can be taken to) claim that there are no genuine instances of transmission failure provided we operate with the right account of the sources of warrant-or-evidence for future reasoning. My aim in this paper is to (begin to) clear the way for instances of transmission failure regardless of the account of the sources of warrant-or-evidence for future reasoning with which one operates. My aim is not to claim there are in fact genuine instances of transmission failure; merely to render it possible, on all – or most – plausible accounts of the sources of warrant-or-evidence for future reasoning.

KEYWORDS: evidence, transmission failure, epistemic circularity

0.1 Consider the following argument:

(MOORE) (1) I have hands.

(2) If I have hands an external world exists.

(3) An external world exists.

Our focus in this paper is on (MOORE), but points will generalise to (reasoning by means of) arguments with a similar structure. (MOORE) is clearly valid and its premises are true. Yet, to many, reasoning by means of this argument – call it (MOORE)-reasoning – seems defective. On one view, it is an example of transmission failure. In section 2 we'll come to precisify this concept. For now let's make do with a very rough, first pass at transmission failure. Philosophers locating the defectiveness of (MOORE)-reasoning in transmission failure, typically take (MOORE)-reasoning to suffer from a form of epistemic circularity: it's only if one already has warrant for (3), that one can have a warrant for (1). If (MOORE)-reasoning suffers from transmission failure in this way, we can say that propounding the argument cannot offer one a new route to a warranted belief in its conclusion.

0.2 What's one's warrant to believe (1)? Put differently: what's one's evidence for (1)? (Suppose the conditional premise (2) is knowable a priori.) These questions are badly formed. There are various different warrants one might have for (1). Lots of things could be one's evidence for (1). Suppose I'm sitting blindfolded,

© LOGOS & EPISTEME, II, 4 (2011): 557-574

unable to see my hands, and Edna whispers gently in my ear: "You have hands." Edna's utterance – on one plausible view – counts as good testimonial evidence for me that I have hands. That's one form of evidence I might have for (1). And there are many others. But let's focus on visual-perceptual evidence that I have hands, and specifically the *mere having* of a visual experience *as of* having hands – the having of which is compatible with my failing to have hands.[1] (Indeed, unless we focus on a particular form of evidence it will be indeterminate whether transmission failure occurs.) With this assumption about the nature of my evidence for (1), the idea that (MOORE)-reasoning involves epistemic circularity has some initial plausibility. For, it might be said, it's only if I antecedently and independently have a warrant for (3) that an experience as of having hands constitutes a warrant for (1). If, by contrast, I had good reason to think that (3) was false and that I was a handless brain-in-a-vat with experiences that were being generated by an evil scientist, then merely having a visual experience as of having hands would not be such good evidence for (1).

If (MOORE)-reasoning does involve epistemic circularity, then following through the reasoning could not provide a new route to a warranted belief in (3). Even if one were to allow that it would provide a *new*, gratuitously circuitous, *route*, still it would not be a way of arriving, *for the first time*, at a position in which one had a warrant to believe (3).

0.3 Some philosophers – viz. Timothy Williamson,[2] John Hawthorne,[3] and Nicholas Silins[4] – (can be taken to) claim that there are *no* genuine instances of transmission failure, provided we operate with the right *account of the sources of warrant-or-evidence*[5] *for future reasoning*.[6] As to sources of warrant-or-evidence,

---

[1] An experience *as of* x is an experience representing x, the having of which is compatible with ~x (and, indeed, compatible with x). This may need to be finessed, depending on one's view on object-dependent experience, to the following: An experience *as of* x is an experience representing x, *and the having of an experience with the same conscious character* is compatible with ~x (and, indeed, compatible with x).

[2] Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000), ch.9.

[3] John Hawthorne, *Knowledge and Lotteries* (Oxford: Clarendon, 2004), 34 n.86: "I find most alleged examples of [transmission failure] unconvincing."

[4] Nicholas Silins, "Transmission Failure Failure," *Philosophical Studies* 126 (2005): 71-102. But see Nicholas Silins, "Basic Justification and the Moorean Response to the Skeptic," in *Oxford Studies in Epistemology: Vol. 2*, eds. Tamar Gendler and John Hawthorne (Oxford: Oxford University Press, 2007), 108-140, for an evolution in his views (128 n.27).

[5] Cf. Peter Klein's notion of "sources of justification" (Peter Klein, "Skepticism and Closure: Why the Evil Genius Argument Fails," *Philosophical Topics* 23 (1995): 213-36.)

[6] We can certainly grant to these epistemologists that there is *an* exceptionless notion of what we might call warrant transmission. We would then say that when *other epistemologists* talk

pre-theoretically, or before commitment to any particular theory, warrant or evidence can come in myriad forms: blood and DNA-samples, hearsay, facts, states of affairs, objects etc. Our principal focus, though, will be on two particular possible sources of warrant-or-evidence: an experience and a premise (i.e. a proposition).[7] As to future reasoning, we may take inference to be our paradigm case thereof (as contrasted, for example, with coming to know that from which one infers).[8] My aim in this paper is to (begin to) clear the way for instances of transmission failure *regardless of* the account of the sources of warrant-or-evidence for future reasoning with which one operates. My aim is not to claim there *are in fact* genuine instances of transmission failure; merely to render it *possible* on *all* – or most – plausible accounts of the sources of warrant-or-evidence for future reasoning.[9]

0.4 Silins claims instances of transmission failure only arise on a particular – putatively flawed – *account* of the sources of warrant-or-evidence for future reasoning.[10] Call this *Silins's claim*. Here's a key commitment of this account. (Silins is going to reject the commitment and the account of which it is a commitment. It may help to keep (MOORE) in mind when reading this passage.):

> [Y]ou might know a certain premise, or indeed be absolutely certain that the premise is true, yet still not possess the premise as a warrant for a conclusion you have competently deduced from the premise. On this view, if your inference provides you with any warrant for believing[11] the conclusion at all, the warrant

about there being instances of transmission failure, this suggests that they are using a more demanding notion of warrant transmission than the exceptionless notion. And one possible more demanding requirement – our focus – would be to exclude epistemic circularity.

[7] Some readers may find it hard to think of a premise in an argument as *warrant-or-evidence*. My parenthetical reference to a proposition may make it less hard. Regardless, I ask for patience on the part of the reader, and trust the idea will be clearer by the paper's end.

[8] And so "future" is not the temporal notion where the contrast is with "past" and "present" reasoning. Instead, in our paradigm case, we suppose a reasoner knows a proposition and ask what sources of warrant-or-evidence such a reasoner has at his disposal should he go on (*in the future*; *subsequently*) to perform an inference. Finally, our focus in this paper is on future *deductive* reasoning.

[9] And, indeed, on *all* – or most – plausible *conceptions of evidential warrant* (a term explained in 1.1 *infra*).

[10] Silins, "Transmission Failure," 84-89. I call the account *fallibilism*- (explained in 1.2 *infra*). Strictly it is an account of the sources of warrant-or-evidence for future reasoning *and a conception of evidential warrant*. Silins's claim receives a sharpening later (see n.47 *infra*), but this version will do in the meantime. Finally, I take it this claim can be fairly ascribed to Silins. Nonetheless, should he resist it, the claim itself is worth independent exploration.

[11] My note: For Silins, "when [he] say[s] that someone's warrant *for believing* that p is that w... [he] means that the person believes that p, and that the person believes that p on the basis of

> provided is the prior warrant for which you believe the premise. Thus even if you know that p, or are absolutely certain that p, it is still not the case that your warrant for believing the conclusion is that p.[12]

Sticking with (MOORE), suppose, as before, my warrant to believe (1) is my having an experience *as of* having hands. Nonetheless, suppose I can come to know, on the basis of this warrant, that I have hands. Suppose also, if I know (1), I am certain that (1) is true. Finally suppose I competently deduce – using the conditional, (2) – (3) from (1). Silins now cuts in that, on this (to-be-rejected) account of the sources of warrant-or-evidence for future reasoning, if my inference provides me with any warrant to believe (3) at all, the warrant provided is my experience as of having hands. Crucially, Silins points out, on this account, even though I know that I have hands, and am certain that I have hands, it's not the case that I can use the fact that I have hands as a source of warrant-or-evidence *for future reasoning*. Specifically, I can't use *the premise* that I have hands – a premise, note, whose truth is *incompatible with* my being in an idealistic world – as a source of warrant-or-evidence to believe there to be an external world. Rather, on this account, all I can use is *my experience* as of having hands, combined with inference – an experience, note, the having of which is *not incompatible with* my being in an idealistic world.[13]

Silins's claim that instances of transmission failure arise only on one (to-be-rejected) account is resisted in this paper. One cannot confine the possibility of transmission failure to this one account of the sources of warrant-or-evidence for future reasoning. So, at least, I shall argue.

---

the relevant [warrant, w]." (Silins, "Transmission Failure," 74.) I do not take this to be orthodox usage, but we can grant it for present purposes without harm.

[12] Silins, "Transmission Failure," 88.

[13] It might be worth quickly pointing out what else a philosopher operating with this account *can* do (all assuming he meets the epistemic standards for knowing (1)). First, he can competently deduce (3) and come to believe (3) on that basis: he can allow that, in the exceptionless sense of what we might call warrant transmission (cf. n.6 *supra*), the experiential warrant for (1) is transmitted to (3). Second, if such a philosopher is a *dogmatist* (see James Pryor, "The Skeptic and the Dogmatist," *Nous* 34 (2000): 517-49), he can allow that, even in a more demanding sense of warrant transmission that excludes epistemic circularity, the experiential warrant for (1) is transmitted to (3). Finally – whether dogmatist of not –, he can allow that, if asked why he believes (3), he can legitimately reply: Well, I have hands – I know that much – and (3) follows from what I know.

1. Conceptions of Evidential Warrant and Accounts of Sources of Warrant-or-Evidence for Future Reasoning

1.1 Consider (where, throughout, '*S*' is a placeholder for a subject and '*p*' for a proposition):

> *Fallibilism*: *S* can know *p* when *S*'s evidence for *p* is compatible with ~*p*.
>
> *Infallibilism*: *S* cannot know *p*, when *S*'s evidence for *p* is compatible with ~*p*.

Focusing on (MOORE), fallibilists can suppose my *having an experience as of having hands* is – or can be – my way of coming to know that I have hands.[14] Infallibilists, by contrast, typically suppose my *seeing that I have hands* is my way of coming to know that I have hands, where my seeing that I have hands is incompatible with my not having hands.[15] Think of fallibilism and infallibilism as particular *conceptions of evidential warrant*: they're particular views of what it takes for a piece of evidence to warrant a belief in a proposition for a subject such that that subject can acquire knowledge. A philosopher's conception of evidential warrant is typically associated with a particular account of the sources of warrant-or-evidence for future reasoning. A fallibilist typically takes an *experience* as of having hands as his source of warrant-or-evidence for future reasoning (and indeed sometimes, perhaps often, for (1) itself). An infallibilist typically takes (knowledge of) *(1) itself* – a *premise* – as his source of warrant-or-evidence for future reasoning (and indeed sometimes, perhaps often, for (1) itself).

1.2 Philosophers who've explored the *possibility* of transmission failure in (MOORE)-reasoning (and in like arguments) – viz. Martin Davies, James Pryor, and Crispin Wright[16] – operate with a fallibilist conception of evidential warrant. Let's – exhaustively and exclusively – partition fallibilism as follows:

---

[14] Such fallibilists will be dogmatists. All dogmatists are fallibilists; but not all fallibilists are dogmatists.

[15] Cf. John McDowell, a *fact disjunctivist* (John McDowell, "Criteria, Defeasibility, and Knowledge," *Proceedings of the British Academy* 68 (1982): 455-79, and "Singular Thought and the Extent of Inner Space," in *Subject, Thought and Context*, eds. Philip Pettit and John McDowell (Oxford: Clarendon Press, 1986), 137-68), and Williamson (*Knowledge and Its Limits*, ch.1) for whom seeing that *p entails* knowing that *p*. Note that a fallibilist *may also*, for example, allow that sometimes, perhaps often, we see that *p* – that is, we *enjoy* evidence incompatible with ~*p*.

[16] See Martin Davies, "Epistemic Entitlement, Warrant Transmission and Easy Knowledge," *Proceedings of the Aristotelian Society: Supplementary Volume* 78 (2004): 213-45, and "Two Purposes of Arguing and Two Epistemic Projects," in *Mind, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*, ed. Ian Ravenscroft (Oxford: Oxford University Press, 2009), 337-83, James Pryor, "What's Wrong with Moore's Argument," *Philosophical Issues* 14

> *Fallibilism-*: *S* can know *p*, when *S*s evidence for *p* is compatible with ~*p*, and when *S* knows *p*, *p* itself is not part of *S*s evidence-base for future reasoning.[17]
>
> *Fallibilism+*: *S* can know *p* when *S*s evidence for *p* is compatible with ~*p*, and when *S* knows *p*, *p* itself is part of *S*s evidence-base for future reasoning.

These theses are constructed by combining a conception of evidential warrant (first conjunct: fallibilism) with an account of the sources of warrant-or-evidence for future reasoning (second conjunct: "evidence-base for future reasoning").[18] Precisely what the second conjunct of fallibilism- forbids is a subtle matter (cf. n.13 supra). I take it, somewhat suggestively, to be the following: The fallibilist- denies that a known premise, p, can itself be used in future reasoning, but this leaves open that future reasoning from a known premise, p, can lead to, for example, further knowledge.[19] Fallibilists are typically fallibilists-.[20] But nothing in fallibilism rules out being a fallibilist+.

1.3 Now consider:

> *Infallibilism+*: *S* cannot know *p* when *S*s evidence for *p* is compatible with ~*p*, and when *S* knows *p*, *p* itself is part of *S*s evidence-base for future reasoning.[21]

(2004): 349-78, and Crispin Wright, "Facts and Certainty," *Proceedings of the British Academy* 71 (1985): 429-72, and "(Anti-)Sceptics Simple and Subtle: Moore and McDowell," *Philosophy and Phenomenological Research* 65 (2002): 330-48.

[17] Note that this leaves open that, though *p* itself is *not* part of *S*s evidence-base *for future reasoning*, it *is* part of *S*s evidence-base *for knowing p*.

[18] A note about the *exhaustivity* of these two theses: I take it that *p* either *necessarily is* or *necessarily is not* part of *S*s evidence-base for future reasoning. Consistently with this, contingent matters could be in play rendering *p*, though part of *S*s evidence-base for future reasoning, *not exploitable* (e.g. on account of epistemic circularity). (Alternatively, one might recast the second conjunct of fallibilism- as "*p* itself *necessarily* is not part of *S*s evidence-base for future reasoning" and the second conjunct of fallibilism+ as "*p* itself is *or can be* part of *S*s evidence-base for future reasoning.")

[19] Where "future reasoning *from* a known premise" is, perhaps, elliptical for "future reasoning *from warrant for* a known premise." Finally, such a fallibilist can allow that, if the believed premise, *p*, is not known, perhaps because it is not true, then reasoning from *p* to a conclusion may be above reproach and yet will not normally lead to knowledge.

[20] E.g. Fred Dretske, notwithstanding he's classifiable as an infallibilist on a different way – present in the literature – of carving things up. Finally, I take it dogmatists can be either fallibilists- or fallibilists+.

[21] There is close to being merely *logical space* for *infallibilism-* (*modulo* "evidence-base" being used in the way I describe). At any rate, it's difficult to discern the motivation for such a thesis. Given this, in the main text I quickly revert to simply talking of infallibilists.

As with our previous two theses, this thesis is constructed by combining a conception of evidential warrant (first conjunct: infallibilism) with an account of the sources of warrant-or-evidence for future reasoning (second conjunct: "evidence-base for future reasoning"). To conclude: How do the fallibilist+ and infallibilist+ differ? Each, after all, allows *p itself* into one's evidence-base for future reasoning. Two features distinguish them. First, most obviously, they operate with different conceptions of evidential warrant: for the former – but not the latter – one can know *p* when one's evidence for *p* is compatible with ~*p*. Second, and less obviously, the latter's allowance of *p* itself into one's evidence-base *may well not be confined* to future reasoning (*p* itself, on versions of this view, is part of one's evidence-base *for knowing p*), while the former's allowance *may well be* so confined (on versions of this view).[22] In part because our chief focus in this paper is on *future* reasoning, and in part because we only seek to secure the *possibility* – rather than genuine instances – of transmission failure, neither difference, though important, turns out to be *fundamental* for our purposes.[23]

We've got a (novel) taxonomy of conceptions of evidential warrant and accounts of the sources of warrant-or-evidence for future reasoning to hand. Now let's try to precisify the concept of transmission failure, and locate it within this taxonomy.

2. Transmission Failure and (In)fallibilism

2.1 Consider this (sufficient) condition for non-transmission:

> *(NT)* **Non-transmission of warrant**
>
> Epistemic warrant is not transmitted from the premises of a valid argument to its conclusion if the putative support offered for one of the premises is conditional on its being antecedently and independently reasonable to accept the conclusion.[24]

---

[22] Cf. nn.15 and 17 supra.

[23] Many issues I've opened up in this section require more detailed treatment. For example: What does it take – knowledge? warrant(ed belief)? truth? – for a proposition to be included in one's evidence? Is (possession of) propositional evidence for *p* compatible with ~*p*? On some views your evidence will include the proposition that *p* when you know *p*, but you could have had that evidence even if you warrantedly, yet falsely, believed that *p*. Moreover, my discussion of the presented theses has proceeded in hedged terms with, for example, "often," "typically," and "may well" in place of "necessarily" at many junctures (here and in what follows). I don't see that this is avoidable, however.

[24] Davies, "Epistemic Entitlement," 221.

Let's spell out *one* view of how (MOORE)-reasoning *might* suffer from transmission failure.[25] Let's take the fallibilist-. Suppose the fallibilist- claims[26] that:

> *(AW)* In order for one's visual experience as of having hands to act as warrant for
>
> (1) one must, antecedently, have warrant for (3).[27]

If *(AW)* is true, one's warrant for (1) will depend on an antecedent warrant for (3). So (MOORE)-reasoning will suffer from epistemic circularity and *(NT)* will be triggered. Suppose that one *in fact has* antecedent warrant for (3), so that one's visual experience does count as warrant for (1). Then, according to *(NT)*, the warrant for (1) that is constituted by one's experience as of having hands is not transmitted across the conditional – (2) – to serve as warrant for (3).[28]

2.2 What about our fallibilist+?[29] Consider him debating with our fallibilist-: If you're allowing – as a good fallibilist should – that I can come to know *p* notwithstanding my evidence for *p* is compatible with ~*p*, and allowing that I do indeed come to know (1) on the basis of such *non-entailing* evidence, why are you hamstringing me from using *(1) itself* as evidence in my future reasoning? If you were a sceptic about our knowledge of quotidian propositions, I'd vehemently disagree with you – being a good fallibilist myself –, but I could at least make sense of your prohibition on my use of (1) as evidence in future reasoning: I couldn't so use (1) as I wouldn't – ex hypothesi – know (1). But given – as we can happily

[25] Cf. Davies, "Two Purposes," for another (complementary) way.

[26] Cf. Wright, "Facts and Certainty," "(Anti-)Sceptics," and "Warrant for Nothing (And Foundations for Free?)," *Aristotelian Society Supplementary Volume* 78 (2004): 167-212.

[27] Rejection of *(AW)* is a defining feature of dogmatism. In this section we focus on fallibilists not flatly rejecting *(AW)* – that is, fallibilists allowing that (MOORE)-reasoning might involve epistemic circularity, and so might suffer from transmission failure. Indeed given the paper's overall focus is not to claim there *are in fact* genuine instances of transmission failure, but merely to render it *possible* on *all* – or most – plausible accounts of the sources of warrant-or-evidence for future reasoning, we do not definitively adjudicate on *(AW)* (or its cognate, *(AW*)* – see 3.2 *infra*) at all in this paper.

[28] This failure of transmission is, trivially, not a failure of transmission in the exceptionless sense (cf. n.6 *supra*). It is failure of transmission in the more demanding sense that requires that there not be epistemic circularity. And it seems that this remains the situation even if we regard the warrant for (1) as being *constituted* by *both* the experience as of having hands *and* the antecedent warrant for (3).

[29] Cf. Silins, "Transmission Failure."

agree as fallibilists – I know (1), your exclusion of *it* – *that premise* – from my evidence-base for future reasoning is unmotivated.[30]

Our fallibilist+ makes a good case. So let's grant him – *pro tem* – victory over his fallibilist- foe. Our fallibilist+ may have won this particular battle, but not yet the war. Fallibilist- is a dogged soul. Suppose he – accepting defeat over the inclusion of *(1) itself* in one's evidence-base for future reasoning – marshals *(AW)* and claims that fallibilist+'s (MOORE)-reasoning thereby suffers transmission failure.

If, however, fallibilism+ is true, it's no longer clear *(AW)* leads to (MOORE)-reasoning suffering (complete)[31] transmission failure. Why not? After all, isn't it still the case – thanks to *(AW)* – that *(NT)* is triggered? The wrinkle is that it's no longer clear that *(NT)* is *applicable*, *given the truth of fallibilism+*. And that's because of *(NT)*'s mere talk of "*support…*for one of the premises," and not, additionally, talk of *premises themselves.*[32] *(NT)* seems to, *pace* fallibilism+, preclude *(1) itself* from counting amongst one's evidence-base for future reasoning. In other words, *(NT)* doesn't seem to capture adequately the account of the sources of warrant-or-evidence for future reasoning with which a fallibilist+ operates. Take this to be what I'm driving at in suggesting *(NT)* isn't *applicable* to fallibilism+ (and, *mutatis mutandis*, in subsequent uses of "applicable"). Put differently: the non-transmission of one's *warrant for (1)* – which, by *(NT)*, is one's *support for* (1) – leaves open that *(1) itself* (which one ex hypothesi knows) transmits across the conditional ((2)) to provide warrant for (3). In sum, I think, for all we've said, our fallibilist+ has won another (interim) battle.

2.3 Before responding to our fallibilist+, let's tie together the fallibilist+ and the infallibilist (i.e. the infallibilist+ – cf.n.21 *supra*). To be sure each philosopher – the fallibilist+ and the infallibilist – operates with a different conception of evidential

---

[30] Both of these fallibilists should agree that, provided that one knows (1), one will do well to believe (3), since it follows from (1) by a palpably valid inference. (Of course, if one gets good evidence against (3) then the palpably valid inference may have to be run in the opposite direction.) But, because of possible epistemic circularity, there is more to be said about why one does well to believe (3). The argument via (1) and (2) appears to involve a detour and, in order to understand one's warrant to believe (3) after the inference, one needs to articulate the nature of the antecedent warrant – that is, one's warrant to believe (3) even before the inference.

[31] As will become clear, here (and in what follows), when I talk of "complete" transmission failure I am bracketing any exceptionless notion of warrant transmission (cf. n.6 *supra*) and focusing on a more demanding notion. If there is, in this sense, "complete" transmission failure, it means *all*, and not just *some*, of one's sources of warrant-or-evidence for future reasoning fail to transmit. Germanely to this, I explicitly expand what I call the *standard* account of *warrant* in 3.2 *infra*.

[32] This is evidence philosophers operating with *(NT)* are (closet) fallibilists-.

warrant. But, for the purposes of *this* transmission failure debate, this difference is less important than one's account of the sources of warrant-or-evidence for future reasoning.[33] Of *more* importance, then, is whether or not, when a subject comes to know (1) (or a like proposition), *(1) itself* becomes part of that subject's evidence-base for future reasoning. And on this the fallibilist+ and the infallibilist are united: it does. We thus must alter standard taxonomies. Typically, in debates pertaining to evidence, fallibilists and infallibilists are taken to be the two camps ripe for comparison.[34] But, in this transmission failure debate it's best to group fallibilists- in one corner (by dint of their *refusal to allow* (1)-type propositions into one's evidence-base for future reasoning), and fallibilists+ and infallibilists together in the other corner (by dint of their *allowing* (1)-type propositions into one's evidence-base for future reasoning). We thus might hope that a good response – in the case at hand – to the fallibilist+ will also count as a good response to the infallibilist.

3. Modifying (NT) (and other germane theses)

3.1 The challenge is to see if we can construct a motivated (sufficient) condition for non-transmission of warrant which – unlike *(NT)* – is applicable to fallibilism+ and infallibilism, and which, if the condition is satisfied, precludes (1) itself from transmitting across the conditional ((2)) to provide warrant for (3).[35] A natural first response to putting the challenge this way would be to ask: What does it mean for *(1) itself* to fail to transmit across the conditional ((2)) to provide warrant for (3)? After all (MOORE) is an instance of *modus ponens*. And let's assume a plausible knowledge-closure principle, say: If S knows P and S knows P entails Q, then S knows (or at least is in a position to know) Q. Given this (and given (2) is knowable a priori), if I know (1), I know (or am in a position to know) (3). What room is there for a motivated principle – a modification of *(NT)* – granting this,

[33] Our excursus into conceptions of evidential warrant was still useful. First, standard taxonomies foreground conceptions of evidential warrant: so this excursus helps locate our taxonomy in relation to standard ones. Second, while there are no entailments in play from one's conception of evidential warrant to one's account of the sources of warrant-or-evidence for future reasoning (or conversely), we might think that the former *helps explain* the latter.

[34] Or a divide is effected between forms of *internalism* (typically associated with fallibilism) and *externalism* (typically associated with infallibilism). Effecting such a divide will not serve present purposes.

[35] Ideally this condition would also be applicable to *fallibilists*-. But this seems unrealistic. Nonetheless, a desideratum of parsimony – viz. *ceteris paribus*, construct as few theses as possible – operates in what follows.

yet allowing that (1) itself can fail to transmit across the conditional ((2)) to provide warrant for (3)?

3.2 First, let's distinguish between a closure question and a transmission question.[36] The germane closure question asks: Is it possible to know (1), know that (1) entails (3), yet not be in a position to know (3)? The germane transmission question asks: Is it possible to know (1), know that (1) entails (3), yet not *thereby* – that is, in virtue of recognising the entailment from (1) to (3) – be in a position to know (3)? Or, differently: Is it possible to know (1), know that (1) entails (3), yet not be in a *first time* position to know (3)? Let's answer the closure question: "No," yet remain agnostic about the transmission question.[37] I take it that, even answering the closure question: "No," conceptual room is left open to deny that (1) itself transmits.

Second, let's expand what I'll call the *standard* account of *warrant*. Standardly, one's warrant is one's warrant *for this or that proposition*, and does not include *the proposition itself.* But we're not bound by standard usage. Indeed we're currently granting the fallibilist+ and infallibilist their claims that premise (1) becomes part of one's evidence-base for future reasoning, once one knows (1). So we might modify *(NT)* by making explicit that warrant includes any known premises.[38] Here's a first pass:

> *(NT*)* **Non-transmission of warrant**
>
> Epistemic warrant – where warrant includes any known premises – is not transmitted from the premises of a valid argument to its conclusion if the putative support offered for one of the premises – or knowledge of any of the premises themselves – is conditional on its being antecedently and independently reasonable to accept the conclusion.[39]

---

[36] Cf. Davies, "Two Purposes."

[37] For philosophers answering the closure question: "Yes," cf. Fred Dretske ("Epistemic Operators," *Journal of Philosophy* 67 (1970): 1007-23, "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971): 1-22, and "The Case Against Closure," and "Reply to Hawthorne," in *Contemporary Debates in Epistemology*, eds. Matthias Steup and Ernest Sosa (Malden: Blackwell, 2005), 13-25, 43-6), Robert Nozick, *Philosophical Explanations*. (Cambridge: Belknap, 1981), and Mark Heller, "Relevant Alternatives and Closure," *Australasian Journal of Philosophy* 77 (1999): 196-208. For remarks on how one's operative account of the sources of warrant-or-evidence for future reasoning affects one's answer to the closure question, see Klein, "Skepticism and Closure."

[38] "Warrant" is hereinafter used in this broad sense unless stated, or indicated, otherwise.

[39] Objection: If knowledge of a premise is conditional on antecedent warrant for the conclusion then this will be because epistemic warrant for the premise is conditional on antecedent warrant for the conclusion. So it is unclear why we need to introduce *(NT*)*. Reply: Insofar as

Recall that, to show (MOORE)-reasoning suffers transmission failure by *(NT)*, our fallibilist- marshalled thesis *(A W)*. And we can construct a thesis paralleling *(A W)* – call it *(A W*)* – which makes room for (complete) transmission failure in (MOORE)-reasoning by *(NT*)*:

> *(A W*)* In order for one's warrant for (1) – where warrant includes any known premises – to so act as warrant, one must, antecedently, have warrant for (3).

If *(A W*)* is true, and (as a central case) if one *in fact has* antecedent warrant for (3), (MOORE)-reasoning suffers transmission failure by *(NT*)*. Is, then, this pair of theses – *(NT*)* and *(A W*)* – motivated?[40]

4. Motivating our Modifications

4.1 Let's start with *(A W*)*. As with *(A W)*, I don't want to marshal arguments purporting to secure *(A W*)*'s truth. Instead, here, I simply want to present a way of formulating or expressing a defense of these (modified) theses. Recall, our overall strategy is not to secure *genuine instances* of transmission failure; it's merely to render it *possible*, on *all* – or most – plausible accounts of the sources of warrant-or-evidence for future reasoning. So, how might one go about expressing a defense of *(A W*)*? Let's approach this question indirectly. We'll first consider how one might express a defense *(A W)* and see if that helps us in discerning a way to express a defense of *(A W*)*. Recall *(A W)*:

> *(A W)* In order for one's visual experience as of having hands to act as warrant for (1) one must, antecedently, have warrant for (3).

One (likely) way of formulating or expressing a defense of this thesis would be by making use of notions contained within an account of a relation of *epistemic antecedence* (or *priority*):

> *(EA)* One's warrant for accepting p is antecedent to one's warrant for accepting q iff one's reasons for accepting p do not *presuppose or rest on* one's reasons for accepting q.[41]

---

this objector is adopting my expanded account of warrant, I agree with the premise but not the conclusion. Only *(NT*)* allows us to make explicit the pivotal relation of epistemic antecedence between a *(known) premise* and a *standard warrant* (see sections 4 and 5 *infra*).

[40] So my focus in this section (and those to come) is on *known* premises. What happens when, say, the premise is false but warrantedly believed (cf. n.23 *supra*)? I bracket this question, but see, *inter alia*, Peter Klein, "Useful False Beliefs," in *Epistemology: New Essays*, ed. Quentin Smith (Oxford: Oxford University Press, 2008), 25-61.

Someone defending *(A W)* would likely claim that one's visual experience acting as warrant for (1) presupposes or rests on one's reasons for accepting an external world exists. We don't need to *agree* with that bald statement offered in defense of *(A W)*.[42] Indeed we here offer no reasons for *(A W)*'s (or *(A W*)*'s) truth. We just need to understand the claim so formulated or expressed.

4.2 That's *(A W)*. Now what about *(A W*)*? We might formulate or express a defense of *(A W*)* by making use of notions contained within an account of the following relation of epistemic antecedence:

> *(EA*)* One's warrant for accepting p is antecedent to one's warrant for accepting q – where warrant includes any known premises – iff one's reasons for accepting p do not *presuppose or rest on* one's reasons for accepting q – where one's reasons include any known premises.

We're supposing, recall, that *(1) itself* – which we ex hypothesi know – is part of our evidence-base for future reasoning. In assessing (MOORE)-reasoning armed with notions from *(EA*)*, we're (in part) asking: Does my knowing that (1) presuppose or rest on my reasons for accepting (3)?

Let's make an assumption – an assumption favourable to fallibilists+ in particular. To understand why we're making this assumption we need to understand a claim some (prominent) infallibilists make. Some infallibilists claim that *all and only* what one knows is one's evidence.[43] So for this type of infallibilist our supposition that (1) itself is *part of* one's evidence-base for future reasoning, and that one's evidence-base for future reasoning *includes* (1) itself, is a slightly misleading supposition: for them, (knowledge of) (1) itself *exhaustively constitutes* one's evidence-base for future reasoning (prior to an act of inferring (3)).[44] Any

---

[41] Cf. Pryor, "The Skeptic and the Dogmatist," 525. Should the left-hand-side of *(EA)* strictly be: It's not the case that one's warrant for accepting q is epistemically antecedent (or prior) to one's warrant for accepting p? Couldn't the right-hand-side be satisfied and one's warrant for p be epistemically *neither-prior-to-nor-posterior-to* one's warrant for q? Finally, note that we can also, I take it, say: One's warrant for p_1 is antecedent to one's warrant for p_2 iff one's warrant for p_2 is *dependent* on one's warrant for p_1 and one's warrant for p_1 is *not dependent* on one's warrant for p_2.

[42] Nor do we need to find this notion of one reason presupposing or resting on another perfectly perspicuous; it's left intuitive.

[43] See Williamson, *Knowledge and Its Limits*, ch.9.

[44] An infallibilist *might* claim that one's evidence for (1) consists of knowledge of (1) *and knowledge* of one's having an experience as of having hands. If he does so, the points I make in relation to the fallibilist+ here can be marshalled, *mutatis mutandis*, with respect to this infallibilist. But this is – for good reason – currently an unpopular way to be an infallibilist, so I'll ignore it in what follows. An infallibilist will typically only claim that one's evidence for

pragmatic implication that there must or may be more to one's evidence-base for future reasoning than (1) itself should be cancelled. So I hereby cancel any such implication by explicitly noting that "part" does not mean "proper part," and "includes" does not mean there need be any more evidence. For our fallibilist+, meanwhile, *all* that one knows is in one's evidence-base for future reasoning, but not *only* what one knows: one still, in addition to (knowledge of) (1) itself, has one's visual experience as of having hands as part of one's evidence-base for future reasoning. This, at least, is one plausible way of carving up evidence for a fallibilist+.[45]

What's the purpose of this detour? Recall, the strategy is to explore using notions from *(EA*)* to formulate or express a defense of *(AW*)*, and to claim *(AW*)* opens up the possibility of transmission failure for fallibilists+ and infallibilists by *(NT*)*. Given this, we don't want the fallibilist+[46] to be inexorably lumbered with (complete) transmission failure in (MOORE)-reasoning because *part of his warrant for (1)* – the part he shares with the fallibilist-, viz. a visual experience as of having hands – fails to transmit to (3).[47] To allow for this would be to slur over the important difference between the fallibilist- and the fallibilist+. To preclude this from happening, we make our assumption, viz. *some* of one's warrant can fail to transmit in an argument without *all* of one's warrant failing to transmit.[48] Maybe this assumption seems too obvious to need stating. But note that, without this assumption, if we suppose that *(AW*)*'s truth can be secured (perhaps with use of notions from *(EA*)*), this inevitably leads to a fallibilist+ being committed to (MOORE)-reasoning suffering transmission failure by *(NT*)*. And that's because

(1) consists of knowledge of one's having an experience as of having hands in the *bad case* – that is, when ~(1). But our principal focus in this paper is not on bad cases.

[45] There's logical space for a fallibilist+ to claim that his knowledge that (1) *trumps* or *wipes out* any non-entailing evidence he has for (1). But I find this a peculiar way of carving things up. To complete the picture, for our fallibilist- neither all that one knows, nor only what one knows, is in one's evidence-base for future reasoning.

[46] Or the infalliblist (cf. n.44 *supra*); but the fallibilist+ is the philosopher driving our present concern.

[47] Recall that the fallibilists+ allows that when *S* knows *p*, *p* itself is part of *S*'s evidence-base for future reasoning. It is thus not a flawed account of the sources of warrant-or-evidence for future reasoning by the lights of Silins's claim (cf. 0.4 *supra*). But it should be clear that we would not have answered Silins's claim by showing that the warrant the fallibilist+ can share with the fallibilist- may fail to transmit. To make the dialectic clearer we may wish to sharpen Silins's claim thus: instances of *complete* transmission failure only arise on a particular – putatively flawed – account of the sources of warrant-or-evidence for future reasoning (cf. n.31 *supra*).

[48] *Mutatis mutandis* for *reasons*: some of one's reasons for (1) may presuppose or rest on one's reasons for (3) without other of one's reasons for (1) presupposing or resting on one's reasons for (3).

our trio of theses – *(EA*)*, *(A W*)*, and *(NT*)* – don't themselves discriminate between different warrants one might have for (1) (and for (3)). Given the two types of warrant on which we're focusing – viz. an experience and a premise – are lumped together, if one type fails to transmit in (MOORE)-reasoning – viz. the experience –, *(NT*)* is sufficiently coarse-grained (a bad thing) that it delivers the result that (MOORE)-reasoning suffers from transmission failure, period. This is an unwelcome result: our modified (transmission failure) theses – *(EA*)*, *(A W*)*, and *(NT*)* – should be able to discriminate the salient differences between the fallibilist- and the fallibilist+. We need theses which *leave room* – in line with our assumption – for one type of warrant to *fail to transmit* in (MOORE)-reasoning (e.g. here, the experience), and another type of warrant *to transmit* (e.g. here, the premise). Rather than further refining, and complicating, our modified theses *themselves*, I leave the task of avoiding the foregoing unwelcome result to *sensitive application* of the modified theses.

4.3 Let's take stock. We might reconstruct the foregoing dialectic thus. We set out fallibilism-. We then set up three theses – *(EA)*, *(A W)*, and *(NT)* – which seemed motivated, and captured adequately the fallibilist-'s account of the sources of warrant-or-evidence (for future reasoning). We used notions from *(EA)* to descry one way to express a defense of *(A W)*. We then noted that, assuming *(A W)* and fallibilism-, we had transmission failure in (MOORE)-reasoning by *(NT)*. Fine. We then noted that it wasn't clear the account of the sources of warrant-or-evidence (for future reasoning) presupposed by these three theses was the same account with which fallibilists+ and infallibilists operate. Next we set up three modified theses – *(EA*)*, *(A W*)*, and *(NT*)* – to see if we could construct a motivated transmission failure thesis which operated with an account of the sources of warrant-or-evidence for future reasoning friendly to fallibilists+ and infallibilists. But while our three modified theses operated with a (disjunctive) account of the sources of warrant-or-evidence (for future reasoning) not unacceptable to fallibilists+ and infallibilists, they were vulnerable to two objections from the two camps: a *pragmatic implication* objection (from infallibilists, which we cancelled) and a *coarse-grained* objection (from fallibilists+, which we deflected).

In sum, our original three theses – *(EA)*, *(A W)*, and *(NT)* – are serviceable for analysis of fallibilism- (but not fallibilism+ and infallibilism). Our three modified theses – *(EA*)*, *(A W*)*, and *(NT*)* – are serviceable for analysis of fallibilism+ and infallibilism (but not fallibilism-). So we have two sets of three theses, which sets jointly cover all – or most – plausible accounts of the sources of warrant-or-evidence (for future reasoning).

5. Might Introducing (EA*) be More Difficult than Introducing (EA)?

5.1 There may be especial problems involved with introducing our *modified* theses. Let's focus on a putative problem arising over introducing *(EA*)* (which doesn't arise when introducing *(EA)*). How exactly can *a premise* – a source of warrant-or-evidence for future reasoning, by fallibilism+ and infallibilism – fail to transmit? I accept this explanatory challenge. Consider a fallibilist+ or infallibilist debating with a proponent of our modified theses, and reasoning thus: I can sort of make sense of *support for* a premise – say, an experience – failing to transmit across entailments or inferences. I've heard talk of an *abstract space of warrants* and I can see that some warrants might be *epistemically antecedent* to others in that abstract space of warrants. So if I'm trying to non-question-beggingly engage the external world sceptic, I can see that my warrant for (1) could fail to be epistemically antecedent to my warrant for (3). But I can't make sense of relations of epistemic antecedence between a *premise* ((1)) and a *warrant* for a premise ((3)). The former is a proposition I, ex hypothesi, know; the latter is warrant for a proposition I fail – prior to inference – to know. To talk of a relation of epistemic antecedence between a premise and warrant for a premise is close to a category error: it's introducing an epistemic relation between two radically different things. And while, for example, knowledge is an epistemic relation between a (knowing) *subject* and a *proposition* (at a time) – two very different relata if ever there were –, there's something distinctively problematic about introducing a relation of *epistemic antecedence* between two relata of radically different kinds. To introduce a relation of epistemic antecedence, don't the relata have to be of the same kind?[49] This, I take it, is a forceful objection which could be made by a fallibilist+ or infallibilist to *(EA*)*, and any resultant use of *(NT*)* via *(AW*)*.[50]

5.2 We must distinguish between *a premise* and *knowledge of* a premise. Once we do so the objection dissolves. So far I've played fast and loose with what form evidence for future reasoning takes – better: how we should describe such evidence – for a fallibilist+ or infallibilist. I've switched between *(1) itself* and *knowledge of* (1). I take it these are, for our purposes – we suppose I *know* (1) –,

[49] Cross-categorial relations are routinely posited in philosophy. An example from metaphysics: the truthmaker relation has been posited between a (non-propositional) truthmaker and a (true) proposition (see David Armstrong, *Truth and Truthmakers* (Cambridge: Cambridge University Press, 2004)). So the objection in the main text is a specific objection focused on the relation of epistemic antecedence.

[50] A trio of theses such as these is *a* (prominent) way to argue for transmission failure – it may not be the *only* such way (cf. 2.1 *supra*).

two different descriptions of the same piece of evidence. And we can note that epistemologists who operate with the former description typically take (1) itself to be in one's evidence-base for future reasoning *because* one (ex hypothesi) knows (1). I propose we describe (part of) the fallibilist+'s and the infallibilist's evidence-base for future reasoning as *knowledge of (1)* (rather than as *(1) itself*).[51] Once this is done there's no especial problem about seeing how a relation of epistemic antecedence might hold. We simply ask: Does my knowing (1) presuppose or rest on my warrant for (3)?[52] To be sure, different epistemologists will give different answers to this question. But that's no slur on the intelligibility of the question – indeed it counts in its favour.[53]

A caveat: To require that the fallibilist+ and infallibilist describe this evidence in this way in this transmission failure debate, isn't to require that he *always* so describe his warrant. Nor, relatedly, is it to say that the alternative description of the evidence is deficient.

In sum, one might have difficulty with the intelligibility of the notion of epistemic antecedence, period. But I've shown there's no especial problem of applying the notion of epistemic antecedence to a fallibilist+'s and an infallibilist's account of the sources of warrant-or-evidence for future reasoning. Put differently: if one hasn't a problem with *(EA)*, then one shouldn't have a problem with *(EA*)*

6. Conclusion

6.1 I take myself to have shown that motivated transmission failure conditions can be constructed which are applicable to – capture adequately – all – or most – plausible accounts of the sources of warrant-or-evidence for future reasoning. I haven't shown that there are in fact genuine cases of transmission failure. I have,

---

[51] Why, then, did I play fast and loose with my description of the evidence? First, I want to leave open (cf. n.23 supra) that something short of knowledge of a proposition can suffice for a proposition to be in one's evidence-base for future reasoning. Second, I want to keep vivid, with Silins, the thought that, upon competent inference from a known premise, *the premise itself* can become one's warrant to believe the conclusion.

[52] Pryor does "extend th[e] notion of epistemic priority to knowledge as follows: you count as *knowing p1* antecedently to knowing *p2* just in case you know *p1* and *p2*, and *the justification* on which you base your belief in *p1* is antecedent to the justification on which you base your belief in *p2*" ("The Skeptic and the Dogmatist," 525). But insofar as we are here interested in (at one level of abstraction – cf. n. 38 *supra*) a *cross-categorial* relation this notion will not serve our purposes (cf. n. 39 supra).

[53] Indeed, even the following extreme response, I take to count in its favour: Once I know (1), never mind how, I'll be able to *exploit* (1) as evidence for (3).

though, demonstrated that the fallibilist+ and infallibilist cannot evade questions of transmission failure as easily as they might have hoped. Williamson's, Hawthorne's, and Silins's hard-line against genuine instances of transmission failure is not warranted.[54,55]

[54] It is a further task – one not attempted in this paper – to show why, for example, an infallibilist might plausibly take (MOORE)-reasoning to *suffer* from some kind of epistemic circularity.

[55] Thanks to Martin Davies and Nico Silins for stimulating discussion.

TRACING CULPABLE IGNORANCE

Rik PEELS

ABSTRACT: In this paper, I respond to the following argument which several authors have presented. If we are culpable for some action, we act either from *akrasia* or from culpable ignorance. However, *akrasia* is highly exceptional and it turns out that tracing culpable ignorance leads to a vicious regress. Hence, we are hardly ever culpable for our actions. I argue that the argument fails. Cases of *akrasia* may not be that rare when it comes to *epistemic* activities such as evidence-gathering and working on our intellectual virtues and vices. Moreover, particular cases of *akrasia* may be rare, but they are not exceptional when we consider chains of actions. Finally and most importantly, we can be culpable for our actions even if we do not act from *akrasia* or from culpable ignorance, namely in virtue of our unactivated dispositional beliefs.

KEYWORDS: *akrasia*, culpability, epistemic obligations, ignorance, vicious regress

On May 3rd 1945, Sir Arthur Coningham, commander in the British Tactical Air Force, ordered the attack on three German ships in the Bay of Lübeck. Unbeknownst to him, the Germans had filled these ships with about 10,000 concentration camp survivors. All three ships were sunk. Most of the SS guards survived, but an estimated 7,800 camp survivors died. Was Coningham at least partly culpable for their deaths? The answer, of course, crucially depends on whether his ignorance was culpable or not. If it was not, then it seems unfair to blame him, but if he should have known better, then it seems that he is at least partly blameworthy for the tragedy. When, then, would his ignorance count as culpable? Presumably, if at some earlier time he could have found out that there were prisoners on board but did not investigate the matter sufficiently carefully or failed to listen to certain people who possessed more information than he did.

However, there is a problem here. Imagine that Coningham indeed failed to investigate the matter sufficiently carefully. Then, we may assume that at that earlier time he (falsely) *believed* that he need not gather any additional evidence. But if he falsely believed that, he was ignorant. If he was *inculpably* ignorant, it seems unfair to hold him responsible for acting as he did. If he was *culpably*

© LOGOS & EPISTEME, II, 4 (2011): 575-582

ignorant, then presumably he performed a culpable action[1] in the past which resulted in that culpable ignorance. But we can say the exact same thing about that prior culpable action, and about a culpable action prior to that, and so on. It seems we are on the road to a vicious regress. Clearly, the same trouble can be raised for much more mundane situations in which someone acts from culpable ignorance.

One might think that the point even extends to certain cases in which someone clearly displays evil intentions. Joseph Stalin, presumably, thought that it was a good thing to order the massacre of thousands of Polish officers in the Katyn forest. Since that belief was false, he was ignorant of its wrongness. If he should have known better, then there ought to be some prior culpable act from which his ignorance issued. And so on. The problem seems to arise for any situation in which a person acts in or from ignorance – usually a false belief, but sometimes the absence of any belief – but in which she nevertheless seems culpable for what she does. In all those cases, it seems that, ultimately, we cannot explain why that person's ignorance is culpable. The only exception are cases of *akrasia*, scenarios in which a person does or fails to do something despite occurrently (consciously) believing that doing so is wrong. Cases of clear-eyed *akrasia*, however, seem to be rare: it seems that what we believe and what we do hardly ever come apart in this way. It follows that anybody is hardly ever culpable for some action.

This argument, which I dub the Argument from Culpable Ignorance (ACI), has been presented and discussed by William FitzPatrick, James Montmarquet, and Michael Zimmerman.[2] Before considering how ACI can be met, let me try to be a bit more precise about its structure. I think it can be rendered slightly more formally as follows:

---

[1] In this paper, I confine myself to actions, but *mutatis mutandis* the same could be said about omissions.

[2] See William J. FitzPatrick, "Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge," *Ethics* 118 (2008): 589-601; James A. Montmarquet, *Epistemic Virtue and Doxastic Responsibility* (Boston: Rowman and Littlefield, 1993), vii-viii, 7-9, 45-8; James A. Montmarquet, "Culpable Ignorance and Excuses," *Philosophical Studies* 80 (1995): 41-43; James A. Montmarquet, "Zimmerman on Culpable Ignorance," *Ethics* 109 (1999): 845; Michael J. Zimmerman, "Moral Responsibility and Ignorance," *Ethics* 107 (1997): 410-8; Michael J Zimmerman, *Living with Uncertainty: The Moral Significance of Ignorance* (Cambridge: Cambridge University Press, 2008), 173-93. Gideon Rosen offers an argument similar to ACI; see Gideon Rosen, "Skepticism about Moral Responsibility," in *Philosophical Perspectives: Ethics*, Vol. 18, ed. John Hawthorne (Malden, MA: Blackwell Publishing, 2004), 295-311. The conclusion of his argument, however, is that we can hardly ever *know* when someone has performed a culpable action, since we cannot ascribe *akrasia* with any confidence.

(1) For any person *S*, if *S* is culpable for some action *A*, then *S* does *A* either from *akrasia* or from ignorance. [ass.]

(2) Cases of *akrasia* are highly exceptional. [ass.]

(3) If *S* is culpable for *A*, then in nearly all cases, *S* does *A* from ignorance. [from (1) and (2)]

(4) *S* is culpable for an action *A* done from ignorance only if *S* is culpable for that ignorance. [ass.]

(5) If *S* is culpable for an action *A*, then in nearly all cases, *S* performed *A* from culpable ignorance. [from (3) and (4)]

(6) *S* is culpable for *S*'s ignorance only if *S* is culpable for some past action *B* from which that ignorance issued. [ass.]

(7) If *S* is culpable for an action *A*, then in nearly all cases, *S* is culpable for another culpable action *B* that preceded *A*. [from (5) and (6)]

(8) If *S* is culpable for an action *A*, then in nearly all cases *S* is culpable for an infinitely long chain of culpable actions each of which precedes the other. [from (7)]

(9) There are no infinitely long chains of culpable actions each of which precedes the other. [ass.]

(10) For any person *S* and action *A*, *S* is hardly ever culpable for performing *A*. [from (8) and (9)]

ACI's conclusion is, of course, deeply disturbing: we believe that people are frequently culpable for what they do. Moreover, if culpability for beliefs is due to one's failing to perform certain actions in the past, as the argument presupposes, then it would also follow that we are hardly ever culpable for our beliefs and that would surely be another disturbing conclusion. It is, therefore, important to find a satisfactory way of dealing with the argument. I know of two responses to ACI that have been given in the literature.

According to James Montmarquet and William FitzPatrick, we should reject premise (6): not every instance of culpable ignorance is to be explained by a prior culpable action from which that ignorance issues. There is such a thing as *fundamental* and *underived* culpability for ignorance. More specifically, ignorance is culpable if it issues from insufficient regard of truth and evidence. Insufficiently regarding truth and evidence, like the exemplification of other epistemic vices and virtues, is not to be considered as an action, but as a modality of the belief-forming process, as a way of believing.[3] Perhaps Montmarquet and FitzPatrick are right

[3] See Montmarquet, *Epistemic Virtue*, 41-3, 45-7; Montmarquet, "Culpable Ignorance," 843-5.

that insufficient regard of truth and evidence does not count as an action, but their suggestion will not do. For, clearly, people can be inculpable for insufficient regard of truth and evidence. People may be brainwashed or raised in a society which cares very little about truth and evidence. Insufficient regard of truth and evidence cannot be intrinsically culpable. FitzPatrick seems to be aware of this worry, for he adds that one is culpable for exercising some intellectual vice only if one could reasonably have been expected to do better.[4] Unfortunately, this does not solve the problem either. For, we have not been given an answer to the question of *why* or in *virtue of what* one could reasonably have been expected to do better from an epistemic point of view. We still have not found a source of culpability. Even more importantly, assuming that insufficient regard of truth and evidence is always wrong, it still seems that if people act with insufficient regard of truth and evidence, they either act from *akrasia*, that is, against their better judgement, or from ignorance that acting with insufficient regard of truth and evidence is wrong. Hence, this approach relocates the problem rather than solving it.

Michael Zimmerman, on the other hand, proposes that we accept ACI. On his view, the argument surprisingly shows that we are far less often culpable for our ignorance and for actions performed from ignorance than we might initially think. Thus, we should blame each other only in cases in which an act is done either from *akrasia* or from ignorance which issued from an act done from *akrasia*.[5] This option is clearly highly unattractive: we should be very suspicious of philosophical arguments that call for a substantial revision of one of our socially most important and deeply entrenched practices. We should take this route only if each of the premises of the argument is at least as convincing as the thesis that people are frequently blameworthy for their ignorance and for actions performed from ignorance.

If Montmarquet's and FitzPatrick's response to the argument fails, is there a way to avoid the disturbing conclusion and Zimmerman's strongly deflationary proposal? It is implausible to deny (9), the premise that there are no infinitely long chains of culpable actions. For one thing, even if we *could live* forever, none of us *has lived* forever and, hence, there are no such things as infinitely long chains of culpable actions. (6), as we saw above, is also plausible: if one *should* have known better, then there must be something such that if one had done that thing, one *would* have known better. It is especially plausible because, as William Alston has

[4] See FitzPatrick, "Moral Responsibility," 606-10.

[5] See Zimmerman, "Moral Responsibility," 421-6; Zimmerman, *Uncertainty*, 173-93.

famously argued, we do not have direct control over our beliefs.[6] We do not freely choose to hold some belief. Thus, we are responsible for our beliefs in virtue of our control over such epistemic activities as evidence gathering and working on our intellectual virtues and vices. For instance, we may have an obligation to check whether anyone is lurking behind the target area before firing the gun, or an obligation to read certain articles in medicine before performing a complicated operation. These obligations are such that if we fail to meet them, we risk ignorance. As to (4), ever since Aristotle, philosophers have defended this premise.[7] It just seems unfair and incorrect to blame someone for some action if she did it from blameless ignorance, that is, if it is not the case that she should have known better. Hence, (9), (6), and (4) are unproblematic. The problem with ACI, I will argue, is rather that (1) and (2) are false and that (8) does not follow from (7).

Starting with the latter, let us assume that (7) is true, that is, that if *S* is culpable for *A*, then in nearly all cases she is culpable for another culpable act *B* that preceded *A*. Imagine, for instance, that if *S* is culpable for *A*, then the probability that *A* is preceded by another culpable act *B* is .95. The probability that *B* is preceded by another culpable act *C* is, of course, also .95, so that the probability that *A* is preceded by two culpable acts *B* and *C* is .95 x. 95 = .9025. Clearly, the longer the series of actions, the more likely it is that some action performed from *akrasia* occurs at some point in the series. It may be unlikely that a *particular* act is done from *akrasia*, it is not unlikely that *some act or other* in a chain of actions is done from *akrasia*. But if (8) is false, only a significantly weaker conclusion than (10) can be established.

Second, (2) says that actions from *akrasia* are rare. This may be true for actions *in general*. It does not follow that acting from *akrasia* is rare for *all* particular kinds of actions. More specifically, it does not seem that exceptional when it comes to such *epistemic* actions as gathering evidence and working on our intellectual virtues and vices. Students frequently violate an obligation to prepare an exam, policemen often violate an obligation to gather all the relevant evidence, and many of us fail to become more open-minded in the course of our lives. I think there is a particular reason for this. For *most morally important actions*,

---

[6] See William P. Alston, "The Deontological Conception of Epistemic Justification," in his *Epistemic Justification: Essays in the Theory of Knowledge* (Ithaca: Cornell University Press, 1989), 115-52. This also seems to be the view of Holly Smith, according to whom one is culpably ignorant only if that ignorance is due to some prior benighting act. See Holly Smith, "Culpable Ignorance," *The Philosophical Review* 92 (1983), 547-8.

[7] See Aristotle, *The Nicomachean Ethics*, trans. H. Rackham (Cambridge, MA: Harvard University Press, 2003), 145-7 (III.v.7-8); Gideon Rosen, "Culpability and Ignorance," *Proceedings of the Aristotelian Society* 103 (2003): 64-6; Zimmerman, "Moral Responsibility," 412.

such as sentencing a criminal or buying an expensive car, we have a pretty firm grasp of at least some of the consequences. We easily foresee that the criminal ends up in jail and that the amount of money on my bank account is significantly reduced. With *epistemic actions*, however, things are different. Here, we usually do *not* foresee which particular ignorance will result from that and which actions we will perform from that ignorance. That makes it much easier to violate such obligations, even if we believe that we should not do so (which makes them cases of *akrasia*). We know that we run a certain risk, but since we do not foresee the precise consequences, we are more likely to succumb to the temptation of violating the obligation in question, despite our belief that we should not.[8]

Finally and most importantly, (1) is false: it is *not* the case that if *S* is blameworthy for some action *A*, then *S* does *A* either from *akrasia* or from ignorance. Someone with dispositional beliefs about her circumstances or about the normative status of an action can be blamed for not activating those dispositional beliefs.[9] Imagine that I am babysitting and that my friend tells me that her four year old daughter, Sarah, is seriously allergic to milk. If, an hour later, I completely forget about that and give Sarah a glass of milk, I am clearly blameworthy for that, simply because I had the dispositional, but unactivated (non-occurrent) belief that I should not give Sarah any milk. I could and should have thought about it or remembered it. I clearly did not act from *akrasia*: it is not that I was fully aware of the danger but succumbed to the temptation by weakness of will. Nor was I *ignorant* that Sarah is allergic to milk or *ignorant* that I should not give Sarah any milk.[10] Surely, I *knew* that Sarah is allergic to milk and that

[8] Zimmerman, *Uncertainty*, 190, leaves some room for the idea that what kind of consequences is involved makes a difference to how easy or difficult it is to act from *akrasia*. He fails to notice the crucial point, though, that we fairly easily act from *akrasia* when it comes to epistemic obligations, since in such cases we do not foresee which particular doxastic atitudes we will come to hold as a result of violating or meeting them. This point is crucial, for culpable ignorance will often be due to the violation of such obligations.

[9] The false premise implicit in (1), then, is that only occurrent beliefs can count as reasons for which one acts or should act. For this premise, see Rosen, "Skepticism," 307; Zimmerman, "Moral Responsibility," 421-422; Zimmerman, *Uncertainty*, 190-1. For another example against this principle, see Keith Frankish, "Partial Belief and Flat-Out Belief," in *Degrees of Belief*, ed. Franz Huber et al. (Dordrecht: Springer, 2009), 86.

[10] I will not address the issue of precisely what it is to be ignorant, that is, under which conditions one is ignorant. I have defended a particular answer to this question elsewhere; see Rik Peels, "What Is Ignorance?," *Philosophia* 38 (2010): 57-67, and Rik Peels, "Ignorance Is Lack of True Belief: A Rejoinder to Le Morvan," *Philosophia* 39 (2011): 345-55. In those two articles, I argue that ignorance is lack of true belief rather than lack of knowledge. The truth of the claim that I make here, namely that one is *not* ignorant that *p* if one dispositionally

dispositional belief, perhaps in combination with the absence of excusing circumstances, is sufficient to properly blame me for giving Sarah milk. The same applies to many other situations, such as forgetting one's friend's birthday and violating one's promise by forgetting to take some groceries along on the way home.[11] Not all situations in which one performs a blameworthy action are situations of clear-eyed *akrasia* or ignorance.

One could, of course, bite the bullet and stick with the claim that only *occurrent* beliefs give rise to obligations and, unless certain excuses hold, to blameworthiness. Since this assertion clearly contradicts our intuitions in the examples that I just gave it seems that we should not accept that claim without an argument of some kind. Such an argument is provided by Michael Zimmerman. According to Zimmerman, only occurrent beliefs can render one blameworthy for the violation of an obligation because only occurrent beliefs play a role in the reasons for which one performs an action. And unless something plays a role in the reasons for which one performs an action one cannot in virtue of that thing be blamed for performing the action. Zimmerman qualifies his position in one regard, though. He grants that one can be blameworthy in virtue of one's dispositional beliefs in cases of *routine* or *habitual* actions, cases in which one performs actions for reasons to which one does not advert at the time at which one performs them.[12]

However, I think that there is good reason to reject Zimmerman's claim. Imagine that I am teaching a class on evolutionary theory and, in the course of my lecture, tell the students that Darwin's *On the Origin of Species* was published in 1859. This is something I tell them because I believe it, because I take it to be something that the students ought to know, because I believe that there are students in the room, because I believe that I can transfer knowledge to my students by telling them something, and so forth. But, clearly, I need not consciously consider all these reasons in order for it to be true that my telling the students that Darwin's *On the Origin of Species* was published in 1859 is based on those reasons. It seems that in this case my dispositional (dormant, tacit) beliefs play a crucial role in my reasons for performing this particular action. Something similar seems to apply in cases in which I am blameworthy rather than

---

believes that *p*, does not depend on whether ignorance is lack of true belief or lack of knowledge.

[11] Thus also Angela M. Smith, "Responsibility for Attitudes: Activity and Passivity in Mental Life," *Ethics* 115 (2005): 236; Brian Weatherson, "Deontology and Descartes' Demon," *The Journal of Philosophy* 105 (2008): 552; Susan Wolf, "Asymmetrical Freedom," *Journal of Philosophy* 77 (1980): 164.

[12] See Zimmerman, *Uncertainty*, 190-1. This also seems to be the view of Rosen, "Skepticism," 307.

praiseworthy or neutrally appraisable for what I do.[13] Now, I do not think that there is any natural sense in which the action of my saying that *On the Origin of Species* was published in 1859 is a routine or habitual action. It may be the very first time in my life that I utter this sentence, but, as the example shows, I may still do it for reasons that at that time I am not consciously considering. If Zimmerman were to insist that this action is habitual, then it seems that we would have to say that many of our actions, if not the majority, are habitual. Since Zimmerman allows that in the case of habitual actions we can be blameworthy in virtue of our dispositional beliefs, it would follow that we could still be blameworthy for many of our actions.

In summary, we have traced two ways of acquiring or maintaining blameworthy ignorance. First, one can perform the action from which one's ignorance issued from clear-eyed *akrasia*. As we saw, such *akrasia* is not as rare in the case of evidence gathering and working on our intellectual virtues and vices as it is when it comes to other actions. Second, one can perform an action resulting in ignorance with unactivated dispositional beliefs about one's circumstances or the normative status of that action that one should have activated. It seems highly likely that for a substantial number of our actions that make a difference to what we believe, one or both of these options are realized at that moment or at some earlier moment to which that action is relevantly related. I conclude that the argument fails to establish that we are hardly ever blameworthy for our actions or for our beliefs.[14]

[13] The view that beliefs need not be occurrent in order to guide action is also adopted by Robert Audi, "Acting for Reasons," *The Philosophical Review* 95 (1986): 515, 521, and Eleonore Stump, "Intellect, Will, and the Principle of Alternate Possibilities," in *Perspectives on Moral Responsibility*, ed. John M. Fischer et al. (Ithaca: Cornell University Press, 1993), 254.

[14] I would like to thank Anthony Booth, Herman Philipse, Michael Zimmerman, and the audience at the *85th Joint Session of the Aristotelian Society and the Mind Association* at the University of Sussex (Summer 2011) for their valuable comments on earlier versions of this paper.

PROMISES TO KEEP: SPEECH ACTS AND THE VALUE OF REFLECTIVE KNOWLEDGE

John TURRI

ABSTRACT: This paper offers a new account of reflective knowledge's value, building on recent work on the epistemic norms of speech acts. Reflective knowledge is valuable because it licenses us to make guarantees and promises.

KEYWORDS: knowledge account of assertion, epistemic norms, epistemic value, guaranteeing, reflective knowledge, Ernest Sosa

1. Introduction

Recently epistemologists have been preoccupied with questions about epistemic value and norms.[1] In this paper, I explore one way that these issues intersect, which promises to shed new light on a perennial epistemological issue, namely, the distinction between animal and reflective knowledge. I argue that reflective knowledge is valuable at least in part because it licenses us to make guarantees and promises.

2. Animal and reflective knowledge

Many distinguished philosophers have distinguished between animal knowledge and reflective knowledge.[2] But Ernest Sosa is most closely identified with the

[1] E.g. Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford university Press, 2000), Michael DePaul, "Value Monism in Epistemology," in *Knowledge, Truth, and Duty*, ed. Matthias Steup (Oxford: Oxford University Press, 2001), Jonathan Kvanvig, *The Value of Knowledge and the Pursuit of Understanding* (Cambridge: Cambridge University Press, 2003), Ernest Sosa, *A Virtue Epistemology* (Oxford: Oxford University Press, 2007), John Hawthorne and Jason Stanley, "Knowledge and Action," *Journal of Philosophy* 105 (2008): 571-90, Wayne Riggs, "The Value Turn in Epistemology," in *New Waves in Epistemology*, eds. Vincent Hendricks and Duncan Pritchard (New York: Palgrave Macmillan, 2008), Linda Zagzebski, *On Epistemology* (Belmont, CA: Wadsworth, 2009), Adrian Haddock, Alan Miller and Duncan Pritchard, *The Nature and Value of Knowledge: Three Investigations* (Oxford: Oxford University Press, 2010), John Turri, "The Express Knowledge Account of Assertion," *Australasian Journal of Philosophy* 89 (2011): 37-45.

[2] E.g. Rene Descartes, *Meditations, Objections and Replies*, eds. and trans. Roger Ariew and Donal Cress (Indianapolis: Hackett, 2006 [1641]), Charles Peirce, *Philosophical Writings of*

distinction, having made it a central and enduring plank in his important and influential epistemology, and having done more than anyone to clarify the distinction and demonstrate its philosophical utility.[3] Accordingly I will take Sosa's treatment as the starting point of my discussion, the ultimate goal being to help explain the value of reflective knowledge.

Animal knowledge is simple first-order knowledge unaided by explicit or implicit reflection on our cognitive powers or how the current circumstances affect their operation. It includes many ordinary beliefs formed on the basis of perception, memory and introspection. *Reflective knowledge* requires a true *second*-order belief that your first-order belief amounts to knowledge.[4] It involves, to one degree or another, reflection on our cognitive powers and how the current environment affects their operation.

Sosa gives "pride of place" to reflective knowledge,[5] judging it "an important epistemic desideratum," more worthy than the merely animal.[6] And Sosa is not alone here. As mentioned earlier, Descartes, Peirce, Russell, Sellars and others have adopted a similar bi-level epistemology, with special (in some cases, nearly exclusive) distinction attaching to the higher-level intellectual achievement.[7] But some have questioned the relative importance and worth of reflective knowledge.[8] What makes reflective knowledge distinctive? What sets it apart from plain old knowledge? Why should we care about it?

Peirce, ed. Justus Buchler (New York: Dover, 1955), Betrand Russell, *Human Knowledge: Its Scope and Limits* (New York: Routledge, 1948), 202, 451-2, and Wilfrid Sellars, "The Structure of Knowledge," in *Action, Knowledge and Reality: Studies in Honor of Wilfrid Sellars*, ed. H.N. Castañeda (New York: Bobbs-Merill, 1975), Lecture 1, esp §33.

[3] Ernest Sosa, *Knowledge in Perspective* (Cambridge: Cambridge University Press, 1991); "Reflective Knowledge in the Best Circles," *Journal of Philosophy* 94 (1997): 410-30; "How to Resolve the Pyrrhonian Problematic: A Lesson from Descartes," *Philosophical Studies* 85 (1997): 229-49; "Replies," in *Ernest Sosa and His Critics*, ed. John Greco (Malden, MA: Blackwell, 2004); and *A Virtue Epistemology*.

[4] Sosa, *A Virtue Epistemology*, 32. Elsewhere Sosa defines "human knowledge" as "apt belief aptly noted" ("Replies to Commentators on *A Virtue Epistemology*," *Philosophical Studies* 144 (2009): 137-47, cf. 141), a characterization previously reserved for "reflective knowledge." But context reveals that in defining human knowledge this way, he is just aiming to avoid what he sees as a merely terminological dispute with opponents.

[5] "Replies," 291.

[6] "Replies to Commentators," 144.

[7] Sosa, "How to Resolve the Pyrrhonian Problematic," defends this reading of Descartes.

[8] E.g. Hilary Kornblith, "Sosa on Human and Animal Knowledge," in *Sosa and His Critics*, John Greco "How to Preserve Your Virtue while Losing Your Perspective," in *Sosa and His Critics*, and John Greco, "Virtue, Luck and the Pyrrhonian Problematic," *Philosophical Studies* 130 (2006): 9-34.

Sosa replies to these questions with three points.[9] First, reflective knowledge is more closely connected to intellectual autonomy, which we value. Second, the claim that reflective knowledge is specially important best explains the perennial fascination with certain forms of skepticism, especially the Pyrrhonian sort. Third, reflection and reflective knowledge enhances practical autonomy, "the whole person's" ability to control her conduct, rather than her conduct being the mechanical effect of cognitive modules such as vision or memory. There might be room to contest each of these three points, but that's not my purpose here. Instead I offer an additional, complementary proposal.

Sosa is keenly aware of how deeply knowledge is intertwined with the social.[10] But one important social aspect of knowledge does not figure into his response, namely, its role in licensing speech. In what follows, I will develop this neglected possibility by suggesting that reflective knowledge's special status derives at least partly from its role in licensing important speech acts, which facilitate valuable social activities and relationships.

3. Assertion and the value of knowledge

I take as my starting point the well supported – though by no means uncontested – claim that you may assert something only if you know it. Call this the knowledge account of assertion, or 'KA' for short.[11]

---

[9] Sosa, "Replies," 291-2.

[10] See e.g. *Knowledge in Perspective*, 26, 48-9, 275-6.

[11] For defense of KA, see Peter Unger, *Ignorance: A Case for Skepticism* (Oxford: Oxford University Press, 1975), ch. 6, Williamson, *Knowledge and Its Limits*, ch. 11, Keith DeRose, "Assertion, Knowledge and Context," *Philosophical Review* 111 (2002), 167-203, John Turri, "Prompting Challenges," *Analysis* 70 (2010): 456-462, Turri, "The Express Knowledge Account of Assertion," and Matthew Benton, "Two More for the Knowledge Account of Assertion," *Analysis* 71 (2011): 684-87. For criticism of KA, see Matthew Weiner, "Must We Know What We Say?" *Philosophical Review* 114 (2005): 227-51, Igor Douven, "Assertion, Knowledge, and Rational Credibility," *Philosophical Review* 115 (2005): 449-85, Jennifer Lackey, "Norms of Assertion," *Noûs* 41 (2007): 594-625, Christopher Hill and Joshua Schechter, "Hawthorne's Lottery Puzzle and the Nature of Belief," *Philosophical Issues* 17 (2007): 102-22, Janet Levin, "Assertion, Practical Reason, and Pragmatic Theories of Knowledge," *Philosophy and Phenomenological Research* 76 (2008): 359-384, Jonathan Kvanvig, "Assertion, Knowledge, and Lotteries," in *Williamson on Knowledge*, eds. Duncan Pritchard and Patrick Greenough (Oxford: Oxford University Press, 2009), Jessica Brown, "Knowledge and Assertion," *Philosophy and Phenomenological Research* 81 (2010): 549-566, and Rhys McKinnon, "How Do You Know that 'How Do You Know?' Challenges a Speaker's Knowledge?" *Pacific Philosophical Quarterly* (forthcoming).

As far back as Plato's Meno, philosophers have wondered why knowledge is more valuable than mere true belief. If a true belief that this is the road to Larissa will get you to Larissa just as well as knowledge that this is the road to Larissa, Plato wondered, then why is knowledge better than mere true belief? Following Duncan Pritchard, we can call this the *primary value problem*.[12] A different question is what Pritchard calls the *secondary value problem*: why is knowledge "more valuable than any proper subset of its parts"?[13] For instance, if justified true belief is necessary but not sufficient for knowledge, then why is knowledge more valuable than justified true belief? Furthermore, many thinkers believe that knowledge has, as Pritchard puts it, a "distinctive value," meaning that it is better than true belief, and justified true belief, not just in degree, but also in kind. Accordingly Pritchard identifies the tertiary value problem, which is to explain "why knowledge has not just a greater degree but also a different kind of value than whatever falls short of knowledge."[14]

KA helps solve all three problems at once. Knowledge, and no epistemic status short of it, licenses assertion, which speaks to the primary and secondary problems. The difference between having and lacking permission is a difference in kind, not degree, which speaks to the tertiary problem. But even setting aside Pritchard's tripartite classification of problems, KA enables a simple and straightforward explanation of why we value knowledge. Assertion is the primary means of expressing and sharing information, which is essential to the welfare of rational social beings like us. So if knowledge is the norm of assertion, then its value derives at least partly from its role in licensing this vital social activity. And given how important assertion is, surely KA enables us to, as Pritchard enjoins, explain why "the long history of epistemological discussion has focused specifically on the stage in [the] continuum of epistemic value that knowledge marks rather than some other stage (such as a stage just before the one marked out by knowledge, or just after)."[15]

---

[12] Duncan Pritchard, "The Value of Knowledge," *Harvard Review of Philosophy* 16 (2009): 2-19, cf. p. 3.

[13] Compare Kvanvig, *The Value of Knowledge and the Pursuit of Understanding*.

[14] Pritchard, "The Value of Knowledge," 4.

[15] Pritchard, "The Value of Knowledge," 4. This last consideration fails, however, to meet one further constraint that Pritchard thinks an explanation of knowledge's value might strive to respect. The further constraint is that knowledge is "somehow *precious*, in the sense that its value is not merely a function of its practical import," but is instead "non-instrumentally valuable" ("The Value of Knowledge," 4). However, I am not persuaded that we should accept this last constraint. Whereas there is a robust intuitive basis for claiming that knowledge is *somehow* intellectually better than mere true belief, and (once apprised of the Gettier

4. Alethic speech acts

Now let's consider more carefully what sort of speech act assertion is. Doing so will help us understand how it relates to other speech acts and thereby give us some insight into their epistemic norms, which will in turn suggest a distinctive source of reflective knowledge's value.

Other things being equal, insofar as an assertion is true, it is good qua assertion; insofar as it is false, it is bad *qua* assertion.[16] In virtue of this, let's say that assertion aims at truth. Other speech acts also aim at truth, such as guessing, conjecturing and guaranteeing. Call speech acts aimed at truth *alethic speech acts*.

Alethic speech acts differ in two important, closely related ways. First, some place more credibility on the line than others. Guessing extracts little if any of your credibility. Let a pure blind guess be a guess where your total evidence is indifferent among any and all relevant options. Pure blind guessing extracts no credibility and so carries no epistemic requirement. You are required, at most, only to not guess against your evidence.[17] Conjecturing extracts more credibility than guessing, asserting more than conjecturing, and guaranteeing more than asserting. Our evidence for sorting alethic speech acts in this way is both phenomenological and practical. On the one hand, we feel the difference between saying "I guess that Q," "Q," and "I guarantee that Q." We're prone to feel more anxiety as we move from guessing to asserting, or from asserting to guaranteeing. On the other hand, we're entitled to hold people to greater account when they guarantee that something is true than when they merely assert it, and the same is true when they assert it compared to when they merely guess it.[18] If I guarantee that McCain will win, whereas you merely assert that McCain will win, and Jones simply guesses that McCain will win, then when it turns out that McCain doesn't win, I lose more credibility than you do, and you more than Jones. Moreover, in the case of a pure blind guess, we're not entitled to hold someone to account at all. If you say "This is a pure blind guess: Q," and it turns out that Q is false, then it

problem) better than mere justified true belief, I don't likewise think there is a robust intuitive basis for claiming that knowledge is distinctively, non-instrumentally valuable. That is a *highly* theoretical claim, which stands in need of argumentation. Thanks to Duncan Pritchard for discussion on this point.

[16] Williamson, *Knowledge and Its Limits*, 244ff.

[17] Suppose you're asked to guess whether Q or R. Arguably, if your evidence on balance indicates that one of the options is more likely, then you ought to guess that option, and if you guess otherwise, then the resulting guess would to that extent be bad.

[18] Plausibly the practical and phenomenological points are related. The practical point probably explains why we feel more anxiety as we move up the scale of alethic speech acts. Interesting as this is in its own right, I won't pursue the matter further here.

would be completely out of line for me to place less credibility in your future word as a result. To approach the matter slightly differently, consider that we feel more confident purchasing a product when the manufacturer guarantees that it will work, than when the manufacturer simply says that it will work. And we in turn would feel more confident when the manufacturer says that it will work, than when the manufacturer guesses that it will work.

Second, the more credibility an alethic speech act extracts, the stricter the epistemic norms governing it.[19] Guessing requires virtually nothing by way of evidence or epistemic standing. Conjecturing requires that you have at least some evidence favoring Q, and perhaps that Q be the most probable alternative given your evidence. Henry Jekyll suggests as much when he writes to Mr. Utterson: "I must here speak by theory alone, saying not that which I know, but that which I suppose to be most probable."[20] As already noted, knowledge is the norm of assertion. Some alethic speech acts are more emphatic than assertion, and they extract more credibility than does assertion. Call such speech acts *ultra-assertive*. Guaranteeing is ultra-assertive. Just as conjecturing has a stricter epistemic norm than guessing, and asserting a stricter norm than conjecturing, we should expect ultra-assertive speech acts to have stricter epistemic norms still. So guaranteeing requires more than knowledge. Call a state or requirement more demanding than knowledge *ultra-epistemic*. One ultra-epistemic state available to warrant ultra-assertive speech acts is knowledge of knowledge, a possibility we turn to next.

5. Promising and the value of reflective knowledge

J.L. Austin claimed that *asserting that you know Q* counts as *guaranteeing that Q*.[21] Wilfrid Sellars argued that "I know that Q" means the same thing as "Q, and I have reasons good enough to guarantee that Q."[22] I disagree with Austin and Sellars on the nature of the connection between asserting that you know and guaranteeing, though I do agree with them that there's clearly an important intuitive connection between them. But what is the connection? If knowing that you know were the epistemic norm of guaranteeing, then we could explain the

[19] John Turri, "Epistemic Invariantism and Speech Act Contextualism," *Philosophical Review* 119 (2010): 77-95.

[20] See the final section of "Henry Jekyll's Full Statement of the Case," in Robert Louis Stevenson, *The Strange Case of Dr. Jekyll and Mr. Hyde* (Mineola, NY: Dover, 1991 [1886]).

[21] J.L. Austin, "Other Minds," *Proceedings of the Aristotelian Society, Supplementary Volumes* 20 (1946): 148-187.

[22] Wilfrid Sellars, "Epistemic Principles," in *Action, Knowledge, and Reality*, ed. H. Castañeda (Indianapolis: Bobbs-Merrill, 1975), and reprinted in *Epistemology: An Anthology*, eds. Ernest Sosa, Jaegwon Kim, Jeremy Fantl, and Matthew McGrath (Malden, Mass.: Blackwell, 2008).

intuitive connection as follows. When you assert that you know Q, you represent yourself as having the authority to assert that you know Q. Given that knowledge is the norm of assertion, this amounts to representing yourself as knowing that you know Q. So if knowing that you know is the norm of guaranteeing, then by asserting that you know Q, you represent yourself as having the authority to guarantee Q.[23]

Call *the KK account of guaranteeing* the view that you may guarantee something only if you know that you know it. The KK account and the definition of reflective knowledge together entail that reflective knowledge is the norm of guaranteeing.

Now we're positioned to explain reflective knowledge's special status, in much the same way we earlier explained knowledge's value. Reflective knowledge is important because it enables us to engage in the important social practice of guaranteeing. A promise is plausibly regarded as a special type of guarantee, perhaps distinguished by being properly offered only when the promisor and promisee are somehow intimate.[24] Guarantees and promises are essential to accomplishing our goals in some contexts, and are integral to the health and maintenance of important relationships. Sometimes we find ourselves faced with a skeptical or hesitant fellow whose help we need, but who won't be satisfied with anything less than a proper guarantee that things will turn out a certain way. Merely asserting that they'll turn out that way isn't enough to reassure him. And many are the times when a promise helps reassure a frightened child, timorous friend or worried spouse.

Note a further dimension of my proposal. Guarantees come in degrees.[25] You can guarantee, emphatically guarantee and absolutely guarantee. Emphatically or absolutely guaranteeing extracts more credibility than merely guaranteeing. And it is of course natural to suppose that the more emphatic the guarantee, the more robust the epistemic requirement licensing it. Reflective knowledge likewise comes in degrees.[26] You can reflectively know something better or worse. The

[23] I defend this line of reasoning in "Knowledge Guaranteed," forthcoming in *Noûs*.

[24] The OED defines 'promise' (n.) like so, "A declaration or assurance made to another person (usually with respect to the future), stating a commitment to give, do, or refrain from doing a specified thing or act, or **guaranteeing** that a specified thing will or will not happen" (emphasis added). Here I'm thinking of *propositional* promising (i.e. promising *that*), rather than *infinitive* promising (i.e. promising *to*).

[25] This is true of conjecture, assertion and swearing as well.

[26] Sosa, *A Virtue Epistemology*, and "Replies to Commentators." Its degree is a function of a number of things, including the levels of reflective ascent, the scope and depth of the

gradability of reflective knowledge can thus match that of a guarantee's strength: the stronger the guarantee, the better the reflective knowledge required to license it.

Even setting aside the KK account, context certainly often requires us to do more than merely assert something to persuade our intended audience: we are challenged to stand by our words, or have our authority to make an assertion disputed. Having reflective knowledge – knowing that you know – positions you to properly handle such challenges. And so long as reflective knowledge suffices here, or even normally enables a successful response, then even if it isn't strictly necessary, it would be valuable indeed and certainly something we theorists could profitably study.

This last suggestion is at least broadly related to Sosa's views on reflective knowledge's place in properly understanding and responding to skepticism. But my proposal gets traction even prior to reflection on skepticism, because the challenges I speak of are a common part of ordinary life, and so provide a broad and robust vindication of reflective knowledge's value.

6. Conclusion

In sum, reflective knowledge is special in no small part because it licenses important ultra-assertive speech acts, such as guaranteeing and promising. I conclude that we've located a plausible source of reflective knowledge's special status.[27]

reflection involved, the variety and severity of the challenges to which it enables a response, and the reliability of the cognitive dispositions exercised in reflection.

[27] For helpful feedback and conversation, I thank Matthew Benton, Ian MacDonald, Rhys McKinnon, Duncan Pritchard, Ernest Sosa and Angelo Turri. This research was supported by the Social Sciences and Humanities Research Council of Canada.

DEFENDING INTEREST–RELATIVE INVARIANTISM

Brian WEATHERSON

ABSTRACT: I defend interest-relative invariantism from a number of recent attacks. One common thread to my response is that interest-relative invariantism is a much *weaker* thesis than is often acknowledged, and a number of the attacks only challenge very specific, and I think implausible, versions of it. Another is that a number of the attacks fail to acknowledge how many things we have independent reason to believe knowledge is sensitive to. Whether there is a defeater for someone's knowledge can be sensitive to all manner of features of their environment, as the host of examples from the post-Gettier literature shows. Adding in interest-sensitive defeaters is a much less radical move than most critics claim it is.

KEYWORDS: knowledge, interest-relativity, defeaters, evidence

Defending Interest-Relative Invariantism

In recent years a number of authors have defended the interest-relativity of knowledge and justification. Views of this form are floated by John Hawthorne[1], and endorsed by Jeremy Fantl and Matthew McGrath,[2] Jason Stanley[3] and Brian Weatherson.[4] The various authors differ quite a lot in how much interest-relativity they allow, but what is common is the defence of interest-relativity.

These views have, quite naturally, drawn a range of criticisms. The primary purpose of this paper is to respond to these criticisms and, as it says on the tin, defend interest-relative invariantism, or IRI for short. But I don't plan to defend every possible version of IRI, only a particular one. Most of the critics of IRI have assumed that it must have some or all of the following features.

1. It is harder to know things in high-stakes situations than in low-stakes situations.
2. There is an interest-sensitive constituent of knowledge.

[1] John Hawthorne, *Knowledge and Lotteries* (Oxford: Oxford University Press, 2004).

[2] Jeremy Fantl and Matthew McGrath, "Evidence, Pragmatics, and Justification," *Philosophical Review* 111 (2002): 67–94, doi:10.1215/00318108-111-1-67, *Knowledge in an Uncertain World* (Oxford: Oxford University Press, 2009).

[3] Jason Stanley, *Knowledge and Practical Interests* (Oxford: Oxford University Press, 2005).

[4] Brian Weatherson, "Can We Do Without Pragmatic Encroachment?" *Philosophical Perspectives* 19 (2005): 417– 443, doi:10.1111/j.1520-8583.2005.00068.x.

3. IRI stands and falls with some principles connecting knowledge and action, such as the principles found in John Hawthorne and Jason Stanley's "Knowledge and Action."[5]

My preferred version of IRI has none of these three features.[6]

First, it says that knowledge changes when the odds an agent faces change, not when the stakes change. More precisely, interests affect belief because whether someone believes p depends inter alia on whether their credence in p is high enough that any bet on p they actually face is a good bet. And interests affect knowledge largely because they affect belief. Raising the stakes of any bet on p does not directly change whether an agent believes p, but changing the odds of the bets on p they face does change it. In practice raising the stakes changes the odds due to the declining marginal utility of material goods. So in practice high-stakes situations are typically long-odds situations. But knowledge is hard in those situations because they are long-odds situations, not because they are high-stakes situations.

So my version of IRI says that knowledge differs between these two cases.

> **High Cost Map**: Zeno is walking to the Mysterious Bookshop in lower Manhattan. He's pretty confident that it's on the corner of Warren Street and West Broadway. But he's been confused about this in the past, forgetting whether the east-west street is Warren or Murray, and whether the north-south street is Greenwich, West Broadway or Church. In fact he's right about the location this time, but he isn't justified in having a credence in his being correct greater than about 0.95. While he's walking there, he has two options. He could walk to where he thinks the shop is, and if it's not there walk around for a few minutes to the nearby corners to find where it is. Or he could call up directory assistance, pay $1, and be told where the shop is. Since he's confident he knows where the shop is, and there's little cost to spending a few minutes walking around if he's wrong, he doesn't do this, and walks directly to the shop.
>
> **Low Cost Map**: Just like the previous case, except that Zeno has a new phone with more options. In particular, his new phone has a searchable map, so with a few clicks on the phone he can find where the store is. Using the phone has some very small costs. For example, it distracts him a little, which marginally raises the

[5] John Hawthorne Jason Stanley, "Knowledge and Action," *Journal of Philosophy* 105 (2008): 571-90.

[6] It is a tricky exegetical question how many of the three features here must be read into defences of IRI in the literature. My reading is that they do not have to be read in, so it is not overly original of me to defend a version of IRI that does away with all three. But I know many people disagree with that. If they're right, this paper is more original than I think it is, and so I'm rather happy to be wrong. But I'm going to mostly set these exegetical issues aside, and compare different theories without taking a stand on who originally promulgated them.

> likelihood of bumping into another pedestrian. But the cost is very small compared to the cost of getting the location wrong. So even though he is very confident about where the shop is, he double checks while walking there.

I think the Map Cases are like the various cases that have been used to motivate interest-relativity[7] in all important respects. I think Zeno knows where the shop is in High Cost Map, and doesn't know in Low Cost Map. And he doesn't know in Low Cost Map because the location of the shop has suddenly become the subject matter of a bet at very long odds. You should think of Zeno's not checking the location of the shop on his phone-map as a bet on the location of the shop. If he wins the bet, he wins a few seconds of undistracted strolling. If he loses, he has to walk around a few blocks looking for a store. The disutility of the loss seems easily twenty times greater than the utility of the gain, and by hypothesis the probability of winning the bet is no greater than 0.95. So he shouldn't take the bet. Yet if he knew where the store was, he would be justified in taking the bet. So he doesn't know where the store is. Now this is not a case where higher stakes defeat knowledge. If anything, the stakes are lower in Low Cost Map. But the relevant odds are longer, and that's what matters to knowledge.

Second, on this version of IRI, interests matter because there are interest-sensitive defeaters, not because interests form any kind of new condition on knowledge, alongside truth, justification, belief and so on. In particular, interests matter because there are interest-relative coherence constraints on knowledge. Some coherence constraints, I claim, are not interest-relative. If an agent believes □p, that belief defeats her purported knowledge that p, even if the belief that p is true, justified, safe, sensitive and so on. It is tempting to try to posit a further coherence condition.

> **Practical Coherence:** An agent does not know that p if she prefers φ to ψ unconditionally, but prefers ψ to φ conditional on p.

But that is too strong. For reasons similar to those gone over at the start of Hawthorne,[8] it would mean we know nearly nothing. A more plausible condition is:

> **Relevant Practical Coherence:** An agent does not know that p if she prefers φ to ψ unconditionally, but prefers ψ to φ conditional on p, for any φ,ψ that are relevant given her interests.

---

[7] Such as the Bank Cases in Stanley, *Knowledge and Practical Interests*, or the Train Cases in Fantl and McGrath, "Evidence, Pragmatics, and Justification."

[8] Hawthorne, *Knowledge and Lotteries*.

When this condition is violated, the agent's claim to knowledge is defeated. As we'll see below, defeaters behave rather differently to constituents of knowledge. Some things which could not plausibly be grounds for knowledge could be defeaters to defeaters for knowledge.

Relevant Practical Coherence suffices, at least among agents who are trying to maximise expected value, to generate an interest-relativity to knowledge. The general structure of the case should be familiar from the existing literature. Let p be a proposition that is true, believed by the agent, and strongly but not quite conclusively supported by their evidence. Let B be a bet that has a small positive return if p, and a huge negative return if ¬p. Assume the agent is now offered the bet, and let φ be declining the bet, and ψ be accepting the bet. Conditional on p, the bet wins, so the agent prefers the small positive payout, so prefers ψ to φ conditional on p. But the bet has a massively negative expected return, so unconditionally the agent does not want it. That is, unconditionally she prefers φ to ψ. Once the bet is offered, the actions φ and ψ become relevant given her interests, so by Relevant Practical Coherence she no longer knows p. So for such an agent, knowledge is interest-relative.

This suggests our third point. This version of IRI does not take IRI to be a consequence of more general principles about knowledge and action. It simply says that there exist at least one pair of cases where the only relevant difference between agents in the two cases concerns their interests, but one knows that p and the other does not.[9] I happen to think that most of the general principles that philosophers have used to try to derive IRI are false. But since IRI is much weaker than those principles, that is no reason to conclude IRI is false.[10]

The existence of interest-relativity is then quite a weak claim. There are plenty of stronger claims in the area we could make. I prefer, for instance, a version of IRI where being offered bets like B defeats knowledge that p even if the agent does not have the preferences I ascribed above. (That could be because she isn't trying to maximise expected value, or because she's messed up the expected value calculations.) But knowledge could be interest-relative even if I'm wrong about those cases.

So I've set out a version of IRI that lacks three features often attributed to IRI. I haven't argued for that theory here – I do that at much greater length in

[9] And this is true even though p is not a proposition about their interests, or something that is supported by propositions about their interests, and so on.

[10] I will consider, and tentatively support, one principle stronger than IRI in the final section. But the key point is that these general principles are not needed to defend IRI.

"Knowledge, Bets and Interests."[11] But I hope I've done enough to convince you that the theory is both a version of IRI, and not obviously false. In what follows, I'll argue that the theory is immune to the various challenges to IRI that have been put forward in the literature. This immunity is, I think, a strong reason to prefer this version of IRI.

1. Experimental Objections

I don't place as much weight as some philosophers do on the correlation between the verdicts of an epistemological theory and the gut reactions that non-experts have to tricky cases. And I don't think the best cases for IRI relies on such a correlation holding. The best case for IRI is that it integrates nicely with an independently supported theory of belief, and that it lets us keep a number of plausible principles without drifting into skepticism.[12] But still, it is nice to not have one's theory saying exorbitantly counterintuitive things. Various experimental results, such as the results in May et al.[13] and Feltz and Zarpentine,[14] might be thought to suggest that IRI does have consequences which are counterintuitive, or which at least run counter to the intuitions of some experimental subjects. I'm going to concentrate on the latter set of results here, though I think that what I say will generalise to related experimental work. In fact, I think the experiments don't really tell against IRI, because IRI, at least in my preferred version, doesn't make any unambiguous predictions about the cases at the centre of the experiments. The reason for this is related to my insistence that we concentrate on the odds an agent faces, not the stakes she faces.

Feltz and Zarpentine gave subjects related vignettes, such as the following pair. (Each subject only received one of the pair.)

> **High Stakes Bridge:** John is driving a truck along a dirt road in a caravan of trucks. He comes across what looks like a rickety wooden bridge over a yawning thousand foot drop. He radios ahead to find out whether other trucks have made it safely over. He is told that all 15 trucks in the caravan made it over without a problem. John reasons that if they made it over, he will make it over as well. So, he thinks to himself, "I know that my truck will make it across the bridge."

---

[11] Forthcoming in a volume on knowledge ascriptions edited by Jessica Brown and Mikkel Gerken. Penultimate draft available at http://brian.weatherson.org/papers.shtml.

[12] These points are expanded upon greatly in "Knowledge, Bets and Interests."

[13] Joshua May, Walter Sinnott-Armstrong, Jay G. Hull, and Aaron Zimmerman, "Practical Interests, Relevant Alternatives, and Knowledge Attributions: an Empirical Study," *Review of Philosophy and Psychology*, forthcoming, doi:10.1007/s13164-009-0014-3.

[14] Adam Feltz and Chris Zarpentine, "Do You Know More When It Matters Less?" *Philosophical Psychology* 23 (2010): 683–706, doi:10.1080/09515089.2010.514572.

> **Low Stakes Bridge**: John is driving a truck along a dirt road in a caravan of trucks. He comes across what looks like a rickety wooden bridge over a three foot ditch. He radios ahead to find out whether other trucks have made it safely over. He is told that all 15 trucks in the caravan made it over without a problem. John reasons that if they made it over, he will make it over as well. So, he thinks to himself, "I know that my truck will make it across the bridge."[15]

Subjects were asked to evaluate John's thought. And the result was that 27% of the participants said that John does not know that the truck will make it across in Low Stakes Bridge, while 36% said he did not know this in High Stakes Bridge. Feltz and Zarpentine say that these results should be bad for interest-relativity views. But it is hard to see just why this is so.

Note that the change in the judgments between the cases goes in the direction that IRI seems to predict. The change isn't trivial, even if due to the smallish sample size it isn't statistically significant in this sample. But should a view like IRI have predicted a larger change? To figure this out, we need to ask three questions.

1. What are the costs of the bridge collapsing in the two cases?
2. What are the costs of not taking the bet, i.e., not driving across the bridge?
3. What is the rational credence to have in the bridge's sturdiness given the evidence John has?

Conditional on the bridge not collapsing, the drivers presumably prefer taking the bridge to not taking it. And the actions of taking the bridge or going around the long way are relevant. So by Relevant Practical Coherence, the drivers know the bridge will not collapse in Low Stakes Bridge but not High Stakes Bridge if the following equation is true. (I assume all the other conditions for knowledge are met, and that there are no other salient instances of Relevant Practical Coherence to consider.)

$$\frac{C_H}{G + C_H} > x > \frac{C_L}{G + C_L}$$

where G is the gain the driver gets from taking a non-collapsing bridge rather than driving around (or whatever the alternative is), C_H is the cost of being on a collapsing bridge in High Stakes Bridge, C_L is the cost of being on a collapsing bridge in Low Stakes Bridge, and x is the probability that the bridge will collapse. I assume x is constant between the two cases. If that equation holds, then taking the bridge, i.e., acting as if the bridge won't collapse, maximises expected utility in

[15] Feltz and Zarpentine, "Do You Know More," 696.

Low Stakes Bridge but not High Stakes Bridge. So in High Stakes Bridge, adding the proposition that the bridge won't collapse to the agent's cognitive system produces incoherence, since the agent won't (at least rationally) act as if the bridge won't collapse. So if the equation holds, the agent's interests in avoiding C_H creates a doxastic defeater in High Stakes Bridge.

But does the equation hold? Or, more relevantly, did the subjects of the experiment believe that the equation hold? None of the four variables has their values clearly entailed by the story, so we have to guess a little as to what the subjects' views would be.

Feltz and Zarpentine say that the costs in "High Stakes Bridge [are] very costly – certain death – whereas the costs in Low Stakes Bridge are likely some minor injuries and embarrassment."[16] I suspect both of those claims are wrong, or at least not universally believed. A lot more people survive bridge collapses than you may expect, even collapses from a great height.[17] And once the road below a truck collapses, all sorts of things can go wrong, even if the next bit of ground is only 3 feet away. (For instance, if the bridge collapses unevenly, the truck could roll, and the driver would probably suffer more than minor injuries.)

We aren't given any information as to the costs of not crossing the bridge. But given that 15 other trucks, with less evidence than John, have decided to cross the bridge, it seems plausible to think they are substantial. If there was an easy way to avoid the bridge, presumably the first truck would have taken it. If G is large enough, and C_H small enough, then the only way for this equation to hold will be for x to be low enough that we'd have independent reason to say that the driver doesn't know the bridge will hold.

But what is the value of x? John has a lot of information that the bridge will support his truck. If I've tested something for sturdiness two or three times, and it has worked, I won't even think about testing it again. Consider what evidence you need before you'll happily stand on a particular chair to reach something in the kitchen, or put a heavy television on a stand. Supporting a weight is the kind of thing that either fails the first time, or works fairly reliably. Obviously there could

[16] Feltz and Zarpentine, "Do You Know More," 702.

[17] In the West Gate bridge collapse in Melbourne in 1971, a large number of the victims were underneath the bridge; the people on top of the bridge had a non-trivial chance of survival. That bridge was 200 feet above the water, not 1000, but I'm not sure the extra height would matter greatly. Again from a slightly lower height, over 90% of people on the bridge survived the I-35W collapse in Minneapolis in 2007.

be some strain-induced effects that cause a subsequent failure,[18] but John really has a lot of evidence that the bridge will support him.

Given those three answers, it seems to me that it is a reasonable bet to cross the bridge. At the very least, it's no more of an unreasonable bet than the bet I make every day crossing a busy highway by foot. So I'm not surprised that 64% of the subjects agreed that John knew the bridge would hold him. At the very least, that result is perfectly consistent with IRI, if we make plausible assumptions about how the subjects would answer the three numbered questions above.

And as I've stressed, these experiments are only a problem for IRI if the subjects are reliable. I can think of two reasons why they might not be. First, subjects tend to massively discount the costs and likelihoods of traffic-related injuries. In most of the country, the risk of death or serious injury through motor vehicle accident is much higher than the risk of death or serious injury through some kind of crime or other attack, yet most people do much less to prevent vehicles harming them than they do to prevent criminals or other attackers harming them.[19] Second, only 73% of these subjects in this very experiment said that John knows the bridge will support him in Low Stakes Bridge. This is rather striking. Unless the subjects endorse an implausible kind of scepticism, something has gone wrong with the experimental design. But if the subjects are implausibly sceptical, then we shouldn't require our epistemological theory to track their gut reactions. (And if something has gone wrong with the experimental design, then obviously can't be used as the basis for any objection.) So given the fact that the experiment points broadly in the direction of IRI, and that with some plausible assumptions it is perfectly consistent with that theory, and that the subjects seem unreasonably sceptical to the point of unreliability about epistemology, I don't think this kind of experimental work threatens IRI.

2. Knowledge By Indifference and By Wealth

Gillian Russell and John Doris[20] argue that Jason Stanley's account of knowledge leads to some implausible attributions of knowledge, and if successful their objections would generalise to other forms of IRI. I'm going to argue that Russell

[18] As I believe was the case in the I-35W collapse.

[19] See the massive drop in the numbers of students walking or biking to school, reported in Sandra A. Ham, Sarah Martin, and Harold W. Kohl III, "Changes in the percentage of students who walk or bike to school-United States, 1969 and 2001," *Journal of Physical Activity and Health* 5 (2008): 205–215, for a sense of how big an issue this is.

[20] Gillian Russell and John M. Doris, "Knowledge by Indifference," *Australasian Journal of Philosophy* 86 (2009): 429–437, doi:10.1080/00048400802001996.

and Doris's objections turn on principles that are prima facie rather plausible, but which ultimately we can reject for independent reasons.[21]

Their objection relies on variants of the kind of case Stanley uses heavily in his *Knowledge and Practical Interests* to motivate a pragmatic constraint on knowledge. Stanley considers the kinds of cases we used to derive IRI from Relevant Practical Coherence. So imagine an agent who faces a choice between accepting the status quo, call that φ, and taking some giant risk, call that ψ. The giant risk in this case will involve a huge monetary loss if ¬p, and a small non-monetary gain if p. Stanley says, and I agree, that in such a case the agent doesn't know p, even if their belief in p is true, well supported by evidence, and so on. Moreover, he says, had ψ not been a relevant option, the agent could have known p. I agree, and I think Relevant Practical Coherence explains these intuitions well.

Russell and Doris imagine two kinds of variants on Stanley's case. In one variant the agent doesn't care about the material loss associated with ψ ∧ ¬ p. As I would put it, although their material wealth would decline precipitously in that case, their utility would not, because their utility is not tightly correlated with material wellbeing. Given that, the agent may well prefer ψ to φ unconditionally, and so would still know p. Russell and Doris don't claim this is a problem in itself, but they do think the conjunction of this with the previous paragraph is a problem. As they put it, "you should have reservations ... about what makes [the knowledge claim] true: not giving a damn, however enviable in other respects, should not be knowledge-making."[22]

Their other variant involves an agent with so much money that the material loss is trifling to them. Since the difference in utility between having, say, eight billion dollars and seven billion dollars is not that high, perhaps they will again prefer ψ to φ unconditionally, so still know p. But it is, allegedly, counterintuitive to have the knowledge that p turn on the agent's wealth. As Russell and Doris say, "[m]atters are now even dodgier for practical interest accounts, because money turns out to be knowledge making."[23] And this isn't just because wealth can purchase knowledge. As they say, "money may buy the instruments of knowledge ... but here the connection between money and knowledge seems rather too direct."[24]

---

[21] I think the objections I make here are similar in spirit to those Stanley made in a comments thread on Certain Doubts, though the details are new. The thread is at http://el-prod.baylor.edu/certain_doubts/?p=616.

[22] Russell and Doris, "Knowledge by Indifference," 432.

[23] Russell and Doris, "Knowledge by Indifference," 433.

[24] Russell and Doris, "Knowledge by Indifference," 433.

The first thing to note about this case is that indifference and wealth aren't really producing knowledge. What they are doing is more like defeating a defeater. Remember that the agent in question had enough evidence, and enough confidence, that they would know p were it not for the practical circumstances. As I said in the introduction, practical considerations enter debates about knowledge in part because they are distinctive kinds of defeaters. It seems that's what is going on here. And we have, somewhat surprisingly, independent evidence to think that indifference and wealth do matter to defeaters.

Consider two variants on Gilbert Harman's 'dead dictator' example.[25] In the original example, an agent reads that the dictator has died through an actually reliable source. But there are many other news sources around, such that if the agent read them, she would lose her belief. Even if the agent doesn't read those sources, their presence can constitute defeaters to her putative knowledge that the dictator died.

In our first variant on Harman's example, the agent simply does not care about politics. It's true that there are many other news sources around that are ready to mislead her about the dictator's demise. But she has no interest in looking them up, nor is she at all likely to look them up. She mostly cares about literature, and will spend her day reading old novels. In this case, the misleading news sources are too distant, in a sense, to be defeaters. So she still knows the dictator has died. Her indifference towards politics doesn't generate knowledge – the original reliable report is the knowledge generator – but her indifference means that a would-be defeater doesn't gain traction.

It might be objected here that the agent doesn't know the dictator has died because there are misleading reports around saying the dictator is alive, and she is in no position to rebut them. But this is too high a standard for knowledge. There are millions of people in Australia who know that humans are contributing to global warming on purely testimonial grounds. Many, perhaps even most, of these people would not be able to answer a carefully put together argument that humans are not contributing to global warming, such as an argument that picked various outlying statistics to mislead the reader. And such arguments certainly exist; the conservative parts of the media do as much as they can to play them up. But the mere existence of such arguments doesn't defeat the average person's testimonial knowledge about anthropogenic global warming. Similarly, the mere existence of misleading reports does not defeat our agent's knowledge of the

[25] Gilbert Harman, *Thought* (Princeton: Princeton University Press, 1973), 75.

dictator's death, as long as there is no nearby world where she is exposed to the reports.[26]

In the second variant, the agent cares deeply about politics, and has masses of wealth at hand to ensure that she knows a lot about it. Were she to read the misleading reports that the dictator has survived, then she would simply use some of the very expensive sources she has to get more reliable reports. Again this suffices for the misleading reports not to be defeaters. Even before the rich agent exercises her wealth, the fact that her wealth gives her access to reports that will correct for misleading reports means that the misleading reports are not actually defeaters. So with her wealth she knows things she wouldn't otherwise know, even before her money goes to work. Again, her money doesn't generate knowledge – the original reliable report is the knowledge generator – but her wealth means that a would-be defeater doesn't gain traction.

The same thing is true in Russell and Doris's examples. The agent has quite a bit of evidence that p. That's why she knows p. There's a potential practical defeater for p. But due to either indifference or wealth, the defeater is immunised. Surprisingly perhaps, indifference and/or wealth can be the difference between knowledge and ignorance. But that's not because they can be in any interesting sense 'knowledge makers,' any more than I can make a bowl of soup by preventing someone from tossing it out. Rather, they can be things that block defeaters, both when the defeaters are the kind Stanley talks about, and when they are more familiar kinds of defeaters.

3. Temporal Embeddings

Michael Blome-Tillmann[27] has argued that tense-shifted knowledge ascriptions can be used to show that his version of Lewisian contextualism is preferable to IRI. Like Russell and Doris, his argument uses a variant of Stanley's Bank Cases.[28] Let O be that the bank is open Saturday morning. If Hannah has a large debt, she is in a high-stakes situation with respect to O. In Blome-Tillmann's version of the example, Hannah had in fact incurred a large debt, but on Friday morning the creditor waived this debt. Hannah had no way of anticipating this on Thursday. She has some evidence for O, but not enough for knowledge if she's in a high-stakes

[26] Thanks here to an anonymous referee.

[27] Michael Blome-Tillmann, "Contextualism, Subject-Sensitive Invariantism, and the Interaction of 'Knowledge'-Ascriptions with Modal and Temporal Operators," *Philosophy and Phenomenological Research* 79 (2009): 315–331, doi:10.1111/j.1933-1592.2009.00280.x.

[28] In the interests of space, I won't repeat those cases yet again here.

situation. Blome-Tillmann says that this means after Hannah discovers the debt waiver, she could say (1).

(1) I didn't know O on Thursday, but on Friday I did.

But I'm not sure why this case should be problematic for any version of IRI, and very unsure why it should even look like a reductio of IRI. As Blome-Tillmann notes, it isn't really a situation where Hannah's stakes change. She was never actually in a high stakes situation. At most her perception of her stakes change; she thought she was in a high-stakes situation, then realised that she wasn't. Blome-Tillmann argues that even this change in perceived stakes can be enough to make (1) true if IRI is true. Now actually I agree that this change in perception could be enough to make (1) true, but when we work through the reason that's so, we'll see that it isn't because of anything distinctive, let alone controversial, about IRI.

If Hannah is rational, then given her interests she won't be ignoring ¬ O possibilities on Thursday. She'll be taking them into account in her plans. Someone who is anticipating ¬ O possibilities, and making plans for them, doesn't know O. That's not a distinctive claim of IRI. Any theory should say that if a person is worrying about ¬ O possibilities, and planning around them, they don't know O. And that's simply because knowledge requires a level of confidence that such a person simply does not show. If Hannah is rational, that will describe her on Thursday, but not on Friday. So (1) is true not because Hannah's practical situation changes between Thursday and Friday, but because her psychological state changes, and psychological states are relevant to knowledge.

What if Hannah is, on Thursday, irrationally ignoring ¬ O possibilities, and not planning for them even though her rational self wishes she were planning for them? In that case, it seems she still believes O. After all, she makes the same decisions as she would as if O were sure to be true. But it's worth remembering that if Hannah does irrationally ignore ¬ O possibilities, she is being irrational with respect to O. And it's very plausible that this irrationality defeats knowledge. That is, you can't be irrational with respect to a proposition and know it. Irrationality excludes knowledge. In any case, I doubt this is the natural way to read Blome-Tillmann's example. We naturally read Hannah as being rational, and if she is rational she won't have the right kind of confidence to count as knowing O on Thursday.

There's a methodological point here worth stressing. Doing epistemology with imperfect agents often results in facing tough choices, where any way to describe a case feels a little counterintuitive. If we simply hew to intuitions, we risk being led astray by just focussing on the first way a puzzle case is described to

us. But once we think through Hannah's case, we see perfectly good reasons, independent of IRI, to endorse IRI's prediction about the case.

4. Problematic Conjunctions

Blome-Tillmann offers another argument against IRI, that makes heavy use of the notion of having enough evidence to know something. Here is how he puts the argument. (Again I've changed the numbering and some terminology for consistency with this paper.)

Suppose that John and Paul have exactly the same evidence, while John is in a low-stakes situation towards p and Paul in a high-stakes situation towards p. Bearing in mind that IRI is the view that whether one knows p depends on one's practical situation, IRI entails that one can truly assert:

(2) John and Paul have exactly the same evidence for p, but only John has enough evidence to know p, Paul doesn't.[29]

And this is meant to be a problem, because (2) is intuitively false. But IRI doesn't entail any such thing. We can see this by looking at a simpler example that illustrates the way 'enough' works.

George and Ringo both have $6000 in their bank accounts. They both are thinking about buying a new computer, which would cost $2000. Both of them also have rent due tomorrow, and they won't get any more money before then. George lives in New York, so his rent is $5000. Ringo lives in Syracuse, so his rent is $1000. Clearly, (3) and (4) are true.

(3) Ringo has enough money to buy the computer.

(4) Ringo can afford the computer.

And (5) is true as well, though there's at least a reading of (6) where it is false.

(5) George has enough money to buy the computer.

(6) George can afford the computer.

Focus for now on (5). It is a bad idea for George to buy the computer; he won't be able to pay his rent. But he has enough money to do so; the computer costs $2000, and he has $6000 in the bank. So (5) is true. Admittedly there are things close to (5) that aren't true. He hasn't got enough money to buy the computer and pay his rent. You might say that he hasn't got enough money to buy the computer given his other financial obligations. But none of this undermines (5).

[29] Blome-Tillmann, "Contextualism, Subject-Sensitive Invariantism," 328-9.

Now just like George has enough money to buy the computer, Paul has enough evidence to know that p. Paul can't know that p, just like George can't buy the computer, because of his practical situation. But that doesn't mean he doesn't have enough evidence to know it. He clearly does have enough evidence, since he has the same evidence John has, and John knows that p. So, contra Blome-Tillmann, IRI doesn't entail this problematic conjunction.

In a footnote attached to this, Blome-Tillmann offers a reformulation of the argument.

> I take it that having enough evidence to 'know p' in C just means having evidence such that one is in a position to 'know p' in C , rather than having evidence such that one 'knows p.' Thus, another way to formulate (2) would be as follows: 'John and Paul have exactly the same evidence for p, but only John is in a position to know p, Paul isn't.'[30]

Now having enough evidence to know p isn't the same as being in a position to know it, any more than having enough money to buy the computer puts George in a position to buy it. So I think this is more of a new objection than a reformulation of the previous point. But might it be a stronger objection? Might it be that IRI entails (7), which is false?

(7) John and Paul have exactly the same evidence for p, but only John is in a position to know p, Paul isn't.

Actually, it isn't a problem that IRI says that (7) is true. In fact, almost any epistemological theory will imply that conjunctions like that are true. In particular, any epistemological theory that allows for the existence of defeaters which do not supervene on the possession of evidence will imply that conjunctions like (7) are true. For example, anyone who thinks that whether you can know that a barn-like structure is really a barn depends on whether there are non-barns in the neighbourhood that look like the structure you're looking at will think that conjunctions like (7) are true. Again, it matters a lot that IRI is suggesting that traditional epistemologists did not notice that there are distinctively pragmatic defeaters. Once we see that, we'll see that conjunctions like (7) are not surprising at all.

5. Holism and Defeaters

The big lesson of the last few sections is that interests create defeaters. Sometimes an agent can't know p because adding p to her stock of beliefs would introduce either incoherence or irrationality. The reason is normally that the agent faces

[30] Blome-Tillmann, "Contextualism, Subject-Sensitive Invariantism," 329n23.

some decision where it is, say, bad to do φ, but good to do φ given p. In that situation, if she adds p, she'll either incoherently think that it's bad to do φ although it's good to do it given what is (by her lights) true. Moreover, the IRI theorist says, being incoherent in this way blocks knowledge, so the agent doesn't know p.

But there are other, more roundabout, ways in which interests can mean that believing p would entail incoherence. One of these is illustrated by an example alleged by Ram Neta to be hard for interest-relative theorists to accommodate.

> Kate needs to get to Main Street by noon: her life depends upon it. She is desperately searching for Main Street when she comes to an intersection and looks up at the perpendicular street signs at that intersection. One street sign says 'State Street' and the perpendicular street sign says 'Main Street.' Now, it is a matter of complete indifference to Kate whether she is on State Street–nothing whatsoever depends upon it.[31]

Let's assume for now that Kate is rational; dropping this assumption introduces mostly irrelevant complications. That is, we will assume Kate is an expected utility maximiser. Kate will not believe she's on Main Street. She would only have that belief if she took it to be settled that she's on Main, and hence not worthy of spend- ing further effort investigating. But presumably she won't do that. The rational thing for her to do is to get confirming (or, if relevant, confounding) evidence for the appearance that she's on Main. If it were settled that she was on Main, the rational thing to do would be to try to relax, and be grateful that she had found Main Street. Since she has different attitudes about what to do simpliciter and conditional on being on Main Street, she doesn't believe she's on Main Street.

So far so good, but what about her attitude towards the proposition that she's on State Street? She has enough evidence for that proposition that her credence in it should be rather high. And no practical issues turn on whether she is on State. So she believes she is on State, right?

Not so fast! Believing that she's on State has more connections to her cognitive system than just producing actions. Note in particular that street signs are hardly basic epistemic sources. They are the kind of evidence we should be 'conservative' about in the sense of Pryor.[32] We should only use them if we

---

[31] Ram Neta, "Anti-intellectualism and the Knowledge-Action Principle," *Philosophy and Phenomenological Research* 75 (2007): 182, doi:10.1111/j.1933-1592.2007.00069.x.

[32] James Pryor, "What's Wrong with Moore's Argument?" *Philosophical Issues* 14 (2004): 349–378, doi:10.1111/j.1533-6077.2004.00034.x.

antecedently believe they are correct. So for Kate to believe she's on State, she'd have to believe the street signs she can see are correct. If not, she'd incoherently be relying on a source she doesn't trust, even though it is not a basic source.[33] But if she believes the street signs are correct, she'd believe she was on Main, and that would lead to practical incoherence. So there's no way to coherently add the belief that she's on State Street to her stock of beliefs. So she doesn't know, and can't know, that she's either on State or on Main. This is, in a roundabout way, due to the high stakes Kate faces.

Neta thinks that the best way for the interest-relative theorist to handle this case is to say that the high stakes associated with the proposition that Kate is on Main Street imply that certain methods of belief formation do not produce knowledge. And he argues, plausibly, that such a restriction will lead to implausibly sceptical results. But that's not the only way for the interest-relative theorist to go. What they could, and I think should, say is that Kate can't know she's on State Street because the only grounds for that belief are intimately connected to a proposition that, in virtue of her interests, she needs very large amounts of evidence to believe.

6. Non-Consequentialist Cases

None of the replies yet have leaned heavily on the last of the three points from the introduction, the fact that IRI is an existential claim. This reply will make heavy use of that fact.

If an agent is merely trying to get the best outcome for themselves, then it makes sense to represent them as a utility maximiser. But when agents have to make decisions that might involve them causing harm to others if certain propositions turn out to be true, then I think it is not so clear that orthodox decision theory is the appropriate way to model the agents. That's relevant to cases like this one, which Jessica Brown has argued are problematic for the epistemological theories John Hawthorne and Jason Stanley have recently been defending.[34]

[33] The caveats here about basic sources are to cancel any suggestion that Kate has to antecedently believe that any source is reliable before she uses it. As Pryor notes, that view is problematic (James Pryor, "The Skeptic and the Dogmatist," *Noûs* 34 (2000): 517–549, doi:10.1111/0029-4624.00277). The view that we only get knowledge from a street sign if we antecedently have reason to trust it is not so implausible.

[34] The target here is not directly the interest-relativity of their theories, but more general principles about the role of knowledge in action and assertion. But it's important to see how

> A student is spending the day shadowing a surgeon. In the morning he observes her in clinic exam- ining patient A who has a diseased left kidney. The decision is taken to remove it that afternoon. Later, the student observes the surgeon in theatre where patient A is lying anaesthetised on the operating table. The operation hasn't started as the surgeon is consulting the patient's notes. The student is puzzled and asks one of the nurses what's going on:
>
> *Student*: I don't understand. Why is she looking at the patient's records? She was in clinic with the patient this morning. Doesn't she even know which kidney it is?
>
> *Nurse*: Of course, she knows which kidney it is. But, imagine what it would be like if she removed the wrong kidney. She shouldn't operate before checking the patient's records.[35]

	Left	**Right**
Remove Left Kidney	1	-1
Remove Right Kidney	-1	1
Check notes	1 - ε	1 - ε

It is tempting, but I think mistaken, to represent the payoff table associated with the surgeon's choice as follows. Let Left mean the left kidney is diseased, and Right mean the right kidney is diseased.

Here ε is the trivial but non-zero cost of checking the chart. Given this table, we might reason that since the surgeon knows that she's in the left column, and removing the left kidney is the best option in that column, she should remove the left kidney rather than checking the notes.

But that reasoning assumes that the surgeon does not have any obligations over and above her duty to maximise expected utility. And that's very implausible, since consequentialism is a fairly implausible theory of medical ethics.[36]

It's not clear exactly what obligation the surgeon has. Perhaps it is an obligation to not just know which kidney to remove, but to know this on the basis of evidence she has obtained while in the operating theatre. Or perhaps it is an

---

IRI handles the cases that Brown discusses, since these cases are among the strongest challenges that have been raised to IRI.

[35] Jessica Brown, "Knowledge and Practical Reason," *Philosophy Compass* 3 (2008): 1144-1145, doi:10.1111/j.1747- 9991.2008.00176.x.

[36] I'm not saying that consequentialism is wrong as a theory of medical ethics. But if it is right, so many intuitions about medical ethics are going to be mistaken that such intuitions have no evidential force. And Brown's argument relies on intuitions about this case having evidential value. So I think for her argument to work, we have to suppose non-consequentialism about medical ethics.

obligation to make her belief about which kidney to remove as sensitive as possible to various possible scenarios. Before she checked the chart, this counterfactual was false: Had she misremembered which kidney was to be removed, she would have a true belief about which kidney was to be removed. Checking the chart makes that counterfactual true, and so makes her belief that the left kidney is to be removed a little more sensitive to counterfactual possibilities.

	Left	**Right**
Remove Left Kidney	1 - δ	-1 - δ
Remove Right Kidney	-1 - δ	1 - δ
Check notes	1 - ε	1 - ε

However we spell out the obligation, it is plausible given what the nurse says that the surgeon has some such obligation. And it is plausible that the 'cost' of violating this obligation, call it δ, is greater than the cost of checking the notes. So here is the decision table the surgeon faces.

And it isn't surprising, or a problem for an interest-relative theory of knowledge, that the surgeon should check the notes, even if she believes and knows that the left kidney is the diseased one. This is not to say that the surgeon does know that the left kidney is diseased, just that the version of IRI being defended here is neutral on that question.

There is a very general point here. It suffices to derive IRI that we defend principles like the following:

> - Whenever maximising expected value is called for, one should maximise expected value conditional on everything one knows.

Maximising expected value is called for often enough that there exist the kinds of pairs of cases IRI claims exist. That's because in some cases, changing the options facing an agent will make it the case that which live option is best differs from which live option is best given p, even though the agent antecedently knew p.

But that doesn't imply that maximising expected value is always called for. Especially in a medical case, it is hard to square an injunction like "Do No Harm!" with a view that one should maximise expected value, since maximising expected value requires treating harms and benefits symmetrically. What would be a problem for the version of IRI defended here was a case with the following four characteristics.

> - Maximising expected value is called for in the case.

- Conditional on p, the action with the highest expected value is φ.
- It would be wrong to do φ.
- The agent knows p.

It is tempting for the proponent of IRI to resist any attempted counterexample by claiming it is not really a case of knowledge. That might be the right thing to say in Brown's case. But IRI defenders should remember that it is often a good move to deny that the first condition holds. Consequentialism is not an obviously correct theory of decision making in morally fraught situations; purported counterexamples that rely on it can therefore be resisted.

DEBATES

BAILEY ON INCOMPATIBILISM AND THE "NO PAST OBJECTION"

Anthony BRUECKNER, Christopher T. BUFORD

ABSTRACT: In "Incompatibilism and the Past," Andrew Bailey engages in a thorough investigation of what he calls the "No Past Objection" to arguments for incompatibilism. This is an objection that stems from the work of Joseph Keim Campbell and that has generated an Interesting literature. Bailey ends by offering his own answer to the No Past Objection by giving his own argument for incompatibilism, an argument that he claims to be immune to the objection. We have some observations to make regarding what we take to be Bailey's answer to the objection (all of whose details are left to the reader – we attempt to fill this lacuna).

KEYWORDS: incompatibilism, consequence argument, Joseph Keim Campbell's "No Past" objection, Andrew Bailey

In "Incompatibilism and the Past," Andrew Bailey engages in a thorough investigation of what he calls the "No Past Objection" (hereafter NPO) to arguments for incompatibilism (hereafter INC).[1] This is an objection that stems from the work of Joseph Keim Campbell and that has generated a literature on the objection.[2] Bailey ends by offering his own answer to the objection by giving his own allegedly NPO-immune argument for INC. We have some observations to make regarding what we take to be Bailey's answer to NPO, *all* of whose details are left to the reader.

Here is Bailey's statement of NPO: "The premises of … [vanInwagen's] Consequence Argument [as well as those of all the other prominent arguments for INC] are not necessary truths because there needn't be a past." This objection is based on Campbell's example of Adam, who comes into existence at the first moment of time t, and at t Adam, let us suppose, reaches for the famous apple, performing a seemingly free action.

Our take on NPO is as follows. Assume that each of the extant arguments for INC (all canvassed by Bailey) can 'take care of' each allegedly free action relative to which there is a past, showing that on the assumption of determinism, each such act is *not* free. Still, none of the extant arguments for INC can 'take care

[1] Forthcoming in *Philosophy and Phenomenological Research*.

[2] See his "Free Will and the Necessity of the Past," *Analysis* 67 (2007): 105-11.

© LOGOS & EPISTEME, II, 4 (2011): 613-617

of' Adam's seemingly free reaching for the apple at t, for there is no past relative to t. Each of the extant arguments requires, for one reason or another, that there be a past relative to the time of a targeted allegedly free action, in order to show that on the assumption of determinism, the targeted action is not free.

Bailey rejects ways of responding to NPO that lay down a requirement that S performs a freely only if "some *historical* condition is satisfied (a condition that entails that S existed prior to her performing a)." One example: S's having deliberated whether to perform a.[3] Bailey rejects each variant on this strategy, claiming that each "takes on rather special metaphysical commitments." We will return to this objection below.

Bailey formulates what he calls "Another Argument" for INC as a response to NPO. Not only does this argument take care of all the allegedly free acts relative to which there is a past, but, further, it takes care of Adam's reaching for the apple. Another Argument is centered around a proposed necessary condition for S's being 'free with respect to a truth' ('s(t)' stands for a proposition expressing the total state of the universe at time t):

> (NC) If S is free with respect to p, then there is some time t such that the conjunction of s(t) and the laws is compossible with not-p.

Another Argument runs as follows:

> P1. Necessarily, for any subject S, and any truth p, if S is ever free with respect to p, then there is some time t such that the conjunction of s(t) and the laws is compossible with not-p. (premise [using NC])
>
> P2. Necessarily, if determinism is true, then for any time t and truth p, the conjunction of s(t) and the laws entails p. (definition of determinism)
>
> P3. Therefore: necessarily, if determinism is true, then for any time t and truth p, the conjunction of s(t) and the laws is not compossible with not-p. (from P2)
>
> P4. Therefore: necessarily, if determinism is true, for any subject S, and any truth p, S is not ever free with respect to p. (from P1 and P3).

Bailey declares that Another Argument is 'immune to' NPO, but, surprisingly, he does not give a word of explanation as to why. To answer NPO, Bailey's argument must take care of Adam's reaching for the apple at the first moment of time t. The only clue as to how this is supposed to work is given in his setting up of NC:

> Worlds x and y *share a time* just in the case that the complete state of the world at some time … in x is an intrinsic duplicate of the complete state of the world at

[3] See Anthony Brueckner, "Retooling the Consequence Argument," *Analysis* 68 (2008): 10-12.

> some time in y. … if determinism is true, a world sharing *any* time with the actual world (and sharing the actual world's laws) shares all times and all truths with the actual world.

This shows that Bailey (though he does not explicitly say so) is working with vanInwagen's 'backward-forward' formulation of determinism:

> (VI) If p and q are any propositions that express the state of the world at some instants, then the conjunction of p with the laws of nature entails q.[4]

Since VI leaves it open as to temporal ordering of the instants, VI says both that (1) the past plus the laws determine a unique future, and that (2) the future plus the laws determine a unique past. So on VI, if world W and @ share a time t and share @'s laws, then not only do they share all times subsequent to t ('forward determinism'), but, further, they share all times prior to t ('backward determinism').

Let us return to Adam. By NC, for his reaching for the apple to be a free action, we must find a time t' such that the conjunction of s(t') and the laws is compossible with *not-(Adam reaches for the apple at t)* (hereafter *not-R*). On the assumption of VI, no time after t will fill the bill. Take any such time t''. By the 'backward' component of VI, the conjunction of s(t'') and the laws entails R. So no time after t can play the role of the desired t'. The only time left is t itself, the first moment of time. Since s(t) entails R, the conjunction of s(t) and the laws is, trivially, *not* compossible with not-R. So t cannot be the t' that we seek. Thus Adam is not free with respect to R, since NC is not satisfied. So Bailey *can* take care of Adam's allegedly free action at t – on the assumption of VI-style determinism, Adam does *not* freely reach for the apple.

We have several observations about our reconstructed Bailey-style answer to NPO.

1. We do not know whether this is *Bailey's* answer, since he says absolutely nothing about how his Another Argument provides an answer to the objection. But we can see no other way of answering the objection using the materials Bailey has provided.

2. The answer we have reconstructed requires the 'backward' component of VI. Bailey does not say so, but he is committed by his discussion of NC to this 'backward determinism' as well as to the standard 'forward determinism.'

3. In general, if S's action A at t satisfies NC (so that it can be true that S is 'free with respect to the proposition p' according to which S does A at t), then the time t' which is relevant to this satisfaction must be *different from* t. This is because t' must be such that the conjunction of s(t') and the laws is compossible with not-p

[4] See Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983), 65.

(i.e., not-(S performs A at t)). As noted in our reconstruction of Bailey's answer to NPO, it is trivial that s(t) is not compossible with not-p, and so the conjunction of s(t) and the laws is not compossible with not-p.

4. One consequence of observation (3) is that Bailey's judgement about one of his examples is inconsistent with his own NC. Consider a world W in which there is just one moment of time t, in which Adam reaches for the apple – no past relative to t, no future relative to t (this is like Bailey's 'Instantaneous Chooser' case). Bailey discusses a principle that Roberto Loss formulates in an argument for INC that is meant to avoid NPO.[5] According to 'the necessity of the present,' roughly: if p is true at t, then no one has any choice at t about p's being true at t. Bailey seems to miss the intuitive point behind this principle in the course of giving alleged counterexamples to it, one of which involves the world W just described. Loss's point in putting forward his principle was that once you begin to F, it is *too late* to have a choice about F'ing. So if it is true at t that you are F'ing, then you have no choice about its being true at t that you are F'ing. You are already F'ing! Bailey's alleged counterexamples to the necessity of the present all have the form: assume that in the following conditions S has a choice at t about F'ing at t. This seems to us to be question-begging. However, picking up the main thread of this observation, one of Bailey's alleged counterexamples to the necessity of the present involves W. Bailey maintains that there is no bar to holding that Adam's reaching for the apple at the unique moment of time t in W is freely done and, contra the necessity of the present, is an action that Adam has a choice about at t. But as noted in (3) above, Adam's act at t cannot be freely done (and thus presumably cannot be chosen at t by Adam) because NC is violated: since t is the only moment of time in W, there is no t' in W which is such that s(t') and the laws are compossible with not-(Adam reaches for the apple at t).

5. Observation (4) shows that Bailey's NC has the effect of committing him to a sort of historical condition on free action, a commitment, we noted earlier, that he holds to be a problematic cost for an answer to NPO. Bailey's historical condition is that if S does A freely at t (if S is 'free with respect to the proposition' that S does A at t), then either there is a past relative to t, or there is a future relative to t. If this condition is not satisfied, then it immediately follows that NC is not satisfied as well, as shown in (4). Thus S does not do A freely, given the non-satisfaction of the historical condition. If the historical condition *is* satisfied, then there will be a time, or times, which are at least candidates for allowing the satisfaction of NC, since the time, or times, will be distinct from t.

[5] Roberto Loss, "Free Will and the Necessity of the Present," *Analysis* 69: 63-9 (2009).

6. Observation (5) brings out the following peculiarity of Bailey's NC. We say to Bailey: "Imagine a world that begins at t with Adam reaching for the apple at t. Does it make sense to suppose that Adam's action is free?" Bailey responds, "It all depends. Is there any future after t in this world? If no, then Adam's action is unfree, since NC will not be satisfied. If yes, then Adam's action is at least a candidate for being free (if, say, the world in question is indeterministic and NC is thus satisfied)."

We conclude that the answer to NPO that we have reconstructed from the materials provided by Bailey has problematic costs, just as do the other answers canvassed by Bailey in his paper.

ON EPISTEMIC ABSTEMIOUSNESS: A REPLY TO AIKIN, HARBOUR, NEUFELD, AND TALISSE

Alex BUNDY

ABSTRACT: The principle of suspension says that when you disagree with an epistemic peer about p, you should suspend judgment about p. In "Epistemic Abstainers, Epistemic Martyrs, and Epistemic Converts," Scott F. Aikin, Michael Harbour, Jonathan Neufeld, and Robert B. Talisse argue against the principle of suspension. In "In Defense of Epistemic Abstemiousness" I presented arguments that their arguments do not succeed, and in "On Epistemic Abstemiousness: A Reply to Bundy" they argue that my arguments are not successful. I here clarify and defend my arguments.

KEYWORDS: disagreement, epistemic abstemiousness, epistemic martyrdom, epistemic conversion

In "Epistemic Abstainers, Epistemic Martyrs, and Epistemic Converts," Scott F. Aikin, Michael Harbour, Jonathan Neufeld, and Robert B. Talisse (hereafter AHNT) argue against the principle of suspension.[1]

> (PS) If S disagrees with an epistemic peer about p, S is aware of the disagreement, and S is justified in thinking her peer is a peer , then S should suspend judgment about p.[2]

In "In Defense of Epistemic Abstemiousness," I reply to AHNT's arguments. I argue that their arguments do not succeed.[3] In "On Epistemic Abstemiousness: A Reply to Bundy," they respond to my reply; they think that my counter-arguments are unsuccessful.[4] This is my response. It is still my position that neither of the

[1] Scott Aikin, Michael Harbour, Jonathan Neufeld, and Robert B. Talisse, "Epistemic Abstainers, Epistemic Martyrs, and Epistemic Converts," *Logos & Episteme* 1, 2 (2010): 211–9.

[2] This principle is labeled *PS+* in my response on 289. Alex Bundy, "In Defense of Epistemic Abstemiousness," *Logos & Episteme* II, 2 (2011): 287–92.

[3] Alex Bundy, "In Defense."

[4] Scott Aikin, Michael Harbour, Jonathan Neufeld, and Robert B. Talisse, "On Epistemic Abstemiousness: A Reply to Bundy," *Logos & Episteme* II, 3 (2011): 425–8.

arguments they presented in their original paper succeed in showing that PS is false.

The epistemic martyrdom argument

In their first argument AHNT claim that PS is a false principle because following it could lead one to mislead those one disagrees with. In their example that is supposed to illustrate this problem with PS, Mary believes not-p, Alf believes p, each judges to the other to be an epistemic peer, and each is aware of the disagreement about p. It is thus a case where PS applies, and so Betty suspends judgment regarding p. Alf, in turn, becomes more confident in p because he judges that Betty's suspension provides some evidence that his belief that not-p is correct. AHNT argue that Alf should not become more confident in p in this case, and that PS is to blame for this mistake.

My response to this argument was that whether one's adjustment of one's beliefs leads others to be misled is irrelevant to whether that adjustment is epistemically rational.[5] I took it that the notion of rationality in question in the disagreement literature is such that whether a belief is rational supervenes on the evidence that one has, and not on the evidence one has as well as the future consequences for the doxastic attitudes of oneself or others. In their response, AHNT make it clear that they think this is the wrong way to evaluate PS. They think that PS should be evaluated, at least in part, based on its consequences for others, and for one's future epistemic well-being.

I acknowledge that norms like PS can be evaluated in this way. Following Richard Foley, different forms of rationality may be distinguished based on their respective 'ends.'[6] AHNT clearly favor a form of rationality according to which one's own future doxastic states, and the doxastic states of others, are ends. I took the form of rationality of which PS is supposed to be a norm of to concern,

[5] I actually examined two ways of interpreting AHNT's martyrdom argument The first took the argument to say PS is a bad principle because a person's (in the example, Betty's) following PS would lead to others being misled. The second took the argument to say that PS supports the claim that a peer's *shift* from belief that p to suspension of belief regarding p provides evidence for one's own belief that not-p. My objection to the former interpretation was based on the claim that how one's beliefs affect the beliefs of others is irrelevant to evaluating a norm of rationality like PS; my objection to the latter was not based on this claim. AHNT do not state which reading of their argument they intended, but their response is to the former argument, and so presumably that is one of the arguments they intended to put forward.

[6] Richard Foley, *The Theory of Epistemic Rationality* (Cambridge: Harvard University Press, 1987), 7–8.

roughly, the ends of now having true beliefs, and now not having false beliefs. It is not surprising, then, that AHNT found my counter-argument to be unconvincing.

I do not, however, think my response was completely off the mark. The other- and future-regarding notion of rationality AHNT favor is not the one in question in the debates regarding the appropriate way to respond to disagreement with a peer. If it were, then PS would not have the supporters it does. AHNT themselves point out that what motivates Feldman's support of PS is that disagreement with a peer indicates that one's own view is false.[7] If the present accuracy of one's beliefs is not the only end associated with the notion of rationality in play in PS, then it is not immediately clear what the rationale for PS would be, since then evidence that one might be mistaken about a particular proposition would not be sufficient for showing that one's doxastic attitude towards that proposition is unreasonable. Thus, evaluating PS according to its consequences for one's future doxastic states and the doxastic states of others involves a change of subject. I agree with AHNT that a study of norms that take the one's future doxastic states and the doxastic states of others as ends is worthwhile. But I also think that the form of rationality that has to do with the goal of now having true beliefs and not having false beliefs is one worthy of study.

Finally, even if we grant that it is appropriate to evaluate PS in the way AHNT do, their argument does not succeed in showing that PS is false. The reason for this is that in their story, Alf is misled when he comes to believe that Betty changed her belief. And presumably this is because Betty did something to let him know she changed her belief – e.g. she told him she now suspends belief regarding p. It is thus Betty's actions that lead to Alf being misled, not her belief. Betty could easily change her belief, and claim that she did not change her belief. PS, thus, does not lead to Alf's being misled, and PS does not lead to the systematic misleading of others in the way AHNT say it does.

Of course, it could be that Betty's adhering to PS and some other norm would lead to Alf's being misled. If, for instance, Betty had an epistemic duty to always be honest and vocal about her doxastic states, then it could be that PS would, if followed by an epistemically responsible agent, systematically lead others to be misled. If there is such an epistemic duty, then AHNT may have an argument against PS as a norm of other-regarding rationality. But as I suggested above, if PS is taken to be a norm that has to do with securing true beliefs for one's future self and for others, then PS would be unmotivated in any case.

---

[7]Aikin et. al., "Epistemic Abstainers," 217. The cited passage from Feldman is on page 307 of Richard Feldman, "Evidentialism, Higher-Order Evidence, and Disagreement," *Episteme* 6 (2009): 294–312.

The conversion argument

In their second argument against PS, AHNT argue that PS is a mistaken principle because it would lead a person deploying it to be misled whenever in a disagreement with a 'stubborn' peer who does not change his belief in the face of disagreement with a peer. In the case of the disagreement between Betty and Alf, we are supposed to imagine Betty applying PS and suspending belief regarding p, while Alf retains his belief that p. Alf's retaining his belief that p is now supposed to be evidence for Betty for p. And when Betty considers whether it is appropriate to suspend belief regarding p, she will see that Alf – her peer – thinks that it is not; he thinks that one should believe p. So Betty should suspend about whether to suspend when it comes to p. And this is supposed to be further evidence for p. And after Betty reasons along these lines for awhile, with Alf stubbornly retaining his belief that p the whole time, Betty will eventually arrive at the point where the appropriate attitude to have towards p is believing p.

In response to my response to this argument, AHNT claim that I claim that Alf would be 'double-counting' the evidence in this case, and they think that this is a mistake. My response to the conversion argument, however, does not rely on any claims about why Alf is being stubborn, since as far as I can tell, the conversion argument does not rely on what Alf's reasons for being stubborn are. Betty is the one who is supposed to be reasoning herself into a bad epistemic situation via repeated application of PS. So if my counter-argument were that Alf is doing some double-counting, that would be a bad counter-argument; but that was not my argument.

AHNT also suggest that I should hold that Betty would be double-counting the evidence in this case. I think AHNT are taking me to say that both (a) Alf's belief that p, and (b) his retaining the belief that p in the face of disagreement, should count as only one piece of evidence for Betty. And I agree with AHNT that this would be a mistake; these are clearly two different pieces of evidence. But this was not my argument either. In fact, my argument depends on (a) and (b) being treated as distinct pieces of evidence with different import. I argued that following PS, Betty should suspend judgment regarding p when she discovers (a), that Alf believes p. But then I further suggest that when Alf does not suspend judgment regarding p, she acquires (b), which gives her reason to think that Alf is not applying PS, which in turn gives her reason to think that Alf should not be fully trusted when it comes to p. Betty thus could reasonably conclude that Alf is not a peer when it comes to p and so does not have to apply PS to the case. PS thus does not have the implausible consequence that whenever one disagrees with a stubborn peer, one should end up with the peer's belief.

In response to this argument, AHNT suggest that if Betty takes Alf's failure to conform to PS as evidence that Alf is not a peer, then she is treating PS as a rough criterion of peerhood. And they think this would be "against the spirit of the principle," since PS is about taking the opinions of peers seriously, while my suggestion is that Betty can use the purported truth of PS to dismiss Alf as not being her peer. And they are right that I do think that PS, if true, is a criterion of epistemic peerhood. It is a criterion of epistemic peerhood in the sense that as a norm of rationality, a person's failure to conform to it indicates an epistemic failure of the agent, and it is the epistemic successes and failures of an agent that enable one to judge whether someone is one's peer. I think that all norms of rationality are criteria of epistemic peerhood in this sense.

The apparent implausibility of Betty's demoting Alf from peerhood in the way I suggested can be explained away in two ways. First, as I specified peerhood, it was relative to a particular proposition. So when Betty judges Alf to not be a peer, it is only with respect to this one proposition, p. And this is consistent with her continuing to judge him to be overall as smart, well-informed, etc. as she; indeed, it even is consistent with her thinking that Alf is, overall, her epistemic superior. All Betty is doing is judging that Alf's opinion about the truth of p should be taken with a grain of salt; she is not coming to a judgment of Alf's overall epistemic character.

Second, my suggestion that Betty can take Alf's stubbornness as evidence that Alf is not her peer with respect to p does not entail that in every case of disagreement with a supposed peer about p, one may judge that the stubborn peer is her epistemic inferior regarding p. In fact, I admitted that in some cases a peer's stubbornness in such a situation will be evidence that the person is better positioned epistemically, and so is one's epistemic superior when it comes to evaluating whether p. So dismissal of a peer's opinions is not prescribed even when it comes to disagreement about particular propositions.

Perhaps the most serious objection AHNT have to my reply to the conversion argument is that it has the consequence that disagreement between peers who believe each other to be peers and are aware of the disagreement is impossible.[8] Given that there are actual cases just like this, if this is true my argument is clearly mistaken. The impossibility is supposed to follow from the claim that "whenever two parties are locked in a disagreement over p, they will, if they are in fact peers, agree to suspend judgment with respect to p."[9] I do not see how this is supposed to

[8] They just say "disagreement between peers is impossible," without the qualifications I have added; It is clear they take the qualifications to be understood.

[9] Aikin et. al., "On Epistemic Abstemiousness," 427.

follow from PS or my suggestion that PS can be used as one way of evaluating whether someone is a peer. It is true that both peers in such a case should, if PS is true, suspend judgment. Yet if neither suspends judgment they can nevertheless remain peers, since they would both be equally guilty of violating PS.

Finally, AHNT suggest that PS is a mistaken principle, because if there is a disagreement between peers about the truth of PS, then PS says one should suspend judgment about PS. PS thus undermines itself in instances where peers disagree about PS; and presumably this is actually the case – there are epistemic peers who recognize each other as peers who disagree about whether PS is true. And if PS can undermine itself in this way, the thinking goes, it must be a false principle. This is an argument against PS that was not presented in AHNT's original paper. It is a version of an argument presented by Adam Elga in his "How to Disagree About How to Disagree."[10] Elga thinks the argument works. I think it might work as well. So AHNT may have a good reason for thinking that PS is false. The arguments presented in AHNT's original paper, however, do not show that PS is false.

[10] Adam Elga, "How to Disagree About How to Disagree," in *Disagreement*, eds. Richard Feldman and Ted A. Warfield (Oxford: Oxford University Press, 2010), 178–82.

RE-REDUCING RESPONSIBILITY: REPLY TO AXTELL

Trent DOUGHERTY

ABSTRACT: In this brief reply to Axtell, I review some general considerations pertaining to the disagreement and then reply point-by-point to Axtell's critique of the dilemma I pose for responsibilists in virtue epistemology. Thus I re-affirm my reductionist identity thesis that every case of epistemic irresponsibility is either a case of ordinary moral irresponsibility or ordinary practical irrationality.

KEYWORDS: virtue epistemology, evidentialism, reductionism

Guy Axtell raises roadblocks for my reduction of responsibility.[1] In this short piece, I attempt to shore up my argument for my reduction of responsibility, which I encapsulated in the following identity thesis.

> *IT* Each instance of [so-called] epistemic irresponsibility is just an instance of purely non-epistemic irresponsibility/irrationality (either moral or instrumental). (RR, 4)

In this essay, I will defend IT, since IT still seems true to me. The emphasis is on the 'just' and it helps to consider this conditional which IT entails. If X is an instance of (so-called) epistemic irresponsibility, then X is just an instance of either (standard non-epistemic) moral irresponsibility or (standard non-epistemic) instrumental irrationality. That is, we have a standard category of moral ir/responsibility which applies to things like forgetting to mail an important check, drinking too much (non-addictively), and the like. And we have a standard category of practical ir/rationality which applies to things like buying lottery tickets, spending too much on a watch, and the like. Neither of these has anything particularly epistemic about it. My thesis is that every proposed example of so-called epistemic irresponsibility falls into one of these two categories. Furthermore, when the

[1] Guy Axtell, "Recovering Responsibility," *Logos & Episteme* II, 3 (2011): 429-454. I, in turn, was defending evidentialism as set forth by Earl Conee and Richard Feldman's *Evidentialism* (Oxford: Oxford University Press, 2004). See Trent Dougherty, "Reducing Responsibility: An Evidentialist Account of Epistemic Blame," *European Journal of Philosophy*, 2011, doi: 10.1111/j.1468-0378.2010.00422.x.

morally irresponsible action has to do with a belief the irresponsibility doesn't take on some sui generis, emergent 'epistemic' nature any more than forgetting to feed the cat takes on a sui generis, emergent 'feline irresponsibility.' And the same goes for practical irrationality. Being practically irrational with respect to some matter of belief does not result in some sui generis, emergent 'epistemic' irrationality any more than paying too much for a meal takes on some sui generis, emergent 'culinary irresponsibility.' In short, the subject matter of a morally irresponsible or practically irrational act does not result in the emergence of a new category or the blending of two other categories. There are *nothing but* moral irresponsibility or practical irrationality in cases of epistemic irresponsibility. That is the thesis I defended and that Axtell takes aim at.

One slogan under which I could march is the following: "The Ethic of Belief is Ethics." That is, it is a form of *applied ethics*. Some applied ethicist focus on such actions as killing in war or killing in the womb, but the ethicist of belief focuses on such actions as gathering evidence, spending time in reflection, and the like. This is a very important kind of applied ethics. It is one I'm deeply interested in. But it simply confuses matters to think it is a part of epistemology. And there is much more at stake here than a turf war over the term 'epistemology.' By failing to realize that the ethics of belief is just a kind of applied ethics, serious mistakes are made about the nature of epistemic justification, knowledge, and other forms of positive epistemic status.[2]

A general critique – which will inform my response to Axtell – of virtue responsibilism and 'regulative' epistemology more generally (even when it results in really great work[3]) is that it is surely a contingent matter what kinds of acts positively correlate with obtaining epistemic desiderata – truth among the foremost – and so properly the subject of empirical psychology. While this is acknowledged in the case of Roberts and Wood, I believe it can lead to confusion about the nature of epistemic concepts. For example, Axtell brings to bear against me Roger Pouivet's claim that "good intellectual habits ground our pretensions to warranted beliefs, and to knowledge. And habits are properties of persons, not of beliefs."[4] His point is to argue for taking personal justification to be more basic than doxastic or propositional justification. In reply, note that for the claim not to

[2] For example, see Jason Baehr's "Evidentialism, Vice, and Virtue," in *Evidentialism and Its Discontents*, ed. Trent Dougherty (Oxford: Oxford University Press, 2011) (which appears next to Axtell's "From Internalist Evidentialism to Virtue Responsibilism"), along with replies by Conee and Feldman.

[3] Like Robert Roberts and Jay Wood, *Intellectual Virtues: An Essay in Regulative Epistemology* (Oxford, 2007) and Jason Baehr, *The Inquiring Mind* (Oxford, 2011).

[4] Axtell, "Recovering Responsibility," 432.

be a tautology 'good' can't just mean 'truth conductive.' Rather, these habits must display some kind of ideality for which they are praiseworthy. But whether these idealistically praiseworthy features actually *are* truth conductive is a question for empirical psychologists. What it is *not* is an objection to any theory about the *nature* of any epistemic concept, be it evidence, justification, knowledge, or what have you. And that is what evidentialism is: it is a theory about the nature of evidence, justification, and, ultimately, knowledge.

Another quite broad disagreement is the following claim by Axtell. "Evaluation of persons (agents), as contrasted with the evaluation only of beliefs (cognitive states), is properly part of the theory of knowledge."[5] I think this is false. Persons can be evaluated for their actions or their properties. Beliefs are not actions.[6] I can't even locate a referent for an act-like thing we might call a 'believing' (which is a term Chisholm uses). Acts of inquiry are actions, such as walking to the library, checking out a book, reading it, interviewing witnesses, conducting experiments, and the like. And as I have suggested, such acts are just as easily evaluable morally and rationally when they are aimed at forming true beliefs as they are when aimed at finding a girlfriend.

But what properties of the person would we be evaluating as epistemic? Presumably not properties like *going to the library*, *checking out a book*, and the like, for those are just proxies for actions. Perhaps the relevant properties could be dispositions or habits. But when we assess an individual in this way we immediately run into a generality problem. Consider the following dispositions: *to inquire diligently; to inquire diligently on matters non-religious; to inquire diligently on matters horticultural; to inquire diligently on Tuesdays; to inquire diligently after having had coffee; to inquire diligently on matters caffeinated*. How finely do the norms go? And what of other kinds of dispositions, say, *to play games fairly*. Is there a sui generis, emergent kind of ludatory normativity which evaluates such dispositions? To evaluate an individual according to her dispositional properties may or may not make sense. But if it does, I see no reason to postulate any new fundamental normative categories besides the evidential, moral, and pragmatic.

Against the background of these broad conceptual disagreements, let us now turn to Axtell's reply to my specific argument. Axtell agrees that I've presented just the right kind of challenge with my Case of Craig the Creationist,[7] however, he thinks he can answer my challenge to come up with a non-reductionist, non-

[5] Axtell, "Recovering Responsibility," 434.

[6] At least not all of them. See my "Knowledge Happens: Why Zagzebski Hasn't Solved the Meno Problem," *Southern Journal of Philosophy* 49, 1 (2011): 73-88.

[7] Axtell, "Recovering Responsibility," 434.

evidentialist explanation of the cases. He notes that I set up the following trilemma concerning any case of epistemic irresponsibility: Either there are interests at stake or there are not. If there are not, then there is no irresponsibility. If there are, then they are either the individual's interests – in which case it's a failing of practical rationality – or the interests of others – in which case it is a case of moral irresponsibility. I will look at Axtell's objections to the horns of the dilemma he treats.

1. Objections to the Moral Horn of the Dilemma

Axtell offers four objections to this horn of the dilemma. First, Axtell says that what I have done is offer an error theory and that the theory for which it is an error theory needs defense. I reply that this is the wrong way to think about my argument (certainly I never put it that way). Rather, I took it as a datum that in such cases we feel that *something* is amiss in the case, but it's unclear what (it's just not plausible that it's *obvious* that there is this particular kind of sui generis, emergent normativity that no one can seem to pin down). I then offer as a prima facie plausible hypothesis that such cases entangle multiple forms of ordinary normativity. This explains both the feeling that something is amiss and our inability to nail down one property as being at fault. So I was not advancing an error theory but rather a *theory*.

Second, Axtell says that "That every self-regarding consideration is a non-moral consideration, and every other-regarding consideration is a moral one, is not an assumption that I think many ethicists share."[8] But I asserted no such thing, nor does my argument rely on any such assumption. Rather, I offered as the best explanation *of a certain kind of case* that it falls into these categories. Furthermore, I was explicit that these categories can overlap. In the idealized hypothesis I presented I only ever mentioned these categories as exclusive, but frankly I have no dog in the fight as to where to draw the line between the moral and the practically rational. In fact, my own view is that the moral is at best a heuristic notion within the broader class of the practically rational.

Third, Axtell says that I bias the question by casting responsibilism as concerned with belief formation 'behavior' rather than intellectual dispositions and habits. The problem with casting it in terms of belief formation behavior, he says, is that it biases the discussion toward my interpretation that the failings are ethical failings because behavior is what gets ethically evaluated. In reply, I think this dog won't hunt, for it's too clearly true that such acts of inquiry are at the core

[8] Axtell, "Recovering Responsibility," 445.

of intellectual responsibility. Axtell says my case is just the right kind of case and it is itself about a refusal to inquire, so the criticism seems odd. As very pertinent examples, both the Baehr piece mentioned above and the original criticisms of evidentialism addressed by Conee and Feldman in "Evidentialism"[9] involve objections to evidentialism from cases which are based on inadequate inquiry.

Nevertheless, even setting that aside, everything I say can be applied to intellectual dispositions and habits, *mutatis mutandis*. The bad habit of over-feeding the cat does not imply the existence of some sui generis, emergent cat normativity. It's just a bad habit that involves a cat. The habit of under-feeding a mind doesn't imply the existence of some sui generis, emergent epistemic normativity. It's just a bad habit that involves a mind. Some people are quite generally prudent, others quite generally imprudent, others spotty. But it would be profligate to posit an emergent form of normativity for every hole in someone's prudence.

Fourth, Axtell says that my gambit ignores "intellectual standards involving consideration of whether the agent manifested normal intellectual motivations, utilized effective or ineffective cognitive strategies for problemsolving, avoided known cognitive biases and fallacious tendencies in reasoning, etc."[10] I'm puzzled by this response since these things seem paradigmatic issues of morality or practical rationality. Axtell glosses this claim thusly.

> This latter evaluation will be clearly epistemic to the extent that what the presence (or absence) of normal desire for true belief and strategic efforts at inquiry is salient in explaining isn't the agent's blameworthiness in acting upon the belief, but simply why the agent in this instance was or wasn't successful in achieving distinctively epistemic aims such as true belief, etc.[11]

This will take a bit of unpacking. The base claim is that a certain kind of evaluation is clearly epistemic in nature. Let's call the evaluation – which I've claimed are paradigmatically non-epistemic – V, so the key claim is as follows.

> Key Claim: V is clearly epistemic.

The Key Claim is supported by appeal to an alleged fact – let's call it F – with this structure: that X is salient in explaining not Y but, rather, Z. Let's get the values clear.

> X = the presence (or absence) of normal desire for true belief and strategic efforts at inquiry

[9] The essay, originally appearing in *Philosophical Studies* in 1985 is anthologized in Conee and Feldman's *Evidentialism*.

[10] Axtell, "Recovering Responsibility," 446.

[11] Axtell, "Recovering Responsibility," 446.

Y = the agent's blameworthiness in acting upon the belief

Z = why the agent in this instance was or wasn't successful in achieving distinctively epistemic aims such as true belief

F = X is salient in explaining not Y but Z

I confess I do not see why F supports the Key Claim. I any evaluation of Z seems to me to be a matter of practical rationality for the reasons I've stated. That the aim of a venture is epistemic doesn't give rise to a new natural kind of normativity any more than in a case where the aim of the venture is distinctively automotive. In conversation, Robert Roberts has suggested that epistemic ventures are more core to what it is to be human than the matters to which I've compared it. This may well be, but then there are a number of other kinds of distinctively human activity which seem to be to be also at the core of the human: humor, gardening, play, and the like. It seems profligate to posit sui generis forms of normativity for each of these. And it seems a mere philosopher's bias to think that these truly humane matters are of secondary importance to intellectual activity.

2. Objections to the Practical Rationality Horn of the Dilemma

Axtell's objections to the practical rationality horn of the dilemma are less straightforward. I will deal with a number of points piecemeal. First, Axtell presumes "Evidential fit is presumably supposed to supply reflectively 'good reasons' that one can discursively offer as grounds for the belief."[12] The most basic error here is that evidential fit has any kind of purpose. It is a relation that holds among propositions. Perhaps Axtell means the purpose of *seeking beliefs which fit ones evidence* is to offer grounds for the belief. But of course that will vary from individual to individual and from case to case. Some individuals don't give a flying fig what others think. Some care but know too well they are surrounded only by those who don't. Yet others might both care and be surrounded by others who do but who already know what they know. One might seek evidential fit simply for self-enlightenment. It would make perfect sense for someone filled with wonder on a desert island to seek evidence for her hypotheses. I fail to see what any of this has to do with the theory of knowledge, epistemic justification, or rationality.

Even if the purpose of getting beliefs which fit the evidence *was* to answer challenges, it is not necessarily true that "if the agent hasn't been active in updating information when needed in order to maintain rational confidence, or put any effort into inquiring into counter-evidence to their belief once presented with it, then that agent's reasons aren't going to wash when someone asks them

[12] Axtell, "Recovering Responsibility," 447.

why they (still) believe it."[13] This just confuses dialectical considerations with epistemic considerations. Some people through good rhetoric or sheer cunning will be able to convince people on the basis of what they can think up on the moment. This is the standard route to a 'B' paper for undergrads. So I just don't see what these sorts of considerations are going to do to undercut evidentialism, which is principally a theory of epistemic justification.

Next, Axtell asks "But why can't it be both instrumentally and epistemologically evaluable, where true belief is valued intrinsically and true beliefs serve many a practical goal?"[14] In keeping with my theme, just substitute 'horticculturally' for 'epistemologically.' There are ever so many things that have some intrinsic value and serve practical goals. My argument was about fundamental types of normativity. Does Axtell really want ever so many fundamental kinds of normativity?

In the same paragraph, Axtell claims "Efforts to be actively fair-minded are obviously intimately involved in the improvement of agent reliability in areas of contested belief." This is far from clear to me. It has an appearance of common sense, but it is at best contingent. Furthermore, we know from the results of cognitive psychology that many such common sense sounding propositions are false. It may well be that – at least for some – concentrating too hard will actually induce more error. This is more than a mere theoretical possibility. One manifestation of this effort would be to read lots of philosophy. This is not known to be a reliable method of finding the truth.

Here is a claim I simply don't understand.

> [I]f the agent's performance in gathering and weighing evidence is thought to be irresponsible for want of skills rather than sound intellectual motivation, then the sense of 'instrumental' irrationality that could be applied would clearly be consonant with epistemic evaluations rather than representing merely practical or pragmatic irrationality.[15]

It defies my conceptual resources to ascribe irresponsibility do to want of skill. What skill did Craig lack (Craig seems to have been lost somewhere along the way)? How could the story be told along the lines Axtell mentions? After all, that was the challenge: Tell a story of epistemic irresponsibility which is clearly not moral or prudential failing. Failures due to lack of skill might be sad or comical but they can't be cases of any kind of irresponsibility as far as I can see unless the fact of the lack of skill came about via a moral or prudential shortcoming.

---

[13] Axtell, "Recovering Responsibility," 447.
[14] Axtell, "Recovering Responsibility," 448.
[15] Axtell, "Recovering Responsibility," 449.

3. Concluding Remarks

Finally, the phrase "due care and diligence" shows up repeatedly in Axtell's reply. I want to focus on the word 'due.' How does one go about determining how much care and diligence are *due* to a matter. One can accept that all truths have some intrinsic value without endorsing that all matters are *due* one second of thought. So how do we sift them? The only way I know how is by *what's at stake*. One way for much to be at stake in determining whether p for S is to *care deeply* about whether p. Another is for someone's well-being to hang in the balance on whether p. But of course the all-things-considered judgment concerning how much care is due a question depends on an all-things-considered judgment about what's at stake. So in the first case, we have no answer to the question how much care is due until we know what else is at stake for S. And the process of weighing S's various interests and resources to determine just how much care is due the question whether p for S is just what practical rationality is all about. In the second case, depending on just what relation S bears to the individual whose well-being is at stake, S might have a duty to give considerable care to whether p, one that trumps other considerations. In his case, the failure to do due diligence is most simply explained by S's being *morally* irresponsible. But in neither case do we find the need to posit another fundamental form of normativity, and that was the thesis I started out to defend.

REPLY TO LICON ON TIME TRAVEL

Steven D. HALES

ABSTRACT: In this paper I offer a rejoinder to the criticisms raised by Jimmy Alfonso Licon in "No Suicide for Presentists: A Response to Hales." I argue that Licon's concerns are misplaced, and that his hypothetical presentist time machine neither travels in time nor saves the life of the putative traveler. I conclude that sensible time travel is still forbidden to presentists.

KEYWORDS: time travel, presentism, eternalism, metaphysics

Presentists are committed to the simultaneity of reality; for them, everything that exists, exists in the objective now. They avoid quantifying over times other than the present by insisting on essentially tensed facts. Eternalists, on the other hand, hold that all times are equally real, and that 'the present' is an indexical like 'here,' 'this,' or perhaps 'actual.' In "No Time Travel for Presentists," I argue that time travel requires leaving the objective present, which, for a presentist, contains all of reality.[1] Therefore to leave the present is to leave reality entirely; i.e. to go out of existence. Thus presentist 'time travel' is best seen as a form of suicide, not a mode of transportation to a disjoint time. Time travel is impossible for presentists.

Licon's objection to my suicide argument has two key components.[2] The first is the assumption of Humean supervenience for times, and the second is a thought experiment about (what Licon considers) a presentist time machine. The Humean supervenience constraint is that the time of the universe supervenes on the total arrangement of all the matter/energy it contains. If there are two states of the universe with identical matter/energy configurations, then they are temporally identical as well. It is somewhat analogous to the idea that two clocks with their gears and hands in the same positions are set to the same time. Licon seems to be presupposing a non-relativistic, absolute time for the universe, but let that pass. It doesn't really matter for my rejoinder.

Here's Licon's conception of a presentist time machine: "if machine F is capable of rearranging all of the matter and energy in the universe, such that it is

[1] Steven D. Hales, "No Time Travel for Presentists," *Logos & Episteme* I, 2 (2010): 353-360.

[2] Jimmy Alfonso Licon, "No Suicide for Presentists: A Response to Hales," *Logos & Episteme* II, 3 (2011): 455-464.

indistinguishable from a past or future moment, and the identity of indiscernibles holds between the instantiated moment and the past or future moment, then F is a time machine."[3] So if we have a machine that instantaneously rearranges all of the matter/energy in the universe to the precise configuration it had in 1862, then it is once again 1862. Since Licon's machine resets the great clock of the universe from 2012 to 1862 without passing through any times in between, Licon argues that this is a discontinuous leap and constitutes travel in time. My suicide argument was that a presentist time traveler must leave the time at which all of reality is located (i.e. the present), which means leaving reality *tout court*. Licon's putative time machine attempts to avoid this objection by bringing the rest of reality along with it.

My response is fairly simple: first, irrespective of the presentist/eternalist debate, Licon's machine does not describe time travel at all, and secondly, it is still a suicide machine. Here's why.

Not a time machine

According to David Lewis, a requirement for time travel that is neutral with respect to the presentist *vs* eternalist issue is that the personal time of the traveler be discontinuous with external time. In other words, the time traveler's calendar must be different from the calendar of the rest of the world. If a time traveler journeys 1000 years into the future, the traveler does not age 1000 years even though the world outside of the time machine does. The distinction between personal time and external time nicely accommodates the view of relativistic physics that a near-lightspeed spacecraft is a time machine that travels into the future. Due to relativistic effects, a twin aboard the spacecraft ages much more slowly than her twin back on Earth. When the spacecraft returns to Earth, the traveler may have aged a week whereas her twin aged 30 years. The personal time of the traveler was different than the external time of her Earth-bound twin.

Licon's 'time machine' erases the distinction between personal and external time. His machine instantaneously reconfigures the entire universe, so everything is always at exactly the same time. The would-be traveler's calendar is exactly the same as everyone else's. Not distinguishing between personal and external time is a problem because it means that Licon lacks the resources to effectively describe temporal discontinuity.

Suppose we set the controls of Licon's time machine for one month into the future. According to Licon, this means that the entire universe undergoes a somewhat radical reconfiguration of matter/energy. Yet why should we consider

[3] Licon, "No Suicide," 462.

this new state of the universe 'one month in the future'? There was no alternative future history of the universe, no other calendar on which we can show the days that were skipped or sped through by the time machine. The sole history of the universe involves an unusual redistribution of matter/energy at a certain point, but that doesn't mean that anything traveled in time, or jumped one month into the future. The universe was in state A at one moment and in state B the next moment. There is no reason at all to think that state B represents a discontinuity in time. It *is* the next moment in time.

By symmetry, the same objection holds for supposed "backwards time travel". The universe was in state A at one moment, and at some state B the next, which may be qualitatively identical to an earlier state, and still not have traveled backwards in time. In line with Licon's supervenience requirement, this might be best seen as a recreation, or a reenactment, of an earlier time. The universe is not going backwards in time and then stopping at a designated date because there is no sense in which it is skipping over any other times. Neither the universe nor any proper part of it is temporally disjoint with the rest. Licon's scenario is more like Tom Robbins's eternally recurring, manufactured world:

> For Christmas that year, Julian gave Sissy a miniature Tyrolean village. The craftsmanship was remarkable. There was a tiny cathedral whose stained-glass windows made fruit salad of sunlight. There was a plaza and *ein Biergarten*. The *Biergarten* got quite noisy on Saturday nights. There was a bakery that smelled always of hot bread and strudel. There was a town hall and a police station, with cutaway sections that revealed standard amounts of red tape and corruption. There were little Tyroleans in leather britches, intricately stitched, and beneath the britches, genitalia of equally fine workmanship. There were ski shops and many other interesting things, including an orphanage. The orphanage was designed to catch fire and burn down every Christmas Eve. Orphans would dash into the snow with their nightgowns blazing. Terrible. Around the second week of January, a fire inspector would come and poke through the ruins, muttering, "If they had only listened to me, those children would be alive today."[4]

The Tyrolean village resets itself every year, but it does not travel backwards in time. The resetting of its clock just is the village's next moment in time.

In sum, the notion of temporal discontinuity makes sense only if the distinction between personal and external time is preserved. Since Licon's 'time machine' prohibits a distinction between personal and external time, it cannot produce a discontinuity in time. And since it cannot do that, it is not a time machine after all.

---

[4] From Tom Robbins, *Even Cowgirls Get the Blues*, cited in Douglas Hofstadter and Daniel Dennett, *The Mind's I* (New York: Basic Books, 1981), 295.

Still a suicide machine

According to Licon, backwards time travel means that his machine annihilates every configuration of matter/energy and instantaneously replaces it with a configuration that existed earlier. Now, Licon's commitment to Humean supervenience requires that the time of the universe supervene on the arrangement and construction of absolutely everything in the universe, no exceptions. Therefore included in that annihilated configuration is the supposed time traveler and the time machine itself. The time traveler and time machine did not exist in 1862, and Licon cannot make the familiar presentist move of claiming that the time machine changed the tensed facts about the past so that it becomes true that "there was a time traveler in 1862."

If all of the matter/energy in the present moment is instantly rearranged to the exact configuration of matter/energy in 1862, then no one traveled back in time. '1862' is a rigid designator denoting a particular arrangement of matter and energy, and Licon's Humean supervenience constraint entails that 1862 is recreated down to the smallest detail. A recreation of 1862 does not allow for some matter to be differently assembled so that it forms a 'time traveler' from the future on the grounds that such an arrangement would not be 1862. Since no one no one currently alive was alive in 1862, a total rearrangement of every particle in the universe to what it was in 1862 means that everyone alive in 2012 just went out of existence. Licon's time machine is a suicide machine after all.

To conclude, Licon hypothesizes a machine that can instantly rearrange all of the matter and energy in the universe. It does not make sense to suppose that the machine discontinuously skips ahead to a future time, nor can such a machine do more than recreate a past time. Therefore it does not provide travel into either the future or the past. Additionally, it cannot offer passage into the past, since neither the machine nor any passenger can be placed into a specifically ordered arrangement of matter and energy (upon which times supervene) that does not already contain them. Thus Licon's machine is not a time machine. However, since the machine annihilates all of the present configuration of the universe without transporting any survivors, it is a suicide machine. So I do not see that Licon's thought experiment counts against my original argument that presentists are incapable of time travel.

HISTORY OF EPISTEMOLOGY

COMPLEMENTARITY AND ANTINOMY

Teodor DIMA

ABSTRACT: In this study we present some contributions of the logician and philosopher Petre Botezatu (27.02.1911-01.12.1981), who turned the idea of complementarity, formulated by Niels Bohr for the interpretation of the wave-particle structure of the quantum world, into an ordering principle of his work. Thus, he understood general logic as a synthesis in which the style of classical logic is complementary to the style of the 20th century logic. He didn't give up either the mathematical modelling of logical language or the conceptual description through natural language. Thus, *natural operational logic* was created. Then Petre Botezatu assessed the achievements and failures of deduction in order to build the notion of *methodological antinomy*, and formulated five antinomies of axiomatization and five antinomies of formalization. The main purpose of this study is to present and interpret them.

KEYWORDS: complementarity, determination, determinism, natural operational logic

The main purpose of this paper is the presentation and interpretation of some original contributions of the philosopher and logician Petre Botezatu (27.02.1911-01.12.1981), PhD, Professor at the Faculty of History and Philosophy, "Alexandru Ioan Cuza" University of Iaşi, Romania. He was considered "the Romanian logician with the highest number of achievements, who, above all, through his natural, operational, logic, created a whole new domain, classical-symbolical, with multiple applications in education and in scientific thinking and practice."[1] He directed his first researches towards the elaboration of his PhD thesis, *Cauzalitatea fizică şi panquantismul (The Physical Causality and the Panquantism)*, defended in 1945. In other words, texts of general epistemology and philosophy guided his steps in the first ten formative years and then, until the end of his career, his work on logic twinned with metalogical and epistemological interpretations.

However, his PhD thesis remained a manuscript as he was uninterested in publishing it, explaining his reservations through the fact that he should have reconsidered his opinion on *the value of determinism in modern physics*. Yet,

[1] Alexandru Surdu, *Contribuţii româneşti în domeniul logicii în secolul XX* (Bucureşti: Fundaţia "România de Mâine," 1999), 14.

studying some of his main published works, we have noticed that many of the ideas found in that thesis constantly incited Petre Botezatu. Actually, evaluating his own scientific activity in a *Curriculum Vitae* written on February 10, 1981, he considered that in his thesis he had "demonstrated that two fundamental confusions were made during the passionate discussion around Heisenberg's relations: between indetermination and indeterminism and between quantitative causality and qualitative causality. In quantum mechanics there is indetermination, not indeterminism and what has become impossible is the quantitative determination of causal relations, which exist in their qualitative form (various kinds of particle collision). Determinism exists under a new form and the controversies around it can be mainly explained by the conflict between certain forms of realism and idealism."[2]

Petre Botezatu came to these conclusions from the 1940s, after he analysed the main results obtained by the quantum physicists and their philosophical interpretations. Thus, Werner Heisenberg, who was working in Copenhagen under the direction of Niels Bohr, formulated interpretations on Louis de Broglie's wave mechanics under the influence of some positivist and phenomenological ideas that became the creed of the *Wiener Kreis* philosophical school. Here it was considered that physical theory should only rely on quantities whose values can be directly observed and should avoid any representation for which there would be certain physical elements inaccessible to experience. Animated by this 'Copenhagen spirit', which from certain points of view reminded of the spirit of the classical conception of energetism, Heisenberg's *quantum mechanics*, sometimes called *matrix mechanics* as it was mathematized, was presented under the form of a pure formalism which rejected any image of microphysical world, explaining all phenomena observable at the atomic scale by means of mere algebraic calculations.

In this way an opposition was reached, difficult to reconcile: on the one hand, wave mechanics was trying to obtain a representation of the microphysical phenomena within space and time, providing a clear and intelligible image of the wave-particle association, and on the other hand, quantum mechanics was elaborating a formalism capable of providing with accuracy the future evolution of the phenomena experimentally observed. Secondly, quantum physics was *interpreted probabilistically*, without any causal mechanism. The wave from wave mechanics becomes the solution of an equation with partial derivatives, solution made possible by a mathematical instrument, and the particle has no longer a permanent localization in space, but exists in a state of potential in an entire region of space and is statistically distributed between several states of motion. Therefore Niels Bohr introduced the notion of *complementarity*: wave and

[2] Petre Botezatu, *Curriculum Vitae* (manuscript), 3.

particle are two complementary aspects of reality, a reality which otherwise cannot be intelligibly described from the microphysical point of view.

Petre Botezatu adopted the idea of complementarity, presented it in his PhD thesis[3] and turned it into an *ordering principle* of his work. In Chapter II, Part III of his thesis, invoking Niels Bohr's opinions, he detailed specifications that were later on presented and explored by philosophers of physics. Petre Botezatu maintained that the idea of complementarity was a result of "the insufficiency of the concepts created by the common intuition when they are applied in microphysics."[4] Namely, pairs of notions, such as: wave and particle, causality and space-time, which, in the description of microcosm, exclude and complete one another at the same time. In our opinion, the insufficiency was due to the fact that quantum physicists did not operate a *fundamental distinction*: macrophysical phenomena are observed in a natural space-time world, while microphysical phenomena are observed under experimental conditions and their description depends on instruments, on their power and precision.

Prior to 1950 Petre Botezatu became preoccupied with building a *general theory of reasoning*. First, he analysed the logical operations of the Aristotelian syllogistic and he discovered that Aristotle had determined the logical operation subjacent to syllogism according to the actual steps of thinking – *the transfer of a property between two classes related to each other by inclusion* – which represents a typical operation of thought. Hence the idea that it can be generalized so that it operates between other mutually related logical objects. Thus Petre Botezatu defined the *transitive logical operation*. Then, making good use of Kant's and Goblot's ideas on mathematical constructivity, he also defined the *constructive logical operation*: composing an object from other objects. Thus the natural logic was founded.

The new conception was presented in the paper "Teoria raționamentului întemeiată pe structura obiectelor" ("The Theory of Reasoning Based on the Structure of Objects"),[5] held at Alexandru Ioan Cuza University of Iaşi in 1958, and later on, in the same year, in the paper "La logique et les objets," XII International Congress on Philosophy, Venice-Padua (12–18 September 1958)[6]. The study "Les

[3] Petre Botezatu, *Cauzalitatea fizică şi panquantismul*, ed. Teodor Dima (Iaşi: Editura Universității "Alexandru Ioan Cuza", 2002).

[4] Botezatu, *Cauzalitatea fizică*, 161.

[5] In *Analele ştiinţifice ale Universităţii „Al. I. Cuza" din Iaşi. Ştiinţe sociale*, V (1959): 183-198.

[6] In *Atti del XII Congresso Internazionale di Filosofia* V (Firenze: Sansoni, 1960), 77-83.

raisonnements transitifs"[7] followed and after that the volume *Schiță a unei logici naturale* (*Sketch of a Natural Logic*)[8] provided the complete theoretization.

Almost at the same time, the project of a (theoretical) natural logic became of interest to other logicians and epistemologists, especially in France, the homeland of rationalism: Jean Piaget's, *genetic operational logic*,[9] René Poirier's *organic logic*,[10] and Robert Blanché's *reflexive logic*.[11] Petre Botezatu's idea of a natural logic developed independently of the aforementioned authors' achievements, having different roots, as we previously showed. Also, its basis was the idea that *the relation between mathematical logic and traditional logic is a functional one*, which ultimately would be proven by means of metalogical investigations.

*

* *

The fundamental idea of complementarity suggested Petre Botezatu that in methodology every success leads to a failure, that one cannot have unlimited success. "It is in fact a broadened lesson of the complementarity of aspects, and quantum mechanics rendered us familiar to that."[12] In *Valoarea deducției (The Value of Deduction)*, from which we've just quoted, he made the balance of the achievements and failures of deductions in order to formulate epistemological interpretations: the methodological objectives are not all compatible; they moderate each other so that advancing on one direction means retreating from another. Certainly, he commented, no obstacle can prevent us from advancing ever so far on a certain path, but this has a price; that is, sacrifices in another sector. As far as logic is concerned, opposite tendencies can always be balanced by choosing strategies by means of which maximum benefits are associated with minimum losses.

These reflections inspired him the term *methodological antinomy*, in the Kantian sense of simultaneous presence of two contradictory theses that appear to be equally justified. The following opposite statements, for instance, can be held: (1) language is formalizable and (2) language is not formalizable. The antinomy is resoluble but not like in the case of paradoxes, through a theory of levels or types, but by *differentiating the points of view*. A theory is (relatively) formalizable from

[7] In *Acta logica* 1 (1960): 59-81.

[8] Petre Botezatu, *Schiță a unei logici naturale. Logică operatorie* (București: Editura Științifică, 1969).

[9] Jean Piaget, *Traité de logique. Essai de logistique opératoire* (Paris: A. Colin, 1948). Edition reviewed for symbolism and J.-B. Grize's mathematical ideas (Paris, 1972).

[10] René Poirier, *Logique et modalité du point de vue organique et physique* (Paris: Hermann, 1952).

[11] Robert Blanché, *Raison et discours. Défense de la logique réflexive* (Paris: Vrin, 1967).

[12] Petre Botezatu, *Valoarea deducției* (București: Editura Științifică, 1971), 168.

the *synthactical*, i.e. its internal construction, point of view, but it is not relatively formalizable from the *semantic*, i.e. its interpretation, point of view.

As deductive devices diversified and improved, the number of antinomies increased and Petre Botezatu grouped them into *five antinomies of axiomatization* and *five antinomies of formalization*. We will present, explicate and comment them briefly, and our approach will reveal some of the discussions *en vogue* 50 years ago on the achievements and failures of deductive devices. Many of them are still topical.

1. *The* ***antinomy of simplification****: the simplification of bases leads to the complication of construction*. The extension of the axiomatic method revealed the fact that for a given theory, larger or smaller classes of axioms and primitive (undefined) terms can be selected. Among other requirements, logicians wanted the axiomatized theories to be as unified as possible, to have axioms independent from one another, and the consequence was a minimal demonstrative basis: a number of axioms and primitive terms as small as possible. For example, for the propositional logic, numerous axiomatic variants were proposed:[13] Frege, in 1879, used two functors (negation and implication) and six axioms (Lukasiewicz reduced them to five) and Whitehead-Russell, in 1910, used also two functors (negation and disjunction) and six axioms (Bernays reduced them to four). The French logician Jean Nicod, in a study published at Cambridge,[14] showed that if the idea of incompatibility – Sheffer's functor – is taken as the main functor, all Russell's axioms can be reduced to just one. But it has five variables and 43 signs! Nicod himself admitted that such an axiom was uncomfortable for demonstrations. To explain its antinomy, Petre Botezatu took an example given by Alfred Tarski[15] who noted dissimilarities between the 'methodological' value and the 'didactical' value of two axiomatization variants for the arithmetic of real numbers. The first system, which presented the set of real numbers as an ordered Abelian number, had four primitive terms and nine axioms. The system was methodologically superior, being the simplest axiomatic construction of the entire arithmetics. But this formal advantage was accompanied by didactical disadvantages – complex and difficult definitions and demonstrations. The second system characterized the set of real numbers as an ordered body structure and needed six primitive terms and twenty axioms. It had methodological disadvantages – neither the terms nor the

[13] Cf. Joseph Dopp, *Notions de logique formelle* (Louvain: Publications universitaires de Louvain, 1965), 261-275.

[14] Jean Nicod, "A Reduction in the Number of the Primitive Propositions of Logic," *Proceedings of Cambridge Philosophical Society* 19 (1917): 32-41.

[15] Alfred Tarski, *Introduction à la logique* (Paris: Gauthier-Villars, 1960), ch. X.

axioms met the independence criterion. Instead, the system had didactical advantages: important chapters of real number arithmetics were simple to elaborate by elementary reasonings. The antinomy occurred between the concern for the independence of axioms and primitive terms and the facility of demonstrations. Therefore in axiomatization practice, the independence criterion is given less attention, being considered the one that ensures the elegance of construction, even if it is closely related to the condition of system non-contradiction.[16]

2. *The* ***antinomy of strength****: a system's increasing strength is accompanied by the degradation of its metatheoretical quality*. Adopting Jean Ladrière's explanations, Petre Botezatu showed that a system is stronger if it can be interpreted under the form of a broader theory. Thus, in arithmetics axiomatization, Tarski's system is stronger than Peano's because the first covers the arithmetics of real numbers, while the second only the arithmetics of natural numbers. But the construction of ever stronger systems has the shortcoming of satisfying less the metatheoretical conditions: non-contradiction, completeness, categoricity, decidability. *What is gained in broadness is lost in formal perfection*. For example, propositional logic is a modest axiomatic system, but with precious metatheoretical qualities: it is non-contradictory, complete and decidable. However, it has little strength: it does not include the notions of predicate, function, quantifier, etc. If the system is enriched with the necessary terms in order to get a stronger logic, its formal properties start to faint. Thus, the enlarged predicate calculus is neither complete (Gödel) nor decidable, and its non-contradiction cannot be proved. It was believed that the obstacles were due to the imperfection of the used formalisms and that they could be overcome by other formalisms. Ultimately, limitation theorems were demonstrated: Gödel's (1931) and Church's (1936). Extended, generalized and unificated by subsequent theorems, they proved that the limits expressed requirements of formalisms. Non-Gödelian systems (Myhill, Church), which increase the system's strength almost to the limitation theorem, were elaborated. But they were either insufficiently equipped (Myhill's system doesn't have the negation operator and the universal quantifier) or too disconnected from the common intuition of certain operations. Once again Gödel was right – only the elementary systems have the property of decidability. This result was corroborated by Church's theorem – a stronger logic, such as predicate restrictive logic is not resoluble even if partial solutions can exist, hence the conclusion that a logical machine able to solve all the problems in mathematics can never exist

[16] Cf. E.W. Beth, *The Foundations of Mathematics* (Amsterdam: North-Holland Publishing Company, 1959), § 32.

(Turing). The antinomy of strength keeps its balance by admitting Tarski's notion of "essentially indecidable theory."

In the following, we present briefly the other three axiomatization antinomies:

3. *The* ***antinomy of purity****: the purification of demonstration involves the insufficiency of the basis*.[17] Petre Botezatu commented *in extenso* this antinomy because it involves the problem of the relation between deduction and intuition, a dualism firmly proclaimed by Descartes. But the deductive method by its own structure, which was enunciated by Aristotle, aims for purity: each step complies with an inference rule applied to a statement that is considered true. For example, *Principia Mathematica* – the logic was constructed first and the mathematics resulted from it. In the process, the theory is enriched with constants and variables, but the deduction rules are the same. The intuitive arguments, judiciously or not accepted in the beginning, are now avoided. This is the plan for the perfect deduction (however, it has been noticed that this is an ideal, i.e., not available in its pure form because the *intuition cannot be completely eliminated*). Several logicians noted that any formal axiomatics is bordered by intuitive domains: at the bottom, by the concrete interpretations given to it through models, one of which being the model that made it possible; on top, by the previous scientific knowledge, which steps in with its indubitable truth and intuitive significance.[18] From formalized mathematics we arrive to logic, from logic, if it has to be formalized, to metalogic, and so on. In order to avoid the *fallacy of infinite regress*, reasoning must be stopped at some level and this is, no doubt, the level of intuition. Then, Petre Botezatu analyzed the subject matter of the ultimate foundation of theories relying on the results obtained at the *International Symposium of Science Methodology*, Warsaw, 1961. Here it was emphasized that the axiomatic method alone cannot verify the axioms.[19] *Formal thinking cannot exhaust the content of intuitive thinking*. Intuition resists formalization and this resistance takes various and unexpected forms. In order to demonstrate the theorem, it is not enough to establish the axioms and primitive terms alone; methods must be chosen that will make possible the deduction from the foundation to the theorem.

---

[17] Botezatu, *Valoarea deducției*, 175-180.

[18] Robert Blanché, *L'axiomatique* (Paris: P.U.F., 1959), 65-66; Roger Martin, *Logique contemporaine et formalisation* (Paris: P.U.F., 1964), 189; E.W. Beth, Jean Piaget, *Epistémologie mathématique et psychologie* (Paris: P.U.F., 1961), ch. V, IX, X; Jean Ladrière, *Les limitations internes des formalismes* (Louvain: E. Nauwelaerts/Paris: Gauthier-Villars, 1957), ch. X.

[19] Cf. *The Foundations of Statements and Decisions*, ed. Kazimierz Ajdukiewicz (Warszawa: PWN-Scientific Publishers, 1965).

4. *The* ***antinomy of exactness****: the exactness of terms requires the idealization of objects*. Petre Botezatu considered that the deductive edification of a theory is not reduced to a mere arborescent organization of its sentences in order to give a unique and unitary system, but *in-depth changes* must be made, in the structure and the content under organization. Einstein, discussing the relation between theory and reality, paradoxically stated that "Since mathematical propositions address reality, they are uncertain, and since they are uncertain, they do not address reality."[20] Indeed, mathematical propositions describe the properties of ideal objects. When applied to reality, the formal object and the empirical object must coincide; for example, a geometric circle and a physical circle; this is very difficult to accomplish. Hence, philosophers of science proposed various solutions. N.R. Campbell spoke about a 'dictionary' supposed to translate the concrete terms into the abstract ones. P.W. Bridgmann asked for operational 'rules' and Rudolf Carnap, for 'correspondence rules.'[21] Petre Botezatu considered that such devices only conceal the difficulty, they do not get rid of it. Robert Blanché concluded that any axiomatics can be read in two different ways: an abstract, rational and formal way, and a concrete, empirical and material way.[22] Thus, geometry splits into pure, axiomatic geometry, without intuitive content, where truth means non-contradiction, and applied geometry, which is intuitive and deals with physical laws. The debates are still going on. Botezatu described various solutions to show that the opposition between exact and inexact concepts is a major achievement of contemporary logic. In this way the distance between theoretical and empirical thinking is better understood, but the antinomy is still there: the gain in exactness is the loss in expression.

5. *The* ***antinomy of abstractization****: the abstractization of structures involves the undetermination of theory*. Petre Botezatu was referring to the capacity of instruments to penetrate into the intimacy of complex phenomena in order to create new scientific theories: relativity, quantum theory, theory of heredity, cibernetics, etc. Due to the advance of the abstractization process, these theories ignore the particular properties of objects and then isolate some relations, leaving others aside. The selection and various combination of relations were fruitful approaches as they rendered obvious the fact that very different scientific fields can be unified, such as logic and electronics, measure theory and probability theory, etc. The structures are *multivalenced* and *plural*. Once a set of relations is selected, the abstractization process continues and the number of fundamental

[20] Albert Einstein, *Mein Weltbild* (Berlin, 1955), 119.
[21] Rudolf Carnap, *Philosophical Foundation of Physics* (New York: Basic Books, 1966), § 24.
[22] Blanché, *L'axiomatique*, § 28.

relations decreases gradually but new forms are created. Thus, *theories progressively fail to correspond with reality*. For example, logic. Despite the multitude of formalisms, researches and achievements, logic still fails to express appropriately the natural line of thinking. Therefore, the process of structuring the theories represents an indubitable success and a failure at the same time. In a mathematical structure we can include more, but cover each sector less. *The gain in extension is the loss in intension*.

*

* *

Petre Botezatu considered that axiomatization antinomies are primordial as compared to formalization antinomies because the deductive method resorts to axiomatization in most of the cases. However, formalization represents the superior level deduction can reach, but here as well the progress in one direction means the withdrawal from another.

6. *The* ***antinomy of rigurosity****: the exactness of the demonstration imposes the complication of the demonstration*. Formalization requires that all the elements needed to create a new teory be mentioned, such as *rules for the formation* of correct formulas, *rules for the tranformation* of one formula into another and *rules for the definition* of terms. Thus the derivation of theorems from axioms becomes entirely formal: one sequence of signs is turned into another even if the meanings are unknown. *Full proof* is obtained in this way, i.e., at least apparently, nothing relies on intuition any more. But full proof needs more, that is to go back from theory to theory until it reaches the basic logical theory. Now there will be no more elements of subjectivity and a superior level of rigurosity is achieved. But Petre Botezatu considered that this procedure had a practical disadvantage, not negligible at all: if any thought is formalized, the demonstration becomes excessively extended and complicated: "a gain in exactness and methodological accuracy comes with a loss in clarity and inteligibility."[23] On the other hand, the risk of error is still not overcome. "The longer the demonstrations, the higher the risk of errors, and the increase in the number of elementary operations, as long as they are not performed by a machine, is rather more dangerous than useful."[24] Hence the paradoxical situation in which mathematicians often avoid formalization, although they created it. Botezatu concluded that, given the current level of performance, *the operations cannot be certain and simple at the same time*.

[23] Tarski, *Introduction à la logique*, 118, cf. Botezatu, *Valoarea deducției*, 188.
[24] Martin, *Logique contemporaine*, 187.

7. *The* ***antinomy of totalization****: closing a system means transcending the system*. Aristotle himself stated that a syllogism is correct only if "no exterior term is needed for the consequence to be necessary."[25] Axiomatization and formalization continue this necessity: a formalism either is sufficient or denies itself. But it was proved that *self-sufficiency is unrealizable* because first of all it is impossible to create a language formalized within itself; a language (*object language*) is always created within another language (*metalanguage*), which is given. For example, the arithmetic language is created by means of the English language. Of course, a metalanguage can also be formalized using another metalanguage, in several formalization steps, but this has to stop somewhere and there the language will be informal. Botezatu concluded that "we cannot dispose of the intuitive in order to create the formal."[26] On the other hand, it is common knowledge that Gödel proved that one of the undecidable propositions that affect stronger systems (containing arithmetic at least, e.g. *Principia Mathematica*) is the very one that states the consistency of the system. Hence, *the noncontradiction of a system cannot be proved through the means of the system itself*. So, a formalism cannot become a closed system. Two indispensable external references always accompany it: a reference to an informal basic language and a reference to a superior formalism. We may say that a given formal system grows on an intuitive substructure and continues in a stronger formal system and these connections cannot be cut when the problem of justification arises.[27]

8. *The* ***antinomy of consistency****: The consistency of a system involves the incompleteness of the system*. Petre Botezatu correlated this antinomy with the previous one and derived it from Gödel's theorems. It was demonstrated that a formal system containing arithmetics, if it is noncontradictory, it contains undecidable propositions, such as Proposition *G* which states its own indemonstrability. But another troublesome aspect appears: Proposition *G*, although indemonstrable, is true, a fact which can be proved by metatheoretical means: it states that integers have a certain mathematical property that is well-defined and belongs to each number. Because the system contains a property that is undecidable and true at the same time, *the system of axioms is incomplete*. In other words, the system of axioms is not strong enough and not all theorems can be derived from it. Moreover, even if new axioms would be added to the system so that Proposition *G* could be demonstrable, another undecidable formula can be created, and then another one, and so on. So, if a formal system is consistent, it is

[25] Aristotel, *Analitica primă*, I, 1, 24 b.
[26] Botezatu, *Valoarea deducției*, 190.
[27] Botezatu, *Valoarea deducției*, 191.

incomplete and remains incomplete. From the epistemological point of view, it can be concluded that full formalization is not the most beneficial way to organize a deductive theory; human spirit is free and overcomes the barriers of formalization, learning from practical and intellective experience.[28]

9. *The* ***antinomy of interpretation****: the formalization of a system entails the relativization of interpretations*. A formal system is designed to interpret a certain theory, which in fact was originally its model. For example, Peano's system was created in order to give an interpretation to the arithmetics of real numbers. In other words, when creating a formalism, one starts from just one point, a model, and by means of that model many other points are reached: *a multitude of interpretations are created, all based on the same formal pattern*. At the epistemic level, there are structural similarities and connections between set theory, number theory, geometry and logic, which allow the theorems to be extended from one field to another. However, there exist a *lack of precision* which can be explained in the following way. An inconsistent (contradictory) system has no models, while a consistent system can have several models. If all the models of a system are isomorphic, that is structurally identical, the system is categorical or monomorphic; otherwise, it is noncategorical or polymorphic. The plurality and the irregularity of models generate difficulties when trying to determine the concepts. For example, Peano's axioms should be categorical and able to characterize univocally the natural numbers. But Thoralf Skolem (1933) proved that the sequence of numbers cannot be characterized by a finite system of axioms that would distinguish it from other sequences. Skolem created ordered sets of integer functions, sets that satisfy the axioms and yet belong to another type of order than the sequence of natural numbers. In 1950, Leon Henkin proved that any consistent formal system containing the theory of integers has irregular models and therefore cannot be categorical. Petre Botezatu gave the most conclusive example: set theory. Thoralf Skolem had already proved, in 1920, that it was impossible to create an absolutely categorical system. The theorem proved that any theory of the first-order predicate logic had an uncountable model even if it was conceived for a countable model, which it actually had. But the set theory belongs to this category, which means that, if it has a model, it certainly has an uncountable model as well. Cantor established a precise demarcation between the countable and the uncountable; now the boundary is disappearing. Analyzing the interpretations given to this state of facts, Botezatu made clear that, following axiomatization, formalization as well is not to be reduced to a mere translation from one language into another. Structural changes occur in the process and perhaps the most significant one is the

[28] Botezatu, *Valoarea deducției*, 192.

loss of univocity. Depending on the standpoint, this can be an advantage or a disadvantage.[29]

10. *The* ***antinomy of expressiveness****: the semantization of a system involves the inexpressiveness of that system*. Because systems are created in view of interpretations, the syntactic construction is backed up by a semantic one. But in formal systems the formulas have only the property of being derivable (demonstrable) or not, while in the language of interpretations the statements can be true or false, hence truth is a semantic concept when it involves the sign-object relation and it is also a fundamental epistemological notion. Hence the following question arises: is it possible to introduce the concept of truth into the formalism of a system? Alfred Tarski (1936), preoccupied with this problem, proved that, given a strong enough formalism, *the theory of 'true' semantic predicate cannot be formalized in that system*, otherwise a Liar-type paradox would arise. In 1948 he proved that the same limitation is required for other semantic concepts as well, such as *definissability*. Botezatu's conclusion was that in strong formal systems metatheoretical (semantic) notions can be created, which means that the formal system has been transcended. It follows that formal systems are never self-reflexive whenever they have to be both comprehensive and expressive.

*

* *

These *methodological antinomies* formulated by Petre Botezatu express the idea that we cannot create a formal system posessing all the necessary qualities: purity and comprehesiveness, consistency and completeness, decidability, etc. We can advance in one direction if we withdraw from another. Or we can moderate our ambitions and then we may fulfil more objectives. In any field of study there are scholars who attempt to build a great system, complete and unique, but only partial and perfectible system are ever created, and this ensures the progress of knowledge. From the cognitive point of view, Petre Botezatu noted the *alternation between adjustment and assimilation*. Formalization is a process of *deductive assimilation*: the scientific objects are transfigured in order to be integrated into the sequence of deduction. Interpretation, that is the transition from the relations between signs to the relations between objects, is a process of *adjustment*. In other words, the construction and improvement of formal systems oscillate between these two opposite tendencies. Depending on the objectives, one tendency will prevail over the other.[30]

[29] Botezatu, *Valoarea deducției*, 194.

[30] Botezatu, *Valoarea deducției*, 198-199.

The idea of methodological antinomy continued to preoccupy Petre Botezatu. For example, he tried to explain the growing up crisis of higher education formulating the *antinomy of accessibility* between *intuitive* and *structural* and the *antinomy of structuring*, that is knowledge structuring facilitates knowledge consolidation but makes knowledge renewal difficult. He presented them at his last public conference, held in the autumn of 1981 during "The Days of Iaşi University." He pointed out that in the process of scientific teaching, the objectives of methodics cannot be fulfilled all at once.[31]

The two antinomies relied on two *fundamental laws*[32]: 1. *The law of the gap between the scientific-technical revolution and the change in mentality*: the adaptive effect is considerably delayed. New habits and a new perspective take time and need favourable circumstances in order to be structured and consolidated. 2. *The law of information distortion*: the accuracy in conveying the message is directly proportional to the cultural level of the senders and indirectly proportional to their emotional tension. Petre Botezatu warned of the fact that "Ignorance makes it difficult, if not impossible, for the scientific message to be received with accuracy. On the other hand, the emotional turmoil has the same negative effect."[33] These two nomic relations generate antinomical situations. Two of them were formulated by Petre Botezatu, as previously showed.

The subject matter of antinomies has also preocupied us. Certainly, the initial impulse came from Petre Botezatu,[34] and then we added the ***antinomy of certainty***: *"certainty in experimental research limits the applicability of inductive methods*. The study of real phenomena requires that logic's formal frame of reference becomes more flexible and enhanced. But the logical loss is the gnoseological gain."[35] This conflict can be avoided by moderating the two natural characteristics of thinking: the need for abstractization, formalism, deduction, satisfied by logic, and

[31] Cf. Teodor Dima, *Privind înapoi cu deferență* (București: Editura Academiei Române, 2006), 189.

[32] Petre Botezatu, "*Homo Logicus*," in his *Interpretări logico-filosofice* (Iași: Junimea, 1982), 334. This paper was published for the first time in *Analele ştiinţifice ale Universităţii „Al. I. Cuza" din Iaşi*, Philosophical Sciences XXV, III b (1979): 57–72.

[33] Botezatu, "*Homo Logicus*."

[34] In a text evoking the personality of Petre Botezatu, we confessed that the idea of dealing with the problem of antinomies came to our mind when Petre Botezatu's volume, *Valoarea deducției* (1971), was published. Here the author discusses the ten methodological antinomies. In that period we were elaborating our PhD thesis. See Teodor Dima, *Pseudo-Jurnal din Iaşi, strada Sărăriei, numărul 174*, in *Symposion* V, 1(9) (2007): 173–183.

[35] Teodor Dima, *Metodele inductive* (București: Editura Ştiinţifică, 1975), 129-130.

the need for concreteness, empirism, induction, closer to the spirit of experimental sciences.[36]

Bogdan Olaru, appreciating our contribution to the study of antinomies, observed that "Teodor Dima wrote on several occasions on antinomies and Romanian authors who dealt with the subject. Mainly, there are three constants of these researches: 1) the analysis of the interpretations of the Kantian antinomies given by some Romanian philosophers; 2) the assessment of Romanian contributions to the of antinomies of thinking; 3) his own interpretation of Kant's antinomies."[37]

Joining the analysis of the "temperate audacities" of reason, Bogdan Olaru admits there is a counterpart of theoretical antinomies in moral philosophy, considering them *moral dilemmas*; there are instances when two opposing theses, with the same subject but different predicates (*it is good/it is bad* from the moral point of view), are seemingly equally well-founded. Often, the solution to a moral dilema is to admit there is not just one solution, unless there is agreement upon the fact that this is an interim solution or a solution indissolubly dependent on restrictions imposed by the context. He argued that "emergency exit through the formulation of some postulates is the least convincing. Only by appeasing the impulse to settle the dilemma by resort to *a priori* postulates one can be given the best chance to understand the nature of formal conflicts. And the chance to find that they are real, not apparent."[38]

In conclusion, continuing and developing by original detailings the preference of modern and contemporary thinking for an interdisciplinary approach to rationality, tempted to transcend and even elude logicity, Petre Botezatu proved that the aporias, dilemmas and antinomies of knowledge are given new expressions once they are formulated, interpreted and commented.

[36] We presented the antinomy of certainty for the first time in public on 23rd February 1973, when we defended our PhD thesis at the University of Bucharest.

[37] Bogdan Olaru, „Prolegomene la o critică a raţiunii temperate sau: despre câştigurile practice ale antinomiilor teoretice," in *Revista de Filosofie* LVIII, 1–2 (2011): 103–116. A shorter version of the paper was published in the volume *De Dignitate Philosophiae*, ed. Cătălina-Daniela Răducu et al. (Iaşi: Terra Nostra, 2009), 307–326.

[38] Olaru, „Prolegomene la o critică a raţiunii temperate," 116.

NOTES ON CONTRIBUTORS

Anthony Brueckner is Professor of Philosophy at the University of California, Santa Barbara. His main areas of specialization are epistemology, philosophy of language, metaphysics, and philosophy of mind. His areas of interest include the metaphysics of free will and the metaphysics of death. He is the author of *Essays on Skepticism* (2010) and of numerous articles among which: "Johnsen on Brains in Vats," (*Philosophical Studies*, 2006) and "Justification and Moore's Paradox," (*Analysis*, 2006). Contact: brueckne@philosophy.ucsb.edu.

Christoper T. Buford is Lecturer at the University of Akron. His areas of specialization are epistemology, philosophy of mind, personal identity and his areas of competency are bioethics, logic, and history of modern philosophy. He is the author of: "Reply to Baumann on factivity and contextualism" (with Anthony Brueckner, *Analysis*, 2010), "Baker on the Psychological Account of Personal Identity" (*Acta Analytica*, 2009), and "Thinking Animals and Epistemology" (with Anthony Brueckner, *Pacific Philosophical Quarterly*, 2009). Contact: cb72@uakron.edu.

Alex Bundy is Ph.D. candidate at the University of California, Santa Barbara, Department of Philosophy. While his main area of specialization is epistemology, his academic interests include ethics, philosophy of language, metaphysics, and philosophy of mind. He is the author of "On Epistemic Permissivism" (with Anthony Brueckner, *Synthese*, 2011) and of "In Defense of Epistemic Abstemiousness (*Logos & Episteme*, 2011). Contact: bundy@umail.ucsb.edu.

Adrian Costache is postdoctoral fellow with The Romanian Academy, Iaşi Branch and lecturer in the Department of Didactics of the Human Sciences, "Babeş-Bolyai" University, Cluj-Napoca. He holds a PhD from "Babeş-Bolyai" University for a thesis entitled *Between Hermeneutics and Deconstruction. A Critical Approach to the Question of Understanding* (defended in 2010). His area of specialization include philosophical hermeneutics and deconstruction and his current fields of research are the philosophy of history, the neo-metaphysics of Gilles Deleuze, and the didactics of the human sciences. Contact: johnadriancostache@gmail.com.

Teodor Dima is Member of the Romanian Academy, and Director of the „Gheorghe Zane" Institute for Economic and Social Research. His main contributions are in epistemology, philosophy of science, and logic. His books include: *Metodele inductive* (*The Inductive Methods*) (1975), *Între prudență și eroare. Eseu asupra naturii și dinamicii științei* (*Between Prudence and Error. Essay on the Nature and Dynamics of Science*) (1978), *Explicație și înțelegere* (*Explanation and Understanding*) (vol I. 1980, vol. II 1994), and *Privind înapoi cu deferență* (*Looking Backwards with Deference*) (2006). Contact: tudoreldima2008@yahoo.com.

Trent Dougherty is Assistant Professor of Philosophy at Baylor University. His areas of specialization are epistemology/rationality theory (evidence/reasons), philosophy of religion, and philosophy of language. His areas of competence include philosophy of science, metaphysics and mind, and decision theory. He is the author of *The Problem of Animal Pain: A Theodicy for All Creatures Great and Small* (under contract with Palgrave MacMillan) and the editor of *Evidentialism and Its Discontents* (2011) and *Skeptical Theism: New Essays* (under contract with Oxford University Press). Contact: Trent_Dougherty@Baylor.edu.

Axel Gelfert is Assistant Professor of Philosophy at the National University of Singapore. His areas of research are history of social epistemology, ethics and epistemology of testimony. His areas of competence are philosophy of science and technology, epistemology, history of Western philosophy. He is the author of "Expertise, Argumentation, and the End of Inquiry" (*Argumentation*, 2011), "Steps to an Ecology of Knowledge: Continuity and Change in the Genealogy of Knowledge" (*Episteme*, 2011), and "Indefensible Middle Ground for Local Reductionism About Testimony" (*Ratio*, 2009). Contact: phigah@nus.edu.sg.

Stephen R. Grimm is Assistant professor of Philosophy at Fordham University. His main areas of research are in epistemology, the philosophy of science, and value theory. He is also interested in the history of philosophy. He is the author of "Understanding" (*The Routledge Companion to Epistemology*, eds. Sven Bernecker and Duncan Pritchard, 2010), "Epistemic Normativity" (*Epistemic Value*, eds. Adrian Haddock, Alan Millar, and Duncan Pritchard, 2009), and "Knowledge, Practical Interests, and Rising Tides" (*The Point and Purpose of Epistemic Evaluation*, eds. John Greco and David Henderson, Oxford University Press, forthcoming). Contact: sgrimm@fordham.edu.

Steven Hales is Professor at the Department of Philosophy, Bloomsburg University. His areas of specialization include epistemology and metaphysics. He is the author of *Relativism and the Foundations of Philosophy* (2006), *Nietzsche's Perspectivism* (with Rex Welshon, 2000), and editor of many volumes such as: *A Companion to Relativism* (2011), *Delight in Thinking: An Introductory Reader* (with Scott C. Lowe, 2006), and *Analytic Philosophy: Classic Readings* (2001). Contact: hales@bloomu.edu.

Paul Humphreys is Professor of Philosophy at the University of Virginia. His areas of specialization are philosophy of science, metaphysics, and epistemology. His current interests include the metaphysics and epistemology of emergence, computational science, empiricism and realism, and strategy. He is the author of *Extending Ourselves: Computational Science, Empiricism, and Scientific Method* (Oxford, 2004) and the editor (with Mark Bedau) of *Emergence: Contemporary Readings in Science and Philosophy* (2007). Contact: pwh2a@virginia.edu.

Mark McBride is a Postdoctoral Fellow at the National University of Singapore (NUS), Faculty of Law. His areas of specialization are epistemology, philosophy of mind, ethics, and legal philosophy. His publications include "Reply to Pardo: Unsafe Legal Knowledge?" (*Legal Theory*, 2011), "A Puzzle for Dogmatism" (*Logos & Episteme*, 2011), and "A Challenge for New Defenders of the Doctrine of Double Effect" (*Journal of Ethics & Social Philosophy*, 2010). Contact: markmcbride1978@gmail.com.

Rik Peels is a PhD student at the University of Utrecht, the Netherlands. His areas of competence are philosophy and theology and his main areas of interest are epistemology, the ethics of belief, and philosophy of religion. He is the author of "Why Responsible Belief Is Blameless Belief," (with Anthony Booth, *The Journal of Philosophy*, 2010) "What Is Ignorance?" (*Philosophia*, 2011), and "Epistemic Desiderata and Epistemic Pluralism," (*Journal of Philosophical Research*, 2010). He also edited (with Henk Geertsema and Jan van der Stoep) *Philosophy Put to Work: Contemporary Issues in Art, Society, Politics, Science and Religion* (2008). Contact: rik.peels@phil.uu.nl.

John Turri is Assistant Professor at University of Waterloo. His areas of interest are epistemology, philosophy of language, philosophy of mind, metaphysics, and ethics. He is the editor of *Virtuous Thoughts: The Philosophy of Ernest Sosa* (Springer, forthcoming), *Epistemology: A Guide* (Wiley-Blackwell, forthcoming) and the author of many articles among which: "A New and Improved Argument for a Necessary Being" (*Australasian Journal of Philosophy*, 2011), "Believing for a

Reason" (*Erkenntnis*, 2011) and "Mythology of the Factive" (*Logos & Episteme*, 2011). Contact: jturri@uwaterloo.ca.

Brian Weatherson is Associate Professor of Philosophy at Rutgers University. His areas of specialization are epistemology and philosophy of language and his areas of competence are metaphysics, philosophy of mind, and logic. He is the author of "Deontology and Descartes' Demon" (*Journal of Philosophy*, 2008), "The Bayesian and the Dogmatist" (*Proceedings of the Aristotelian Society*, 2007), and "Can we Do Without Pragmatic Encroachment?" (*Philosophical Perspectives*, 2005). Contact: brian@weatherson.org.

LOGOS & EPISTEME:
AIMS & SCOPE

Logos & Episteme is a quarterly open-access international journal of epistemology that appears in printed and on-line version in March, June, September, and December. Its fundamental mission is to support and publish various current reflections and researches that aim at investigating, analyzing, interpreting or philosophically explaining the human knowledge in all its aspects, forms, types, dimensions or practices.

For this purpose, the journal will publish articles, reviews or discussion notes focused as well on problems concerning the general theory of knowledge, as on problems specific to the philosophy, methodology and ethics of science, philosophical logic, metaphilosophy, moral epistemology, epistemology of art, epistemology of religion, social or political epistemology, epistemology of communication. Studies in the history of science and of the philosophy of knowledge, or studies in the sociology of knowledge, cognitive psychology, and cognitive science are also welcome.

The journal intends to promote all methods, perspectives and traditions in the philosophical analysis of knowledge, from the normative to the naturalistic and experimental, and from the Anglo-American to the continental or Eastern.

The journal accepts for publication texts in English, French and German, which satisfy the norms of clarity and rigour in exposition and argumentation.

NOTES TO CONTRIBUTORS

1. Accepted Papers

The journal accepts for publication articles, discussion notes and book reviews submitted exclusively to *Logos & Episteme* and not published, in whole or substantial part, elsewhere. The editors of *Logos & Episteme* reserve the right to refuse all future papers belonging to the authors who fail to observe this rule.

2. Submission Address

Please submit your manuscripts electronically at: logosandepisteme@yahoo.com. Authors will receive an e-mail confirming the submission. All subsequent correspondence with the authors will be carried via e-mail. When a paper is co-written, only one author should be identified as the corresponding author.

3. Paper Size

The articles should normally not exceed 12000 words in length, including footnotes and references. Articles exceeding 12000 words will be accepted only occasionally and upon a reasonable justification from their authors. The discussion notes must be no longer than 3000 words and the book reviews must not exceed 4000 words, including footnotes and references. The editors reserve the right to ask the authors to shorten their texts when necessary.

4. Manuscript Format

Manuscripts should be formatted in Rich Text Format file (*rtf) or Microsoft Word document (*docx) and must be double-spaced, including quotes and footnotes, in 12 point Times New Roman font. Where manuscripts contain special symbols, characters and diagrams, the authors are advised to also submit their paper in PDF format. Each page must be numbered and footnotes should be numbered consecutively in the main body of the text and appear at footer of page. For all references authors must use the Humanities style, as it is presented in The Chicago Manual of Style, 15th edition. Large quotations should be set off clearly, by indenting the left margin of the manuscript or by using a smaller font size. Double quotation marks should be used for direct quotations and single quotation marks

should be used for quotations within quotations and for words or phrases used in a special sense.

5. Official Languages

The official languages of the journal are: English, French and German. Authors who submit papers not written in their native language are advised to have the article checked for style and grammar by a native speaker. Articles which are not linguistically acceptable may be rejected.

6. Abstract

All submitted articles must have a short abstract not exceeding 200 words in English and 3 to 6 keywords. The abstract must not contain any undefined abbreviations or unspecified references. Authors are asked to compile their manuscripts in the following order: title; abstract; keywords; main text; appendices (as appropriate); references.

7. Author's CV

A short CV including the author`s affiliation and professional address must be sent in a separate file. All special acknowledgements on behalf of the authors must not appear in the submitted text and should be sent in the separate file. When the manuscript is accepted for publication in the journal, the special acknowledgement will be included in a footnote on the first page of the paper.

8. Review Process

The reason for these requests is that all articles, with the exception of articles from the invited contributors, will be subject to a strict blind-review process. Therefore the authors should avoid in their manuscripts any mention to their previous work or use an impersonal or neutral form when referring to it. The review process is intended to take no more than six months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors. Processing of papers in languages other than English may take longer. The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers. The editors reserve their right to ask the authors to revise their papers and the right to require reformatting of accepted manuscripts if they do not meet the norms of the journal.

9. Acceptance of the Papers

The editorial committee has the final decision on the acceptance of the papers. Papers accepted will be published, as far as possible, in the order in which they are received and they will appear in the journal in the alphabetical order of their authors.

10. Responsibilities

Authors bear full responsibility for the contents of their own contributions. The opinions expressed in the texts published do not necessarily express the views of the editors. It is the responsibility of the author to obtain written permission for quotations from unpublished material, or for all quotations that exceed the limits provided in the copyright regulations. The papers containing racist and sexist opinions assumed by their authors will be rejected. The presence in texts of sexist or racist terms is accepted only if they occur in quotations or as examples.

11. Checking Proofs

Authors should retain a copy of their paper against which to check proofs. The final proofs will be sent to the corresponding author in PDF format. The author must send an answer within 3 days. Only minor corrections are accepted and should be sent in a separate file as an e-mail attachment.

12. Reviews

Authors who wish to have their books reviewed in the journal should send them at the following address: Institutul de Cercetări Economice şi Sociale „Gh. Zane" Academia Română, Filiala Iaşi, Str. Teodor Codrescu, Nr. 2, 700481, Iaşi, România. The authors of the books are asked to give a valid e-mail address where they will be notified concerning the publishing of a review of their book in our journal. The editors do not guarantee that all the books sent will be reviewed in the journal. The books sent for reviews will not be returned.

13. Property & Royalties

Articles accepted for publication will become the property of *Logos & Episteme* and may not be reprinted or translated without the previous notification to the editors. No manuscripts will be returned to their authors. The journal does not pay royalties. Authors of accepted manuscripts will receive free of charge two copies of the issue containing their papers.

14. Permissions

Authors have the right to use their papers in whole and in part for non-commercial purposes. They do not need to ask permission to re-publish their papers but they are kindly asked to inform the Editorial Board of their intention and to provide acknowledgement of the original publication in *Logos & Episteme*, including the title of the article, the journal name, volume, issue number, page number and year of publication. All articles are free for anybody to read and download. They can also be distributed, copied and transmitted on the web, but only for non-commercial purposes, and provided that the journal copyright is acknowledged.

Bun de tipar: 2011 · Apărut: 2011 · Format 17 × 24 cm

Iaşi, str. Grigore Ghica Vodă nr. 13
Tel. Dif.: 0788.319462; Fax: 0232/230197
editura_ie@yahoo.com; www.euroinst.ro

Volum elaborat în cadrul Proiectului POSDRU/89/1.5/S/56815.
Se distribuie GRATUIT.