Logos &
Episteme

an international journal
of epistemology

TABLE OF CONTENTS

ARTICLES

FROM HISTORICAL CHANGE TO HISTORICAL KNOWLEDGE: DIRECTIONS OF A NEW EPISTEMOLOGY OF THE HUMAN SCIENCES[1]

Adrian COSTACHE

ABSTRACT: The present paper endeavors to trace the sketch of a possible epistemology of the human sciences. In this sense it begins with the determination of the object of knowledge in the human sciences through a careful examination of the reality of history and of the human world. Then, considering the peculiarity of the domain of the human sciences the paper proceeds to show that their object of knowledge is best understood as "event" in the sense of Gilles Deleuze and Alain Badiou. And, in the end, it circumscribes two modes of knowledge of this object of the human sciences understood as event.

KEYWORDS: knowledge in the human sciences, event, Gilles Deleuze, Alain Badiou

1. The Basic Framework of the Epistemological Problem in the Human Sciences

The basic framework of any possible epistemological inquiry can be traced easily. The epistemological problem is anchored by the poles of the knower and the to-be-known and is posed in the space opened by the questions: "What is there to know?" and "How can one know what there is to know?"

Between the poles of the knower and the to-be-known though there is a profoundly dissymmetrical relation. If the knower can be a priori determined as subject, for only a subject can undertake a process of knowledge, the to-be-known remains completely undetermined (X). That is why any additional determination of the epistemological problem in this general form can be done only by posing it in a concrete manner, by saturating the X to-be-known in a determined epistemological context.

Depending on how one determines the X to-be-known, though, the epistemological problem is subjected to certain modulations bringing along a variable distribution of the importance, the signification and logical order in which the questions constituting it are to be approached.

From the point of view of the natural sciences for which the X to be known presents itself as a substantial, material entity standing as an object in front of the subject, the first question – "What is there to know?" – becomes insignificant, the epistemological problem reducing itself to "How can we know what there is to know?"

From the point of view of the human sciences, on the other hand, we are confronted with the opposite situation. From this point of view the central question is "What is there to know?." From its answer one can derive more or less directly the answer of the other.

That in the human sciences or, following the German tradition, the sciences of the spirit (*Giesteswissenschaften*) the epistemological problem gravitates around the object to be known is evident from the very beginning in that this is not an "object" *per se*. One might be tempted to say that it is actually a subject or, if we were to follow Hegel (in connection with whose thought the first systematic reflections in the domain of the human sciences have appeared), the spirit (*l'esprit*, *die Geist*). Of course, this temptation is one to which we should avoid falling prey. Suffice it to broaden our perspective for a moment to understand that the world of the humanities comprises also the worlds that have been but are not anymore. Otherwise put, the world researched by the human sciences is a world constituted also of the world of the past.

The world of the past though is never given in the present as pure spirit only insomuch as it is handed down through oral transmission. Which confronts us with a simple alternative: either we accept that the only thing that can be handed down from the past is the oral tradition, or the understanding of the "object" of study of the human sciences simply as spirit is a reductive one and, due to this, unsustainable. Thus we will have to tie the spirit to a material basis thereby offering it an additional degree of objectivity.

Dilthey, the methodologist of the Historical School and, probably, the most important epistemologist of the human sciences of the 19th century recognizes this from the very first page of his *Formation of the Historical World in the Human Sciences*. He writes:

> Besides the natural sciences, a group of conceptual cognitive results emerged naturally from the task of life itself. These results are linked to one another by their common object. History, political economy, the sciences of law and of the state, the study of religion, literature, poetry, architecture, music, of

> philosophical world-views and systems, and finally, psychology are such sciences. All these sciences refer rather to the same grand fact: the human race – which they describe, narrate, and judge, and about which they form concepts and theories. What one customarily separates as physical is undivided in this fact of the human sciences. It contains the living nexus of both. We ourselves belong to nature, and nature is at work in us, unconsciously, in dark drives. States of consciousness are constantly expressed in gestures, looks, and words; and they have their objectivity in institutions, states, churches, and scientific institutes, History operates in these very contexts. Of course, this does not exclude the possibility that the human sciences employ the distinction between the physical and the psychical whenever their purposes require it. But then they must remain conscious that they are working with abstractions, not with entities and that these abstractions are valid only within the limits of the point of view within which they are projected.[2]

By recognizing the dual constitution of the "object" of study of the human sciences – on the one hand, natural/ material and, on the other, spiritual – we are finally on the right track for determining the X to-be-known. But we have not reached our destination. For if we take another look at the sphere of history we will see, on the one hand, that the rapport between nature/matter and spirit constituting the X to-be-known is not as simple as it looks at first sight and, on the other, that the reality of history brings about the need for a more precise determination of the X to be known.

Let's tackle these matters in order. Earlier, when we were determining the "object" of the human sciences as a dual entity, constituted as nature/matter and spirit it seemed that, although necessary, the first term is rather a "frame" for the second. The past, though, shows us that things are not precisely so. For the spirit makes history, it leaves its traces upon the times to come, becoming thus worthy of interest for the historians to the same degree as nature. In such case, though, the spirit becomes the accessory term.

We find the best example in this sense in Michel Foucault's *History of Madness*[3] which shows that what brings about the "great confinement" and the transition from what Foucault calls the "Classical Age" (which is more or less identical to what traditional historiography calls "Modernity") is the black plague sweeping Europe since the second half of the 14th century. A natural fact, the plague, restructures from the ground up the lives of the people, leading to the instauration of completely new relations between the individual and his peers and

---

[2] Wilhelm Dilthey, "The Formation of the Historical World in the Human Sciences," in *Selected Works. Vol. III*, eds. Rudolf A. Makkreel and Frithjof Rodi (Princeton & Oxford: Princeton University Press, 2002), 101–102.

[3] See Michel Foucault, *Histoire de la folie* (Paris: Gallimard, 1972), 56–91.

a completely new rapport with the natural environment and with the transcendent. The plague will bring about new rules for the contact between people and will decide the appropriate contexts for such contact. It will impose strict rules for interacting with the domestic animals offering means of sustenance. And thirdly, it will restructure the self-understanding of the individual in relation to the divine, for the plague can be regarded both as a divine punishment among others and as the first moments of the Apocalypse.

This new manner of relating to peers, the environment and the transcendent gains the function of a model and is taken up once again almost spontaneously as soon as the social context (an economic crisis) requires it. Otherwise put, as soon as the social context requires it, on the one hand, the growth of the work force and of the productivity of labor and, on the other hand, the reduction of the social costs, the social rapports forged during the plague years are re-enacted once again. Just like the plagued from before, now the mad, the beggars, the petty thieves, the prostitutes, the pariah get thrown out of the cities and put in jail.

On the other hand, the fact that historical reality requires a more precise determination of the X to-be-known than as a dual entity constituted as nature/matter and spirit becomes apparent in the way in which historical research is done beginning with the second half of the 20th century.

Once again Foucault provides us with the privileged example. Already in the Introduction to *The Archeology of Knowledge* he observes:

> For many years now historians have preferred to turn their attention to long periods, as if, beneath the shifts and changes of political events, they were trying to reveal the stable, almost indestructible system of checks and balances, the irreversible processes, the constant readjustments, the underlying tendencies that gather force, and are then suddenly reversed after centuries of continuity, the movements of accumulation and slow saturation, the great silent, motionless bases that traditional history has covered with a thick layer of events. […] Beneath the rapidly changing history of governments, wars, and famines, there emerge other, apparently unmoving histories.[4]

> About the same time, in the disciplines that we call the history of ideas, the history of science, the history of philosophy, the history of thought, and the history of literature (we can ignore their specificity for the moment), in those disciplines which, despite their names, evade very largely the work and methods of the historian, attention has been turned, on the contrary, away from vast

---

[4] Michel Foucault, *The Archeology of Knowledge*, trans. A. M. Sheridan Smith (London & New York: Rutledge, 2002), 3–4.

unities like 'periods' or 'centuries' to the phenomena of rupture and discontinuity.[5]

At first sight one could be tempted to see here the sign of a methodological superficiality of the historical disciplines. But, as Foucault shows, things are not that simple. For if beginning with the second half of the 20th century the past becomes a discontinuous series for the different particular histories while for the history as such a continuous flux this is the direct result of the endeavors of the researchers in this group of disciplines to question the document taken as the support of the X to-be-known in the sphere of history.

As Foucault remarks:

> [The] document was always treated as the language of a voice since reduced to silence, its fragile, but possibly decipherable trace. Now, through a mutation that is not of recent origin, but which has still not come to an end, history has altered its position in relation to the document: it has taken as its primary task, not the interpretation of the document, nor the attempt to decide whether it is telling the truth or what is its expressive value, but to work on it from within and develop it: history now organizes the document, divides it up, distributes it, orders it, arranges it in levels, establishes series, distinguishes between what is relevant and what is not, discovers elements, defines unities, describes relations.[6]

Of course, if the document can be ordered, redistributed, organized in series, etc. this is because the historical reality itself which it tries to describes can be treated so. The possibility of all these operations exists precisely because historical reality can present itself either as a continuous flux or a discontinuous series.

Hence the supplementary determination of the X to-be-known in the human sciences by the exigency to account for both continuity and discontinuity in the passing of time. With this though, the epistemological problem of the human sciences becomes more complicated than it was. For which concept applies to the X thus determined?

As an answer to this question *The Archeology of Knowledge* offers us the concept of statement:

> [A] statement is always an event that neither the language (*langue*) nor the meaning can quite exhaust. It is certainly a strange event: first, because on the one hand it is linked to the gesture of writing or to the articulation of speech, and also on the other hand it opens up to itself a residual existence in the field of a memory, or in the materiality of manuscripts, books, or any other form of recording; secondly, because, like every event, it is unique, yet subject to

---

[5] Foucault, *The Archeology of Knowledge*, 4.

[6] Foucault, *The Archeology of Knowledge*, 7.

repetition, transformation and reactivation; thirdly, because it is linked not only to the situations that provoke it, and to the consequences that it gives rise to, but at the same time, and in accordance with a quite different modality, to the statements that precede and follow it.[7]

Foucault's concept of statement seems to be perfectly capable to account for both the continuity and discontinuity in the passing of time. But it remains completely silent when it comes to showing how are these articulated upon one another. Taking language as a starting point for determining the X to-be-known in the human sciences and maintaining himself within its domain Foucault does not ask himself either how do the ruptures in the flow of time come about or how they can instate new continuous fluxes. As the last passage quoted shows us, for him, the fact that the event is at the same time unique and repeatable is purely and simply given.

This is the second and last exigency with which we are confronted by historical reality in determining the X to-be-known in the human sciences, an exigency born out of a paradoxical experience, common to each and every one of us in everyday life. This exigency is the most difficult to satisfy. For each and every one of us sees how everything changes day by day, but despite this nothing ever happens. Just as we see that although nothing seems to be taking place, out of the sudden things are completely different than they were.

On the one hand, we are witnessing the monotonous succession to power of the different political parties, to the continuous worsening of the living conditions, the accelerated degrading of the environment and so many other causes and sufficient conditions for a break or what could be called a "cardinal change" in the course of history, yet nothing happens.

On the other hand, we see how the days, months and years go by "silently," all the changes taking place being what could be called "ordinal changes," changes caught up in the logic of things, completely foreseeable and stripped of any element of novelty. But, despite all these, out of the sudden a revolution, something like the "Arab Spring" or "Occupy Wall Street" is taking place.

Charles Péguy gets the point exactly:

For years and years, for ten, fifteen, twenty years, for thirty years you struggle with a certain problem and you cannot give any solution to it, and you struggle with a certain evil and you cannot bring any remedy. And an entire people struggles. And entire generations struggle. And out of the sudden one turns its back. And the face of the whole world changed. Neither the same problems are still posed (others will be), nor the same difficulties will present themselves, nor

---

[7] Foucault, *The Archeology of Knowledge*, 31.

the same maladies are still considerable. Nothing has happened. Yet everything is different. Nothing has taken place. Yet everything is new. Nothing has taken place. And all that is old ceases to exist and all that is old has become foreign.[8]

There are thus three conditions to be satisfied by the concept of X to-be-known in the human sciences:

(i) to refrain from attributing an ontological priority to any of its two constitutive elements. For the X isn't first and foremost either nature or spirit. As we have seen, history and through it, the human world can be made by both;

(ii) to prove capable to account for both the continuity and the breaks in the flow of time;

(iii) to prove capable to account for the way in which continuity and discontinuity get articulated with one another thus producing both ordinal and cardinal changes.

Armed with these three conditions it is high time to get back to our question: which concept can satisfy them?

2. The X To-be-known as Event

We would like to answer the above question straightforwardly by saying: if properly understood, the concept able to satisfy the exigencies imposed by the X to-be-known in the human sciences is the concept of event. When considered closely one can see that from the psychic to the economic and social processes, from the processuality of thinking to that of writing, otherwise put, from psychology to economy and sociology, from philosophy to the theory of literature – this entire group of disciplines deals exclusively with what takes place in its reflexive field, with events.

This answer can be discerned between the lines of Foucault's *Archeology of Knowledge*, right in the concept of statement proposed. For if we distance ourselves from the dimension of language, that which brings about the problem with the determination of the X to-be-known in the human sciences as statement, what we are left with is precisely the idea of event, unique but, at the same time, subject to repetition, transformation and reactivation, tied to the situations provoking it and the effects it itself provokes.

At the same time though, this concept has been anticipated by the philosophical hermeneutics of the 20th century initiated by Martin Heidegger and continued by Hans-Georg Gadamer. For Gadamer tradition is the true "object" of the human sciences and this is always given to us as an event.

---

[8] Charles Péguy, *Clio*, 266. (Translation is mine.)

But if philosophical hermeneutics is able to anticipate this answer but not to arrive at it as such this is because it falls prey to the same trap Foucault fell in his *Archeology of Knowledge*, i.e., that of situating the event in the proximity of language and trying to think it in terms of language. For Gadamer the true event is the event of understanding, which is the exclusive performance of language. In *The Continuity of History and the Existential Moment* he writes explicitly:

> [W]hen something encounters us within the tradition in such a way that we understand it, then that itself is an event. And something happens when one, so to speak, accepts a word from the tradition, when one allows a word to speak to him.[9]

Through this Gadamer and philosophical hermeneutics in general are confronted with the impossibility of recognizing what we have called cardinal change, that is any revolution happening in the course of history. This impossibility is clearly demonstrated by *Truth and Method*:

> 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.[10]

Precisely this is the reason why, at the beginning of this section when we were identifying the X to-be-known in the human sciences as event, we were formulating the precaution: "if properly understood." We believe that we can find such a "proper" understanding of the concept of event in the works of Gilles Deleuze and Alain Badiou who are in complete accord in this regard, despite the distance that separates them, a distance carefully highlighted by Badiou in his *Deleuze. «La clameur de l'Être»*.[11]

2.1. The Event in Deleuze – The Nomadic Systematization of Singularities

In developing his conception of the event Deleuze follows closely the basic insights of the materialist ontology of the Stoics. For the Stoics the event is situated in strict opposition to the concept of body.[12] As it is well known, for the

---

[9] Hans-Georg Gadamer, "The Continuity of History and the Existential Moment," *Philosophy Today* 16,. 3-4 (1972): 237.

[10] Hans-Georg Gadamer, *Truth and Method*, trans. Joel Weinsheimer and Donald G. Marshall (London & New York: Continuum, 2004), 282–283.

[11] See Alain Badiou, *Deleuze. «La clameur de l'Être»* (Paris: Fayard/Pluriel, 2010), especially 7–m 15.

[12] We will not insist here on the Deleuzian concept of event for we have already provided a careful treatment of it in Adrian Costache, "Real Events – Ideal Events: A Deleuzian Approach to the Concept of Historical Event," *European Journal of Science and Theology* 8, 3 (2012).

Stoics only bodies exist, the universe being nothing else than a great mixture of bodies. In this mélange each body causes and is caused by all the others and so each body is at the same time active and passive by rapport to all the others. The events on the other hand are the incorporeal effects of this mélange on the surface of the bodies.

If only the bodies are properly said to exist, of the events we must say that they subsist. If bodies are active and passive, events are impassive. And whereas bodies exist in a time defined as Chronos, the time proper to the events is that species of eternity constituting itself through the ceaseless avoidance of the present the Greeks called Aion. Because of this Deleuze distinguishes between the event proper and its realization in a certain space at a particular time as between "two courses of events," one of them "ideal" and the other one "real" and "accidental" (LS 53).

The strict identification of the bodies as causes and of the events as effects and the dichotomy instituted between them seems to prohibit the identification of Deleuze's concept of event as the X to-be-known in the human sciences. For, on the one hand, it appears to be deprived of the power to leave any mark on the bodies that produce it, and, on the other, it proves incapable to determine other events in its turn. Although as incorporeal effects events can never become themselves causes *per se*, for Deleuze, they can become "quasi-causes" determining through counter-actualization both its spatio-temporal realization and other events.

In order to see how the quasi-causality and counter-actualization work we should turn our gaze toward what in *The Logic of Sense* Deleuze calls the "static ontological genesis"[13] of the event. For Deleuze this genesis is tied to the emission of a series of remarkable points in a transcendental field and is the result of an "immanent principle of auto-unification through a nomadic distribution."[14] What this means to say is that from a static point of view the event is nothing else than arbitrary auto-unification and systematization of some of the points appearing in a transcendental field. By the transcendental field of the event Deleuze understands the pre-individual and impersonal plane on which something is given to somebody. The heterogeneous series of remarkable points is simply a series of punctual unities that can be remarked such as to green, to cool, to make noises, to count to five hundred, etc.

---

[13] Gilles Deleuze, *The Logic of Sense*, trans. Mark Lester and Constantin V. Boundas (New York: Columbia University Press, 1990), 109.
[14] Gilles Deleuze, *The Logic of Sense*, 102.

Adrian Costache

Deleuze though is a positivist in the same sense in which Foucault was declaring himself to be a "happy positivist"[15] in the *Archeology of Knowledge*, i.e., someone who does not recognize negativity and negation. In *Difference and Repetition* he writes:

> The negative is an epiphenomenon. Negation, like the ripples in a pond, is the effect of an affirmation which is too strong or too different. Perhaps two affirmations are necessary in order to produce the shadow of negation as *Nachfolge*.[16]

> Negation is difference, but difference seen from its underside, seen from below. […] Negation results from affirmation: this means that negation arises in the wake of affirmation or besides it, but only as the shadow of the more profound genetic element – of that power or 'will' which engenders the affirmation and the difference in the affirmation.[17]

Precisely this is why, for Deleuze, nothingness in the common metaphysical sense of this concept does not exist. Or, better put, nothingness cannot be thought in privative terms. For Deleuze too nature "abhors a void":

> There is a non-being, yet there is neither negative nor negation. There is a non-being which is by no means the being of the negative, but rather the being of the problematic.[18]

By not recognizing the negative and negation, though, and by reversing the old principle *omnis determination negatio* as *omnis determinatio afirmatio* we can understand that along with and through the auto-unification of some of the remarkable points of the series as this or that particular event, the other remarkable points in the transcendental field do not disappear into nothingness, but will be subjected in their turn to another arbitrary systematization thus bringing about a new event.

The first event does not produce the second *per se*, but without the second event taking place would not have existed. Precisely the first auto-unification of the remarkable points of the series is what, to put it like this, forces a new exercise of the principle of nomadic distribution thus leading to the second event. The counter-effectuation Deleuze talks about is precisely this. And due to the fact that through it a new event occurs, we can say that it was quasi-caused by the first. Here is the exact way in which while maintaining their status of effects produced

---

[15] Michel Foucault, *The Archeology of Knowledge*, 141.

[16] Gilles Deleuze, *Difference and Repetition*, trans. Paul Patton (New York: Columbia University Press, 1994), 54.

[17] Deleuze, *Difference and Repetition*, 55.

[18] Deleuze, *Difference and Repetition*, 202.

by the bodily causes the incorporeal events are able to mark the other events and the bodies producing them.

Before seeing whether this rather bizarre concept of event manages to satisfy the epistemological exigencies imposed by the X to-be-known in the human sciences we will have to see if it can really be applied to the historical world. If have to admit, the abstraction of Deleuze's concept of event seems completely estranged from the concreteness of the reality of history.

For this we will have to turn our eyes to the ontology of the multiple proposed by Alain Badiou in *Being and Event*. Even though at first sight this strategy might seem a detour from the course of our investigation the recourse to Badiou's work is justified by a number of reasons:

> (i) Badiou offers us a philosophy of the event based on an ontological doctrine similar to that of Deleuze;
> (ii) due to the structural identity between the two philosophers' conception of the event;
> (iii) because of the fact that, unlike Deleuze, Badiou approaches explicitly the problem of the correspondence between the abstract concept of event he proposes and the intuitive idea, i.e., the phenomenon as it is given to us in intuition.

2.2. The Event in Badiou – The Additional Signifier

Here too we will have to start with the ontological background of the concept. Just like for Deleuze everything begins with an emission of singular points, for Badiou the beginning is to be found in a multiple presenting itself. Such a multiple though – which is the generic form of presentation of being-qua-being or, otherwise put, the mode of being of what is -, is a pure multiple, constituted in its turn of other multiplicities. That is why, the one the Greeks were attributing first and foremost the status of being, strictly speaking is not. The one in merely an operation, the "count-as-one" through which the multiple is structured as a situation. In *Manifesto of Philosophy* Badiou writes:

> In the interests of brevity, let us call 'situation' a state of things, any presented multiple whatsoever.[19]

And in *Being and Event* he adds:

> Granted the effectiveness of presentation, a situation is the place of taking-place, whatever the terms of the multiplicity in question.[20]

---

[19] Alain Badiou, *Manifesto of Philosophy*, trans. Norman Madarasz (Albany/New York: State University of New York Press, 1999), 36.

Adrian Costache

A close look at this concept of situation introduced by Badiou shows that it is not only similar but, actually, structurally identical with the Deleuzian transcendental field. And this from two points of view: there is first of all a functional and, second of all, a structural identity between these two concepts.

The functional identity with the transcendental field comes to light through the very mode in which the concept of situation is defined. For, as we have seen, it too is nothing else than the "place" of the occurrence of the event. In its turn, the structural identity announces itself through the fact that both concepts share the same "aspect" in a two-fold sense. On the one hand due to the fact that both are a minimal structuring or systematization of a multiplicity. For, as we have seen, the transcendental field takes the form of a simple series of remarkable or singular points. And, on the other hand, due to the fact that the multiplicity constitutive to both is perfectly heterogeneous in its nature.

In *Being and Event* Badiou distinguishes between two types of multiplicities: (i) *natural multiplicities* which, given the homogeneity, come to be represented as subsets[21] of the situation in which they are presented and thus, its constitutive elements. And (ii) *singular multiplicities* whose elements, due to their heterogeneity, cannot be organized as subsets and are never represented in the situation. For Badiou the first type of multiplicity is specific to natural situations, subject to the law of determination and in which nothing really happens, every change that appears being dictated and completely predictable starting from the prior states of the situation. On the other hand for Badiou, just like for Deleuze, singular multiplicities must be reserved to the situations open to the occurrence of what Badiou calls "evental sites" and, thus, to the happening of an event.

> It is rational to think the ab-normal or the anti-natural, that is, history, as an omnipresence of singularity – just as we have thought nature as an omnipresence of normality. The form-multiple of historicity is what lies entirely within the instability of the singular; it is that upon which the state's metastructure has no hold.[22]

> I will term situation in which at least one evental site occurs *historical*. I have chosen the term 'historical' in opposition to the intrinsic stability of natural situations.[23]

---

[20] Alain Badiou, *Being and Event*, trans. Oliver Feltham (London & New York: Continuum, 2005), 24.

[21] In the mathematical sense of set theory to which Badiou sends constantly.

[22] Badiou, *Being and Event*, 174.

[23] Badiou, *Being and Event*, 177.

Given this functional and structural identity between the elements of genesis and the constitutive moments of the event in the two philosophers, the profound similarity that exists between the Badiou and the Deleuze's philosophy of the event should not surprise anyone.

In *Being and Event* Badiou defines the event in the following manner:

> Take, in a historical situation, an evental site X.
> I term event of the site X a multiple such that it is composed of, on the one hand, elements of the site, and on the other hand, itself.
> The inscription of a *matheme of the event* is not a luxury here. Say that S is the situation and $X \in S$ (X belongs to S, X is presented by S) the evental site. The event will be written e_x (to be read 'event of the site X'). My definition is then written as follows:
> $e_x = \{x \in X, e_x\}$
> That is, the event is a one-multiple made up of, on the one hand, all the multiples which belong to its site, and on the other hand, the event itself.[24]

All this comes to say that the event is produced by the multiples presented in the historical situation but is completely different than it. It exists through these multiples but subsists independently of them. In a deliberate Deleuzian vocabulary, the event "hovers"[25] like a double over the evental site and the situation in which it occurs, being though completely indifferent to it and totally independent from it. For, as Badiou shows, this event constituted of the multiple of the evental site and itself has a supernumerary nature, appearing as a supplement to the situation given. A supplement that

> … can neither be named, nor represented by referring to the resources of the situation (its structure, the established language naming its terms, etc.). It is inscribed by a singular naming, the bringing into play of an *additional signifier*.[26]

Considering the paradoxical formulation the event receives in Badiou's thought the question whether it really finds a correspondent in historical reality becomes even more pressing than it was in Deleuze. For this possibility seems even more unsustainable. Maybe precisely this is the reason why, unlike Deleuze, Badiou approaches it explicitly in *Being and Event* immediately after defining the event in the above manner.

In order to answer this question Badiou will adopt an intuitive strategy. He will take recourse to a concrete historical event and will try to show that it can

---

[24] Badiou, *Being and Event*, 179.
[25] Deleuze, *The Logic of Sense*, 100.
[26] Badiou, *Manifesto of Philosophy*, 36.

really be decomposed down to the fundamental terms of his conception of the event.

> Take the syntagm 'the French Revolution.' What should be understood by these words? One could certainly say that the event 'the French Revolution' forms a one out of everything which makes up its site; that is, France between 1789 and, let's say, 1794. There you'll find the electors of the general Estates, the peasants of the Great Fear, the sans-culottes of the towns, the members of the Convention, the Jacobin clubs, the soldiers of the draft, but also, the price of subsistence, the guillotine, the effects of the tribunal, the massacres, the English spies, the Vendeans, the *assignats* (banknotes), the theatre, the Marseillaise, etc. The historian end up including in the event 'the French Revolution' everything delivered by the epoch as traces and facts. This approach, however – which is the inventory of all the elements of the site – may well lead to the one of the event being undone to the point of being no more than the forever infinite numbering of gestures, things and words that co-existed with it. The halting point for this dissemination is *the mode in which the Revolution is a central term of the Revolution itself*; that is, the manner in which the conscience of the times – and the retroactive intervention of our own – filters the entire site through the one of its evental qualifications.[27]

Obviously, any historical event can be decomposed into such a series of elements which, with enough perseverance, can be unfolded to infinity. That is why any historical event can be made of such a multiple only insomuch as it is also made of itself.

Through the intuitive character of Badiou's concept of event we arrive at the intuitive character of Deleuze's concept. The structural identity between these terms shows that to the multiple constitutive of the event for Badiou it corresponds a series of emissions of singular points in a transcendental field. And to the auto-unification of the series as this or that event under the influence of the principle of nomadic distribution it corresponds the additional signifier given to the event and without which it could never become what it is.

Now, in light of the intuitive character of Deleuze and Badiou's understanding of the event it becomes easier to see that it manages to satisfy all the exigencies imposed in order to be able to be taken as X to-be-known in the human sciences. The first exigency established was to be able to account for both nature/matter and spirit as agents of history and the human world and not to grant any of these terms an ontological priority over the other. Upon close inspection it becomes manifest that the Deleuzian concept of event satisfies this requirement *ab initio*. For, translating the matter in these terms, the Deleuzian event has an ideal,

---

[27] Badiou, *Being and Event*, 180.

spiritual being resulting from material nature. But this requirement is not really a problem for Badiou's concept of event either. For, as we have seen through the example discussed, the elements of the multiple constituting the evental site are both spiritual and material.

But also the second and the third exigencies are just as easy satisfied. These exigencies were asking of the concept whereby the X to-b-known in the human sciences to be able to account for the occurrence of both the continuity and the discontinuity in the passing of time and the possibility of their articulation. Insomuch as the event is defined as a minimal systematization or structuring of a given multiplicity or a multiple of multiplicities we can understand that its occurrence equals to the introduction of a discontinuity in a continuum of the series or the situation *and* the constitution of a continuity through discontinuation of the flow of the given series or situation.

The question that imposes itself upon us now, the last question to be asked in order for our sketch for a new epistemology of the human sciences to be complete is that adjacent to "What is there to know?" Namely, "How can one know the X determined as event?"

In order to answer this question Deleuze begins by noting that there are actually two modes of knowing an event.

> In a great work of philosophy, Péguy explains that there are two ways of considering the event. One consists in going over the course of the event, in recording its effectuation in history, its conditioning and deterioration in history. But the other consists in reassembling the event, installing oneself in it as in a becoming, becoming young again and aging in it, both at the same time, going through all its components or singularities.[28]

Of course, if there are two modes of considering the event this is because for Deleuze and Badiou, in a way, the event itself is given in two modes. As we have seen, in Deleuze we have, on the one hand, the event itself, in its impassive purity and, on the other hand, the event embodies through its spatio-temporal realization. In Badiou we have the event as the additional signifier of a situation and the multiple of the situation structured as event.

For Deleuze the first mode of considering the event is the historical one, the one peculiar to the science of history and the other historical human sciences because

---

[28] Gilles Deleuze and Félix Guattari, *What Is Philosophy?* trans. Hugh Tomlinson and Graham Burchell (New York: Columbia University Press, 1994), 111.

> What History grasps of the event is its effectuation in states of affairs or in lived experience, but the event in its becoming, in its specific consistency, in its self-positing as concept, escapes History.[29]

In this latter mode of presenting itself the event opens itself up to philosophical knowledge.

> … the object of philosophy is not to contemplate the eternal or to reflect history but to diagnose our actual becomings.[30]

In historical knowledge, the ideal event, what can be thought from the singular points given in the transcendental field becomes the archeological principle guiding the digging out and the ordering of the traces left by the production of the event upon the bodies that produced it. Or, in Badiou's vocabulary, the event as additional signifier becomes the order word whereby the multiples presented in a situation start to signify something.

In philosophical knowledge on the other hand, quite the contrary, the spatio-temporal realizations of the event become just as many ramps for a leap towards what can be thought through the remarkable points given in the transcendental field, towards experimenting as many of their virtual combinations. As Deleuze shows,

> To think is to experiment. […] Without history experimentation would remain indeterminate and unconditioned, but experimentation is not historical. It is philosophical.[31]

As we can see, there is an ontological priority of the pure event and a chronological anteriority of its spatio-temporal realization in Deleuze. The pure event transforms itself and reaches its eternal truth with each and every one of its realizations and with all it is also better and better known.

At the end of our inquiry we would like to bring to light the profoundly revolutionary meaning of this view upon the matter of knowledge in the human sciences. In our opinion, in these dense passages we find two of the most important theses for the epistemology of the human sciences formulated in the second half of the 20th century.

The first is that historical and philosophical knowledge are in the end competing forms of knowledge, complementing each other and raising equal claims upon one and the same field of knowledge. With this philosophy is brought once again with its feet on the ground from its ivory tower. For the shared domain

---

[29] Deleuze and Guattari, *What Is Philosophy*, 110.
[30] Deleuze and Guattari, *What Is Philosophy*, 111.
[31] Deleuze and Guattari, *What Is Philosophy*, 111.

and the overlapping explanations – the philosophical theories explaining different social phenomena, the psychological approach of philosophical questions, etc. – considered by classical epistemology as cases of transgression usually resolved through the exclusion of philosophy become the normal case and the common situation.

And the second thesis is that historical knowledge is not or, better put, should not let itself be guided by what is already known or what can already be thought about the event studied, but by what can be virtually thought about it. Historical knowledge is not grounded in tradition and is not acquired through the terms of an already given view of the world. Even though it is a science of the past history is grounded on the future.

With this we are as far away as possible from the classical epistemology of the human sciences.

EXPLICATING A STANDARD EXTERNALIST ARGUMENT AGAINST THE KK PRINCIPLE

Simon D'ALFONSO

ABSTRACT: The KK principle is typically rejected in externalist accounts of knowledge. However, a standard general argument for this rejection is in need of a supportive explication. In a recent paper, Samir Okasha argues that the standard externalist argument in question is fallacious. In this paper I start off with some critical discussion of Okasha's analysis before suggesting an alternative way in which an externalist might successfully present such a case. I then further explore this issue via a look at how Fred Dretske's externalist epistemology, one of the exemplifying accounts, can explain failure of the KK principle.

KEYWORDS: knowledge, externalism, reliabilism, KK principle, Fred Dretske

According to various versions of the KK (Knowing that One Knows) principle, if one knows some proposition p then they know or are at least in a position to know that they know that p. This principle is often taken to be a dividing factor between internalist and externalist accounts of knowledge, with the former typically endorsing the principle and the latter typically rejecting it. A general explanation of one of the standard externalist arguments for rejecting the KK principle, exemplified by the reliabilist family of knowledge accounts, is given in the following passage:

> For, if warrant may be external to our cognitive perspective, then there is no special reason to expect those who know that p to be in a position to know that their belief that p is warranted. This can be seen this more clearly by focusing on the reliabilist theory of knowledge. If one's belief that p is produced by a reliable process that one knows nothing about, then one may have no way of knowing that this belief constitutes knowledge, and thus no way of knowing that one knows that p.[1]

Thus according to this externalist position, an epistemic agent can know that p without knowing that their belief that p was reliably produced (the externalist condition). But if knowledge is defined as reliably produced true belief, then knowing that they know that p implies knowing that their belief that p was

---

[1] David Hemp, "The KK (knowing that one knows) principle," *Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/kk-princ/#H1, last accessed 6 September 2013.

Simon D'Alfonso

reliably produced. As suggested by Samir Okasha[2] though, relying on this conflict to reject the KK principle proves to be problematic and the standard argument against the KK principle that uses it is fallacious.

Okasha represents the generic externalist line of reasoning in question with the following formal reasoning in epistemic logic:

(1) Kp (assumption)
(2) $\neg K[Bpisreliable]$ (assumption)
(3) $Kp\equiv[Bp\wedge p\wedge Bpisreliable]$ (externalism definition)
(4) $Kp\Rightarrow KKp$ (KK principle, assumed for *reductio*)
(5) KKp (from (1), (4))
(6) $K[Bp\wedge p\wedge Bpisreliable]$ (from (3), (5))
(7) $K[Bpisreliable]$ (from (6), assuming knowledge distributes across conjunction)
(8) $\neg[Kp\Rightarrow KKp]$ (from (1), (2), (3), (4), (7), by *RAA*)

As he points out though, this argument is fallacious, as the derivation of (6) from (3) and (5) involves substitution within an intensional context; specifically, substituting the externalist definition of knowledge for the second occurrence of 'K' in the expression 'KKp.' Might there be a way however of deriving (6) and rejecting the KK principle that does not involve this problematic move?

Appealing to the closure of knowledge under known implication ($Kp\wedge K(p\Rightarrow q)\vdash Kq$) offers one plausible way to derive (6), provided that a further assumption is made regarding knowledge of the definition of knowledge. The reasoning then becomes:

(1) Kp (assumption)
(2) $\neg K[Bpisreliable]$ (assumption)
(3) $K(Kp\equiv[Bp\wedge p\wedge Bpisreliable])$ (knowledge of externalist definition)
(4) $KKp\Rightarrow K[Bp\wedge p\wedge Bpisreliable]$ (from distribution on (3))
(5) $Kp\Rightarrow KKp$ (KK principle, assumed for *reductio*)
(6) $K[Bp\wedge p\wedge Bpisreliable]$ ((from (1), (4), (5)))
(7) $K[Bpisreliable]$ (from (6), assuming knowledge distributes across conjunction)
(8) $\neg[Kp\Rightarrow KKp]$ (from (2), (7), by *RAA*)

Despite this possibility, Okasha claims that

this weaker principle is insufficient to get us from (3) and (5) to (6); since the equivalence $Kp\equiv[Bp\wedge p\wedge Bpisreliable]$, and thus the implication

[2] Samir Okasha, "On a flawed argument against the KK principle," *Analysis* 73 (2013): 80–86.

$Kp \Rightarrow [Bp \land p \land Bp\,is\,reliable]$, is presumably not something that the typical subject knows or believes. At most, a handful of reliabilist epistemologists might know that this implication holds.[3]

Whilst this might be the case, it must be stressed that we generally take a rejection of the KK principle here to mean that it fails to hold in certain cases. This position is not anti-KK; that is, it is not committed to denying that there can be cases where KKp follows from Kp. It thus suffices to show that there can be certain cases within this argument setting where Kp holds and KKp does not. Now, even if $K(K \equiv [Bp \land p \land Bp\,is\,reliable])$ is only true in cases where the K is the knowledge operator of an externalist epistemologist who knows the definition of knowledge, this suffices to show that there are cases where the KK principle does not hold. The result is that the KK principle would fail in cases where the knowledge operator is that of an externalist epistemologist who knows the externalist definition of knowledge!

In fact, knowledge of the definition is not required. We could start off with the weaker $KKp \equiv K[Bp \land p \land Bp\,is\,reliable]$. If we dissect this equivalence further, we come up with the following three implications:

(1) $KKp \Rightarrow KBp$

(2) $KKp \Rightarrow Kp$

(3) $KKp \Rightarrow K[Bp\,is\,reliable]$

(1) and (2) are uncontroversial axioms in an epistemic-doxastic logic. It all really boils down to formula (3). If (3) is accepted then the KK principle would fail to hold as per the reasoning outlined above.

But even if $K(K \equiv [Bp \land p \land Bp\,is\,reliable])$ were to be the case, it could be claimed that "externalists typically reject even this weaker closure principle ... [so] ... it seems unlikely that a commitment to closure is behind the inference from (3) and (5) to (6)."[4] This point however is problematically cursory and demonstrates a failure to be careful in distinguishing between a rejection of the universality of a principle and an outright rejection of the principle. Rejecting closure does not mean denying that such an inference ever holds. Rather, it means that in certain circumstances and for certain reasons, Kp and $K(p \Rightarrow q)$ can be true whilst Kq is false, for some propositions p and q. Going by the externalist accounts I am familiar with, there is no reason to suggest that they would deny closure in such a case.

---

[3] Okasha, "On a flawed argument," 83.

[4] Okasha, "On a flawed argument," 83.

Simon D'Alfonso

Beyond the analysis thus far, might we be able to present this externalist point against the KK principle in a different way, one that is not susceptible to the problem above? I think that there is; in fact, it is how I first analysed this externalist position on the KK principle.

In the argument above, the premise $\neg K[Bp is reliable]$ is used to form the final contradiction. This premise however should be replaced with another, one that forms part of an iterative application of the externalist definition of knowledge, with Kp being the proposition that is known. According to this application, KKp if and only if:

1. BKp

2. Kp

3. $B[Kp] is reliable$

Thus the KK principle question can be put as

Kp⇒(BKp∧Kp∧B[Kp]isreliable)?

Of these three conditions listed above, we take the first two to hold given Kp. Condition 2 (Kp) trivially follows from the consequent. Condition 1 (BKp) has been used as an argument against the KK principle, for if the KK principle is valid then Kp implies that BKp and thus knowledge would be ruled out for agents that are not capable of introspection or forming second-order beliefs (Kelp *et al* 2011).[5] But let us set that aside and confine ourselves to knowledge and agents for which $Bp⇒BBp$ and $Kp⇒BKp$ holds. This leaves us with condition 3 (B[Kp] is reliable), which can fail to follow from Kp. The externalist line of reasoning thus becomes:

(1) Kp	(assumption)
(2) $\neg(B[Kp] is reliable)$	(assumption)
(3) $Kp \equiv [Bp \wedge p \wedge Bp is reliable]$	(externalism definition)
(4) $KKp \equiv [BKp \wedge Kp \wedge B[Kp] is reliable]$	(substituting Kp for p in (3))
(5) $Kp ⇒ KKp$	(assumed KK principle)
(6) KKp	(from (1) and (5))
(7) $BKp \wedge Kp \wedge B[Kp] is reliable$	(from (4) and (6))
(8) $B[Kp] is reliable$	(conjunct of (7))
(9) $\neg[Kp ⇒ KKp]$	(*RAA* from (2) and (8))

[5] Christoph Kelp and Nikolaj J.L.L. Pedersen, "Second-Order Knowledge," in *The Routledge Companion to Epistemology*, ed. Sven Bernecker and Duncan Pritchard (London: Routledge, 2011), 586–596.

So in order for one to know that they know that p in this externalist framework, their belief that they know that p must be reliable. But this does not necessarily follow from Kp and so the KK principle fails to hold.

Now, this discussion has all been in terms of a generic notion of externalist reliability. In actuality the array of externalist accounts are going to have their own specific externalist correlate and how exactly it can be the case that Kp whilst $\neg(B[Kp]\,is\,reliable)$, is something for a given externalist account to explain. In order to exemplify this externalist line of reasoning and also for the sake of generally discussing the issue, I shall now attempt to fit this analysis around Dretske's prominent externalist account of knowledge.

One of the pioneers of externalism and relevant alternative accounts in epistemology,[6] Dretske gave a complete embodiment of these ideas via his informational account of knowledge.[7] In short, according to this account knowledge is defined as information-caused belief. Information carrying and belief causing signals provide the externalist component and a veridical conception of information ensures that such beliefs are true. For Dretske:

> A signal r carries the information that s is F = The conditional probability of s's being F, given r (and k), is 1 (but, given k alone, less than 1)

A set of relevant alternatives is used to determine the range of possibilities over which this probability assessment is made. Dretske's classic zebra example[8] provides an amusing and simple way to relate these ideas. In this scenario, one is at a zoo where they go to the zebra section and see what they believe to be a zebra. Now, we can say that every time one receives a 'zebra' visual signal in certain circumstances then that means there is a zebra before them and so the required probability is 1. Therefore this visual signal carries the information that there is a zebra and any resulting belief that there is a zebra is knowledge. But for this judgement on the information-carrying status of the signal to be made, certain theoretically possible but non-actual irrelevant alternatives, such as ones in which the creature before the zoo goer is a cleverly disguised mule painted to resemble a zebra or a virtual zebra in a simulation, are not considered. Thus it is within the set of relevant alternatives, which say, consist of standard zoo scenarios, that a 'zebra' visual signal carries the information that there is a zebra.

---

[6] See Fred Dretske, "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971): 1–22, Tim Black, "Contextualism in Epistemology," *Internet Encyclopedia of Philosophy*, 2006, http://www.iep.utm.edu/contextu/#SH3a, last accessed 6 September 2013.

[7] See Fred Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT Press, 1981).

[8] See Fred Dretske, "Epistemic Operators," *Journal of Philosophy* 67 (1970): 1007–1023.

This all means that one can see a zebra at the zoo and know that there is a zebra before them without being able to ascertain that the creature is not a painted mule or a virtual zebra simulation. Such irrelevant alternatives determine a set of conditions: the signal carries the information that 'zebra' given that 'painted mule' is not the case, 'virtual zebra' is not the case, etc. That these conditions obtain is not something the knower has to know though and as the following passage from Dretske captures, it is this that can lead to failure of the KK principle:

> modest contextualism (and, hence, externalism) provides an illuminating explanation of why KK fails. It fails because factual knowledge, according to modest contextualism, depends for its existence on circumstances of which the knower may be entirely ignorant. So the knower can know that P without knowing (as required by KK) that he knows that P.[9]

So the zebra example is a candidate for a situation where the KK principle fails; whilst one knows that there is a zebra before them (Kz) they need not be in a position to know this (KKz). How could this work exactly?

With the strategy I have in mind, the gist involved here is that whilst one can get the required information and know some proposition within the bounds of certain relevant alternatives, the inability to be certain within a wider context and rule out certain irrelevant alternatives precludes their second-order knowledge of that proposition.

In judging $Kz \Rightarrow KKz$, we treat Kz as a proposition that falls within the scope of the outer K operator. Given this, KKz would require the satisfaction of the following three conditions:

1. Kz

2. BKz

3. The belief that Kz being caused by the information that Kz (reliability equivalent).

Conditions 1 and 2 are straightforward enough and were covered in discussion above. Condition 3 is the interesting condition and the one that fails under this externalist framework. In order to explain why it fails here, we need to explain how the information that z is present whilst the information that Kz is not. Such an explanation is available given that the set of relevant alternatives against which the information that Kz is judged is different to and greater than the set of relevant alternatives against which the information that z is judged. This is

---

[9] Fred Dretske, "Externalism and modest contextualism," *Erkenntnis* 61 (2004): 176.

related to the other infamous feature of externalist epistemologies such as Dretske's, namely the failure of knowledge under known implication. One can know that something is a zebra and know that something being a zebra implies that it is not a mule ($\neg m$) without knowing that it is not a disguised mule. This is because the set of relevant alternatives for the proposition $\neg m$ includes 'disguised mule', something that cannot be ruled out with a visual zebra signal.

For a belief that Kz to be caused by the information that Kz would require a signal that carries sufficient information coupled with an agent's reliable mechanism for forming beliefs about their beliefs as knowledge. If we take it as a given that the latter will be consistent across all of the relevant alternatives, then it is the insufficiency of the visual zebra signal that leads to a lack of meta-knowledge. This insufficiency is due to the fact that as with the case of closure and $\neg m$, the proposition Kz is judged against a more demanding set of relevant alternatives than z, a set in which possibilities such as disguised mule are included. Since the signal does not carry the information that z relative to this more demanding set of relevant alternatives and does not carry the information that it carries the information that z, then it is not enough for Kz either. In this way, we can also say that the visual zebra signal does not suffice to carry the information that Kz because there are relevant scenarios (like the painted mule one) where the 'zebra' visual signal occurs and Bz results but $\neg z$ and therefore $\neg Kz$.

In light of this discussion, it is worth briefly noting that we might roughly portray the logical connection between closure and the KK principle with the following arguments:

- $Kp \land (\exists Q)(K(p \Rightarrow Q) \land \neg KQ) \vdash \neg KKp$

- $KKp \vdash \neg (\exists Q)(K(p \Rightarrow Q) \land \neg KQ)$

with another argument to consider being what I term KK closure:

- $KKp \land K(p \Rightarrow q) \vdash Kq$

A point to be made from all of this is that the approach taken towards relevant alternatives will determine whether and how these principles hold. If one is an attributive[10] or 'radical' contextualist as Dretske puts it,[11] then they posit an interpretation of the relevant alternatives idea whereby each of the premises involved in the evaluation of a knowledge/information argument shares the same set of relevant alternatives. In such a case, the KK principle and closure under

---

[10] Patrick Rysiew, "Epistemic Contextualism," *The Stanford Encyclopedia of Philosophy*, 2011, http://plato.stanford.edu/archives/win2011/entries/contextualism-epistemology/, last accessed 6 September 2013.

[11] Dretske, "Externalism and modest contextualism."

known implication will be valid. It is for relevant alternatives theorists such as Dretske, for whom the set of alternatives can differ between propositions,[12] that these principles are rejected.

It is also the case for Dretske that the information carried by a signal for an agent is in part determined by what they already know and what relevant alternatives they are in an epistemic position to rule out. With regards to the formalised arguments above, it is here that we can also see an important connection between $\neg K[B\,is\,reliable]$ and $\neg B[Kp]\,is\,reliable$. If the zoo goer in the zebra-mule scenario does not know that their belief that z is reliable in the sense that they are not in a position to rule out irrelevant alternatives such as 'mule', then their belief that Kz will not be reliable in the sense that it is not caused by the information that Kz. Thus we might posit the following principle: $\neg K[B\,is\,reliable] \Rightarrow \neg B[Kp]\,is\,reliable$.

As long as the standards for Kp are higher than the standards for p, the KK principle can fail. But if one comes to know p by meeting the standards of Kp, then unless some reason other than the externalist one covered here intervenes both Kp and KKp will be true. For example, in the standard zebra-mule scenario suppose that the relevant alternatives for z consist of all standard zoo scenarios and the relevant alternatives for Kz consist of all scenarios in which any animal is in the enclosure. Within these parameters whilst a 'zebra' visual signal does not suffice for KKz, if the method used to determine the type of animal was a DNA test instead, then the information carried by this result would be enough for both Kz and KKz.

Thus this particular externalist idea of KK principle failure relies on a variation of relevant alternative sets for the first-level proposition p and the second-level proposition Kp. In certain applications both sets could be limited to alternatives, such as those in the zebra example in the previous paragraph, that have a distinguishing set of information and can in practice or principle ultimately be uniquely determined. In this way KKp could be met. On the other hand, if the relevant alternatives for Kp include extreme skeptical alternatives such as brain-in-a-vat scenarios, then KKp will never be met (unless we find a way to obtain information that rules out such scenarios!).

---

[12] Steven Luper, "The Epistemic Closure Principle," *Stanford Encyclopedia of Philosophy*, 2010, http://plato.stanford.edu/entries/closure-epistemic/#CloFaiRelAltApp, last accessed 6 September 2013.

MISFIRED SLINGSHOTS: A CASE STUDY ON THE CONFUSION OF METAPHYSICAL AND SEMANTIC CONSIDERATIONS

Andrew McFARLAND

ABSTRACT: Most philosophers today will acknowledge the pitfalls of confusing metaphysical and semantic issues. Many are also familiar with the classic semi-formal argument that has come to be known as 'the Slingshot' and the various philosophical ends to which this argument has been deployed. The combination of the argument's relatively simple theoretical machinery and its wide range of applications make it ripe for abuse. The slingshot was originally conceived as a semantic argument about designation; what it suggests, but does not prove, is that the closest analogue to singular term reference for any expression is that expression's semantic extension. In order to derive more metaphysically robust conclusions, however, many classical deployments of the argument make use of several methodologically suspicious tactics. By cataloguing the more frequent abuses of the argument, we may remind ourselves of a valuable philosophical lesson.

KEYWORDS: the Slingshot argument, facts, common nouns, semantic extension, referent, metaphysics

1. Introduction

There is a familiar argument whose formal presentation is due originally to Alonzo Church[1] and independently to Kurt Gödel[2] that has been used for a number of philosophical purposes. In its first form, in his review of Carnap's *Introduction to Semantics*,[3] Church presented the argument as a rigorous proof against the view that sentences refer to propositions.[4] Others, like Donald Davidson[5] and W.V.

---

[1] Alonzo Church, "Review of *Introduction to Semantics* by Rudolf Carnap," *Philosophical Review* 52 (1943): 298–304, Alonzo Church, *Introduction to Mathematical Logic: Volume 1* (Princeton: Princeton University Press, 1956), 24–25.

[2] Kurt Gödel, "Russell's Mathematical Logic," in *The Philosophy of Bertrand Russell*, ed. Paul Arthur Schillp (Evanston and Chicago: Northwestern University Press, 1971). (Originally published in 1944).

[3] Rudolf Carnap, *Introduction to Semantics* (Cambridge: Harvard University Press, 1959). (Originally published in 1942).

[4] Church, "Review of *Introduction to Semantics*."

Andrew McFarland

Quine,[6] have employed similar arguments in attempts to undermine various philosophical theses. If successful these arguments would have startling results. First, all true sentences corefer (and so too with all false sentences); second, if sentences refer to facts, then there is but one "Eleatic" fact; and third anyone who has a true belief, believes everything that is true (and similarly anyone who believes anything false, believes everything false) or finally four, that all true sentences are necessarily true. Unsurprisingly, this argument has gone by several different names: the Frege-Church-Gödel argument – as the argument has been thought to have its roots in the work of Frege – collapsing arguments, and perhaps most famously the Slingshot.

The most comprehensive work to date on slingshot-arguments is Stephen Neale's *Facing Facts*,[7] which discusses some of the more prominent versions of slingshots as well as their philosophical significance, paying special attention to a version Gödel outlines in discussion of Russell's theory of descriptions. Ultimately, Neale concludes with Gödel that in order to avoid the argument's intended conclusion of a metaphysical "collapse" of all facts into one, one must "give up either (i) an intuitive and straightforward Fregean Principle of Composition or (ii) the idea that definite descriptions are expressions that purport to stand for things."[8] Since giving up compositionality would appear too high a price to pay, one can give up the view that definite descriptions refer. In other words, if one adopts a Russellian treatment of definite descriptions, a theory for which there is, as Neale maintains, independent motivation to accept, then one has the means to avoid Eleatic metaphysical collapse.[9]

[5] Donald Davidson, "The Logical Form of Action Sentences," in Donald Davidson, *Essays on Actions and Events*, 2nd edition (Oxford: Clarendon Press, 2001) (Originally published in 1967), "True to the Facts," and "Truth and Meaning," both in Donald Davidson, *Inquiries into Truth and Interpretation*, 2nd edition (Oxford: Clarendon Press, 2001) (Originally published in 1969, respectively 1967).

[6] W.V. Quine, *Word and Object* (Cambridge: MIT Press, 1960), W.V. Quine, "Reference and Modality," in his *From a Logical Point of View*, 2nd ed. (Cambridge: Harvard University Press, 1961) (Originally published in 1953), W.V. Quine, "Three Grades of Modal Involvement," in his *The Ways of Paradox* (Cambridge: Harvard University Press, 1975), 177–187 (Originally published in 1953).

[7] Stephen Neale, *Facing Facts* (Oxford: Oxford University Press, 2001).

[8] Neale, *Facing Facts*, 128.

[9] Recall that Russell gives definite descriptions of the form, $\ulcorner$the $\phi\urcorner$, a quantificational analysis. Thus a Russellian analysis of "the author of *The Brothers Karamozov*" will have the following structure:

$$(\exists x)[\mathrm{K}x \,\&\, (\forall y)(\mathrm{K}y \rightarrow x{=}y)]$$

Although much ink has been spilled discussing various deployments of the slingshot, this now classic argument is widely abused. The Slingshot was so-dubbed by Barwise and Perry[10] due to its relatively simple philosophical machinery, and its apparent giant-slaying abilities. What's more, the combination of its simplicity and its quasi-formal character make the slingshot ripe for abuse. Careful scrutiny, however, reveals several highly methodologically dubious strategies, which I will detail below, that significantly weaken prominent slingshot deployments. The Slingshot suggests, but does not prove, that the closest analogue to singular term reference for any expression is that expression's semantic extension. In order to derive more metaphysically substantive conclusions, many famous slingshots (i) provide unsound arguments; (ii) build substantive metaphysical premises into their assumptions; or (iii) invoke an abductive assumption at the end, which is then expected to carry serious theoretical weight. This last move, (iii), is especially interesting since few (if any) seem to acknowledge its role, despite the fact that both Frege and Church explicitly invoke it. Moreover, the move can be invoked in a less metaphysically suspicious way, i.e. the way that Frege and Church use it.

My plan will be as follows. In section 2, in order to understand the original Fregean-inspired motivations behind the slingshot, it will be useful to examine it in its earliest forms; doing this will require briefly going back to Frege, since Church, Gödel, Quine and Davidson, like so many were inspired by his work. In section 3 I shall examine several contemporary slingshots, one proposed by Quine,[11] one by Donald Davidson,[12] and a more recent one discussed by Nathan Salmon.[13] Section 4 is devoted to a brief discussion of various principles of substitution. Finally, in section 5, I will catalogue a number of mistakes commonly employed in slingshot-style argumentation with an aim to reminding us of a larger philosophical lesson: to use caution when drawing metaphysical conclusions from linguistic arguments. Although this lesson has been taught before, slingshots in

---

This can be read as saying that there is one and only one individual who authored *The Brothers Karamozov*, where 'K' stands for the predicate 'is author of *The Brothers Karamozov*.' The reason why Gödel thought the slingshot's conclusion could be avoided with a Russellian theory is that this quantificational analysis doesn't treat descriptions as genuinely referring expressions. And since the slingshot relies on the assumption that one may substitute constituent coreferring singular terms *salva designate*, such a substitution will count as illegitimate by Russell's lights.

[10] Jon Barwise and John Perry, "Semantic Innocence and Uncompromising Situations," *Midwest Studies in the Philosophy of Language* VI (1981): 387–403.

[11] Quine, "Reference and Modality."

[12] Davidson, "True to the Facts," Davidson, "Truth and Meaning."

[13] Nathan Salmon, *Reference and Essence*, 2nd ed. (Amherst: Prometheus Books, 2005).

the cases I discuss are illustrative examples of the mistake of confusing language and metaphysics.

2. Historical Background: Frege, Church, and Gödel[14]

2.1. *Sinn* and *Bedeuntung* and Frege's "Slingshot"

The extent to which there is anything one can call a formalized slingshot in the work of Frege is somewhat controversial.[15] Regardless of whether we can correctly attribute to Frege a slingshot argument, it will be worthwhile – especially since both Church and Gödel took themselves to be articulating Frege's implicit reasoning – if we examine his motivations.

Recall that Frege's solution to his now eponymous puzzle hinged on distinguishing between the *sense* of a singular term from its *reference*, and that by positing these two distinct semantic values for the class of singular terms, it would appear quite natural to extend this distinction to other expressions as well.[16] For

---

[14] I shall use the terms "referent" and "designation," along with their plural versions, interchangeably throughout.

[15] As one (re)reads *On Sense and Reference*, one does get a feeling of sorts that a slingshot-style argument is precisely what Frege had in mind, although the question of whether there is an actual slingshot argument in that essay is a question for Frege scholarship. Answering this question definitively will have no effect on the outcome of the current essay.

[16] The distinction between *sense* and *reference* makes its first appearance in "Function and Concept" (1891), and was later developed in more detail in "On Sense and Reference" (1892). In the former, the distinction gets invoked in connection with discussions of mathematical statements concerning the identity sign, "=." At the time, there was debate about how to interpret the identity symbol for an expression like "2 + 5 = 3 + 4," where some favored the view that this sentences expresses an *equality*, but not a strict *identity*. Frege explicitly disagrees with this assessment, claiming that the expressions flanking the identity sign both designate one and the same thing, the number seven, though the thing signified is presented, or picked out differently by the two expressions. According to Frege, those favoring the equality interpretation of the identity sign confuse the "sign and thing signified." (Gottlob Frege, "Function and Concept," in *Translations from the Philosophical Writings of Gotlobb Frege*, eds. Peter Geach and Max Black (Oxford: Basil Blackwell, 1966), 22) It is as if "one wanted to regard the sweet-smelling violet as different from *Viola odorata* because the names sound different. Difference of sign cannot by itself be a sufficient ground for difference of the thing signified." (Frege, "Function and Concept," 22). Thus Frege makes a distinction between the number being picked out – in this example the number seven – and the way that that thing is determined by an expression like "2 + 5." The former Frege calls the *reference* (*Bedeutung*). The latter, the "mode of presentation," he calls the *sense* (*Sinn*). For more on the history of this particular topic, see Kevin C. Klement, *Frege and the Logic of Sense and Reference* (New York: Routledge, 2002).

example, does the sense/reference distinction hold for larger expressions, say, for entire declarative sentences? In response, Frege says that a whole sentence may be regarded as a name and that each sentence contains what Frege calls "a thought [*Gedanke*]."[17] Tabling the question of what exactly Fregean thoughts are for a moment, one might then pose the question of whether a thought is the sense or the reference of a sentence. Here, it is worth quoting Frege in full:

> Let us assume for the time being that the sentence has reference. If we now replace one word of the sentence by another having the same reference, but a different sense, this can have no bearing upon the reference of the sentence. Yet we can see that in such a case the thought changes; since, e.g., the thought in the sentence 'The morning star is a body illuminated by the Sun' differs from that in the sentence 'The evening star is a body illuminated by the Sun.' Anybody who did not know that the evening star is the morning star might hold the one thought to be true, the other false. The thought, accordingly, cannot be the reference of the sentence, but must rather be considered as the sense.[18]

This passage is important for several reasons: first, we're given a couple of crucial assumptions for Frege's argument about the referents of sentences. Second and more importantly, these two assumptions are also invoked in Church's slingshot appearing in 1943.[19] Third, using these assumptions, Frege argues for the conclusion that the thought of a sentence cannot be its referent, but must be the sense. Let's name the two assumptions just mentioned as follows:

> *SR*: Sentences are referring expressions

> *SUB*DES: The referent of a compound referring expression – not containing devices like quotation marks or "believes that" – is preserved when a component referringexpression is replaced by another with the same referent.

The first assumption reflects Frege's tentative proposal to accept that sentences have references, while the second assumption reflects important rules of substitution for singular terms. SUBDES is therefore a generalized principle of substitution for compound expressions, an analogue to similar principles regarding the substitution of coreferring singular terms. The qualification in SUBDES is important both from a logical perspective and an historical one, for even Frege is careful to remark on the importance of the principle's restricted application for nonextensional contexts: "Exceptions are to be expected when the whole sentence or its part is direct or indirect quotation; for in such cases, as we have seen, the

---

[17] Gottlob Frege, "On Sense and Reference," in *Translations from the Philosophical Writings of Gotlobb Frege*, 62.

[18] Frege, "On Sense and Reference," 62.

[19] Church, "Review of Introduction to Semantics."

words do not have their customary reference."[20] One other notable remark about Frege's motivations (and so too with Church) is that Frege sought a *theoretical analogue* to singular term reference for the class of expressions larger than proper names. Calling this an analogue is appropriate, since Frege recognized the counterintuitive nature of positing sentence referents when he asks us to "assume for the time being that the sentence has a reference."[21]

Now let us examine Frege's reasoning for thinking that the referent of the sentence must be either the *True* or the *False*. Informally, it seems that Frege's argument goes something like this.[22]

(a) *SR*: Sentences are referring devices. Sentences *qua* singular terms may be regarded as proper names, and like proper names have both a sense and a reference.

(b) *Compositionality*: the referent of a complex referring expression is a function of the referents of its parts, whereas the sense of a complex referring expression is a function of the senses of its parts.[23]

(c) *SUBDES*: Given (b), the referent of a compound referring expression – free from devices like quotation or "believes that" – is preserved when a component referring expression is replaced by another with the same referent.

(d) The referent of a sentence cannot be the thought (or proposition) it expresses since the thought (proposition) expressed by 'The morning star is a body illuminated by the Sun' differs from the thought expressed by 'The evening star

---

[20] Frege, "On Sense and Reference," 65.

[21] Frege, "On Sense and Reference," 62.

[22] The following argument is a reconstruction developed from Anthony C. Genova, "How Wittgenstein Avoids the Slingshot," *Journal of Philosophical Research* (2001): 1–22 and through personal correspondence with him.

[23] Following common parlance I use the term 'compositionality' in discussion of the Fregean notion. Care should be taken to avoid confusion with other, broader characterizations of compositionality in the philosophy of language and mind. For example, what's often called "The Principle of Compositionality" gets articulated in the following way: the meaning of a complex expression is a function of its meaningful constituents [morphemes] and its syntactic structure." The similarity between this broad notion of compositionality and the idea of Frege's is evident, but the former discusses the relationship between expressions and meanings more generally, while the latter specifically concerns itself with the determination of the two semantic values of *sense* and *reference* and their relationship between respective bits of language. See Zoltán Gendler Szabó, *Problems of Compositionality* (New York: Garland Publishing, 2000) and Jerry Fodor, *The Compositionality Papers* (Oxford: Oxford University Press, 2002) for discussions of the Principle of Compositionality in the philosophy of language and the philosophy of mind respectively.

is a body illuminated by the Sun'. This is so given (c) and because 'The morning star' and 'The evening star' are coreferential.

(f) Whatever the referent of a sentence is, it must remain the same across coreferential substitutions – this includes subsentential expressions as well as full sentences themselves. Sentences with the same truth-value will be substitutable across all extensional contexts.

(g) *IBE_TV*: consequently, since the only semantically relevant thing about sentences that remains unchanged across substitution of coreferring expressions (in extensional contexts) is the truth-value, what else but the truth-value could be the referents of sentences?

The remaining discussion in "On Sense and Reference" focuses on various examples testing Frege's thesis about the referents of sentences: the rather controversial initial assumption that sentences are similar enough to names to warrant applying the sense/reference distinction; an assumption about substitution to preserve reference given a reasonable principle of compositionality; and importantly even the makings of something similar to an abductive premise, which I label '*IBE_TV*.' We can see that most of the assumptions required for a slingshot-style argument to proceed may plausibly be found within Frege's reasoning. It is important to bear these Fregean considerations in mind when examining Church's argument.

2.2. Church's Slingshot[24]

Let's turn to Church's argument appearing in his 1943 review of Carnap's *Introduction to Semantics* (appearing in the same year). This slingshot was a response to Carnap's break with the Fregean view that sentences designate truth-values, opting instead for the alternative thesis that sentences designate propositions. Much of the argument, as should now be apparent, is inspired by Fregean considerations, in particular the recognition that Frege sought a theoretical analogue to singular term reference when constructing his own argument. Church's argument requires four relatively simple assumptions.

---

[24] My reconstruction of the Church argument is a version of what Tyler Burge calls a "standardized form," though Burge notes that the argument "has a number of interesting variants, and … even more uses" (Tyler Burge, "Frege on Truth," in *Frege Synthesized*, eds. Leila Haaparanta and Jaakko Hintikka (Boston: D. Reidel Publishing Company, 1986), 108). However, later Burge notes that it is unlikely that Frege was giving an "elliptical" version of the Church-Gödel argument, since (by Burge's interpretation) Frege "invokes the normative foundations of logic and the normative roots of the primacy of sentences in logical theory in arguing for this conclusion … The Church-Gödel argument makes no such appeal …" (Burge, "Frege on Truth," 109).

SUB$_{DES}$: The referent of a compound referring expression – free from nonextensional devices like quotation marks or "believes that" – is preserved when a component   referring expression is replaced by another with the same referent.

SR: Sentences are referring expressions

ST: A definite description ⌜the ϕ⌝ refers to the only individual that satisfies the formula ϕ, if there is exactly one such individual and refers to nothing otherwise.

LED: Referring expressions that are logically equivalent to one another refer to the same thing.

I'll quickly remark on the two new assumptions. One may regard *ST* as an analogue of the Fregean assumption that sentences are complex names, but in the current case the expressions with which we are concerned are definite descriptions. Note first that this assumption is in keeping with Frege's treatment of descriptions as referring devices, but second that the assumption is contrary to a Russellian descriptional theory – one that treats descriptions as quantificational rather than referential devices. Now consider the assumption I label *'LED.'* Let us say that two singular terms α and β are logically equivalent if and only if ⌜α = β⌝ is logically true.[25] By this principle, logically equivalent referring expressions, whether names or definite descriptions, corefer.

Now consider the proof.

1. Assume SR and consider any two arbitrarily chosen sentences with the same truth-value, *S* and *S′* (e.g. *S* can be "Washington D.C. is the capital of the United States," while *S* may be "Aristotle founded the Lyceum.") (Note: the symbol '∅' stands for the null set).

(a) S

(b) $\{x: x = x \,\&\, \text{not-}S\} = \varnothing$

(c) $\{x: x = x \,\&\, \text{not-}S'\} = \varnothing$

(d) S'

2. (a) and (b) refer to the same thing (by *LED*)

3. (c) and (d) refer to the same thing (by *LED*)

---

[25] For example, consider 'The president's dog' and 'The president's self identical dog,' both of which are logically equivalent; by *LED* these are also coreferential. David Kaplan remarks that this is a "seemingly gratuitous assumption" (David Kaplan, *Foundations of Intensional Logic*, Ph.D. thesis (University of California, Los Angeles, 1964), 13), though some have disagreed. While rejecting *LED* is one way to avoid the slingshot's conclusion, I wish to focus on other ways the slingshot can misfire, though my inclination is that the burden is on challengers of *LED*.

4. (b) and (c) corefer (by *SUB_DES* and *ST*)

5. So (a) and (d) refer to the same thing (1, 2, and 4)

6. So all true sentences refer to the same thing (1 – 5 and Universal generalization)

We could next run a similar argument, but this time instead of true sentences for *S* and *S′* we would use the falsities "New York is the capital of the United States" and "Plato founded the Lyceum" respectively. The result would be that all false sentences corefer. Thus, two quick, relatively simple slingshots get us the following *semantic* thesis: all true sentences refer to the same thing, while all false sentences refer to the same thing.

Notice that in order to draw a metaphysical conclusion about what the referents of sentences *are* – say the Fregean position that sentences refer to truth-values – one will need a further premise, one perhaps similar to the one briefly mentioned above, along the following lines:

> *IBE_TV*: The best explanation to account for the result of the slingshot is that all true sentences refer to a single unique entity *the true*, while all false sentences refer to a single unique entity, *the false*.

Without *IBE_TV* the slingshot is restricted to a more modest semantic thesis, namely that coextensional expressions are also codesignative, while remaining silent on the ontological question of what the referents of true and false sentences are. And though the this premise itself seems pretty reasonable, as far as the slingshot itself is concerned, one might take the argument to give equal support to the claim that all true sentences refer to say the number one, while all false sentences refer to the number zero or the null set.[26]

2.3. Gödel's Slingshot

Gödel's slingshot,[27] like Church's, explicitly makes use of three assumptions similar to the one's invoked by both Church and Frege. Those assumptions are as follows:

> *Referential Compositionality* – "the signification of a composite expression, containing constituents which have themselves a signification, depends only on the signification of these constituents (not on the manner in which this signification is expressed)"

---

[26] To this end one might construct a similar abductive alternative: *IBE_NO*: The best explanation to account for the result of the slingshot is that all true sentences refer to a single unique entity *the number one*, while all false sentences refer to a single unique entity, *the null set*.

[27] Gödel, "Russell's Mathematical Logic."

> $LED^*-\ulcorner\phi\alpha\urcorner$ and $\ulcorner\alpha$ is the object which has the property ϕ and is identical with $\alpha\urcorner$ (in Gödel's words) "means the same thing." [sic][28]

> SR^*: "Every Proposition 'speaks about something', i.e., can be brought to the form $\phi(\alpha)$. Furthermore one would have to use the fact that for any two objects a, b, there exists a true proposition of the form $\phi(a, b)$ as, e.g., $a \neq b$ or $a = a . b = b$".[29]

Though Gödel's wording deviates from Church's language, many of his assumptions do the same theoretical work. First, the assumption above labeled "Referential Compositionality" presumably does the work of Church's substitution assumption. Second, Gödel's assumption LED^*, though seemingly worrisome, is likely best interpreted along the lines of LED, that logically equivalent referring expressions refer to the same thing.[30] Third, I take SR^* as a syntactic articulation doing the work of SR above. Fourth, just as Church requires an assumption about the uniqueness of referents for definite descriptions, so too with Gödel. Finally, just as Church's slingshot needs a further premise to get to an ontological conclusion about truth-values, so does Gödel, though Gödel's language is careful to conclude only that "all true sentences have the same signification (as well as all the false ones)."[31] Gödel leaves readers to piece through the argument for themselves (as the hints of the argument are made only in a footnote), though the reasoning runs the same as the Church deployment. Let us now turn to more contemporary slingshot deployments. As we will see, these slingshots draw far more robust metaphysical conclusions.

---

[28] Gödel, "Russell's Mathematical Logic," 122. Neale remarks that "[LED^*] is less worrying than Gödel's wording might suggest" (Neale, *Facing Facts*, 130), though he does not go on to say why he thinks this. Later, Neale writes, "An examination of the main text … might suggest that [Gödel] intends 'signify the same thing.' Whatever Gödel's intention, for the purposes of this argument I shall attribute to him, it is both sufficient and necessary that if descriptions are singular terms that simply stand for things, then $\ulcorner\phi\alpha\urcorner$ and $\ulcorner\alpha$ is the object which has the property ϕ and is identical with $\alpha\urcorner$] stand for the same fact" (130). I disagree with Neale's latter characterization, mainly because of the use of the phrase "stand for the same fact," for Gödel never uses the term "fact." I find it far more plausible to interpret the phrase "means the same thing," along the lines of Neale's earlier proposal as "signify the same thing," since this is Carnap's characterization of synonymy (see Carnap, *Introduction to Semantics*, 55) and it's reasonable to think that Gödel, having published his article originally two years later in 1944, was familiar with this fact.

[29] Gödel, "Russell's Mathematical Logic," 122.

[30] Cf. footnote 28 above.

[31] Gödel, "Russell's Mathematical Logic," 122.

3. Quine, Davidson, and Common-Noun Slingshots[32]

There are important distinctions between the slingshots presented by both Church and Gödel and those I am about to discuss. One key difference is one of presentation, as both Quine and Davidson present their arguments in the form of a *reductio*. Quine's targets are purportedly "opaque contexts," i.e. intensional contexts involving words like "necessarily" and "possibly," and hyperintensional contexts involving propositional attitude terms like "believes," "wishes," "wants," etc.[33] Davidson's various slingshot deployments serve several different purposes, the most widely discussed being his famous argument against traditional correspondence theories of truth that make use of facts as truth-makers for sentences. Let's begin with the assumptions Church uses in his version of the argument – SUB_{DES}, SR, ST, and LED – and turn first to discussion of Quine.[34]

3.1. Quine's Slingshots

Quine has several slingshots, one in "Reference and Modality" in the first edition of *From a Logical Point of View*, and in his "Three Grades of Modal Involvement," while another occurs in 1960 in *Word and Object*. All versions of the argument may be interpreted to the same end: to cast doubt on the intelligibility of ostensibly nonextensional contexts. I shall focus on the one from 1953 in "Reference and Modality."

---

[32] One other slingshot receiving little to no discussion in the literature may be found in footnote 71 of David Kaplan's essay on demonstratives from 1977 (David Kaplan, "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals," in *On Sense and Direct Reference*, ed. Matthew Davidson (Columbus: McGraw-Hill, 2007). This slingshot deployment like those of Quine and Davidson takes the form of a *reductio* with the aim of showing that certain substitution moves are not legitimate for *pseudo de re* contexts.

[33] Quine appears to use the term "intensional" to describe both modal and hyperintensional contexts alike, though overall it appears he is concerned with contexts that appear to result in some failure of substitution. I will simply use the expression "nonextensional" when referring to such contexts.

[34] Krüger, like Barwise and Perry ("Semantic Innocence and Uncompromising Situations"), agrees that Davidson's argument does not succeed in refuting the correspondence theory of truth. However, Krüger denies that slingshot deployments represent "… a unified tradition launched by Frege … whose representatives have supposedly all lost their 'semantic innocence' …" (Lorenz Krüger, "Has the Correspondence Theory of Truth Been Refuted? From Gottlob Frege to Donald Davidson," in *Why Does History Matter to Philosophy and the Sciences? Selected Essays*, eds. Thomas Sturm, Wolfgang Carl, and Lorraine Daston (Berlin: Walter de Gruyter, 2005), 202). (Originally published in 1995).

More generally, Quine claims that his slingshot allows him to make a more "sweeping observation," that "*any* mode of statement composition other than the truth functions, is referentially opaque."[35] However, in order for this conclusion to be plausible, Quine needs two variants on the principles of Frege and Church:

SUB$_{TV}$: The truth-value of an expression is preserved when a component expression is replaced by another with the same truth-value.

LES: Logically equivalent expressions may be substituted in all contexts *salva veritate*.

The other assumption needed is *ST*, which for Quine's purposes does not depart significantly from the one stated above. Quine's aim is to show that *SUB$_{TV}$* does not extend to expressions containing non-truth-functional operators that still allow *LES*. Restated in Quine's own words, his aim is to show that non-truth-functional expressions are "referentially opaque." It is important to notice that *SUB$_{TV}$* omits the extensional restriction, as this will be a bone of contention I discuss later in this essay.[36] So let Φ be a purportedly non-truth-functional expression such as 'necessarily' or 'possibly,' and assume for *reductio* that *SUB$_{TV}$* applies to sentences containing Φ, for then it will follow that Φ is truth-functional, a presumably unacceptable result. So let S and S' be any two arbitrarily chosen sentences alike in truth-value, say 'Hesperus is Hesperus' and 'Quine was born in Akron.' Suppose further that Φ is 'necessarily.' Then if $\Phi(S)$, then $\Phi(S')$, or if 'Necessarily, Hesperus is Hesperus,' then 'Necessarily, Quine was born in Akron.'

1. $\Phi(S)$ Assumption

2. $\Phi(\{x\colon x = x \,\&\, S\} = \varnothing)$ 1., *LES*

3. $\Phi(\{x\colon x = x \,\&\, S'\} = \varnothing)$ 2., *SUB$_{TV}$*, *ST*

4. $\Phi(S')$ 3., *LES*

Presumably this conclusion is an unacceptable result unless, as Quine states, "the context represented by 'Φ' is referentially opaque."[37] Referential opacity, according to Quine, is problematic since problems arise when one attempts to quantify into nonextensional contexts.[38] Thus, Quine's reasoning appears to be

---

[35] Quine, "Reference and Modality," 159. Emphasis is Quine's.

[36] See Section 5.

[37] Quine, "Reference and Modality," 159.

[38] The argument for why quantifying into nonextensional contexts is unintelligible is hinted at in Quine, "Reference and Modality," but is spelled out in more explicit detail in W.V. Quine, "Quantifiers and Propositional Attitudes," *Journal of Philosophy* 53 (1956): 177–187.

that if we treat purportedly nonextensional expressions as truth-functional, then a slingshot delivers unacceptable results, while if we accept nonextensional expressions as indeed nonextensional, we risk the incoherence of quantifying into them.[39]

As we saw above, in order for this variant of the slingshot to work, Quine must modify the assumptions used by Church and Frege. Although Quine's principles are quite close to the versions employed by Frege and Church, these are concerned with truth-preservation rather than preservation of reference. Thus, Quine's slingshot, with its two variants of the principles discussed earlier, *SUB*TV and *LES*, its *reductio* form, and the conclusion it draws, represents a significant departure from the original reasoning employed in Frege, Church, and Gödel. We will return to these differences later. For now, let us turn to discussion of Davidson.

3.2. Davidson

To my knowledge there are at least three separate slingshots in the work of Davidson, all of which share an affinity with Quine's formulations in taking the characteristic *reductio* form. The first two appear in 1967. One appears in "Truth and Meaning"[40] where Davidson uses the slingshot as a way to show that expressions cannot refer to their meanings since the slingshot would show that all expressions with the same semantic extension end up having the same meaning, a clearly and unacceptable result. The other in "The Logical Form of Action Sentences"[41] raises an objection to Reichenbach's[42] analysis of the logical form of action sentences. According to Davidson, if one adopts Reichenbach's proposal, a quick slingshot shows that there is but one event, and so by *reductio* Reichenbach's view must be false. Finally, perhaps the most famous of the three is the slingshot in "True to the Facts" where Davidson uses the slingshot to object to Correspondence theories of truth that make use of facts as the truth-makers of sentences. If sentences designate facts, Davidson maintains, then all true sentences designate the same fact. Since this Eleatic conclusion is unacceptable, the initial

---

[39] The Kronecker δ version from Quine (*Word and Object*, 148–149) is employed much to the same effect as the earlier 1953 version, but this time using propositional attitude contexts instead. So, where Φ is a propositional attitude operator such as 'believes that,' the result would be that the subject of the sentence ends up believing everything. Unless of course 'believes that' is referentially opaque.

[40] Davidson, "Truth and Meaning," 19.

[41] Davidson, "The Logical Form of Action Sentences," 117–118.

[42] Hans Reichenbach, *Elements of Symbolic Logic* (New York: The Macmillan Company, 1947).

presupposition about sentences designating facts must be false, *reductio ad absurdum*.[43]

Davidson begins by asking when statements with the following form hold:

(3) the statement that *p* corresponds to the fact that *q*

His response is as follows:

> Certainly when "*p*" and "*q*" are replaced by the same sentence: after that the difficulties set in. The statement that Naples is farther north than Red Bluff corresponds to the fact that Naples is farther north than Red Bluff, but also it would seem, to the fact that Red Bluff is farther south than Naples (perhaps these are the same fact). Also to the fact that Red Bluff is farther south than the largest Italian city within thirty miles of Ischia. When we reflect that Naples is the city that satisfies the following description: it is the largest city within thirty miles of Ischia, and such that London is in England, then we begin to suspect that if a statement corresponds to one fact, it corresponds to all.[44]

Next Davidson turns to spelling out in more formal fashion the reasoning employed in the above excerpt. The two assumptions provided are as follows:

(4) the sentences that replace 'p' and 'q' are logically equivalent

(5) 'p' differs from 'q' only in that a singular term has been replaced by a coextensive singular term

Although these suppositions as stated are not themselves principles, it's easy enough to turn them into principles similar enough to the ones already discussed. Thus (4) appears to do the work of the assumption above *LED*, namely that logically equivalent referring expressions corefer, while it's likely that (5) does the work of *SUB_DES*. Though not mentioned, presumably Davidson also needs something like *ST*, and the assumption that descriptions of the form ⌜the fact that ϕ⌝ are designators. Now consider any two arbitrarily chosen sentences that have the same truth-value, *S* and *S′*, and consider the following slingshot:

1. the fact that *S*

2. the fact that $(\iota x)(x = \text{Diogenes} \ \& \ S) = (\iota x)(x = \text{Diogenes})$

3. the fact that $(\iota x)(x = \text{Diogenes} \ \& \ S′) = (\iota x)(x = \text{Diogenes})$

---

[43] While the first slingshot in "Truth and Meaning" is interesting within the context of that essay, since the conclusion reached is far more plausible than Davidson's other uses of the argument, namely that there must be more to meanings than extension, discussion on this version can for the most part be ignored. What's more, this version is put forth as a formalized version of Frege and Church's arguments. For my purposes, however, it's the questionable cases that are of interest.

[44] Davidson, Donald "True to the Facts," 41–42.

4. the fact that S

Davdison's reasoning then proceeds in typical slingshot fashion: 1 and 2 Are codesignative by *LED*; 3 is codesignative with 2 by *SUB_{DES}* and *ST*; 3 and 4 are codesignative by *LED*. Thus, 1 and 4 are codesignative. If we universally generalize with the result, we get the desired Eleatic conclusion: all descriptions with the form ⌜the fact that φ⌝ designate the same fact.

This concludes the discussion on what we might call the "Classic" slingshot deployments. However, before I turn to the final section, there is one more suspect slingshot to discuss, one proposed by Nathan Salmon[45] concerning the referents of commons noun phrases.

3.3. A Slingshot for Common Nouns[46]

Along these lines, one might construct a slingshot argument in the vein of Church and Gödel to argue for the claim that any two arbitrarily chosen common nouns that happen to have the same extension refer to the very same thing. If sound, the argument would strongly suggest that any two common nouns that happen to have the same semantic extension, e.g. 'Tyrannosaurus rex" and "dodo," refer to the same thing.

The argument requires the same assumptions from the original Church version above, though with two minor changes, *v-SUBS* and *CNR*. The former is simply our good old original *SUB_{DES}* reformulated to accommodate common nouns. The latter just states that common noun phrases are referring expressions. Salmon also employs a device that turns an open sentence into a common noun phrase using the phrase "thing which is such that."[47]

---

[45] Salmon, *Reference and Essence*.

[46] This sub-section considers argument one might propose to support the view that Platonic natural kinds are individuated by their metaphysical extensions, a project I discuss in future work. The argument for this view, one I'm calling the *extensional view* of kind individuation, takes the conclusion from a common noun slingshot: that for any arbitrarily selected pair of common nouns with the same semantic extension, those two nouns corefer. Generalizing to other common noun pairs yields the conclusion that common nouns with the same semantic extension corefer, which may then be used to formulate an argument for the individuation of kinds. The specifics of such an argument needn't be articulated in much detail here since my intent is to nip any such argument in the bud by showing how a slingshot of this form fails.

[47] Salmon, *Reference and Essence*, 50–51.

Andrew McFarland

Assumptions[48]

v-SUBS: The referent of a compound referring common noun phrase – free of devices like quotation or "believes that" – is preserved when a component referring expression is replaced by another with the same referent.

CNR: Common noun phrases are referring expressions

ST: A definite description $\ulcorner$the $\phi\urcorner$ refers to the only individual that satisfies the formula ϕ, if there is exactly one such individual and refers to nothing otherwise.

LED: Referring expressions that are trivially logically equivalent to one another refer to the same thing.

Preliminaries

Let v and v' be any two (arbitrarily chosen) common nouns that have the same extension (e.g. '*T-rex*' and 'dodo' or 'molecule of water' and 'molecule of H_2O').

The Argument

1. Assume *CNR* and consider:
    (a) v
    (b) thing x such that $(\imath n)$[if x is a v, then n = 1) & (if x is not a v, then n = 0)] = 1
    (c) thing x such that $(\imath n)$[if x is a v', then n = 1) & (if x is not a v', then n = 0)] = 1
    (d) v'
2. (a) and (b) refer to the same thing by *LED*.
3. (c) and (d) refer to the same thing also by *LED*.
4. $\ulcorner(\imath n)$[if x is a v, then n = 1) & (if x is not a v, then n = 0)]$\urcorner$ and $\ulcorner(\imath n)$[if x is a v', then n = 1) & (if x is not a v', then n = 0)]$\urcorner$ are coreferential by *ST* since both are set equal to 1.
5. So (b) and (c) have the same referent [4., *SUB_DES*].
6. So (a) and (d) have the same referent [2, 3, 5].
7. So any common noun phrases that apply to the very same things (that have the same extension) have the same referent [1 – 6 and Universal Generalization]

Like other slingshots, this conclusion only delivers a semantic conclusion, one telling us only that for any two arbitrarily chosen common noun phrases that happen to have the same extension, those noun phrases have the same referent;

---

[48] This argument is adapted from Nathan Salmon, *Frege's Puzzle* (Atascadaro: Ridgeview Publishing Company, 1991), and Salmon, *Reference and Essence*, 48–52.

the argument itself tells us nothing about what the referents of those expressions are.

It should be noted that Salmon thinks the argument is unsound since line 5 makes an illegitimate appeal to *v-SUBS*: If the argument were sound, so Salmon reasons, the phrases "neighbor of Shakespeare" and "neighbor of England's greatest playwright" would refer to one and the same thing.[49] However, this seems wrong since "it is easy to imagine circumstances in which there are individuals who are of one kind but not the other."[50] Unfortunately, this is the only explanation given in *Reference and Essence* for rejection of *v-SUBS*. Though we are told rejecting *v-SUBS* is the reason for the trouble, I suspect that flat out rejection of the principle is a little too hasty. In fact, I think the reasoning employed in rejecting the sort of substitution move involved in the common noun slingshot has to do with the context created by the phrase "thing *x* such that." Salmon acknowledges this fact in later essays[51] and it's likely the idea he had in mind (albeit implicitly in his earlier work) in rejecting *v-SUBS*.

There is one final topic to address before moving on to criticisms, namely substitution rules. All slingshots require some form of substitution, whether the substitution is intended *salva designate*, to preserve the referent of the larger containing expression, or *salva veritate*. Such a key move needs a brief discussion since the legitimacy of certain slingshot deployments hinges on which substitution moves are acceptable.

4. Substitution Rules[52]

Here I will discuss various substitution principles commonly found in extensional logic. This will be helpful in reminding readers of Frege's and Church's motivations for developing their substitution analogues, provided one is careful to remember that the Frege-Church-Gödel versions were concerned with substitution *salva designate* rather than truth-preservation. As such, certain

---

[49] What I'm calling "*v-SUBS*" Salmon calls the "*Interchange Principle for Common Nouns*" (Salmon, *Reference and Essence*, 52).

[50] Salmon, *Reference and Essence*, 52.

[51] See for instance Nathan Salmon, "The Very Possibility of Language," in his *Metaphysics, Mathematics, and Meaning: Philosophical Papers, Volume I* (Oxford: Oxford University Press, 2001), which I discuss below in Section 4.

[52] This section borrows significantly from Neale, *Facing Facts*, Chapter 7 and I adopt Neale's symbolization in stating the proceeding rules. There is however one addition, *Nu-Substitution*, which is a formal representation of Salmon's *Interchange Principle for Common Nouns*. One other significant departure from Neale is that I prefer not to discuss these as *inference* rules, but rather rules concerning legitimate substitution, *salva designate*.

principles concerned with referent preservation ought therefore to be regarded as a class of principles separate from typical principles of inference. Inference principles, as any student of introductory symbolic logic will attest, *are* truth-preserving rather than referent preserving. In articulating many of the principles found below, however, I shall speak loosely of inference, though it is important to remember that slingshot substitution principles are but *analogues* to principles of inference.

PSME: Principle of Substitutivity for Material Equivalents

$(\phi \equiv \psi)$

$\Sigma(\phi)$

_____

$\Sigma(\psi)$

We can read this as saying that for any two sentences ϕ and ψ sharing the same truth-value, where $\Sigma(\phi)$ is a true sentence containing at least one occurrence of ϕ (in an extensional context), we can legitimately infer $\Sigma(\psi)$, where $\Sigma(\psi)$ is the result of replacing at least one occurrence of ϕ in $\Sigma(\phi)$ by ψ and vice versa. As is familiar, a linguistic context is extensional just in case it permits the substitution of coextensional expressions such that the truth of the larger containing expression is preserved. Consider:

> a) Quine was born in Ohio and Davidson was born in Massachusetts

The binary connective 'and' (along with the other usual truth-functional connectives) is extensional in the sense that it operates on the extensions of sentences, namely their truth-values. Thus, the truth of (a) is preserved if we substitute for either of the sentences flanking 'and' with another true sentence, e.g. "Kripke was born in Nebraska."

PSST: Principle of Substitutivity for Singular Terms

$\alpha = \beta$ or $\Sigma(\alpha)$

$\Sigma(\alpha)$ $\sim\Sigma(\beta)$

_____   _____

$\Sigma(\beta)$ $\alpha \neq \beta$

We can read this principle as saying that if one has two coextensional singular terms α and β, where $\Sigma(\alpha)$ is a sentence containing at least one

occurrence of α (in an extensional context), then one may infer Σ(β), where Σ(β) results from replacing α in Σ(α) by β, and vice versa. So consider:

(b) Venus revolves around the sun

In (b) we may substitute for the proper name 'Venus' the coreferring proper name 'Hesperus' to obtain

(c) Hesperus revolves around the sun

One bit of contention surrounding this principle is the question of what sorts of expressions are included among the class of singular terms. Some, such as Frege, include definite descriptions in this class, while others like Russell did not. Thus, Frege would allow the move from

(d) Hesperus is Phosphorus

to

(d′) The evening star is Phosphorus

provided of course that (d) and (d′) are not themselves embedded in nonextensional contexts. Russell on the other would not treat (d′) as a singular term, asserting instead that it should be given a quantificational analysis (see footnote 1 above).

PSLE: Principle of Substitutivity for Logical Equivalents

$\phi \Leftrightarrow \psi$

$\Sigma(\phi)$

_____

$\Sigma(\psi)$

Let us say that two expressions ϕ and ψ are logically equivalent if and only if the sentence $\ulcorner \phi \Leftrightarrow \psi \urcorner$ is logically true, where '$\Leftrightarrow$' is to be understood as symbolizing dual entailment. So this principle can be read as saying that if two expressions ϕ and ψ are logically equivalent and $\Sigma(\phi)$ is a true sentence containing ϕ as a constituent (in an extensional context), then $\Sigma(\psi)$ will also be true, where $\Sigma(\psi)$ results from replacing ϕ in $\Sigma(\phi)$ with ψ and vice versa. By this principle we can move from

(e) The President's dog is black

to

(e′) The President's self-identical dog is black

This principle is particularly important for slingshot-style arguments, as many have contested the principle's legitimacy in criticizing slingshot deployments.

ι-SUBS: Iota-Substitution (Principle of Substitutivity for Definite Descriptions)

$$\iota x\phi = \iota x\psi \qquad \iota x\phi = \alpha \qquad \iota x\phi = \alpha$$

$$\Sigma(\iota x\phi) \qquad\quad \Sigma(\iota x\phi) \qquad\quad \Sigma(\alpha)$$

$$\underline{\qquad\qquad} \qquad \underline{\qquad\qquad} \qquad \underline{\qquad\qquad}$$

$$\Sigma(\iota x\psi) \qquad\quad \Sigma(\alpha) \qquad\qquad \Sigma(\iota x\phi)$$

For those who accept definite descriptions as singular terms, these substitution rules will be superfluous since PSST already directly licenses these moves. However, for those who adopt a Russellian treatment of descriptions, PSST will not work. Consider

(f) i. Kripke = the greatest philosopher from Omaha

ii. Kripke authored *Naming and Necessity*

_____

iii. The greatest philosopher from Omaha authored *Naming and Necessity*

It's clear that (f) is a valid argument, but if we were to formalize it we cannot use our principle about singular terms, PSST, to make the move (assuming of course a Russellian theory of descriptions):

(f') i. $k = (\iota x)Ox$ premise

    ii. Nk premise

    iii. $N(\iota x)Ox$ i, ii, PSST

Iota-Substitution, however, resolves this problem; rather than using PSST as the justification at iii, one can appeal to ι-SUBS. This triplet of rules says that first, if the unique individual satisfying the constitutive formula ϕ is the same as the unique individual satisfying the constitutive formula ψ, one can substitute $\ulcorner \iota x\phi \urcorner$ for $\ulcorner \iota x\psi \urcorner$ and vice versa. Second and third, if the unique individual satisfying the constitutive formula ϕ is the same individual denoted by the singular term α, then one may substitute (in extensional contexts) $\Sigma(\alpha)$ for $\Sigma(\iota x\phi)$, where $\Sigma(\alpha)$ is the result of replacing α for $\iota x\phi$ in $\Sigma(\iota x\phi)$ and vice versa.

ν-SUBS: Nu-Substitution (Principle of Substitutivity for Common Nouns)

$\alpha = \beta$ or $\tau(\alpha)$

$\tau(\alpha)$ $\tilde{}\tau(\beta)$

_____   _____

$\tau(\beta)$ $\alpha \neq \beta$

In keeping with the Fregean spirit of maintaining the analogy with singular term reference, one might construct a principle for the interchange of common noun phrases. We can read this principle as saying that if two coextensional singular terms α and β (i.e. $\lceil \alpha = \beta \rceil$ is true), where $\tau(\alpha)$ is a common noun phrase containing at least one occurrence of α (in an extensional context), then one may infer $\tau(\beta)$, where $\tau(\beta)$ results from replacing α in $\tau(\alpha)$ by β, and vice versa.

Earlier I mentioned that Salmon[53] suggests a rejection of *ν-SUBS*. However, a flat out rejection of the principle is too quick. The trouble in the common noun slingshot seems to be caused not by the principle of substitution, but by the phrase "thing *x* which is such that," in particular the "that" operator, which Salmon in later essays maintains is arguably nonextensional.[54] If indeed the "that" operator is nonextensional, then slingshots that make the substitution moves for expressions within the scope of these operators have ignored the caveat that substitution is licensed only for extensional contexts. This move is perhaps the most frequently occurring dubious strategy in slingshot deployments.

5. Slingshot Malfunctions

I now turn to discussion of some surprisingly common slingshot malfunctions.

---

[53] Salmon, *Reference and Essence*.

[54] See for example Salmon, *Frege's Puzzle*, 6 and Salmon, "The Very Possibility of Language," 349). One potential hiccup for Salmon's hypothesis is that contexts typically regarded as opaque, e.g. propositional attitude contexts like "believes that" or modal contexts like "It is necessary that," are opaque because of operators like "believes" or "It is necessary" rather than the "that" operator. The reply is that this seems wrong if we consider expressions like "Russell affirmed Logicism," which omits the occurrence of "that" but remains perfectly grammatical. Further, from the truth of this statement one can infer the following: Russell affirmed that mathematics is reducible to logic. That the "that" operator is nonextensional seems to me a plausible hypothesis. But what about expressions like "it's not the case that"? If anything is an extensional operator, surely this is. Salmon's reply (in e-mail correspondence) is that the expression "the case" is synonymous with "true," and that "It is the case that snow is white" is a stylistic variant of "That snow is white is the case." Similarly, "It is not the case that snow is white" is an variant of "That snow is white is not the case," which express the proposition *that snow is white is not true*.

Andrew McFarland

5.1. The Extensional Malfunction

Arguments without the extensional restrictions on substitution go awry in that there is an illegitimate appeal to a substitution principle *very similar* to a legitimate one, although the principle itself in its unqualified form is strictly speaking false.[55] Consider the original formulation of our substitution principle:

> *SUB~DES~*: The referent of a compound referring expression – *free of devices like quotation or "believes that"* – is preserved when a component referringexpression is replaced by another with the same referent.

Preservation of reference for a larger containing expression when substituting coreferring constituent expressions is preserved only when those constituent expressions are not within the scope of ostensibly nonextensional, or *ungerade*, contexts. As I discussed earlier, even Frege himself was careful to include a clause about the principle's inapplicability for cases involving direct or indirect quotation.[56]

But now consider Quine's slingshots in light of this restriction, both of which involve substitution of coreferring expressions under the scope of what Quine claims are purportedly nonextensional operators such as "it is necessary that" or "believes that." Given the extensional restriction on substitution principles, Quine's slingshots might strike one as odd, but recall that *SUB~TV~* as I formulated it did away with this restriction. Also remember that Quine's aim is to show that accepting *SUB~TV~* along with the other assumptions listed earlier delivers the result that supposed nonextensional operators in fact turn out to be extensional, which is evidently unacceptable. Quine's further result, that supposed nonextensional bits of language are simply incoherent, is obviously not delivered by this slingshot, and must be argued for elsewhere.[57]

But the trouble for the Quinean deployments, as well as other slingshots utilizing the *reductio* model, is that the strength of a *reductio* relies on the assumption that all other premises in the argument are true. So imagine that the

---

[55] Philosophers who propose the argument either without or who fail to adhere to the extensional restriction include Barwise and Perry ("Semantic Innocence and Uncompromising Situations"), Davidson ("The Logical Form of Action Sentences," "True to the Facts," "Truth and Meaning"), McGinn (Colin McGinn, "A Note on the Frege Argument," *Mind* 85, 339 (1976): 422–423), Perry (John Perry, "Evading the Slingshot," in *Philosophy and Cognitive Science: Catergories, Consciousness, and Reasoning*, eds. Andy Clark, Jesús Ezquerro, and Jesús M. Larrazabal (Dordrecht: Kluwer, 1996), Quine (*Word and Object*, *From a Logical Point of View*).

[56] Admitedly, neither Church nor Gödel explicitly include a clause restricting the principle to extensional contexts, although neither attempts to use a slingshot with such devices either.

[57] E.g. In Quine ("Quantifiers and Propositional Attitudes").

Quinean supporter replies that the question of which substitution moves are legitimate principles for slingshot-style argumentation is precisely what is at issue, and that deciding which contexts are applicable to various substitutions in advance stacks the deck unfairly against the Quinean argument. However, if one can reject a premise other than the *reductio* premise, and in our current case it is at least *plausible* that the principle that's false is not the *reductio* premise, but the unqualified restriction principle, the whole argument is significantly weakened.

Let's now turn to the Davidson slingshot about facts. Recall that Davidson's goal was to cast doubt on correspondence theories that made use of facts, and that the slingshot supposedly shows that all facts collapse into one great Eleatic fact. However, Davidson's slingshot also contains a questionable context, viz. "the fact that …," arguably nonextensional due to the occurrence of the "that" operator.[58] Since Davidson's slingshot about facts also employs the characteristic *reductio* form, his argument, like the Quinean deployment, is significantly weakened. For the proponent of facts may just as easily reject Davidson's substitution principle while maintaining that sentences still correspond to facts.

Thus, in order to derive the intended disquieting conclusions from these two slingshots, the arguments must include a premise that is at least arguably contestable. This strategy should strike the reader as highly methodologically suspect.

Suppose, however, an advocate of the *reductio*-style slingshot of Davidson's simply grants that the "that" operator is nonextensional. There is another way to read Davidson's slingshot that need not appeal to a contestable, purportedly nonextensional context like "the fact that." Consider the original Church slingshot and its four assumptions. As a reminder:

> *SUB_DES*: The referent of a compound referring expression – free of devices like quotation or "believes that" – is preserved when a component referring expression is replaced by another with the same referent.

> *SR*: Sentences are referring expressions

> *ST*: A definite description ⌜the ϕ⌝ refers to the only individual that satisfies the constitutive "predicate" (or formula) ϕ, if there is exactly one such individual (and refers to nothing otherwise).

> *LED*: Referring expressions that are trivially logically equivalent to one another refer to the same thing.

---

[58] Cf. Krüger, who says that the trouble is caused not by the occurrence of the 'that' operator per se, but the assumption that facts are extensions (Krüger, "Has the Correspondence Theory of Truth Been Refuted," 208).

However, suppose one were to replace *SR* with a variant that we assume for *reductio*

> *SF*: Sentences correspond to facts

If we now run the slingshot as Church did, the Eleatic conclusion appears to follow, and the critic of facts may now claim that *SF* is the premise causing the trouble, for we now no longer have substitutions that take place within a context like "the fact that."

But what this reconstruction of Davidson's slingshot does is suspicious for two reasons. First, it makes the contentious assumption that the correspondence relation is the same as reference.[59] Second, and more importantly for our purposes, it builds into an ostensibly semantic argument a fairly substantive and highly suspicious metaphysical assumption about facts. This strategy represents a significant deviation from the original Frege-Church-Gödel slingshots.

5.2. Dubious Metaphysical Premises

This malfunction involves building dubious metaphysical assumptions into the argument in order to draw what are often shocking metaphysical conclusions. The original slingshot deployments, those by Frege, Church, and Gödel, by themselves established only the semantic thesis that sentences with the same truth-values have the same designation. A more general conclusion may be drawn with the slingshot and expressions other than sentences, namely that the closest thing to singular term reference for *any* expression will be the expression's extension. Any further metaphysical conclusions can be reached only by adding one or more substantive metaphysical assumptions, for example by way of adding an abductive premise, as with Frege and Church. That assumption may look similar to what I earlier called *IBE_TV*.

---

[59] I flag this as contestable move, though for the purpose of this essay, the point is not all that crucial. I would however like to say a little about why collapsing reference to correspondence is at least arguably illegitimate. Historically, both Frege and Mill are notable for distinguishing between the *referent* of an expression on the one hand, and the *content* on the other. Contemporary advocates of Millianism also make a similar distinction, though for certain sorts of expressions – e.g. demonstratives, indexical expressions, and proper names – content and reference are the same. Slingshots, as we've noted, are arguments concerned with the referent, or designata of expressions, rather than their content. What's more, consider the fact that a theory of content, and so too with the notion of correspondence, seeks to preserve certain structural features between an expression, and what that expression is supposed to represent. This mirroring, isomorphism or paralleling, is not something typically associated with reference.

> *IBETV*: The best explanation to account for the result of the slingshot is that all true sentences refer to a single unique entity *the true*, while all false sentences refer to a single unique entity, *the false*.

As briefly mentioned earlier, the conclusion of the slingshot (if sound) seems equally compatible with a variant of *IBETV* where the referents of sentences are say the number one and zero or the empty set. However, recognizing that the slingshot is an argument concerning the relationship between bits of language and their relations to various semantic values, there is reason to favor *IBETV* over a principle invoking numbers since it's difficult to see the initial semantic relevance of numerical entities to whole sentences.[60]

Similarly, the second reconstruction of Davidson's slingshot builds into the assumptions of the argument a rather substantive answer to what the referents of sentences are, namely facts. But this should strike one as a dubious methodological strategy. I can use the slingshot in this way to prove there's only one fact, one truth, one proposition, but only if I build entities into the argument precisely those assumptions needed for the argument to work.

Presumably the rationale behind Davidson's questionable assumption rests on the idea that a correspondence theorist must say that sentences correspond to facts, and that the relation of correspondence is the same as reference. This ignores the possibility of linguistic expressions relating to objects in other ways. For even Frege notes that while the sense of an expression cannot be the reference, we still might say that a sentence *expresses* its sense and *refers* to its referent. This same Fregean strategy – in effect, the loss of one's so-called 'semantic innocence' – is open to one who would like to avoid the Davidson slingshot as well.

6. A Lesson From Misfired Slingshots

I began with an examination of the historical roots of slingshots with the hope that an understanding of the original motivations of the argument's authors would give us insight into later versions of it. I then examined a variety of slingshot deployments with a special focus on those whose conclusions were of significance for metaphysics. Those metaphysically oriented slingshots I argued made use of several methodologically suspect strategies. These dubious strategies came in two main varieties: first, extensional malfunctions as we saw in Quine, Davidson, and one discussed (but not endorsed) by Salmon; and second those deployments whose metaphysical assumptions were to blame for generating the shocking conclusions.

---

[60] Obviously numbers will be relevant if we're inquiring after the referents of numerals.

These suspicious strategies, I argued, make certain slingshot deployments far less metaphysically significant than they might at first appear. What's more, these misfired slingshots malfunctioned precisely because of the confusion of metaphysics and language.

Still, insofar as slingshots are more than merely linguistic, we cannot conclude from this that there are *no* genuinely metaphysical conclusions to be had from primarily linguistic considerations. I have also not argued that the slingshot is of no historical or philosophical significance. To the contrary, the argument is particularly enlightening in that it (i) gives us very strong (though not decisive) evidence for the claim that the designatum of any expression is its semantic extension and that (ii) certain sorts of linguistic contexts require substitution restrictions on pain of generating absurd results. These, however, are interesting *semantic* results; they tell us very little (if anything) about metaphysics.

SELF: A DYNAMIC APPROACH

Teodor NEGRU

ABSTRACT: According to the classical approach, the self was regarded as a pure unchanging spiritual entity, with a cognitive content which is the consequence of self-awareness that characterises human being. Against this classical conception, the convergence approaches of phenomenology, developmental psychology or neuroscience highlighted the fact that the self is the result of the ongoing dynamics of experiences we have as embodied agents, e.g. the dynamic coupling between the embodied agent and the world, the dynamics of the primal emotions and feelings, as well as the dynamics of neural processes. Hence, the self appears as an embodied self, embedded in a certain context having a pre-reflective character, resulting from the direct coupling of the person with the natural or social environment. In conclusion, according to the contemporary approaches, the self is a multifaceted phenomenon, which should be understood from the perspective of the various dynamic relationships mediated among body, brain, and environment.

KEYWORDS: self, embodiment, ecological self, dynamical system, enactivism, consciousness

1. Embodiment of the Self

According to the Cartesian dualist conception,[1] the self was characterised as having a pure spiritual nature, as it could exist independently from any physical substratum, beyond the laws governing the material world. The role of the self, in this conception, was to provide our mental life with unity and continuity, as it was the unchanging mental entity that does not modify along with mental experiences. In its quality of the centre of man's cognitive life, the self is the source of the first-person perspective, having the privilege of direct access to the individual's mental states, considered private therefore inaccessible to others. To the extent that the self is considered to be present in the unfolding of any mental activity, its existence cannot be doubted, which means that it provides certainty to our existence and knowledge.

---

[1] René Descartes, *Meditations on First Philosophy* (Cambridge: Cambridge University Press, 1996). Originally published in 1641.

Teodor Negru

Against the existence of a permanent substratum which guarantees our personal identity, David Hume[2] argued that, if we had a self, it would have to generate an impression that should be constant throughout our lives. But as such an impression, constant and invariable, does not exist, as we rather have the experience of the transient sequence of sensations, this means that it is wrong to think of self as a substantial, indivisible entity. Such a mistake comes from the fact that our imagination is the one that connects the various sensations received from the environment, creating thus the impression of continuity and unity. According to Hume, our mental life is a bundle of intuitions, beliefs or desires, connected by means of resemblance or causality relationships, which create the impression of identity. In this conception, there is no longer need to postulate a centre of our mental life, as the self is regarded as fiction.

The existence and the necessity of the self to man's mental life are also highlighted by Immanuel Kant.[3] He makes a distinction between man's psychological self, which is the object of his empirical consciousness, changing and varying along according to the diversity of impressions, received by sense organs, and the transcendental self (ego), which appears as pure apperception. According to this approach, the self is the one which unifies the data of experience which were firstly organised by intuition and given meaning by synthesis. The transcendental self (ego) is an *apriori* given that precedes all the experiences of the subject making thus possible the organisation and understanding of the sensible material provided by the senses. Thus, the transcendental unity of self-awareness is the one that provides all experiences and representations with the consciousness of belonging to the same subject, namely a shared identity, representing a pure identity-pole.[4]

The critique of the classical approach of the self was made by phenomenology once it questioned the Cartesian vision of the world, by contesting the idea that the natural world, the objective one, and the knowledgeable subject are radically separated. Claiming that the human being has an originary experience of the world, phenomenology denounces the approach of the world from the scientific perspective as an objectivising manner of understanding things, proposing instead an approach from the first-person point

[2] David Hume, *A Treatise of Human Understanding* (Oxford: Oxford University Press, 2009). Originally published in 1739.

[3] Immanuel Kant, *Critique of Pure Reason* (Indianapolis: Hackett Publishing, 1996). Originally published in 1781.

[4] Dan Zahavi, *Phenomenology of Self*, in *The Self in Neuroscience and Psychiatry*, eds. Tilo Kircher and Anthony David (Cambridge: Cambridge University Press, 2003), 56–75.

of view, which has as its starting point the human body. This approach starts form the distinction operated by Edmund Husserl,[5] between the objective body (*Körper*), the body as it is understood by science, and the lived body (*Leib*), considered the seat of experiences and the means by which we have a direct and implicit experience of the world. The role of the body, in the phenomenological acceptance of the term, is to open a perceptual field where things appear in a new configuration. Thus, owing to its sensorimotor capabilities, the body offers proprioceptive and kinaesthetic information, which is essential to the manner we experience the world. The position of the body, the implicit awareness of my body, present in all intentional act, as well as the information obtained via the movement of the body, open a subjective field where objects are perceived and become available through action.

Accepting the idea that a scientific explanation can provide only an understanding of the body from the perspective of its biological processes, Maurice Merleau-Ponty[6] shows that the relationship between man and the world should be approached from the perspective of the phenomenological body, which is an experiential structure that has a direct relationship with the world. According to Merleau-Ponty's phenomenological conception, the problem of the separation between the subject and the world, which appeared alongside Cartesian dualism, can be solved by accepting the idea that, ever since the beginning, we are in a world to which we are intimately connected. This means that we constitute the world we live in and we are also affected by what happens therein. In other words, the world is inseparable from the body, which means that the relationship between them is not cognitive, representational, but it is a dynamic one given by the skilful body. This is possible owing to the motor intentionality, which, by means of the intentional arc connects the lived body directly to the world, transforming the latter into the condition of realising the ipseity of the subject.

Claiming the existence of a dynamic relationship between the body and the world, phenomenology will contest the primacy of the reflexive consciousness, considered from Descartes onwards the fundament of our mental life, theorising instead the flow of our experiences as an interiority that is given to us in a direct, immediate way. According to Dan Zahavi,[7] phenomenology criticises the idea of self as being a persistent substratum of our mental life by postulating the existence

---

[5] Edmund Husserl, *The Basic Problems of Phenomenology: From the Lectures, Winter Semester, 1910–1911* (Dordrecht: Springer, 2006). Originally published in *Husserliana XIII.*

[6] Maurice Merleau-Ponty, *Phenomenology of Perception* (New York: Routledge, 2005). Originally published in 1945.

[7] Dan Zahavi, "Is the Self a Social Construct?" *Inquiry* 52, 6 (2009): 551–573.

of a pre-reflexive level of consciousness characterised by self-referentiality (Sartre) or self-affection (Henry), which would represent the minimal and primary form of self-awareness. This pre-reflective level, constituting from the immediate given of subjectivity and which no long appears as a transcendental structure or an objectivising flow of consciousness, forms ipseity, selfhood, or the minimal self.

From the experiential perspective, the minimal self represents what the philosophy of the mind called "what is like" or first-person perspective, meaning what makes me perceive experience as in such a way that only I can do it. Thus the minimal self represents what is proper to each experience lived from the perspective of a certain person. It is that mineness to be found in all experiences and which makes them be subjective and not objective data, e.g. information from the third-person perspective. To put it differently, the minimal self is the "*invariant* dimension of first-personal givenness throughout the multitude of changing experiences."[8]

Consequently, contrary to the Cartesian self, that was thought as existing independently from the material world and having a cognitive content, the minimal self is regarded as coming from the dynamics of experience we have as embodied agents existing in a world. This means that the minimal self, as awareness of the immediate experience, is a result of being embedded in a certain context. The consequence of the embodied and embedded approach of the self is its understanding from the perspective of the various dynamic relationships that constitutes the individual as an embodied agent. This means that approaching it involves understanding how the immediate awareness of one's own body contributes to the development of the first-person perspective, the importance of the dynamic relationship between the individual and the context it is embedded in, as well as the contribution the brain and the cerebral processes have to the dynamics of the agent's embodiment and embeddedness.

2. The Dynamics of Embodiment

The embodiment of self was the consequence of highlighting the importance of the body to the development of a minimal consciousness regarded as being prior and to lay at the basis of self-awareness. The phenomenology of the lived body emphasises its structures owing to which, ever since birth, we have a dynamic relationship with the world. Hence, the body schema is described as a phenomenal structure that comprises holistically functions, skills, sensorimotor possibilities and bodily skills, functioning at a sub-personal level without the need

---

[8] Dan Zahavi, *Subjectivity and Selfhood: Investigating the First-Person Perspective* (Cambridge: MIT Press, 2005), 132.

for the explicit consciousness of the body.[9] This means that the body schema does not involve a conscious control over the phenomenal body in order to get information about the world and how to act in this world. Therefore, it was considered as having a dynamic character, meaning that it modifies spontaneously, without involving the reflexive consciousness together with the outer changes that affect position, movement or the organic body functions. At the same time, the dynamic character of the body schema also makes its modification contain a potentially direct response to the changes in the environment. Thus, non-cognitive operations of the body, which are the consequence of the bodily skills, form the basis for the existence of primary awareness, which is prior to the reflexive awareness.

This fact is emphasised by proprioceptive awareness as well,[10] which designates the pre-reflexive awareness resulting not only from the position and the abilities of the body but also from its motor possibilities. Therefore, the motor system plays an important role in the development of proprioceptive self. On the other hand, this system, once the new-born starts moving, generates the proprioceptive sense of movement, which is the basis of the development of an incipient form of phenomenal consciousness and the qualitative mental states at the brain level. On the other hand, based on the motor commands, a feedback is formed which verifies whether the movement comes from the subject's body or from its environment, which involves "a non-observational and pre-reflective differentiation between self and non-self."[11]

Nevertheless, from the phenomenological perspective, proprioperception appears as a feature pertaining rather to the organic body as it is insufficient to explaining the dynamic relationship with the world. The lived body is characterised by the kinaesthetic sense which enables us not only to be aware of our position in a given space, but it also gives a perspective on the possibilities of action provided by the perceptual and motor structure of our body. From this perspective, kinaesthesis should be regarded as being a dynamic sense of the lived body, which ensures the integrity of information come both from outside and inside the body.

---

[9] Shaun Gallagher, *How the Body Shapes the Mind* (Oxford: Clarendon Press, 2005), Shaun Gallagher, *Body Schema and Intentionality*, in *The body and the Self*, ed. José Luis Bermúdez (Cambridge: MIT Press, 1995), 225–244, Merleau-Ponty, *Phenomenology of Perception*.

[10] Gallagher, *How the Body Shapes the Mind*.

[11] Gallagher, *How the Body Shapes the Mind*, 175–176.

Teodor Negru

Regarded from the perspective of the relationship with proprioperception, kinaesthesis is thought by Maxine Sheets-Johnstone[12] to be a form of awareness higher than the former. Proprioperception is the immediate awareness of movement we feel at the level of bodily organs, and which has evolved, along with the development of organisms, into a kinaesthetic awareness. Thus, kinaesthesis represents the awareness of movement, meaning the ontological principle of animation underlying life, where cognition represents a stage in its evolution. Therefore, it is not enough to explain our cognitive activity starting merely from our quality of embodied agents, as it happens in classical phenomenology, but we have to take into account the dynamic dimension as well, which results from the qualitative variation of movement characterising our body. Thus, our body is seen as a tactile-kinaesthetic body, which, in an original way, has a kinaesthetic awareness.

To the extent that movement generates "an overall dynamic with distinctive qualities,"[13] kinaesthesis represents in fact the awareness of this felt qualitative dynamic. Hence, the origin of consciousness should be looked for in the tactile-kinaesthetic awareness which is the result of the spontaneous kinetic acts that characterise us ever since the moment we come into the world. Moreover, besides being the source of the qualitative dimension of our mental life, spontaneous movement is at the origin of our self. By means of movement, our sense of animate forms is created, meaning that we become aware of ourselves as agents of movement, and we constitute ourselves as subjects ("I"), with a capacity to act. In this way, the sense of self comes from the tactile-kinaesthetic body, which, by means of kinaesthesis, organises and provides coherence to the inner and outer sensations it feels. Thus primal animation appears to be the source of our self, by means of which, starting from our experiences, we constitute the world and give a meaning to it.

The analysis of the phenomenological structures of the lived body led to the idea that explaining cognition means the existence of more dynamic cycles where the embodied agent is involved. Thus, besides the dynamics of embodiment, which is the bodily pre-reflective awareness, we should also consider the sensorimotor dynamics that lies at the basis of the dynamic relationship with the

---

[12] Maxine Sheets-Johnstone, *The Primacy of Movement* (Amsterdam: Benjamins, 1999), Maxine Sheets-Johnstone, *Body and Movement: Basic Dynamic Principles*, in *Handbook of Phenomenology and Cognitive Science*, eds. Daniel Schmicking and Shaun Gallagher (Berlin: Springer, 2010), 217–234.

[13] Sheets-Johnstone, *The Primacy of Movement*, 147.

world as well as the dynamics of the inter-subjective interaction, which explains how the self is shaped by the encounter with the others.[14]

The dynamics of the embodiment is given by self-regulation, which is important for our survival as biological beings. Starting from this, the human being is regarded as a autopoietic system, characterised by autonomy (i.e., their goal is only to maintain the internal variables constant), individuality (i.e., the effect of maintaining a constant internal organisation is the acquisition of an identity), internal unity (given by self-production processes and not from outside by the observer), and internalising feedback (the output, which is a response to changes in the environment, turns into input, thus they realise an environment-integrated system).[15] From this perspective, subjectivity is seen as being the result of movement characterising the body ever since birth as well as the sensations generated by the ongoing dynamic of the body.

Sensorimotor dynamics is given by the fact that between the embodied agent and the world there is no linear relationship of the input-output type, but an ongoing dynamic coupling of the sensorimotor cycles. This approach leads to the idea that the human being should be understood as a dynamic system, where the feedback of outputs turn into input, generating a flow of causal dynamic patterns, made of the loops among the sensory system, the motor processes, and the world. From this perspective, the human being appears as a complex dynamic feedback system,[16] made up of both internal dynamic relationships, pertaining to bodily processes, and the exterior ones that take into account the relationship between the body and the world. Therefore, the bodily self no longer appears as a private self, but it is the result of the complexity of dynamic relationships established between the embodied agent and the world.

The idea of circular causality characterising the relationship between the embodied agent and the world relies on an enactive conception on cognition. According to this view, perception, movement, and cognition are closely interdependent which means that within the sensorimotor cycle mediated by the nervous system what "one senses depends directly on how one moves, and how one moves depends directly on what one senses."[17] The integration of sensorimotor patterns into a coherent dynamics is carried out by the sub-personal

---

[14] Evan Thompson, *Mind in Life: Biology, Phenomenology, and the Sciences of Mind* (Cambridge: Harvard University Press, 2007).

[15] Humberto Maturana and Francisco Varela, *Autopoiesis and Cognition* (Dordrecht: Kluwer, 1980).

[16] Susan Hurley, *Consiousness in Action* (Cambridge: Harvard University Press, 1998).

[17] Thompson, *Mind in Life*, 244.

mechanisms of the body schema, which organises perception and action directly, spontaneously, without the involvement of the reflexive consciousness due to bodily skills. Consequently, perception should not be understood in representational terms but as a "skilful bodily activity,"[18] which depends on the sensorimotor knowledge. Similarly, consciousness of the qualitative character of experience is not the result of cerebral processes but it has its origins in the patterns forming from the interdependence between the sensory stimulation and motor skills.

The conclusion that can be drawn from the presentation of what Evan Thompson calls the "dynamic sensorimotor hypothesis"[19] is that understanding the human being as an autopoietic system has as a consequence the need to postulate an autonomous selfhood, as a result of the self-organisation of the body and its dynamic coupling with the world. Moreover, this autonomous self, owing to its enactive relationship with the world, is the result of the dynamics between the perceptive experience and the sensorimotor knowledge. This sensorimotor selfhood appears from the operational closure of the dynamic sensorimotor cycle supported by the nervous system, which "is characterized by an invariant topological pattern that is recursively produced by the system and that defines an outside to which the system is actively and normatively related."[20]

3. From Embodied Self to Embedded Self

The embodiment of self emphasises the existence of a minimal embodied self-awareness, with which we are endowed since birth and which results from the dynamics of the body self-organisation as an autopoietic system. In addition to this, understanding the self involves underlining the importance of the relationship with the context where it occurs, which has a decisive role in the subsequent development of the self. Together with the internal constraints, the body is affected from outside as well, by the external constraints and perturbations, which thus become an important part of the self-production process as a dynamic system, that aims at being integrated in the environment. Thus, the self is not only the result of embodiment but also of the dynamic embeddeness in the natural and social environment.

The dependence between the dynamics of embodiment, on one hand, and, on the other hand, the relationship between the embodied agent and the world is

---

[18] Alva Noë, *Perception in Action* (Cambridge: MIT Press, 2004), 2.
[19] Thompson, *Mind in Life*, 254.
[20] Thompson, *Mind in Life*, 260.

emphasised by the way in which James Gibson[21] understands proprioperception. According to him, proprioperception, as a sense by means of which information about the self are acquired, and exteroception, which refers to the sense by means of which we acquire information about the surrounding world, are complementary. Proprioperception, understood as egoreception, is the results of the dynamics of information that come from all our senses, which are activated by the movements of the body. This means that senses, while revealing information about the external world, provide information about the self, meaning that "one perceives the environment and coperceives oneself."[22] Therefore, at the concrete level, proprioperception and exteroception are but two complementary aspects, one of a subjective nature and the other of an objective nature, of perception.

This I we become aware as we discover the world is the ecological self,[23] which involves the awareness of situating the embodied agent in a certain context. The origin of the ecological self is what Gibson[24] called "visual kinesthesis," that is the awareness produced by the optical flow of our visual system, which once it perceives the information about the space and the objects to be found in the space surrounding us, provides information about the position and the movement of the perceiver. Thus, the embodied agent gains the awareness of differentiating itself from the environment which is static, perceiving itself as a source of movement and cause of the changes in the environment. Moreover, visual proprioperception contributes to consolidating the sense of agency of the perceiver, by means of the role it has in acquiring some motor skills in infancy, being the grounding of the other kinaesthetic systems, as it „comprises an innate feedback loop which informs the organism of its relation with the natural environment."[25]

Besides the optic flow, which appears in various shapes, such as occlusion or looming, to establishing the ecological self other modalities contribute such as vestibular proprioperception, somatic proprioperception, and touch. To these, the quasi-conscious motor intentions are added. They are in a permanent dynamics

---

[21] James J. Gibson, *The Ecological Approach to Visual Perception* (New York: Psychology Press, 2007). Originally published in 1979.

[22] Gibson, *The Ecological Approach*, 126.

[23] Ulric Neisser, "Five Kinds of Self-Knowledge," *Philosophical Psychology* 1, 1 (1998): 35-59.

[24] Gibson, *The Ecological Approach*.

[25] George Buttherworth, "An Ecological Perspective on the Self and its Development," in *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-Experience*, ed. Dan Zahavi (Amsterdam: John Benjamins, 2000), 24.

providing the subject with the sense of agency, as an active agent who lives in a world that is perceived according to their intentions and needs.[26]

Besides the kinesthetic character of its origin, another feature of the ecological self is that "the nature of its ongoing interaction with the environment"[27] is specified by objectively information. Moreover, although it manifests since the earliest infancy, it does not involve self-awareness, which develops in the following age-stages, as new skills are acquired. In other words, the ecological self does not involve the awareness of self-representation as it has an unreflective nature.

Another important characteristic of ecological self, seen as an emergent and changing structure resulting from the interaction with the environment, is the close connection with the affective dimension of the embodied agent.[28] Thus, the minimal embodied self-awareness is considered to be the consequence of feelings and emotions we feel since the very moments of life. The feeling experience is what makes us aware of our own body ever since birth and which "elevates organisms from mere responders to volitional actors,"[29] contributing to perceiving us distinctly from other entities in the world. Thus, newly-born are not only "sensing" the world, but they are also experiencing it, having the awareness to differentiate them from the surrounding world.

Along with the awareness of agency, a process of "objectification of the self"[30] develops by means of which self becomes and object of reflection. This process relies on the co-dependence between the perception of the world and self, as it was theorised by Gibson,[31] according to whom the newly-born, owing to the sensory system they are endowed with since birth, can identify affordances in the environment and starting form this, still owing to the perceptive apparatus, they can get information about self. Hence, their interaction with inanimate objects and their tendency to analyse the consequences of the actions on their own body are in fact means to externalise their own emotions and feelings, with a view to analysing them by means of perceptive sense organs. By this process of "early

---

[26] Ulric Neisser, "The Self Perceived," in *The Perceived Self: Ecological and Interpersonal Sources of Self-Knowledge*, ed. Ulric Neisser (Cambridge: Cambridge University Press, 1993), 3–24.

[27] Neisser, "Five Kinds," 41.

[28] Philippe Rochat, "The Self as Phenotype," *Consciousness and Cognition* 20 (2011): 109–119.

[29] Rochat, "The Self as Phenotype," 110.

[30] Philippe Rochat, "Early Objectification of the Self," in *The Self in Infancy: Theory and Research*, ed. Philippe Rochat, Advances in Psychology Book Series 112 (Amsterdam: Elsevier Science, 1995), 53-72.

[31] Gibson, *The Ecological Approach*.

objectification of self," the infant transforms the internal affective dynamics of self, by projecting these emotions and feelings onto objects, in an external relationship, of perceptive analysis of such dynamic aspects of self. Thus, the internal dynamics of emotions is transformed into an external dynamics where these are experimented as characteristics of external objects.

To sum up, according to Rochat, the origin of ecological self should be looked for in the perception of body effectivities, which is the grounding of our ability to make plans in order to achieve a goal. This perceptual ability, which is innate, makes possible for the bodily skills to adjust to and to integrate in the requirements of particular situation in which the agent is. This means that the ecological self "is an emergent property of any biological system that does not merely respond to stimuli, but acts, explores, and invents new means to achieve functional goals."[32] The discovery of the causal efficiency of the agent in the world cannot be separated from the exploration of feelings, which are the consequence of the embodiment of the newly-born. Thus, an equally important role in the early development of self-knowledge is played by the exploration of the emotional experience dynamics of the newly-born, by projecting them onto the objects in the surrounding world onto which they exert the agent skills.

Last but not least, an important role in the emergence of the embedded self is played by the experience of living in society. The interaction with the others represents a dynamic cycle where the "dynamical process of 'self-othering'"[33] contributes to the emergence of an interpersonal self. This type of self is described by Neisser[34] as resulting not from the inferential understanding of some isolated aspects of the interaction with others, but from the direct perspective of the unreflective ongoing relationship which is established within the intersubjective interaction. Just like the ecological self, which it manifest with, the interpersonal self has a kinesthetic nature, being specified and developed by the flow of expressive gestures of the others. Nonetheless, the interaction with the others is considered to be at the origin of the emergence of some skills, otherwise impossible to develop, such as the body image, which is the result of the intersubjective and inter-modal interaction between proprioperception and the perception of the other's facial expression in response to our actions.[35] Consequently, the dynamics of intersubjectivity is regarded as the source of a

---

[32] Rochat, "Early Objectification," 59.
[33] Thompson, *Mind in Life*, 251.
[34] Neisser, "Five Kinds."
[35] Gallagher, *How the Body Shapes the Mind*, 73.

sense of self, which comes together with the awareness of the context where the agent is embedded.

4. The Neurological Dynamics of Self

The importance of the brain to cognition is given not merely by the fact that the latter's development depends on higher cognitive processes, but also on the contribution it has to the emergence of the embodied self. In other words, the brain is no longer regarded as the seat of abstract thinking, but as playing an important role both in the dynamics of embodiment and in the dynamic interaction between the body and the environment.

Thus, Varela and Thompson[36] speak of three "cycles of operation" that would be characteristic to higher primates, to which the brain has a decisive contribution. The first cycle, of organismic regulation, refers to the connection between the neural processes and the haemodynamic ones of the internal organs. Due to the fact that the haemodynamic processes determine inevitably emotional states, the latter become "inescapable affective backdrop of every conscious state."[37] Hence, the organic self is regarded as the result of the dynamics of the affective processes, being characterised, due to the changing feature of feelings, by discontinuity. The next cycle, the one of the sensorimotor connection, carries out the coupling between organism and environment whereby the organism becomes a situated agent. The neural patterns that emerge temporarily ensure the coordination between the sensory and the motor systems, determining their reciprocal conditioning. The last cycle, the one of the intersubjective interaction, which relies on the "mirror neurons," carries out the dynamic coupling with the others and the understanding of the meaning of the others' actions as well.

Admitting the fact that there are more dynamic relationships that characterise the embodied agent leads to underlining the role of the brain in there mediation and to the emergence of a sense of self within this process. Thus, according to Todd Feinberg, self, defined as "unity of consciousness in perception and action that persists in time"[38] is the result of three dissociated hierarchical systems: the anatomically central/medial interoself system, exterosensorimotor system, and the integrative self system which mediates between the first two. The interoself generates a sense of self based on the internal homeostatic and affective

---

[36] Evan Thompson and Francisco Varela, "Radical Embodiment: Neural Dynamics and Consciousness," *Trends in Cognitive Sciences* 5, 10 (2001): 418–425.

[37] Thompson and Varela, "Radical Embodiment,"424.

[38] Todd E. Feinberg, "The Nested Neural Hierarchy and the Self," *Consciousness and Cognition* 20 (2011): 4.

processes. The exterosensorimotor system contributes to the emergence of self by experiencing the external stimuli as subjective feelings that do no come from the body but from outside the body. While the integrative system, which is consists of the heteromodal association cortices, presents as a "convergence zone," where the internal needs of the body are adjusted to the external environment conditions.

Besides the role to mediate among the various cycles of the embodied agent, another important feature of the brain in constituting the self and the consciousness is its dynamic nature. According to the findings in neuroscience, at the grounding of the brain functioning lies not only the local causal interaction among the neurons located in neighbouring regions but long-range connections among its parts, no matter how far are ones to the others. In this context, the brain appears to be a dynamic organ where reciprocal loops are established among its regions, which that lead to carrying out large-scale dynamics that results in performing various mental processes.

Such an approach is "the dynamical core hypothesis"[39] according to which consciousness is the result of the ongoing interaction among several groups of neurons located in different regions of the brain, which form a functional cluster in hundreds of milliseconds. This functional cluster, which relies on the capacity of thalamocortical system to integrate neural loops, undergoes permanent change, allowing other groups of neurons to join the process. Thus, consciousness no longer appears to be located in a certain part in the brain, but rather to be a distributed neural process, determined by the integration of neuronal groups into a dynamic core. Notwithstanding, self is not considered as having an unreflective nature resulting from the primary consciousness, but it is considered as requiring higher-order consciousness, being defined as „consciousness of consciousness."[40] According to this view, we can talk about a self only in the case of beings endowed with language, which as a system characterised by syntax and semantics, makes symbolic thinking and conceptual memory possible, both important to the emergence of higher-order consciousness.

The idea that the origin of self should be looked for in the higher mechanisms of cognition was proved wrong by the emphasis on the importance of feelings in self-regulating and self-awareness of any biological being. According to such a theory, biological organisms should be regarded as complex systems, which aim at preserve their balance despite both the internal and external perturbations

---

[39] Gerald M. Edelman and Giulio Tononi, *A Universe of Consciousness: How Matter Becomes Imagination* (New York: Basic Books, 2000).

[40] Edelman and Tononi, *A Universe of Consciousness*, 194.

and constraints.[41] To this purpose, they develop self-monitoring and control mechanisms whose role is to react to any type of change. The consequence of the dynamics between the control mechanisms, necessary to maintain the equilibrium of the organism, and the internal and external perturbations, is the emergence of feelings, hence, the emergence of subjectivity and consciousness. Thus, consciousness appears as a biological phenomenon, which depends on the ongoing relationship between the embodied subject and the context they are situated, being the result of the process of resistance to variance which characterises the body. It results from the above that "self is shaped and maintained only as a result of ongoing actions and reactions,"[42] which means that the origin of consciousness should be looked for in the dynamics of feelings, which generate a minimal level of self-awareness, which does not involve self-reflection or the ability to manipulate symbols.

The importance of the affective dimension of biological beings to the emergence of the sense of self is also emphasised by Jaak Panksepp.[43] He considers that we should look for the origin of consciousness in something prior to the cognitive structures of the brain that are sufficiently developed at birth and that require their maturation in time. Thus, he speaks of a primitive form of consciousness, the affective consciousness, which is the result of the dynamics of affective feelings, developed from the sensations perceived by the sense organs, homeostatic processes, as well as the responses the brain provides to external or internal stimuli, which together are indispensable to any organism in order to survive. Starting from here, feelings are regarded as being located behind the unfolding conscious life, each of them being represented at the level of the brain by a certain type of neural circuit, which generates particular neurodynamics. The representation of basic emotions by the brain allows for their integration with the action neural schemata, which are the result of bodily motor processes. The interaction between sensorimotor action circuits and emotional circuits generates our first representation as organisms that exist and act in a world, which represents the "primal self." Therefore, the neural origin of primal self is identified

---

[41] Antonio Damasio and David Rudrauf, "The Biological Basis of Subjectivity: A Hypothesis," in *Self-Representational Approaches to Consciousness,* eds. Uriah Kriegel and Kenneth Williford (Cambridge: MIT Press/Bradford Books, 2006), 423–464.

[42] Damasio and Rudrauf, "The Biological Basis," 441.

[43] Jaak Panksepp, "The Periconscious Substrates of Consciousness: Affective States and the Evolutionary Origins of the Self," in *Models of the Self*, eds. Shaun Gallagher and Jonathan Shear (Thorverton: Imprint Academic, 1999), 113–130, Jaak Panksepp, "Affective Consciousness," in *The Blackwell Companion to Consciousness*, eds. Max Velmans and Susan Schneider (Oxford: Blackwell Publishing, 2007), 114–129.

in the periaqueductal gray (PAG) region that exists ever since the beginnings of the development of the brain and that ensures the integration "of a diversity of basic emotional systems (fundamental value schema), various simple sensory abilities (perceptual schema), and primitive but coherent response systems (action schema)."[44]

Similarly, Antonio Damasio[45] considers that the grounding of the sense of self is the dynamics of the body feelings. The proto-self is the result of primal feelings recorded in our nervous system; therefore it does not involve a centralised neural structure as it is unconscious. It presents itself as a collection of neural models undergoing continuous change and reflecting the state of the body at a certain moment.

Based on this proto-self the core self emerges, which, besides sensorimotor images, has a cognitive content as well. The core self emerges as a result of the dynamics of successive changes in the body, which at the cognitive level generates a representation of the surrounding objects and how they affect the body. To put differently, the core self involves several images: an image of the organism (determined by the successive changes in the body, which result in changing the proto-self), an image of the object which affects the sense organs of the body, and an image of the organism's response (which has an affective content). Thus, by becoming aware of the external objects, we become aware of us and of our capacity of agency and ownership. Consequently, the self is regarded as a "dynamic process"[46] of integration of neural processes which originate in the dynamics of body representations at the brain level.

5. Conclusion

The conclusion one can draw from the convergent approaches of phenomenology, psychology or neuroscience is that the self can no longer be regarded, in the Cartesian manner, as having a pure unchanging spiritual nature with a cognitive content, which is the result of self-awareness that characterises the human being. The self comes rather from the ongoing dynamics we experience as embodied agents as well as from the dynamic relationships existing between the embodied agent and the world. Hence, the self appears as an embodied self embedded within a certain context, having a pre-reflective character, which is the result of the

---

[44] Panksepp, "The Periconscious Substrates," 116.

[45] Antonio Damasio, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (San Diego: Harcourt, 1999), Antonio Damasio, *Self Comes to Mind: Constructing the Conscious Brain* (New York: Pantheon Books, 2010).

[46] Damasio, *Self Comes to Mind*, 173.

direct coupling of the person with the natural or social environment. Another distinctive feature of embodied and embedded self is its affective dimension, which is highlighted by the importance of the dynamics of the emotions and primal feelings at its development. Last but not least, an important contribution to the development of the self is that of the brain as a dynamic organ, whose role is to represent and integrate the different sensations coming both from within and from without the organism. To conclude, the self is a multifaceted phenomenon that should be understood from the perspective of the various dynamic relationships that are established among body, brain, and environment.

DEBATE

WHY EPISTEMIC PERMISSIONS DON'T AGGLOMERATE – ANOTHER REPLY TO LITTLEJOHN

Thomas KROEDEL

ABSTRACT: Clayton Littlejohn claims that the permissibility solution to the lottery paradox requires an implausible principle in order to explain why epistemic permissions don't agglomerate. This paper argues that an uncontentious principle suffices to explain this. It also discusses another objection of Littlejohn's, according to which we're not permitted to believe lottery propositions because we know that we're not in a position to know them.

KEYWORDS: agglomeration, epistemic permission, lottery paradox

According to the permissibility solution, the lottery paradox can be solved if epistemic justification is assumed to be a species of permissibility.[1] The paradox arises from the following three claims, which seem individually plausible but jointly inconsistent:

(1-J)   For each ticket, I'm justified in believing that it will lose.

(2-J)   If, for each ticket, I'm justified in believing that it will lose, then I'm justified in believing that all the tickets will lose.

(3-J)   I'm not justified in believing that all the tickets will lose.

If justification is taken to be a species of permissibility, then the first two claims are ambiguous. They both come out true if we disambiguate them as follows:

(1-Narrow) $\mathbf{Pe}B t_1$ & $\mathbf{Pe}B t_2$ & … & $\mathbf{Pe}B t_n$.

(2-Wide)   If $\mathbf{Pe}[B t_1$ & $B t_2$ & … & $B t_n]$, then $\mathbf{Pe}B[t_1$ & t_2 & … & $t_n]$.

(In the symbolism, n is the number of tickets in the lottery; '$\mathbf{Pe}\phi$' stands for 'It is permissible for me that ϕ'; '$B\psi$' stands for 'I believe that ψ'; and, for $1 \le i \le n$, 't_i'

---

stands for 'Ticket number i will lose.') The third claim of the paradox is unambiguously true; we can rephrase it in terms of permissibility as

(3-Unamb) ~$\mathbf{Pe}B[t_1 \ \& \ t_2 \ \& \ \dots \ \& \ t_n]$.

The rationale for (1-Narrow) is that I'm permitted to believe what is sufficiently probable on my evidence:

(High)   If the probability that it is the case that p is sufficiently high on my evidence, then I'm permitted to believe that p.[2]

Given that, for a particular ticket, the probability that it will lose is sufficiently high on my evidence, it follows from (High) that I'm permitted to believe that it will lose; repeating this reasoning gives us all of the conjuncts of (1-Narrow). The rationale for (2-Wide) is a plausible closure principle for (epistemic) permissibility: if I'm permitted at once to believe this, to believe that, etc., then I'm permitted to have a single belief whose content is the conjunction of the contents of the former beliefs.

The crux of the permissibility solution is that (1-Narrow) doesn't entail the antecedent of (2-Wide) and thus doesn't entail the negation of (3-Unamb), because epistemic permissions don't agglomerate. That is, I might be permitted to believe this, permitted to believe that, etc., without being permitted to have all those beliefs together. (Agglomeration is different from the closure principle from the previous paragraph: the former is about the scope of the permissibility operator, while the latter is about the scope of the belief operator.)

While acknowledging that permissions don't agglomerate in the non-epistemic case, Clayton Littlejohn demands an explanation of why epistemic permissions fail to agglomerate. He holds that a principle similar to the following one is required by the permissibility solution:

(Risk-DJ)   If the probability of acquiring an error-containing belief set would get too high by adding the belief that p to your belief set, you cannot justifiably believe p.[3]

He also holds that (Risk-DJ) has implausible consequences.

Whether or not (Risk-DJ) has implausible consequences, the permissibility solution needs nothing like (Risk-DJ) in order to explain why epistemic permissions don't agglomerate. All that is required in addition to what the

---

[2] See Kroedel, "The Permissibility Solution to the Lottery Paradox," 106 and Clayton Littlejohn, "Lotteries, Probabilities, and Permissions," *Logos & Episteme* 3 (2012): 512.

[3] Clayton Littlejohn, "Don't Know, Don't Believe: Reply to Kroedel," *Logos & Episteme* 4 (2013): 234.

permissibility solution itself provides is the following principle, which Littlejohn himself agrees is in the spirit of the permissibility solution:

(Low)   If the probability that it is the case that p on my evidence is sufficiently low, then I'm not permitted to believe that p.[4]

To see that no additional principle other than (Low) is needed to explain why epistemic permissions don't agglomerate, assume that the probability that it is the case that q on my evidence is sufficiently high, as is the probability that it is the case that r, while the probability that it the case that both q and r is sufficiently low on my evidence. (Claims q and r may or may not be about lottery tickets.) From (High), we get that I'm permitted to believe that q and that I'm permitted to believe that r. In sum:

(4) $\mathbf{Pe}Bq$ & $\mathbf{Pe}Br$.

If we apply the closure principle that provided the rationale for (2-Wide) to our case, we get that I'm permitted to believe that q and r if I'm permitted separately to believe that q and to believe that r:

(5) If $\mathbf{Pe}[Bq$ & $Br]$, then $\mathbf{Pe}B[q$ & $r]$.

By assumption, the probability that it is the case that both q and r is sufficiently low on my evidence. Substituting 'q & r' for 'p' in (Low), it follows that I'm not permitted to believe that q and r:

(6) ~$\mathbf{Pe}B[q$ & $r]$.

By modus tollens, from (5) and (6) we get that I'm not permitted both to believe that q and to believe that r:

(7) ~$\mathbf{Pe}[Bq$ & $Br]$.

Claims (4) and (7) yield an instance of non-agglomeration for epistemic permissibility: I'm permitted to believe that q and permitted to believe that r, but I'm not permitted both to believe that q and to believe that r.

Notice that, even if we assume that I do in fact believe that q,

(8) Bq,

it does *not* follow that I'm not permitted to believe that r after all (which, together with our assumption that it's highly probable that r, would contradict (High)).

---

[4] See Littlejohn, "Lotteries, Probabilities, and Permissions," 512; my formulation of (Low) differs slightly from Littlejohn's, but in irrelevant respects.

Claim (7) is equivalent to the claim that it's obligatory for me that if I believe that q, then I don't believe that r:[5]

(9) $\mathbf{Ob}[Bq \supset {\sim}Br]$.

But claims (8) and (9) by themselves don't license an inference to the conclusion that I'm obligated not to believe that r,

(10) $\mathbf{Ob}{\sim}Br$,

which would be equivalent to the conclusion that I'm not permitted to believe that r,

(11) ${\sim}\mathbf{Pe}Br$.

The conclusion would follow only if we assumed the principle of factual detachment. That principle is implausible, however, and Littlejohn himself doesn't endorse it.[6]

Littlejohn makes a second objection, drawing on the claim that no matter how probable it is on my evidence that my ticket will lose, I don't know that it will lose. Indeed, it seems that, if I'm sufficiently reflective, I know that I'm not even in a position to know that it will lose. Littlejohn holds that this rules out that I'm permitted to believe that my ticket will lose. More generally, he endorses the following principle:[7]

(Knowledge)   If I know that I'm not in a position to know that p, then I'm not permitted to believe that p.

The principle (Knowledge) conflicts with (High), as is witnessed by propositions that are sufficiently probable on my evidence but of which I know that I'm not in a position to know them, such as the proposition that my ticket will lose. Littlejohn argues for (Knowledge) by claiming that believing that p despite knowing that I'm not in a position to know that p would be "deeply irrational."[8]

Instead of arguing from (Knowledge) against (High), however, one can argue from (High) against (Knowledge). Prima facie, (High) seems no less plausible than (Knowledge). And proponents of an account of epistemic justification in terms of permissibility are likely to find it congenial to conceive of rational

---

[5] See Kroedel, "The Permissibility Solution to the Lottery Paradox," 108. In the additional symbolism in (9) and (10), '$\mathbf{Ob}\phi$' stands for 'It is obligatory for me that ϕ.'

[6] See Littlejohn, "Don't Know, Don't Believe," 235.

[7] Littlejohn, "Don't Know, Don't Believe," 236.

[8] Littlejohn, "Don't Know, Don't Believe," 236.

acceptability in terms of permissibility as well.[9] Thus, given (High), they have a principled reason to reject Littlejohn's claim that believing a lottery proposition while knowing that one isn't in a position to know it would be irrational, "deeply" or otherwise.[10, 11]

---

[9] See Kroedel, "The Lottery Paradox, Epistemic Justification and Permissibility," 59.

[10] For further discussion of rational belief in the absence of knowledge, see Aidan McGlynn, "Believing Things Unknown," *Noûs* 47 (2013): 385–407.

[11] Thanks to Beau Madison Mount for helpful comments and suggestions.

WHY ASSERTION AND PRACTICAL REASONING ARE POSSIBLY NOT GOVERNED BY THE SAME EPISTEMIC NORM

Robin McKENNA

ABSTRACT: This paper focuses on Martin Montminy's recent attempt to show that assertion and practical reasoning are necessarily governed by the same epistemic norm ("Why Assertion and Practical Reasoning Must be Governed By the Same Epistemic Norm," *Pacific Philosophical Quarterly,* 2013). I show that the attempt fails. I finish by considering the upshot for the recent debate concerning the connection between the epistemic norms of assertion and practical reasoning.

KEYWORDS: epistemic norms, assertion, practical reasoning, speech acts

Introductory remarks

Many think that knowledge is the epistemic norm of both assertion and practical reasoning:

KRA: One must: assert that p only if one knows that p.
KA: One must: act on the belief that p only if one knows that p.[1]

This prompts one to wonder whether there's any connection between the epistemic norm of assertion and the epistemic norm of practical reasoning. The question is whether one would expect the norms to coincide, not whether the norms just happen to coincide. One reason why one would expect the norms to coincide would be if one could derive KRA from KA, or KA from KRA, or both KA

---

[1] See, for instance, John Hawthorne and Jason Stanley, "Knowledge and action," *Journal of Philosophy* 105, 10 (2008): 571–590, Timothy Williamson, *Knowledge and its limits* (New York: Oxford University Press, 2000), Ch. 11, and Timothy Williamson, "Knowledge, context, and the agent's point of view", in *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, eds. Gerhard Preyer and Georg Peter (New York: Oxford University Press, 2005), 91–114. Two clarifications: First, I ignore complications concerning how best to formulate KRA and KA. Second, while I focus on the necessity directions of both norms everything I say will go for the sufficiency directions, and therefore the biconditional versions, too.

and KRA from some third thing. Another reason would be if one had good abductive reasons for expecting the norms to coincide.[2] I'll call the thesis that one would expect the norms of assertion and practical reasoning to coincide 'commonality.'[3]

Martin Montminy[4] argues that KA, together with what he calls the "manifestation norm of assertion," entails KRA. By this he means that any assertion which satisfies the manifestation norm of assertion and is based on known premises (therefore satisfying KA) thereby satisfies KRA. In this paper I argue that the derivation fails. That is, I show that there are assertions which satisfy the manifestation norm of assertion and are based on known premises yet don't satisfy KRA. I, first, outline the derivation and, second, explain why it fails. I finish by discussing the upshot for commonality.

The derivation

Montminy follows a distinguished tradition within speech act theory by taking speech acts in general, and therefore assertion in particular, to be individuated via the attitude they express.[5] A natural suggestion is that assertions express belief:

> BRA: One must: assert that p only if one believes that p.[6]

The idea is not that, in order to assert that p, one must believe that p. That would make insincere assertion – asserting that p while believing that not-p – impossible. Rather, BRA is a condition for sincere or non-defective assertion. One's assertion is sincere or non-defective only if one believes that which one asserts. However, as Montminy notes,[7] BRA seems to rule that assertions such as the following aren't defective:

---

[2] The target of this paper, Martin Montminy, "Why Assertion and Practical Reasoning Must be Governed By the Same Epistemic Norm," *Pacific Philosophical Quarterly* 94, 1 (2013): 57–68, takes the first route.

[3] I take this label from Jessica Brown, "Assertion and Practical Reasoning: Common or Divergent Epistemic Standards?" *Philosophy and Phenomenological Research* 84, 1 (2012): 123–157.

[4] Montminy, "Why Assertion and Practical Reasoning Must be Governed By the Same Epistemic Norm."

[5] See Kent Bach and Robert M. Harnish, *Linguistic Communication and Speech Acts* (Cambridge: MIT Press, 1979), Ch. 3 for an attempt to give a full taxonomy of the various speech acts based on this idea.

[6] Kent Bach, "Applying Pragmatics to Epistemology," *Philosophical Issues* 18, 1 (2008): 77.

[7] Montminy, "Why Assertion and Practical Reasoning Must be Governed By the Same Epistemic Norm," 59.

> DIRECTIONS: Two philosophers, Professor Brown and Professor Green, are debating the quickest way to get to the restaurant. Professor Brown asserts that the quickest way is to turn right. Professor Green doesn't get on at all well with Professor Brown and is disposed to contradict whatever he says. So she asserts that the quickest way is to turn left, not right. However, it just so happens that Professor Green actually believes that the quickest way is to turn left, although that belief played no role in her asserting that the quickest way is to turn left.

Montminy thinks that cases like DIRECTIONS motivate adopting an alternative to BRA, viz.:

> MRA: One must: assert that p only if one's assertion manifests one's belief that p.

Again, the idea is not that, in order to assert that p, one must manifest one's belief that p. That would make it impossible to assert that p while not manifesting one's belief that p. Rather, MRA is a condition for non-defective assertion. But what's involved in assertion manifesting belief? Montminy writes:

> When a speaker manifests her belief that p in asserting that p, her belief is a motivating reason for her assertion: that the speaker believes that p would explain (at least in part) why she asserted that p.[8]

When one asserts that p one manifests a number of beliefs, viz. those beliefs that, taken together, explain why one asserted that p. For example: imagine that Professor Brown believes both that the quickest way to get to the restaurant is to turn right and that his audience wants to know where the restaurant is, and he asserts that the quickest way is to turn right based on these beliefs. In doing so he manifests both the belief that quickest way is to turn right and the belief that his audience wants to know where the restaurant is.[9] From this account of manifestation, together with MRA, it follows that Professor Green's assertion in DIRECTIONS is defective. It is specified that Professor Green's belief that the quickest way is to turn left played no role in her asserting that the quickest way is to turn left.

Montminy[10] argues that we can derive a knowledge norm of assertion from MRA and KA (the knowledge norm of practical reasoning):

---

[8] Montminy, "Why Assertion and Practical Reasoning Must be Governed By the Same Epistemic Norm," 59.

[9] Montminy's considered view seems to be that the norm of assertion requires that one knows both that what one asserts is true and that every other belief that one manifests in asserting it is true (Montminy, "Why Assertion and Practical Reasoning Must be Governed By the Same Epistemic Norm," 65).

[10] Montminy, "Why Assertion and Practical Reasoning Must be Governed By the Same Epistemic Norm," 62.

1. One's assertion that p must manifest one's belief that p (MRA).
2. Assertions manifest beliefs when they're based on those beliefs.
3. Actions must be based on beliefs that count as knowledge (KA).
4. Assertions are a type of (linguistic) action.
5. Assertions must be based on beliefs that count as knowledge.
6. One's assertion that p must be based on one's knowledge that p.
7. One's assertion that p must manifest one's knowledge that p.

The conclusion can be restated as follows:

KMRA: One must: assert that p only if one's assertion manifests one's knowledge that p.

Note that KMRA differs slightly from the original KRA. This is only to be expected, given that the aim is to individuate assertion in terms of the attitude it expresses or manifests. If one thinks that the norm of assertion is knowledge, and one thinks that assertion is individuated in terms of the attitude it manifests, then one will think that assertion manifests knowledge.[11]

Why the derivation fails

In this section I argue that the derivation fails. I'll start by showing that, on Montminy's intended reading of what is involved in assertion manifesting belief, the derivation is unsuccessful. I'll then argue that the natural amendment to his reading won't help.

Given the quoted passage above, Montminy's intended reading seems to be this:

MANIFESTATION: S's assertion that p manifests the belief that p iff S's belief that p explains in part why S asserted that p.

Now consider this variant on DIRECTIONS:

DIRECTIONS*: Two philosophers, Professor Brown and Professor Green, are debating the quickest way to get to the restaurant. Professor Brown has asserted that the quickest way is to turn right. Professor Green believes that the quickest way is to turn right. However, while she knows that her grounds for believing this are fairly good, she also knows that they don't suffice for knowledge. Usually, Professor Green only asserts what she thinks she knows. But she also knows that they'll get to the restaurant eventually, whether they turn left or right, and she has no appetite for getting into an argument with Professor Brown. Professor Green reasons as follows:

---

[11] For a similar view see John Turri, "The Express Knowledge Account of Assertion," *Australasian Journal of Philosophy* 89, 1 (2011): 37–45.

1. I believe that the quickest way is to turn right, albeit on grounds that don't suffice for knowledge.
2. Usually, I only assert what I think I know.
3. However, I know we'll get there regardless, and I have no appetite for getting into an argument.
4. So I'll assert that the quickest way is to turn right. [12]

According to MANIFESTATION, Professor Green's assertion that the quickest way is to turn right manifests her belief that the quickest way is to turn right. That belief, together with her belief that they'll get to the restaurant eventually whether they turn left or right, explains why she asserts that the quickest way is to turn right. So Professor Green's assertion satisfies MRA.[13] Further, her assertion is based on known premises. She knows that she believes (on grounds that don't suffice for knowledge) that the quickest way is to turn right, and she knows that they'll get to the restaurant regardless. So Professor Green's assertion satisfies KA. However, Professor Green's assertion doesn't satisfy KMRA because, as she herself knows, she doesn't know that the quickest way is to turn right. So the derivation fails. An assertion that p can manifest the belief that p and be based on known premises yet not be known.[14]

The problematic step in the derivation is from (5) to (6):

5. Assertions must be based on beliefs that count as knowledge.

6. One's assertion that p must be based on one's known belief that p.

Professor Green's assertion that the quickest way is to turn right must be and indeed is based on beliefs that count as knowledge (her belief that she believes on grounds that don't suffice for knowledge that the quickest way is to turn right, and her belief that they'll get there regardless). But it doesn't follow that Professor

---

[12] Why are steps (1) and (2) required for Professor Green to reach her decision? Imagine that her reluctance to assert what she realises she doesn't know is so great that she needs the re-assurance of knowing that she has good grounds for what she asserts, albeit grounds that fall short of knowledge.

[13] Objection: Professor Green's assertion manifests the belief that she believes that the quickest way is to turn right, not the belief that the quickest way is to turn right. So it violates MRA. Response: While Professor Green uses her belief that she believes that the quickest way is to turn right as a premise in her practical reasoning (rather than just 'the quickest way is to turn right'), it seems right to say that her decision to assert is still based, in part, on her belief that the quickest way is to turn right. If she didn't have the first-order belief then she wouldn't have decided to assert. She can't have the second-order belief without the first-order belief, and the second-order belief is a vital step in her decision to assert.

[14] Note that DIRECTIONS* also shows that an assertion can satisfy BRA (the belief norm of assertion) and be based on known premises yet not be known.

Green's assertion that the quickest way is to turn right must be based on her knowledge that the quickest way is to turn right. All that KA requires is that one's assertions, like one's actions in general, must be based on known beliefs. It doesn't require that one's assertion that p must be based on one's known belief that p.

This shows that, on Montminy's intended reading of what is involved in assertion manifesting belief, the derivation fails. But there's a fairly natural amendment that one might think will save the derivation:

> MANIFESTATION*: S's assertion that p manifests the belief that p iff S's belief that p non-deviantly causes S to assert that p.[15]

There are all sorts of ways in which one's belief that p can deviantly cause one to assert that p. For example, say that I believe that the bank is round the corner, intend to deceive my interlocutors by asserting that the bank isn't round the corner, but misspeak and assert that the bank is round the corner. So perhaps Professor Green's belief that the quickest way is to turn right causes her to assert that the quickest way is to turn right, but in a deviant way. To make this plausible, note that it's specified that Professor Green usually only asserts what she thinks she knows. This case, in which she asserts yet knows that she lacks knowledge, is an exception to the rule. Exceptions to a rule are deviant.

Consider another variant on DIRECTIONS:

> DIRECTIONS**: Two philosophers, Professor Brown and Professor Green, are debating the quickest way to get to the restaurant. Professor Green believes that the quickest way is to turn right. However, while she knows that her grounds for believing this are fairly good, she also knows that they don't suffice for knowledge. However, Professor Green is habitually fairly lax in her assertions; she often asserts things she hasn't got particularly good grounds for believing. Further, she knows this about herself. Professor Green reasons as follows:
>
> 1. I believe that the quickest way is to turn right, albeit on grounds that don't suffice for knowledge.
>
> 2. I usually assert based on grounds that don't suffice for knowledge.
>
> 3. So I'll assert that the quickest way is to turn right.

Again, Professor Green's assertion is based on known premises, and it's specified that she habitually asserts even when she realises that she lacks knowledge. It's hard to see how her belief that the quickest way is to turn right could have caused her assertion that the quickest way is to turn right in a non-deviant way. So the amended MANIFESTATION* won't save the derivation.

---

[15] Montminy himself suggests this reading at one point (Montminy, "Why Assertion and Practical Reasoning Must be Governed By the Same Epistemic Norm," 59 fn. 7).

Whither commonality?

Where does this leave commonality (the idea that we should expect there to be a common epistemic norm of assertion and practical reasoning)? I mentioned in the introductory section that there are two reasons one might have to endorse commonality. The first would be if one could derive the norm of assertion from the norm of practical reasoning, the norm of practical reasoning from the norm of assertion, or the norm of assertion and practical reasoning from the norm of some third thing. I'm skeptical about the prospects for carrying out any of these derivations.[16] The second would be if one had good abductive grounds for expecting the norms to coincide. I think this is a far more promising route, and I'll finish by discussing one possible ground.

It seems clear that we often use assertions as reasons for action.[17] If I assert that the restaurant is round the corner to my audience then, assuming we're in a normal sort of situation, I expect my audience to act on my assertion, and my audience expects to be able to act on my assertion. If we assume that knowledge is the norm of practical reasoning, this provides some support for thinking that knowledge is the norm of assertion. I wouldn't expect my audience to act on my assertion if I didn't expect them to take me to know that the restaurant is round the corner, and my audience wouldn't expect to be able to act on my assertion if they didn't expect me to assert only if I know that the restaurant is round the corner. I certainly don't think one could derive KRA from KA on this basis.[18] Rather, the idea would be that the best way to make sense of both my expectation and my audiences' expectation would be if knowledge were the norm of assertion as well as the norm of practical reasoning.

While I think this sort of strategy is promising, a word of warning. Both KRA and KA are exceptionless norms. Knowledge isn't the norm of most assertion or most practical reasoning, knowledge is the norm of all assertion and all practical reasoning. But all we have are the observations that we *often* use assertions as reasons for action, and that *in normal situations* speakers expect audiences to act on their assertions, and audiences expect to be able to act on speakers' assertions. While this might suggest that knowledge is often the norm of

---

[16] Brown ("Assertion and practical reasoning") provides a battery of arguments against a range of possible derivations.

[17] McKinnon makes this observation, and suggests that it provides abductive support for commonality (Rachel R. McKinnon, What I Learned in the Lunch Room about Assertion and Practical Reasoning," *Logos & Episteme* 3, 4 (2012): 568).

[18] Brown ("Assertion and practical reasoning," 144–155) argues convincingly that such a derivation would fail.

assertion as well as the norm of practical reasoning, or that in normal situations knowledge is the norm of assertion as well as the norm of practical reasoning, it's unclear why it suggests that knowledge is always the norm of assertion as well as the norm of practical reasoning.[19]

REVIEWS

David Christensen and Jennifer Lackey, eds., *The Epistemology of Disagreement: New Essays* (Oxford, New York: Oxford University Press, 2013)

Reviewed by Eugen Huzum

As you probably already presumed after reading the title of this review, the book presented here is a collection of essays investigating the epistemological problems raised by the phenomenon of disagreement (a topic in which, for good reason, an increasing number of epistemologists are interested lately). If I am not mistaken, it is only the third collection of this type, after David Christensen's special issue of *Episteme* on the epistemology of disagreement (in 2009) and Richard Feldman and Ted Warfield's *Disagreement* (Oxford University Press, 2010). Therefore, *The Epistemology of Disagreement: New Essays* is a much welcomed and needed publication. Moreover, I do not think that I am at all exaggerating in saying that it is also a must-have (or, at least, a must-read) book for any epistemologist or student interested in epistemology in general, and in the epistemology of disagreement in particular. Now, let me point out the main reasons which support this very confident statement.

The first reason which pleads for my assertion is the fact that the authors whose texts are included in *The Epistemology of Disagreement* are very well known and respected epistemologists, some of them with already strong contributions to the disagreement literature. To be more precise, the papers reunited in the book are signed by John Hawthorne and Amia Srnivasan ("Disagreement Without Transparency: Some Bleak Thoughts"), Thomas Kelly ("Disagreement and the Burdens of Judgment"), Brian Weatherson ("Disagreements, Philosophical and Otherwise"), David Christensen ("Epistemic Modesty Defended"), Stewart Cohen ("A Defense of the (Almost) Equal Weight View"), Bryan Frances ("Philosophical Renegades"), Sanford Goldberg ("Disagreement, Defeat, and Assertion"), Ernest Sosa ("Can There Be a Discipline of Philosophy? And Can It Be Founded on Intuitions?"), Robert Audi ("Cognitive Disparities: Dimensions of Intellectual Diversity and the Resolution of Disagreements"), Jonathan L. Kvanvig ("Perspectivalism and Reflective Ascent"), and Jennifer Lackey ("Disagreement and Belief Dependence: Why Numbers Matter").

Second, as promised in the title, with only one exception (the paper of Ernest Sosa), all of the essays published in this book are new (and, I must add, very skilful and engaging) contributions on the literature of disagreement.

Third, *The Epistemology of Disagreement* offers a comprehensive and fair treatment to all "traditional" topics and debates in the epistemology of disagreement (and especially the debate between "conciliatory" and "steadfast" theorists of disagreement or the problem of disagreement in and about philosophy).

Fourth, this volume also expands in new directions the epistemology of disagreement, by exploring some new dimensions, concepts or problems regarding (or relating with) this common and challenging – both practically and philosophically – phenomenon (such as the problem of cognitive disparity or the intersections between the epistemology of disagreement and the fallibilistic approach to rationality). As such, the book surely responds to its main objective, which, in the words of its editors, is to "help deepen and expand our understanding of some epistemic phenomena that are central to any thoughtful believer's engagement with other believers." (p. 3)

Finally, fifth, David Christensen and Jennifer Lackey offer a very clear and competent introduction to the main topics and debates of the book. Additionally, the essays reunited in *The Epistemology of Disagreement* are very well written and not excessively technical. In consequence, I think that it is quite safe to say that this book will be a valuable instrument not only for experienced epistemologists, but also for students and, in general, for those approaching for the first time the main questions and debates in the epistemology of disagreement.

Philippe Huneman, ed., *Functions: Selection and Mechanism* **(Dordrecht: Springer, 2013)**

Reviewed by Ciprian Jeler[1]

Scholars might nowadays be tempted to think that writing and publishing books might have become somewhat obsolete. In an academic environment where articles seem to have become all-important, publishing books might seem like an almost superfluous enterprise. However, this book, edited by Philippe Huneman and dealing with the meaning that the concept of function has within different scientific domains – from different branches of biology to psychology, neurosciences or within the study of human artifacts and design –, provides a good example as to why books or collective volumes such as this one continue to be important. It is under this light that I will discuss the book here.

The first point to make about the significance of this book is a historical one. A volume of original essays – i.e. unpublished previously – dedicated specifically to the notion of functions is bound to offer a retrospective view about the renewed philosophical interest that this notion has received since the 1970's. Even more so if we take into account the fact that some of the contributing authors (like William Wimsatt or Larry Wright) are among those who have largely been responsible for this renewed interest, by their seminal papers at the beginning of the 1970's. So, whether we look at the contributions of these authors or those of more recent philosophers that engage in new perspectives about functions, this volume provides a good overview of the controversies that have been carried around the notion(s) of function in different fields during the past four decades.

The second point to make about the significance of this volume is a referential one. During the last four decades, the notion of function has been philosophically analyzed in strict connection with the scientific practices and theories of various fields. The interest of publishing a collective volume about this notion today therefore also stems, as Philippe Huneman's introduction shows, from the recent developments within these specific fields. In biology, we might refer to the recent perspectives about what is necessary for the notion of natural selection (is heritability a necessary condition for selection as such? is the connection between fitness and reproduction self-evident?), to the recent accent

[1] Interdisciplinary Research Department – Human and Social Sciences, 'Alexandru Ioan Cuza' University of Iaşi.

put on the multiple levels of selection, to the "extension" of the classical evolutionary theory that is demanded by researchers in evolutionary developmental biology or in ecology; in other fields we might look towards the mechanistic view of science (in philosophy of science), towards theories in cognitive sciences (like "situated cognition") that challenge classical views or towards new insights about the meaning of "realization" in the philosophy of mind. All of these recent developments, as Huneman shows in his introduction, might not easily accommodate older notions of 'function' and are indicating that a re-thinking of this concept might be necessary. The papers published in this volume are therefore called to respond to these new developments in various disciplines, and this certainly constitutes the main merit of the volume.

The last point about the significance of this book that I want to make is a symptomatic one. It is a point about what this book seems to me to reveal *without necessarily intending to do so*. The book is divided in four sections: the first two are dedicated to biology, while the last two sections are dedicated to psychology, philosophy of mind and technology. The interesting symptom that I think becomes obvious when reading the entire volume – as opposed, for example, to reading separate articles in different journals – is an asymmetry between the two main parts of the book. In the first two sections, virtually all the philosophers of biology that sign the papers advocate one form or another of pluralism about functions (the second section is specifically dedicated to pluralism, but the first one isn't, so this shows an interesting regularity). On the other hand, the papers in the last two sections – whether they are dedicated to problems of cognitive sciences or philosophy of mind or to problems in philosophy of technology – vary a great deal more with respect to the positions assumed and to the meanings they attribute to the notion of functions in general. Of course, one could say that the classical etiological theory of functions sparked by Larry Wright in 1972 and later developed by other philosophers that link it directly to natural selection (thus transforming it into what is known as the selected-effects theory of functions) offers a more natural common ground for philosophers of biology. However, the way in which all the contributors to this volume seem to combine – even though, of course, the forms of this 'combining' *do* vary – this theory with versions of the equally classic "causal role" (or "systemic") theory of functions (initiated by Cummins in 1973) is intriguing. I will not try to discuss here the reasons for such a similarity: it might be determined by the scientific field itself that they are philosophically dealing with (biology) or it might be determined by historical reasons. However, given the fact that some of these philosophers defend radically different views about other biological points (Robert Brandon, for example, who

defends a propensionist notion of fitness is highly critical towards the 'statistical' view of natural selection defended by Denis Walsh), this similarity or this partial convergence needs to be noted. It is in the light of this asymmetry between a smaller variation in positions in the first two sections and a greater variation in the last two sections that I will now briefly turn to each contribution.

Fittingly, the volume begins with a contribution by William Wimsatt. He restates his 1972 general form for attributions of function but, outlining some of the consequences of his more recent work on generative entrenchment and robustness, he shows that there is a limit to the differential selection account of functions and a causal-role functionalist might be right to resist it (especially when we situate ourselves in a macroevolutionary setting or an evolutionary developmental biology setting). The causal-role account of function – where it is needed and applicable – is however not right to give up a selectionist account of functions in general, but only a differential selection account.

Denis Walsh's beautiful contribution contrasts causal/mechanical explanations with emergent teleological explanations. In certain fields (human agency, for example), they both provide, he argues, complete and autonomous explanations of the same phenomena, and his reasoning – stemming from Aristotle – is bases upon the acceptance of an "explanatory emergence" that is altogether different from the notion of ontological emergence. He then goes on to show that recent developments in evo-devo (evolutionary developmental biology) appear to indicate goal-directed capacities of organisms that we cannot account for in a causal/mechanical setting without explanatory loss. An emergent teleological explanation would therefore be needed to complement the mechanistic explanation.

Jean Gayon's contribution turns the weaknesses of the selective (or selective-effects) theories of function into their strengths by showing that even if we might have troubles to argue that oxygen or species do have a function within this framework, this renders selective theories of function more dependant on the specific scientific theories on which they are grounded, and hence, possibly, more reliable and more credible. However, in proposing that the apparent shortcoming of the selective theories of functions might be avoided by ascribing functions only to activities (and not to structures, such as oxygen, for example), Jean Gayon too is moving, in my view, towards a conciliation with the systemic or causal-role theory of functions.

Frédéric Bouchard's interesting contribution is based on the notion of ecosystem evolution and on a more controversial thesis that the evolution of ecosystems might be described as natural selection on the persistence of

ecosystems, where the notion of persistence only partly involves reproduction (namely, of the biotal components therein). The intricacy between the reproducing biotal components (whose functions could thus be accounted by a selected-effects theory of functions) and non-biotal elements (whose functions are only accountable for in a causal-role mode, especially in cases of new ecosystem formation by landslides, hurricanes etc.) compels us, he claims, to resort to a realist explanatory pluralism, where both notions of functions need to be pertinent at the same time.

A similar idea is defended by Robert Brandon, by way of a more general demonstration as to why both historical and non-historical concepts and questions are needed both in geology and in evolutionary biology, and this implicates the use of both notions of functions at the same time.

A different version of pluralism – embedded in a "weak realism" – is defended in Philippe Huneman's contribution. The selected-effects theory of function, he argues, is not capable of distinguishing between different properties of one trait in the specific cases where these properties yield equal fitness values. A different method for discriminating between the functions one trait might be said to have is then needed and, he shows, there actually are a number of *different* methods of this sort employed by evolutionary biologists. But these different methods might yield different results, hence the choice of the method will depend on the explanatory interest of the scientist. This entails the position that ascribed functions do belong to biological structures (hence, it is a realist position), but that they are also partly determined independently of the selected-effects theory of functions, so they might be in part explanation-relative (hence, the "weak" realism). This partial explanatory-relative character thus forces us to admit – at least in more complicated cases with multiple possible functional ascriptions – a combination of an etiological and a systemic theory of functions.

Having thus briefly described the contributions dedicated specifically to evolutionary biology (and various connected fields), we will now turn to the last two sections of the book.

Carl Craver's paper is based on examples from neuroscience and on the new mechanical philosophy that he is (co)advocating, and it intends to show the multiple roles that teleological descriptions play in the search for mechanisms. Etiological notions of functions, while useful, can only play a certain part in explaining the very specific and complex phenomena that are studied in neuroscience. It needs to be complemented by functional descriptions that frame constitutive explanations (a function is therefore the factor that organizes and guides our description of how a higher-level mechanism is composed of lower-

level ones) and that help situate the focal mechanism within higher-level mechanisms. In the end, while making use of the notion of etiological function, Craver actually advocates a systemic view of functions that would be compatible with mechanistic philosophy.

A different – if not contrary – view seems to be offered by Carl Gillett. In his bold – while somewhat technical – contribution, he shows that the assumptions of what he calls the Standard Picture of functionalism in the last decades are not acceptable. Basing his argument on an example from neuroscience, he shows that the notion of second-order property in philosophy of mind is not justified, and that properties of the higher-level can be viewed as having causal powers in their own right since their lower-level "realizers" have *qualitatively different* properties that together result in the higher-level realized property. This might be paralleled with the point made in philosophy of biology (for example, by Samir Okasha) that higher-level selection is certainly reducible to lower-level processes, but it is explanatorily autonomous as long as it is not reducible to lower-level *selection*.

Françoise Longy's contribution tries to show that a realist etiological theory of functions should be used in order to explain both biological phenomena as well as the functions of human artifacts. She shows why the idea that the functions of human artifacts are strictly mind-dependant is unfounded and outlines the way in which the introduction of the notions of probability and real kinds might help us provide the necessary conditions for a general and ontological etiological theory of functions applicable both to biological structures and human artifacts.

On the other hand, Pieter Vermaas and Wybo Houkes take the exactly opposite route and try to generalize the "ICE-theory of functions" they have designed over the last few years specifically to account for human artifacts. They analyze multiple ways in which such a generalization could be done in order to encompass biological phenomena, and their conclusion is that only an *epistemic* way is available for such a generalization, namely that of defining functions of biological phenomena not as belonging to these phenomena themselves, but as belonging to the scientific practices that relate to these phenomena and try to explain them. A general application of their ICE-theory of functions, they argue, can be done only if we give up the idea that real functions are present in the biological phenomena themselves. This is certainly a strong and maybe not the most appealing of positions.

The volume is concluded by a very informative epilogue written by Larry Wright, in which he compares teleological biological phenomena with human agency in order to show that they both exhibit an intricacy between a functional

pattern and functional virtues, and that a clear demarcation between these two notions is needed in the future if we are to better grasp the notion of function. This fitting final essay of the paper certainly makes us look forward to reading his forthcoming book on "the concept of a reason".

The brief outline of the chapters of the book that I've just given here, guided as it was by the asymmetry between its initial and its final sections, has insisted a lot on the way the contributions relate to the two main theories of function that have divided the field during the last decades. In this respect, the sketchy presentation I've given certainly doesn't do justice to the quality of these contributions and to how thought-provoking some of them are. But in the end, these last aspects are the ones that make this book a very good read for anyone working in philosophy of biology, philosophy of mind or philosophy of technology. Moreover, the way the notion of function permeates different research fields and connects (or at least partially connects) different elements and notions – from natural selection to agency, from physiologic or neurophysiologic phenomena to human artifacts – certainly makes this book a rewarding read for any philosopher.

Stephen Hetherington, ed., *Epistemology: The Key Thinkers* (Continuum International Publishing Group, 2012)

Reviewed by Cătălina-Daniela Răducu

Which is the most proper way for a student to get familiar with the content of a philosophical domain? One possible answer is offered by the *Key Thinkers* series published by the Continuum International: it provides clear, concise and accesible guides to the key thinkers in each of the central topics in philosophy. The volumes edited so far in this series: *Aesthetics* (edited by Alessandro Giovanelli), *Ethics* (edited by Tom Angier), *Philosophy of Mind* (edited by Andrew Bailey), *Philosophy of Language* (edited by Barry Lee), *Philosophy of Religion* (edited by Jeffrey J. Jordan) and *Philosophy of Science* (edited by James Robert Brown) are all offering comprehensive overviews of the major thinkers who have contributed to the historical development of the philosphical areas they are concerned with.

It is also the case of the volume I intend to review here. *Epistemology: The Key Thinkers*, edited by Stephen Hetherington, provides a survey of the major contributions in the evolution of epistemology, answering the question formulated by the editor in the first chapter of the book: since the history of epistemology is one of its thinkers, which thinkers have mattered most? And if we study their ideas, concepts, theses, will we be able to unlock the past (or future, I might add) of epistemology? I think Stephen Hetherington has found not only the right question but also the right answer that could help not only students but also specialists in epistemology and the general public make a clear image of the evolution of epistemology, presenting its past and making predictions about its future.

The volume includes ten studies, all of them written by well-known specialists in the field: "Plato`s Epistemology," by Nicholas D. Smith (James F. Miller Professor of Humanities, Lewis and Clark College, USA); "Aristotle on Knowledge," by Robert Bolton (Professor of Philosophy, Rutgers University, USA) and Alan Code (Professor of Philosophy, Stanford University, USA); "Ancient Scepticism," by Gisela Striker (Walter C. Klein Professor (Emerita) of Philosophy and of the Classics, Harvard University, USA); "The Epistemology of Descartes," by Desmond M. Clarke (Member of the Royal Irish Academy, and Professor (Emeritus) of Philosophy, University College, Cork, Ireland); "Locke, Berkeley, Hume: Epistemology," by P. J. E. Kail (University Lecturer in the History of Modern Philosophy, St. Peter's College, University of Oxford, UK); "Kant and Kantian Epistemology," by Melissa McBay Merritt (Lecturer in Philosophy,

University of New South Wales, Australia) and Markos Valaris (Lecturer in Philosophy, University of New South Wales, Australia); "American Pragmatism: Fallibilism and Cognitive Progress," by Christopher Hookway (Professor of Philosophy, University of Sheffield, UK); "Wittgenstein on Knowledge," by Paul Snowdon (Grote Professor of Mind and Logic, University College London, UK); "Quine, Goldman and Two Ways of Naturalizing Epistemology," by Ram Neta (Associate Professor of Philosophy, University of North Carolina, Chapel Hill, USA) and "In Gettier's Wake," by John Turri (Assistant Professor of Philosophy, University of Waterloo, Canada).

The studies are intended to, and they do provide, in my opinion, good, accesible and comprehensive presentations of the main themes approached by the thinkers that have contributed to the development of epistemology in its three main stages over the past two-and-a-half millenia, as they were viewed by the editor: Ancient epistemology – Plato: first epistemology (Chapter 2), Aristotle: empiricism and science (Chapter 3), Sceptical doubts (Chapter 4); Modern epistemology – Descartes: metaphysics and science (Chapter 5), Locke, Berkeley, Hume: observation and its limits (Chapter 6), Kant: experience and categories of thought (Chapter 7); and recent epistemology – American pragmatism: inquiry and fallibilism (Chapter 8), Wittgenstein: sense and certainty (Chapter 9), Quine and Goldman: naturalism and science (Chapter 10), Gettier: fallibility and defining knowledge (Chapter 11).

As Stephen Hetherington says, although the volume does not intend to exhaust all the potentially significant ideas in epistemology, its chapters offer „a solid philosophical base from which to explore whatever else within epistemology likewise deserves further attention" (p. 25).

But the volume does not limit to that. In the final chapter, "Epistemology`s Future Here and Now," Stephen Hetherington attempts to take a glimpse into the future of epistemology by investigating its present and retaining a few ideas that are presently being proposed within epistemology and that might have the potential to influence its future and be included in an update of his own book two hundred years hence: proper functioning, virtue epistemology, knowledge and practice, contextualism, and the idea of knowledge-first. These are all new approaches in epistemology that might be a part of its future.

Hetherington admits, at the same time, that „when pondering various epistemological ideas which could, for all we yet know to the contrary, become influential in later years, we may have been reminded of a few older philosophical thoughts" (p. 253). He mentions, in this respect, Linda Zagzebsky`s version of virtue epistemology, practicalism or Timothy Williamson`s conception of

knowledge, each of them linking epistemology with some ancient ideas of philosophy. This is another way of thinking about the future of epistemology: I agree with Stephen Hetherington that some philosophical ideas may deserve to be repeated; also, refining and reinvigorating of them could be a part of epistemology`s future.

And I agree with Stephen Hetherington in another important respect: that there could never be a Final or Complete epistemology, that epistemology could not run out of ideas, because, as he says in the final lines of his book: „Epistemology clarifies *and* puzzles. It enlightens *and* darkens. It confirms *and* challenges. Even outstanding epistemology might do any of this. Epistemology's most mind-catching thinkers calm us *and* surprise us" (p. 237).

And, finally, I do agree with Stephen Hetherington that his volume is evidence of that. Good evidence, indeed.

NOTES ON THE CONTRIBUTORS

Adrian Costache is 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.

Simon D'Alfonso is a recent doctorial graduate from the University of Melbourne. His main areas of research are logic, epistemology and the philosophy of information and computation. His most recent publication is "The Logic of Knowledge and the Flow of Information" (*Minds and Machines*, 2013). Web: theinformationalturn.net. Contact: me@simondalfonso.id.au.

Eugen Huzum, editor-in-chief of *Logos & Episteme*, is Senior Researcher III at "Gh. Zane" Institute for Economic and Social Research, Romanian Academy, Iasi Branch. His main research interests are in political philosophy, epistemology, metaphilosophy, and the intersections between these three domains of philosophical reflection. His most recent publication in english is "Is the Pursuit of Truth the Primary Task of a Political Philosopher?" (*European Journal of Science and Theology*, 2012). Web: https://ices-ro.academia.edu/EugenHuzum. Contact: eugenh76@yahoo.com.

Ciprian Jeler is a researcher at the "Interdisciplinary Research Department – Human and Social Sciences" of the "Al. I. Cuza" University of Iaşi, Romania. His main areas of research are philosophy of biology and contemporary French philosophy. His recent publications include "Do Elliott Sober's Arguments for Group Selection Really Account for the Causal Nature of Natural Selection?" (*Filozofia*, 2013) and "'Résister à la résistance:' l'implacable et la ressource" (*Annales bergsoniennes VI*, eds. Shin Abiko, Arnaud François, and Camille Riquier, PUF, Paris, 2013). Contact: ciprianjeler@yahoo.com.

Thomas Kroedel is a lecturer in philosophy at the Humboldt University of Berlin. His main areas of research are epistemology, metaphysics, and philosophy of mind. His recent papers include "The Permissibility Solution to the Lottery Paradox – Reply to Littlejohn" (*Logos & Episteme*, 2013), "Counterfactuals and the Epistemology of Modality" (*Philosophers' Imprint*, 2012), "The Lottery Paradox, Epistemic Justification and Permissibility" (*Analysis*, 2012), and "Dualist Mental Causation and the Exclusion Problem" (*Noûs*, forthcoming). Contact: thomas.kroedel@hu-berlin.de.

Andrew McFarland is a doctoral candidate in philosophy at the University of Kansas, and a Chateaubriand Fellow at the IHPST (Paris 1 - Panthèon-Sorbonne). His main areas of interest are in metaphysics, philosophy of language, and philosophy of biology. Contact: andrewlmcfarland@gmail.com

Robin McKenna is a postdoctoral research fellow at the University of Geneva. His main areas of research are in epistemology and philosophy of language, but he also has substantial interests in metaethics and philosophical methodology. His recent publications include "Disagreeing about 'ought'" (forthcoming in *Ethics*), "Normative scorekeeping" (forthcoming in *Synthese*) and "'Knowledge' ascriptions, social roles and semantics" (published in *Episteme*, 10, 4, 2013). Contact: rbnmckenna@gmail.com or robin.mckenna@unige.ch.

Teodor Negru is a PhD candidate at Radboud University Nijmegen, The Netherlands. His main areas of research in cognitive science and philosophy of mind are self, free will, agency and mental causation. He has been published in a wide variety of journals such as *Trans/form/ação, Filosofia Unisinos, European Journal of Science and Theology, Cultura: International Journal of Philosophy of Culture and Axiology*. Contact: theonegru@yahoo.com.

Cătălina-Daniela Răducu is Ph. D. Researcher at the Institute of Economic and Social Research "Gheorghe Zane" of the Romanian Academy. She is also executive editor of *Logos & Episteme*. Her main fields of interest are social and political philosophy (especially gender and feminism), epistemology and analytic philosophy. She is the author of *"Ontological Pluralism" and the Idea of "Commonsense" in the Perspective of the Analytic Philosophy* (Lumen, 2007, in Romanian). Contact: cathy_raducu2001@yahoo.com.

LOGOS & EPISTEME: AIMS & SCOPE

Logos & Episteme is a quarterly open-access international journal of epistemology that appears in printed and on-line version at the end of 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 one copy 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.