EAJP

Inaugural Issue

Edited by
Xiao Zhang, Haojun Zhang,
Andrea Baldini

哲学　　　철학　　　　哲学
ジャーナル　저널　　　杂志

East Asian Journal of Philosophy

EΛP

The *East Asian Journal of Philosophy*:

A New Journal for
a New Idea of
Philosophical Dialogue

In the last decade or so, we have often heard that philosophy is becoming "global." Global philosophy wants to engage with different traditions and approaches in order to solve philosophical problems. Rather than self-imposing disciplinary limits to one's research, a global philosopher aims at searching for new horizons to improve our analysis of philosophical issues and to deepen our understanding of other cultures. In this sense, global philosophy invites us to overcome parochial attitudes such as "Eastern vs Western" or "analytic vs continental," and to put different traditions in dialogue in order to face contemporary philosophical concerns.

By looking at philosophy from this global perspective, with the co-founders of EAJP we realized that an open platform of discussion with a geographical focus on East-Asia did not exist. As a forum for theoretical proposals and discussions both within specific philosophical traditions, and across different traditions, EAJP is open to all schools of philosophical thinking broadly construed. This journal has been in the making for quite a while, and the COVID-19 pandemic caused significant delays in finalizing this project. But on behalf of our Boards, and our Editorial Manager, we are delighted to announce the publication of the inaugural issue of EAJP.

In line with EAJP's inclusive mission, we decided that the journal should be published as a fully open-access online journal. Thanks to the generous sponsorship of the *Center for Ethical Studies of Renmin University of China*, EAJP does not and never will require the payment

of author or reader fees. After getting in touch with several publishers, we opted for Sdvig Press. We decided to publish three issues a year: one issue will be open to all contributions, whereas the other two will be dedicated to specific themes.

This inaugural issue consists of invited articles written by a diverse group of prominent global scholars working in different areas of philosophy. We are extremely thankful to them and for their outstanding contributions. Thanks are also due first to our Editorial Committee, our Editorial Advisory Board, and our Scientific Board for supporting so wholeheartedly this project. Finally, we wish to thank our Editorial Manager, Andrea Altobrando, whose professionalism, attention to detail, and good humor has made this journey a real pleasure.

<div align="right">

Xiao Zhang
Haojun Zhang
Andrea Baldini
Editors-in-Chief
East Asian Journal of Philosophy

</div>

Saul Kripke on Belief Ascription

Mircea Dumitru
University of Bucharest*

ABSTRACT | The paper critically examines the issue of whether failure of attributing propositional attitudes, such as belief ascriptions, is intrinsically failure to substitute codesignative terms (names) within opaque contexts. One open question is whether Fregean "senses" still have any role to play within semantics. This question is more acute in the aftermath of Kripke's (1980), because one of the most remarkable achievements of the arguments Kripke deploys is the demolition of the Frege-Russell tradition in regard to the semantics of names. The descriptive theory of names, the view that names are disguised definite descriptions, and, in general, the Fregean mechanism of referring through senses that are expressed by definite descriptions associated with names have become obsolete. And there has been a growing consensus among logicians and philosophers of language that as long as we stick to direct discourse, to transparent contexts, and to the modal case, the Millian view on proper names that Kripke has advocated so forcefully cannot be plausibly denied. Still, having in view the peculiar consequences that Kripke's Millian view on names may have for opaque contexts, and particularly for the interpretation of the substitutivity of codesignative singular terms (especially names) within propositional attitude ascriptions, one may wonder, and consequently, one can legitimately ask, whether Fregean "senses" don't have still something to offer. Thus, examining and assessing Kripke's (1994) one may legitimately wonder whether there is still a case for Fregean senses within semantics of belief ascriptions. The paper examines aspects of the dialectics which is going on in this debate and points at some open issues which are worth pursuing in the spirit of Kripke's (1994).

---

* *Correspondence:* Mircea Dumitru – Faculty of Philosophy, University of Bucharest, Splaiul Independenței no. 204, 060024 Bucharest, Romania. Email: mircea.dumitru@unibuc.ro

1   Introductory Remarks

As we know, a piece of reasoning such as the following:

> (BC) Lois believes that Superman can fly;
>
> Superman = Clark Kent;
>
> therefore, Lois believes that Clark Kent can fly,

raises the question of why the outcome of applying a sound rule of first-order logic with identity, viz. "= E," within a belief ascription context is an unsound argument. For assuming the story as a fact, and depending upon contingencies regarding Lois' beliefs, she still believes (or at least can believe) that Clark Kent cannot fly, although she believes that Superman can fly, and actually Superman *is* Clark Kent.

Whatever the semantic account of this puzzle, what seems reasonable to resist to is: (i) to say that provided the truth of he premises of (BC) the belief ascriber is entitled to draw the above displayed conclusion, and (ii) to impute to Lois (or in general to whoever is the subject of that belief attribution) contradictory beliefs, if, as it happens, she believes that Clark Kent cannot fly, although Superman *is* Clark Kent, and she believes that Superman can fly.

However, since what is involved in particular in drawing the conclusion of (BC) above seems to be what one can call a principle of substitutivity of coreferential or codesignative names in belief contexts, and since it seems to imply that as a result of its application Lois does have contradictory beliefs, one can construe (BC) as an instance of a *reductio ad absurdum* of that substitutivity principle within belief contexts. Simply put what (BC) shows is that the principle of substitutivity of coreferential names in belief attribution contexts fails.

Then, it seems legitimate to ask that in response to this semantic phenomenon we come up with an explanation of the mechanism that blocks the application of a rule that, save for propositional attitude contexts, quotation contexts and the "so-called" context, is all right.

However, if we look at (BC) as a case in which we cannot ascribe a belief to someone solely on logical grounds precisely because one is not allowed to apply "= E" within belief ascription contexts then one is inclined to think that that substitutivity principle is the heart of the matter when troubles with belief ascriptions occur. And to say this boils down to countenancing the view that in any troublesome belief ascription context a substitutivity principle is essentially involved.

The drawback of such an explanatory route is that it seems to identify *a priori* the class of phenomena of failure of ascribing beliefs on logical grounds with the class of those arguments in which failure of substitutivity of coreferential names

seems to be what is at stake. But as it is well known, Kripke has shown that there are cases in which one has serious problems with attributing to someone beliefs when *no* substitutivity at all is involved. Or at least not in an apparent way.

Thus, it seems to be more cautious not to prejudge the solution to the problem and, accordingly, not to address the problem under the tag "the failure of substitutivity of coreferential names within belief contexts," although "Why we cannot apply such a rule in propositional attitude contexts?"[1] still remains a legitimate question, but to pursue it under the more neutral label of failure of ascribing beliefs solely on logical grounds. This way, we leave open that the latter failure be the outcome of the collapse within beliefs ascription contexts of certain semantical principles, including, but not being reducible to, substitutivity of identicals, whose application outside such contexts is unproblematic.

Thus, keeping in mind the central position that failure of substitutivity of coreferential names within belief ascription contexts has in our topic of belief ascription, one can start addressing the problem of failure of belief ascription solely on logical grounds insisting on the interpretations provided for that failure of applying "= E" in belief ascription contexts. The problem is open, in principle, to several interpretative options of which two are more significant for the semantics of propositional attitude ascriptions.

Either the rule "= E" (also known as Substitutivity of Identicals) fails to apply to such a context, and accordingly any such illegitimate application has as outcome an unsound argument, or else, appearances notwithstanding, the rule is not essentially involved in those cases in which ascriptions of a belief is a *non sequitur* from previously established premises, one of which properly ascribing a belief and the other stating identity between a name which occurs in that ascription and another codesignating name, and thus some other semantic principles might be responsible for failure of ascribing beliefs.

In the former case we have to provide a logical and semantic analysis of that

---

[1] Depending on the construal of Kripke's cases and proposals, one can adopt either a revisionary view according to which, our semantic intuitions notwithstanding, "= E" does apply within belief ascription contexts, but there are other semantic principles whose breaking down within those contexts obscures that fact, or else one can adopt the view that "= E" is illegitimately applied within belief contexts, but nevertheless the core of the phenomenon of failure of belief ascription on logical grounds is not to be identified with failure of substitutivity of coreferential names, on pain of ignoring the existence of puzzling cases of belief ascription contexts in which no substitutivity principle is basically involved. Whatever the interpretation of Kripke's puzzling cases of belief ascription, I believe that the former option, which seems to be embraced by Kripke, still requires an explanation of at least why do we have such a strong intuition that in belief attribution contexts "= E" cannot be applied. It seems to me that we can consistently held that "= E" fails to apply within belief ascription contexts, but nevertheless that the failure of ascribing beliefs on logical grounds is not to be identified with failure of substitutivity of coreferential names within those contexts.

mechanism which is supposed to be responsible for the phenomenon of failure of substitutivity. It is worth noticing that this option doesn't necessarily boil down to the view that "= E" is unsound. The point is rather that a propositional attitude context is such that an application of the rule "= E" within it is upsetting some other assumptions (premises) than those that license a validity-preserving application of the rule. Hence, it is a *mis*application of the rule (perhaps something similar to a fallacious argument), but the rule is all right. Of course, one may held that an application of "= E" *to a propositional attitude context* is unsound. Likewise, we don't challenge the metaphysical principle known as Leibniz's Law, or the Indiscernibility of Identicals that the formal rule implements.

In the latter case, we face the challenge of showing that, contrary to some semantic intuitions that are shaped by compelling evidence, failure of the application of "= E" to premises embedded within a propositional attitude context is not *the* right diagnosis for what happens when from reporting a belief and from stating identity between two codesignative names, one of which occuring into the belief report, we are not entitled to draw any conclusion whatsoever as to the same content of the ascribed belief save for a substitution of one of the two names for the other. Hence, failure of substitutivity is a misnomer for the problem under consideration, since certain other semantic principles that are unproblematic when they are applied in transparent contexts may bring about paradoxical results in an area of ascription of belief.

The remarks that I put forward in this paper are the results of an exploration of both explanatory alternative options. The main point under which that exploration falls is that a unitary explanation of the failure of ascribing beliefs on logical grounds is desirable, and that the reason for that failure may be found, appearance to the contrary notwithstanding, in some structural features of a belief context which are such that they block the use of a substitutivity principle that outside such a context is sound and unproblematic.

Accordingly, one has to show how arguments such as Kripke's are name-sensitive, i.e., are such that, although they are not based on a principle of substitutivity, they occur in a context which bloks the replacement of an occurence of a name (in a belief sentence) with any codesignative name which the subject of that belief attribution is unaware of fitting in her entertained belief, being thus, on a par with similar arguments resting, however, on a substitutivity principle.

The gain of such a strategy seems to be not only in matters methodological, viz. a unitary frame for explaining diverse failures of belief ascriptions on logical grounds, but in motivating a view as to the reasons of that failure, as well. To my understanding, that is a view that should give a proper due to some intension-based concepts, being as such akin to a Fregean explanation of why codesignative

names can't be substituted in propositional attitude, and hence, belief ascription contexts.

I shall begin with a short diagnosis of the problem. After Kripke's *Naming and Necessity*, no discussion of a philosophical topic concerning sense, reference, and attribution of attitude can possibly be the same as before. Although the gist of the doctrine defended in (Kripke 1980) bears heavily on the semantic mechanism of naming and referring in transparent and modal contexts, its revolutionary import has challenged some of the background suppositions of the theories developed for ascription of belief, as well. Particularly, its main tenet that even within approaches of propositional attitude context "[t]he spirit of […] a Millian line (advocated by Kripke in his earlier *Naming and Necessity*) should be maintained as far as is feasible" (Kripke 1994, p. 360).

Consequently, it seems to me that in Kripke's analysis one can trace two main critical arguments against a Frege-Russell theory of belief ascriptions. The first retains the assumption on which the Frege-Russell account rests, viz. that problems with belief ascriptions are caused by failure of substitutivity of codesignative names, and the argument is that a Fregean analysis of belief contexts falls short of an account of the phenomena to be explained.

The second is more revisionary and destructive because it challenges even the assumption which the Fregean analysis is based upon. To this effect Kripke coins certain examples which purport to show that similar problems with belief reports and ascriptions occur when no substitutivity principle is involved, and that some other principles that are benign when used outside belief ascription contexts, viz. a disquotational principle and a translation principle, may be the cause of our failure of attributing beliefs.[2] Thus, those principles themselves and not a substitutivity principle may break down when we come to such kind of contexts.

The point that my paper is making is that cases like those envisaged by Kripke

---

[2] The Disquotational Principle (DP) is a devise of bridging sentences uttered or otherwise entertained by somebody with the content of her beliefs thereby expressed. It comes in two forms: (i) the weaker form states that the sincerely assent, made on reflection, to a sentence 'p' by a normal English (or, for that matter, of any other natural language) speaker who is not confused by whatever lexical error is a sufficient condition for attributing her a belief with that content which 'p' expresses; (ii) the stronger form makes the former sufficient condition a necessary condition as well.

The Principle of Translation (PT) guarantees the preservation of truth for any true sentence under any translation of it into a sentence of any other language, including the very same language, in which case a homophonic translation of that sentence from one idiolect into the idiolect of the translator is allowed.

However, in order that the DP and PT work in cases in which sentences belong to other natural languages than that of the belief ascriber, a Tarskian disquotational principle for truth (usually left tacit), viz. '*p*' is true iff *p*, where '*p*' inside and outside the quotation marks is to be replaced with any sentence of that language, is also needed.

trigger a name-sensitive context which blocks the use of a substitutivity principle, although no appeal to a substitutivity is overtly at stake in the construction of those cases.

Of course, this is to say neither that Kripke's point against substitution as the heart of the matter in belief ascriptions is misguided, nor that he fails to see how his cases can be construed as being essentially the same as cases in which failure of a substitutivity principle is obvious.

It is only fair to pause here and to notice that Kripke himself is completely aware of the parallelism between his "*Londres*/ London" and "Paderewski" cases and that about Jones assenting to "Cicero was bald" and not assenting to "Tully was bald" although Cicero *is* Tully, which is obviously based on a substitutivity principle.[3]

My point, though, is a methodological one, and with a bit of substantive import. Namely that since Kripke's account of cases of failure of belief ascriptions, no matter how instrumental in showing that substitutivity is not the heart of the matter, is not decisive as to what is *that something* which prevents us from drawing conclusions from statements of belief ascriptions, then why not stress as well upon Lois' case where *no failure of such principles as DP and PT is basically involved* more than *the obvious failure of substitutivity*, and, accordingly, why not say that failure of substituting a name with a codesignative name is a salient feature of the contexts in which Kripke' own examples are embedded, coming thus to the idea that failure of substitutivity is *a* paradigmatic (though, maybe, not *the* paradigmatic) case in the understanding of the mechanism of failure of belief ascriptions?

Then, following this line of argumentation, one can try to show how in a context such as Lois' or Peter's, the structural features of the situation which obtains for the subject of the belief ascriptions are such that they prevent her or him, and, consequently, the belief ascriber as well, from proceeding to a substitution of a name which occurs in that context with a codesignative name.

Simply put, Kripke's cases may well not be based on any substitutivity principle, but they are generated within, or are embedded, or at least they trigger a context that blocks substitutivity. And hence, methodologically we are not misguided to address the topic of failure of belief ascriptions through one of the prominent feature of the context in which that failure occurs, viz. through failure of substitu-

---

[3] Kripke is very clear on this matter: "It will intuitively be fairly clear, in these cases, that the situation of the subject is 'essentially the same' as that of Jones with respect to 'Cicero' and 'Tully'. Moreover, the paradoxical conclusions about the subject will parallel those drawn about Jones on the basis of the substitutivity principle, and the arguments will parallel those regarding Jones. Only in these cases, no special substitutivity principle is invoked" (Kripke 1994, p. 364).

tivity of codesignative names within those very contexts.[4]

2   Is Failure of Belief Ascription Intrinsically Failure of Substitutivity?

Whichever way we go, either Frege's or Kripke's, the question of the logical significance that we attach to "= E" in propositional attitude ascription contexts involves a lot more than a formal problem to be dealt with within the confines of a natural deduction system. For it carries its load over topics in the philosophy of language and of logic such as meaning, reference, quantification, and various kinds of modal locutions.

For the purposes of the present paper, one open question that is of a genuine interest is whether Fregean "senses" still have any role to play within semantics. This question is more acute now, in the aftermath of (Kripke 1980), because one of the most remarkable achievements of the arguments he deploys in (Kripke 1980) is the demolition of the Frege-Russell tradition in regard to the semantics of names. The descriptive theory of names, the view that names are disguised definite descriptions, and, in general, the Fregean mechanism of referring through senses that are expressed by definite descriptions associated with names have become obsolete. And there has been a growing consensus among logicians and philosophers of language that as long as we stick to direct discourse, to transparent contexts, and the modal case, the Millian view on proper names that Kripke has advocated so forcefully cannot be plausibly denied.

Still, having in view the peculiar consequences that Kripke's Millian view on names may have for opaque contexts, and particularly for the interpretation of the substitutivity of codesignative singular terms (especially names) within propositional attitude ascriptions, one may wonder, and consequently, one can legitimately ask, whether Fregean "senses" don't have still something to offer. Thus, isn't there a case for Fregean senses within semantics of belief ascriptions (to name just for convenience one of the propositional attitude ascriptions)?

Kripke's arguments to the effect that there are no Fregean senses develop a robust Millian view on names, and consequently, their target is the Frege-Russell descriptive theory of proper names.

According to the Millian view, the sole semantic function of a name is to name/ refer to its bearer. A name is a tag for the object it names.

---

[4] In the analysis and interpretation that I put forward in my paper I also draw on (Forbes 1994) and (LePore and Loewer 1990).

A name doesn't refer to its bearer through the senses that are expressed by some definite descriptions associated to it. It directly refers to its bearer. The mechanism of getting a referent for a name consists in two parts: there is an act of dubbing, in which the name is introduced, and then the referent of the name is passed from one speaker to another through a chain that under normal communicational circumstances leaves the referent unchanged, although it may very well change the meanings that normal speakers associate in their idiolects with that name.

Unlike a definite description, a name doesn't describe the object that uniquely satisfies certain identifying properties. Against Frege's doctrine, Kripke argues that a definite description neither gives the meaning of the name that can be put into correspondence with it, nor fixes its referent.

And to finish this sketch of Kripke's view on names, let's add that unlike definite descriptions that can pick up different objects in different counterfactual situations, a name is a rigid designator that refer to the same object in all the counterfactual situations in which its bearer exists.

No matter how well is defended this Millian view on names within transparent and modal contexts, a strong feeling persists that its consequences with respect to contexts involving belief ascriptions are odd. To mention just one of them: names, unlike definite descriptions, are interchangeable within belief context. For, as Kripke puts it, "[w]hether a given subject believes something is presumably true or false of such a subject no matter how that belief is expressed; so if proper name substitution does not change the content of a sentence expressing a belief, coreferential proper names should be interchangeable *salva veritate* in belief contexts" (Kripke 1994, p. 354).

But the problem is that inter-substitutivity of codesignative proper names in belief contexts far from being truth-preserving seems to be at odds with our semantic intuitions.[5] Lois sincerely assents to "Superman can fly," but not to "Clark Kent can fly" even though Superman is Clark Kent. Thus, she believes that Superman can fly, but she does not believe that Clark Kent can fly.

Now, it is obvious that if a strict Millian view is correct the two propositions referred to by the terms "that Superman can fly," and "that Clark Kent can fly" should be identical. For, as Kripke says commenting upon this consequence of the

---

[5] One note on scope disambiguation is in order here. When I speak about failure of substitutivity of codesignative singular terms within belief contexts, I consistently assume, as it is done in the literature, that belief locutions are read as having large scope, or alternatively, although maybe not equivalently that the beliefs are construed *de dicto*. For if we read them as having small scope, or as being *de re* beliefs, then substitutivity of codesignative singular terms succeeds, and accordingly, the arguments whose conclusions are derived by applying "= E" are valid.

strict Millian view on proper names, if "the linguistic function of a proper name is completely exhausted by the fact that it names its bearer, it would appear that proper names of the same thing are everywhere interchangeable not only *salva veritate* but even *salva significatione*: the proposition expressed by a sentence should remain the same no matter what name of the object it uses" (Kripke 1994, p. 353).

However, adapting a kind of Russellian argument, one can say that the two propositions: "that Superman can fly," and "that Clark Kent can fly" are not the same, unless they have all their properties in common. But the proposition that Superman can fly has at least one property that the proposition that Clark Kent can fly lacks, namely that it is believed by Lois.

Of course, this doesn't necessarily impute any inconsistency to Kripke's view on proper names, because he doesn't advocate in his (Kripke 1980) a principle of substitutivity for propositional attitude ascriptions or epistemic contexts. However, it is arguable that Kripke can't be quite happy and comfortable with his Millian view on proper names since propositional attitude ascriptions and epistemic contexts threaten to show that his refutation of Frege is not as total as he might have wanted.

More interesting, though, it is to figure out whether the supposition on which Kripke successfully built the interchangeability of coreferential proper names in modal contexts does not have as a consequence the countenancing of a Fregean position in regard to belief ascriptions. To make his case regarding the thesis of rigidity of proper names from which it follows the interchangeability in modal context of coreferential proper names compelling, Kripke has to make a sharp distinction between metaphysical modalities and epistemic modalities.

For, to hold that "Superman" and "Clark Kent" name the same person only if the sentence "Superman is Clark Kent" *is necessarily true*, despite the fact that Lois *doesn't know it*, one has to support the view that not all necessary truths are known *a priori*, and hence, that there are necessary truths known *a posteriori*. And what Lois can find out is that she kept referring to the same person using, without being aware about that, two different names. But doesn't this support the Fregean view that in referring to one object one can use different "ways of thinking" (or "modes of presentation") of that object?

Roughly speaking, one sensible conclusion to draw is that in contexts in which *our knowledge* of the truth of our sayings about individuals is sensitive to *our* ways of referring to them, there is a place for the modes of fixing the referents of the names we use. And this, on a reasonable reading, is tantamount to accepting a place for Fregean "senses" as modes of fixing the referents of proper names. The supposed compromise sought here would consists in accepting that a name is a

rigid designator, and that its associated definite descriptions are not synonymous with it, but that they are used nevertheless in order to fix its reference.

What will happen, then, to Kripke's doctrine?

> So it appears that even though, according to my view, proper names would be *modally* rigid – would have the same reference when we use them to speak of counterfactual situations as they do when used to describe the actual world – they would have a kind of Fregean 'sense' according to how that rigid reference is fixed. And the divergence of 'sense' (in this sense of 'sense') would lead to failures of interchangeability of codesignative names in contexts of propositional attitude, though not in modal contexts. Such a theory would agree with Mill regarding modal contexts but with Frege regarding belief contexts. The theory would not be *purely* Millian. (Kripke 1994, p. 356)

This diplomatic attitude that urges a compromise between a Millian and a Fregean view, seems to be undermined, according to Kripke, by some semantic facts that point against the Fregean assumptions of the theory regarding belief contexts. Let's take a brief look at them:

One way of looking at the failure of interchangeability of any two codesignative names within belief contexts seems to support the view that its main source consists in the existence of two different associated descriptions, or modes of determining the reference of the names. Were the subject of a belief attribution able to grasp that the two distinct definite descriptions uniquely identify the same object, she would proceed without further ado to the interchangeability of the two names within the belief context. However, this is not something to find out by pure logic, or by merely introspection. Empirical investigations are required.

This move has two salient features: (i) it provides the Fregean theory with a response to the objection that most names don't have conventional senses shared by all the speakers of a linguistic community; however, the price that this response imposes upon those that embrace it is the view that there is no sense shared by all in the community, but only a community-wide reference; (ii) it fares rather well with the Frege-Russell view that names belong to individuals' idiolects, and their senses are dependent upon certain descriptions associated with them.

Now, leaving aside all the objections that have been raised against this twist of the Frege-Russell theory toward a cluster-of-descriptions theory of names, the most serious objection that Kripke opposes to a Freagen analysis of belief contexts is its failure to "account for the phenomena it seeks to explain" (Kripke 1994, p. 356).

The main Fregean dogma that the sense of an expression determines its reference is at odds with a simple linguistic fact, namely that people sometimes do not know about individuals whom they intend to refer to anything that can be used for uniquely identifying them. So, they cannot associate with a name some definite descriptions that uniquely identify the bearer of the name. However, this doesn't hinder them from using names with a particular determinate reference in their minds.

Thus, to cope with the shortcomings brought about by their lack of specific knowledge about individuals whom they intend to refer to, people usually attach to the names they use indefinite descriptions. Now, as Kripke's "Feynman - Gell-Mann" example shows, the senses, if they can be called so, that indefinite descriptions express are identical. Henceforth from the theory it follows that the referent of one name is the same as the referent of the other. But it is obvious that the speakers cannot say whether or not the two names are codesignative only by looking at the senses that the descriptions express. And actually, it is quite plausible that in cases similar to "Feynman - Gell-Mann" example the names refer to distinct individuals.

I am not sure that Frege would have endorsed the view that an associated *in*definite description is successful in *fixing* the reference of a name, let alone in giving its meaning. However, it seems to me that the fact that a definite description is not forthcoming is not sufficient to block our referring to the bearer of a name, whichever, or whoever is that. And this semantic intuition is exploited by Kripke in his "Feynman - Gell-Mann" example against the premise that we consider now of the Frege-Russell theory, viz. that failure of interchangeability of codesignative names is produced by the difference between the descriptions associated with the names.

Hence, although Fregean senses were promising at one step in the investigation of the ascription of belief, it seems now that they are of no real help for us. "So the apparent failure of codesignative names to be everywhere interchangeable in belief contexts, is not to be explained by differences in the 'senses' of these names" (Kripke 1994, p. 359).

Now, having shown that a Fregean theory of names is not after all very helpful in understanding what is happening when we fail to attribute beliefs, Kripke launches a more destructive attack. This time the target is the very idea that the impossibility of substituting a name for another codesignative name in an intensional context lies at the very root of failure of belief ascriptions, regardless what kind of theory intended as an account of failure of substitutivity is considered.

To this effect, Kripke introduces and discusses several cases in which failure of belief attribution is obvious, but no substitution of a name for another name is

apparent. The series of cases is set up in such a way that with each new case the sense-based account of failure of belief attribution and the identification of this latter problem with failure of substitutivity of codesignative names gradually lose their initial credibility they might have had.

Thus, Pierre's "*Londres*/London"[6] case and its variants to be briefly discussed in the sequel purport to show that we are in a predicament if asked what Pierre believes. And this is a puzzle because from his assent to "*Londres est jolie*" by (DP) and (PT) we can infer that Pierre believes that London is pretty, whereas from his assent to "London is not pretty" by the same two principles[7] we are led to the conclusion that Pierre believes that London is not pretty. And the facts are such that Pierre cannot be blamed for any contradiction whatsoever, for as long as he doesn't know that the city he calls "*Londres*" is the same as the city he calls "London" he can't figure out, using logic only, that his beliefs are inconsistent, and thus, that at least one of them must be false.

As Kripke himself points out, "[Pierre] lacks information, not logical acumen. He cannot be convicted of inconsistency: to do so is incorrect" (Kripke 1994, p. 368).

But, then, one can ask how Pierre's position fails to be the same as Lois'?

For what Pierre is in no position to do is to substitute in the translation into English and French, respectively, of either of the two sentences he endorses the English counterpart for "*Londres*" or the French counterpart for "London." And Pierre can't operate this substitution because he lacks precisely that piece of knowledge that "the cities he calls 'London' and '*Londres*' are one and same ..." (Kripke 1994, p. 368).

Couldn't we say, then, that Lois' and Pierre's position don't differ at all, at least with respect to the possibility of making a substitution? And then, isn't Pierre's "*Londres*/London" case a case of failure of attributing belief due to its occurrence in a situation in which a substitutivity of codesignative names fails? I am inclined to see the things in this way.

But be that as it may, Kripke, however, agrees that these paradoxical conclusions about Pierre parallel Lois' predicament. Only that in Pierre's case, no overt recourse to a substitution principle is made.

And on the other hand, as Kripke repeatedly emphasizes, no other equivalent description of Pierre's situation, which is both possible and not conducive to any paradoxes, will do as a solution to the original puzzle. For the puzzle is: Does

---

[6] Cf. (Kripke 1994, 365–372).

[7] Because, for reasons provided by Quine, a homophonic translation from the idiom of the utterer into that of the interpreter should be taken into account when dealing with an interpreter's attempt at understanding what somebody who belongs to the same linguistic community as hers or his is saying.

Pierre believe or does he not believe that London is pretty?

Once again, according to Kripke, the temptation to see this puzzling situation through Fregean lenses, as an outcome of associating two distinct sets of identifying properties with the same name[8], [9]should be resisted for two main reasons.

First, because even if we concede that this Frege-Russellian explanation shows what Pierre's case really boils down to, it is not an adequate answer to the problem we started with, viz., the behaviour of names in contexts of belief ascriptions.

To my understanding, though, the way Kripke emphasizes the problem we started with, viz., "Does Pierre, or does he not, believe that London (not the city satisfying such-and-such descriptions, but *London*) is pretty?" (Kripke 1994, p. 370) – shows his strong penchant toward assimilating names which occur within contexts of belief attributions with rigid designators. But this is question-begging, because this is precisely the point to be argued: are names in contexts of belief attribution behaving like rigid designators (thereby allowing for a sort of substitutivity principle), or like descriptions (and thus we have at least part of the reason why subtitution fails in such contexts)?

Moreover, Kripke's question what does Pierre believe "not [about] the city satisfying such-and-such descriptions but [about] *London*" (Kripke 1994, p. 370), seems to be more appropriately addressed in a context in which, unlike the context of *de dicto* beliefs which is Kripke's own concern, we are concerned with *de re* beliefs.

Second, according to Kripke, the Frege-Russell approach will not do because even if we grant its basic notion that the puzzle arises from the fact that two differ-

---

[8] A more accurate statement is this: two sets of identifying properties which are expressed by two different definite descriptions are associated respectively to each name from a pair of names which belongs in the class of equivalents of all the pairs of names, the former of which is a given name and the latter of which is one of its translation into another language. Of course, there is something odd in saying that a name has a *translation*, but nevertheless this is the case with names of famous people and places. More often than not the translation consists rather in a phonetic variant of the original name due to peculiarities of pronunciation and to the need of matching sounds from one language into more or less equivalent phonemes of other languages than in a genuine translation of the meanings of names. For, if we endorse Kripke's doctrine on names, these lack any meanings which could be rendered either by a synonymous definite description or by a reference-fixing definite description. However, the qualifier "more often than not" is intended as a reminder of those cases in which those which are currently considered as names were in the beginning, when the baptize took place, nicknames. "Richard Lion Heart" or "Vlad the Impaler," which are names in their own right, ask for a matching of the meaning of their descriptive component part (" ... Lion Heart", and " ... the Impaler") into an equivalent meaning in the target language when they are translated from the original language in which they were coined.

[9] For it seems very naturally to say that had Pierre known that the two distinct sets of identifying properties, which he had learnt in France and in England, respectively, pick up the same city, he would have been in the position to spot an inconsistency within his beliefs using his logical acumen.

ent sets of identifying properties have been associated with "*Londres*" and "London," respectively – an idea that may retain some force regardless the strong arguments against identifying descriptions as being synonymous with, or at least as fixing the reference of, names – and we try to fix the problem by letting only one and the same description be associated with both names, the same type of puzzle will reoccur at another level: the puzzle will involve names which occur at the level of those identifying definite descriptions.

Kripke's point can be interpreted again as being against a substitution based diagnosis of the problem. No constraint along Fregean lines to the effect that Pierre associates the same identifying properties with both "*Londres*" and "London" will do as a solution of the paradox: for, suppose that a definite description phrased in French fixes the referent of "*Londres*" and its translation in English fixes the referent of "London." Now, the same indeterminacy with respect to the referent of a name within a context of belief ascription and of its translation can reoccur, only this time at the level of the names which figure within the uniquely identifying description.[10]

Since the same problem concerning the indeterminacy of the referent of codesignative proper names within contexts of belief attributions can reoccur at any subsequent level at which a "defining" definite description is introduced (because in any such descriptions other names occur, and eventually one would expect demonstratives and indexicals to appear), the hope of a description theorist to provide a Fregean solution to these paradoxical cases is that eventually she can reach an ultimate level where the defining properties are "pure" properties whose expression do not require proper names.[11]

---

[10] For instance if "*Angleterre*" and "*le Palais de Buckingham*" (pronounced "*Bookeengam*") occur in the French definite description which Pierre associates with "*Londres*", then after he learns English by direct method, he can associate with "London" the exact English translation of the French description in which "England" and "Buckingham Palace" will occur throughout in the same places in which "*Angleterre*" and "*le Palai's de Buckingham*" occur in the French description. Now, assuming that Pierre has never seen England and London, but that he was shown some pictures of picturesque parts of the city which is named by a name whose referent is picked up by the French definite description, and then, after he learnt English, he was shown some pictures of unattractive parts of the city which meets the English definite description, then Pierre can consistently believe that London is pretty (because he assents to the French sentence "*Londres est jolie*") and that London is not pretty (because he endorses the sentence "London is not pretty"). And the problem now is the same as before except for the fact that it involves names which occur in the uniquely identifying definite description: Pierre can't figure out that "*Angleterre*" and "England," and "*le Palais de Buckingham*" and "Buckingham Palace" pick up the same referent, respectively.

[11] But in view of the possibility of extending Kripke's puzzling cases to natural kind terms (cf. (Putnam 1975) and (Kripke 1980, 1994)), in the defining descriptions at that presumed ultimate level couldn't occur natural kind terms either.

However, no convincing and plausible proposal as to how such an ultimate level can be reached has been advanced, or, if supposing that it can eventually be reached, how can the properties expressed by such name-free descriptions be still uniquely identifying after all names and related devices for unique referring have been eliminated.

The only way out from this predicament is to endorse a rather extreme philosophical reform which consists in banning the translation of names.

Actually, apart from any point as to the merit of this proposal, to acknowledge that the only solution consists in forbidding translation of names, which in this cases only means to preserve names as they are in their original phonetical guise, fits quite well with my idea that for reasons of not jeopardizing the content of a belief the context within which that ascription of belief is realized should be such that a substitution of a name for a codesignative name is forbidden, unless the subject of that belief attribution is aware of (has knowledge of) the fact that the two names are codesignative.

Although this proposal can be reckoned as being highly implausible in face of our customary practice of learning the names of famous people and places as names within our own language, the proposal, if it is not sentenced before it is judged, looses its *prima facie ad-hoc* character.

For, the practice of translation of names is limited only to names of some famous people and geographical localities, and to ask for a reform, in which we stick to non translated names within contexts of belief ascriptions for the sake of avoiding puzzling cases about belief, is nothing else than to generalize a practice which is in use for the vast majority of names.[12]

Besides, what one gets by translating such a name is a phonetic variant in the target language of that name as it occurs in the original language, and thus, what runs against this proposal is not something pertaining to semantical analysis, but to the weight of tradition.[13]

However, it is not so clear how well this proposal could work. For, suppose that for philosophical reasons concerning reports of beliefs we adopt the policy of non-translatability of names. But once in force, how far does this principle reach? If Putnam's and Kripke's work on the semantics of natural kind terms adequately reveals analogies between proper names and natural kind terms, then one point in case seems to be the similar behavior of both sorts of terms within belief reports. Should we, then, extend a similar policy of non-translatability to

---

[12] As Kripke puts it: "At least the restriction in question merely urges us to mend our ways by doing *always* what we presently do *sometimes*" (Kripke 1994, p. 374).

[13] For the problem raised by the translation of what initially were nickuames, but are currently deemed as names see footnote 7.

natural kind terms? If similar problems ask for similar approaches, then one can't see how to adopt this proposal for one sort of terms, and deny it for the other sort. But this brings the whole idea to an extreme implausibility.

What is typically involved in all these cases of failure of belief ascription is a lack of awareness on behalf of the subject of ascription that two names, or two tokens of the same name in different languages, refer to one individual. For otherwise, if she knew this, she would be able to spot a contradiction within her beliefs by her logical acumen, only.

Is it, then, necessary that in all such cases two distinct tokens of the same name, one of which being the translation of the other, or two distinct tokens of two names occur? Moreover, is it the case that translation from one language into another is essentially involved in Kripke' cases, leaving, thus, open the possibility that the reason of those paradoxically results be that it is something wrong with translation?

Surprisingly not. As the "Paderewski" case shows the same puzzle still arises even if in reporting beliefs we stick to one language only, and even if we confine ourselves to phonetically identical tokens of a single name. In a way "Paderewski" parallels "*Londres* /London" case, and just because no Principle of Translation is needed now,[14] but only the Disquotational Principle, this new version of the puzzle shows more clearly that what is involved in the context of Peter's learning of the referent of the name "Paderewski" is a fundamental ambiguity and that this ambiguity is one source of the puzzle and of our predicament in attributing beliefs to Peter.

For Peter assents to "Paderewski has musical talent," and hence believes that, consecutively to his learning of the name "Paderewski" through a description (definite or indefinite, it doesn't matter) which fixes the referent of that name in the person of the famous pianist.

Then, when he assents to "Paderewski has musical talent," and accordingly he believes that without changing his mind about Paderewski the musician, and without being responsible for any inconsistency, he has introduced the referent of the same name by a referent-fixing description like "the Polish nationalist and Prime Minister."

And because there is no *a priori* guarantee that "Paderewski" names the same individual, for a name in a natural language can denote ambiguously and has more than one denotation[15], it would be simply fallacious for Peter to say and be-

---

[14] Save for those who may make a case for a homophonic translation from Peter's idiolect into the idiolect of the belief ascriber, but even then it is not a translation from one *language* into *another* one.

[15] With respect to this feature, the "two languages" example still works even if we spoke languages in

lieve that whatever holds with respect to the referent of "Paderewski-the-pianist" should also hold with respect to the referent of "Paderewski-the-politician."

To spell the details of this ambiguity think at a certain model which makes true the sentence "Paderewski has musical talent," which Peter assents to, thereby believing it. In the domain of that model we have to put an individual and make "Paderewski" refer to it, and put that individual, as well, in the extension of the predicate "x has musical talent."

Now, "Paderewski has no musical talent" can't be true in the same model, if "Paderewski" uniquely and unambiguously refer to the same individual in the domain. Hence, if we want that Peter do not contradict himself in believing that Paderewski doesn't have musical talent, since we have no recourse to other worlds than the actual world where the individual which "Paderewski" refers to is not in the extension of "x has musical talent," one should fix other parameters of the model such that Peter's beliefs do not get inconsistent.

There may be several options but what seems prominent for the case under discussion is that Peter learns the referent of "Paderewski" which makes the sentence "Paderewski has musical talent," true, say, at a concert. Thereafter, whenever he will utter the sentence "Paderewski has musical talent," he will have in mind *that* "Paderewski" he listened to that concert, and the referent of that Paderewski is fixed by the description "the pianist listened to that concert."

Whereas, Peter learns the referent which he takes for granted that it makes the sentence "Paderewski has no musical talent," true under different circumstances which do not entitle him to believe that the same individual is the referent of "Paderewski" in both sentences that he endorses. Those circumstances are such that they can fix the referent of "Paderewski" though it could be perception again, through a description such as "the Prime Minister of Poland" or "a Polish nationalist."

This seems to be a compelling case for stating that in reporting beliefs names do not refer rigidly, but through description which fix their referent, and this accounts for the fact that names may refer ambiguously just in case one and the same name is associated with two different description which happen not to pick up the same individual.

Coming back to the model which makes the first sentence true, we cannot say that it refutes the second sentence, unless the different descriptions associated with "Paderewski" do pick up the same individual in the domain. And since nobody

---

which all names must denote uniquely and unambiguously. Of course, we never speak languages in which all names denote uniquely, but this being possible obscures the presence of an ambiguity which the construction of the "Paderewski" case rests on.

can know such a fact *a priori* to reason assuming such a supposition would be fallacious.

Let's take a look at the same case from another angle.

Imagine this dialogue between Peter and one of his friends. The latter asks Peter what does he believe about Paderewski, or if we want something more definite, about his musical gift.

Now, supposing that Peter believes that there are two persons with this name, i.e., that the name is used ambiguously and not with a unique referent, he should act upon this belief and proceed first by disambiguating the question. Then, it is very likely that prior to answering his friend's question Peter will ask him: "Paderewski – who? The musician or the politician?"

Now, there are two plausible ramifications of the situation: either Peter's friend is like Peter in not being aware that Paderewski the musician *is* Paderewski the politician, or else Peter's friend is more knowledgeable and he will point out to Peter that "Paderewski" refers in that context to just one person. (Of course, we are not obliged to think that they are using the logicians' jargon.)

Depending on which one is the case, Peter will either reiterate his beliefs about the musician *and* the politician, because for him there might be two different persons, or else will notice that his otherwise prudent attachment to the principle not to infer that one name should refer to one and the same person failed to apply, because this time, the name he learnt in different situation *does* refer to only one person. Accordingly, he will have to give up one of his beliefs, on pain of being accused of logical inconsistency.

So, actually there was no need for disambiguation, but since Peter believed there was, one should retain Peter's use of a description as a referent fixing device for a name which occurs in a belief report. And besides, if his friend is more knowledgeable and wants to spot Peter's mistaken idea that "Paderewski" is used ambiguously, he still needs the referent-fixing description device in order to tune the two tokens of "Paderewski" that Peter is using. Thus, it is very likely that he will say to Peter that "the famous pianist" and "the Prime Minister" pick up the same individual in the person of Paderewski.

Hence, in both situation, viz. either there is an ambiguity and one needs to control it, or there is no ambiguity and one needs to dispel the mistaken impression that there might be one, the usage of a description for fixing the referent of the name is decisive in restoring the communication.

Consequently, although the "Paderewski" case makes the restriction that names should not be translated, if we want to avoid the puzzle, irrelevant, precisely because the puzzle occurs even though the name is phonetically repeated, it doesn't seem to me that it is as successful in showing that the substitutivity issue is en-

tirely alien and irrelevant for the problem raised by the puzzle.

For, if we grant that Peter's assent to contradictory sentences, without thereby being committed to having inconsistent beliefs, is due to his wrong belief that "Paderewski" is used ambiguously, then his failure of seeing that the two sentences *are* contradictory can be accounted for by letting names within contexts of belief attribution behave like descriptions which are asked for fixing their referents.

Accordingly, one token of a name cannot be substituted for another token of the same name, even though they are codesignative, because their referent is fixed by different descriptions, and there is a very sound presumption that the referent of a token might not coincide with the referent of the other.

To risk a substitution in such cases and to get a true conclusion from a premise, or in Lois' case from two premises, is like winning the lottery. It would be to proceed fallaciously because the reasoning would be based on a coincidence, viz. that the two reference-fixing descriptions pick up the same referent for the name to which they are associated.

For this reason, I venture to say that the description-like behavior of names within contexts of belief attribution is the dual of those rare but, however, existing cases in which definite descriptions behave like names; descriptions, like e. g. "The Holy Roman Empire", where the descriptive conditions don't determine the referent, refer rigidly like names do in modal contexts.

I guess we can't say that Kripke has a fully developed theory of belief attribution and opacity. His main purpose in (Kripke 1994) is to provide reasons in favor of the idea that the substitutivity of identicals is not responsible for getting the paradoxical results in arguments involving ascription of belief, and thus, either "=E" is a perfectly legitimate rule in belief contexts, or at least it is not responsible for failure of belief ascriptions.

What "*Londres*/London" example and "Paderewski" example purport to show is that even though a clear case of substitutivity is absent, still there are problems with belief ascriptions. And other principles that lie behind the normal practice of linguistic communication, such as the disquotational principle and the principle of translation, are perhaps responsible for Kripke's puzzle about belief which any theory of belief and names should deal with. The nature of the problem is still elusive, but a hypothesis can be that the cases envisaged in those examples "lie in an area where our normal apparatus for ascription of belief is placed under the greatest strain and may even break down" (Kripke 1994, p. 379).

In the end I want to reiterate and to emphasize some scattered points I made earlier as reactions to, and comments on, Kripke's very ingenious cases and his subtle proposals.

The point that my paper is making is that cases like those envisaged by Kripke trigger a name-sensitive context which blocks the use of a substitutivity principle, although no appeal to a substitutivity principle is overtly at stake in the construction of those cases.

Simply put, Kripke's cases may well not be based on any substitutivity principle, but they are generated within, or are embedded within, or at least they trigger a context that blocks substitutivity.

My reading of this proposal is articulated counterfactually, as follows:

Had the contexts in which "*Londres*/London" and "Paderewski" cases, respectively, occur been such that they would have allowed for a substitution of "*Londres*" for "London" and of "Paderewski-the-musician" for "Paderewski-the-politician" within Pierre's/Peter's idiolect, then no Kripkean puzzle about what does Pierre/Peter believe would have obtained. For Pierre/Peter could have used his logical acumen to spot an inconsistency within his set of beliefs.

Thus, although Kripke's cases don't explicitly rest on a substitutivity principle, they seem to be embedded within contexts which do not allow for such a substitution. In this sense, they could be seen on a par with Lois' case.

Accordingly, one has to show how arguments such as Kripke's occur in contexts which are name-sensitive, i.e., are such that, although they are not based on a principle of substitutivity, they occur in a context which blocks the replacement of an occurrence of a name (in a belief sentence) with any codesignative name which the subject of that belief attribution is unaware of fitting in her entertained belief, being thus on a par with similar arguments resting, however, on a substitutivity principle.

Methodologically, then, we are not misguided to address the topic of failure of belief ascriptions through one of the prominent features of the context in which that failure occurs, viz. through failure of substitutivity of codesignative names within those very contexts.

One very difficult question on which I can speculate only is why contexts of belief ascriptions are name-sensitive?

If by dividing and then eventually conquering the problem we can hope for gaining understanding, then I would like to conjecture about the logical part of the issue, and to defer the discussion on the more speculative aspect concerning the *structure* (*architectonic*) of the content of our beliefs for future discussion.

For the logical facet: I begin with the observation that if both Lois an Pierre/Peter substituted a codesignative name or a translation/phonetic variant of a name, respectively, for a given name or for the original name, respectively, then they would proceed fallaciously.

This suggests that a validity preserving strategy requires at least to refrain

from drawing any conclusion from a belief report in which a name occurs, unless the subject of that belief attribution is aware of the sameness of reference of that name and of another name. Only if this is the case, we could draw a conclusion having the same content with that of the belief report save for the substitution of the codesignative name for the name which occurs in the original belief report.

Apparently, if the content of a belief ascription is not preserved under *that particular way* in which it is believed, i.e., if we don't keep the frame of the content of that belief ascription sealed, and we allow for a substitution of *any* codesignative name for *any* given name within that belief ascription, then the arguments we get thereby are no more truth-preserving.

And this suggests, further, that in order to have a truth-preserving argument in a context of belief ascription a necessary condition is not to jeopardize the ascribed content by substituting a codesignative name for a given name, unless the subject of that belief attribution is aware of (has knowledge of) the fact that the two names are codesignative.

Then, it is legitimately to ask more about the semantics of singular terms, in particular of names, and about their behavior, within contexts of belief attribution.

It seems very likely that cases like those considered in this paper show that we cannot assign a fixed semantic role for names, regardless the context of their occurrence.

This remark boils down neither to the view that names are synonymous to definite descriptions, nor to the view that in all contexts the referent of a name gets fixed through a definite description associated with that name.

Rather, it has only in view that the same way in which there are cases in which descriptions behave like rigid designators, there also are cases in which names behave like descriptions, and that, in particular, a name has a description-like behavior within contexts of belief ascriptions.[16] But then, this is the reason of failure of substitutivity, and this accounts at least in part for cases of failure of attribution of beliefs.

Roughly speaking, one sensible moral to draw is that in contexts where *our knowledge* of truth of our sayings about individuals is sensitive to *our* ways of referring to them, there is a place for the *modes of fixing* the referents of the names we use. And this, on a reasonable reading, is tantamount to accepting a place for Fregean "senses" as modes of fixing the referents of proper names.

The supposed compromise sought here would consist in accepting that in most contexts a name is a rigid designator, but that, nevertheless, in belief ascription

---

[16] For instance, in "Paderewski" case, a reference-fixing description is needed in order to disambiguate the use of the name "Paderewski."

contexts they have a description-like behavior, and precisely for reasons which have to do with disambiguation of the discourse, and for the need of keeping under control the indeterminacy of the referent of that name, a referent-fixing description is associated with it.

My primary motivation for discussing the main tenets of Kripke's theory of proper names in relation with belief attribution has been prompted by his powerful attack against intensional entities and their alleged place within semantics.

However, whereas the arguments deployed in (Kripke 1980) leave almost no hope for the friends of intensions, at least in regard to transparent and modal contexts, Kripke's case against intensions is not so compelling with regard to belief ascription and other propositional attitudes. Here, there are no knock-down argument that render the need for intensional entities superfluous.

Thus, to the extent to which nothing definitive, either positive or negative, can be said about substitutivity within belief ascription contexts, the overall moral we can draw from the discussion of this issue is that in the state of our present knowledge there are no serious reasons to give up the exploration of the failure of attributing beliefs through the topic of failure of substitutivity.

References

Forbes, Graeme. 1994. "Introduction to Opacity and Attribution of Attitude." In *Basic Topics in the Philosophy of Language,* edited by R. M. Harnish, 323–337. United Kingdom: Prentice Hall.

Kripke, Saul. 1980. *Naming and Necessity.* Cambridge, MA: Harvard University Press.

———. 1994. "A Puzzle about Belief." In *Basic Topics in the Philosophy of Language,* edited by R. M Harnish, 352–392. Prentice Hall.

LePore, Ernest, and Barry Loewer. 1990. "A Study in Comparative Semantics." In *Propositional Attitudes: The Role of Content in Logic, Language, and Mind,* edited by C. A. Anderson and J. Owens, 91–111. United States: Cambridge University Press.

Putnam, Hillary. 1975. "The Meaning of 'Meaning'." *Minnesota Studies in the Philosophy of Science* 7:131–193.

Magical Portraits

Cynthia Freeland
University of Houston*

ABSTRACT | Visual portraits can be alluring by being very lifelike. This feature has been a favorite one for artists to explore in film and fiction: it sometimes leads to romance and sometimes to danger and obsession. Whereas Pygmalion and Professor Higgins fall in love with their own creations come to life, in some stories and film portraits, as in the noir classics *Rebecca* and *Laura*, the powerful gaze of women in pictures exerts an uncanny force that extends beyond death. The Harry Potter books and films feature portraits that can move and talk, sometimes in amusing, sometimes educational, ways. My essay meditates on how artists across diverse genres have represented the magical powers of portraits, with a particular focus on one of the more fascinating fictional examples, Oscar Wilde's *The Picture of Dorian Gray*. Wilde's novella offers a complex consideration of relations between art and life. It shows both why portraits can be fascinating and how their magical art can become dark and dangerous.

KEYWORDS | Portraits; Dorian Gray; Pygmalion; Magic; Uncanny

* *Correspondence:* Cynthia Freeland – 513 Agnes Arnold Hall, University of Houston, Houston, TX 77204-3004, USA. Email: cfreeland@uh.edu

Cynthia Freeland

1 Magical Portraits

We often praise portraits for being lifelike. It is particularly meaningful if the person shown in the portrait is someone we have loved, but who is now lost to us. Portraits hold our attention when the people in them look especially real; their eyes seem to look back and follow us around the room.[1] In the world of Harry Potter, the eyes of people in portraits actually do this. The portraits in those stories can move and talk. Fans of the books will know many arcane details about the pictures at Hogwarts School, where buildings are decorated with portraits with subjects who regularly manifest various actions.[2] The Fat Lady beckons and toasts viewers with her wine glass. Sir Cadogan challenges people to a duel and falls off his horse. Even after his death, Albus Dumbledore, speaking from his portrait, offers advice to Harry and to other headmasters. Other characters in pictures curse, warn of upcoming dangers, guard secret passages, and so on. Some of the pictured people can shift location and move between portraits, as the Fat Lady does in Book 1 and Phineas Nigellus does in Book 7. These portrait powers and actions are, naturally, depicted in fun and colorful detail in film versions of the Harry Potter novels.

The explanation for the amazing properties of portraits in Harry Potter, of course, comes down to magic. The sitters' powers derive from the skills of the wizards who made the paintings. The greater the skill of the witch or wizard, the more detailed and lifelike the portrait. Portraits reveal what the artist observed about the people while making their picture. In certain cases, the pictures are especially detailed and active because they are portraits of the school's headmasters. These teachers, presumably persons with great magical skills, have gotten their portraits made at some point and afterward continued to teach their portraits things about themselves and their lives. Most portraits of past headmasters are kept in the rooms of the current headmaster, with some hidden away in a secret closet. Headmasters can visit and confer with the sentient, talking images of their predecessors, and glean advice from them if it is needed.

Students at Hogwarts do not have the same access as teachers to wise figures in portraits, but even a young wizard like Harry can sometimes benefit from the magical powers and actions of portraits of those who have gone ahead. In addition to portrait images, the Harry Potter novels also sometimes depict animated images of people in photographs and mirrors, ones in which a person's reflection (or a ghost) speaks back to them. This does not always work for the best results.

---

[1] I discuss this point and some examples in my article "Moving Picture Portraits" in (Freeland 2020, pp. 97–112).

[2] For further details along with a list of known portraits from the books, their locations and activities, see the section titled ("Portraits" 2021) on The Harry Potter Wiki.

In the first book there is a mirror named "the Mirror of Erised" ("desire" spelled backwards). Harry spends too much time viewing photos of the parents he misses, so Dumbledore hides the mirror away.[3]

2   Pygmalion and Beyond

The theme of artistic creations that magically come alive and begin to move and speak is a very old one in art. In ancient Greek mythology, the artist Pygmalion created an ivory sculpture of a perfect beauty who was more lovely than any real woman. After he prayed to Aphrodite to find a woman like her, the goddess magically brought the beautiful statue to life. The story has been re-told in many other versions, notably in the play of the same name by George Bernard Shaw (1913), and later in its film version, *My Fair Lady* (1964). In these works, the snobby Professor of Languages Henry Higgins bets a friend that he can educate a common flower girl and transform her from a coarse, ill-spoken woman of the streets into an elegant lady who gets accepted in Society at a fancy ball. After a year of intense instruction, his "creation" Eliza Doolittle helps him win his bet. But in the process the teacher loses his heart to his own artistic product. Professor Higgins must then figure out how to woo Eliza given that his treatment of her has been cruel and condescending.

Portraits in history often served as memorials to the dead, but they could also be a living emissary of a real person.[4] Kings and queens sometimes sent their portraits to members of the realm to demonstrate their thanks and promise of patronage. Portraits served romance at other times, being sent abroad in diplomatic enterprises to help arrange or seal a royal marriage. A portrait could even stand in as a proxy for a royal person in a marriage ceremony for far-flung individuals. Artists as distinguished as Holbein and van Eyck were sent abroad on diplomatic missions to paint potential marriage partners for their respective royal patrons, Henry VIII of England and Duke Philip the Good of Burgundy.[5]

There are related fictional instances in which a portrait is the intermediary in a grand narrative of love. A famous Persian legend that was turned into an epic poem by the twelfth-century writer Nizami recounts the story of two young people who fall in love via a portrait. When the princely Khosrow is told of Shirin's remarkable beauty, he longs for her. He sends his painter friend to see Shirin and to show her Khosrow's portrait. She immediately falls in love with the handsome

---

[3] Thanks to Barbara Urdiales and Ariana Peruzzi for these examples.

[4] For more, see my book (Freeland 2010). For insightful criticisms of my view, see (Maes 2015). Helpful sources on the changing nature of portraits in art history are (West 2004) and (Brilliant 1991).

[5] (Meares 2018).

young man in the picture. The two seek to be with each other but run into many challenges and misadventures; the tale ends tragically. Unsurprisingly, episodes from this epic were often illustrated in Persian miniature paintings. These include versions highlighting Shirin's fascinated gaze when it first falls upon the portrait of her future beloved.[6]

The plot lines of such Pygmalion-style stories tend to take two directions. Sometimes the "enlivened" artwork provides the creator with happiness, or at least with interesting challenges and intrigues, as in the stories of Shirin and Khosrow, Pygmalion, and Professor Higgins of *My Fair Lady*. But in other cases, the created being–the portrait or sculpture that comes to life–turns out to be flawed, or even worse, monstrous and threatening. Perhaps the best-known example of this latter path is provided by German writer E.T.A. Hoffmann (1776-1822). His 1816 short tory *The Sandman* is a very disturbing tale in which the main character Nathanael falls in love with the beautiful Olympia. But she turns out to be a wooden doll who has been created by Coppelius, a mysterious and evil alchemist. Coppelius had frightened the young Nathanael on visits to his father, when the boy spied on them and learned that this man carried a collection of real human eyes.

Later in college, Nathanael meets a man named Coppola whom he believes is the evil Coppelius. Coppola is working with a physics professor named Spallanzani, and when Nathanael glimpses the professor's beautiful daughter Olympia through the windows, he is smitten with her. Eventually he meets Olympia and dances with her, falling so deeply in love that he ignores a certain mechanical quality she has. Eventually, he discovers to his horror that she is a wooden clockwork doll who was created by her father along with Coppola. He sees her body ruined and lying on the floor with its eye sockets empty because Coppelius has run off with the eyes. In this story, the magic of creating a lifelike animated being is what we could call bad or black magic. It leads the disturbed young Nathanael to madness and eventual suicide.[7]

Stories by Hoffmann's American contemporary Edgar Allan Poe (1809-1849) often have similarly eerie, tragic plots. Poe wrote a story about a portrait coming to life and haunting an artist, "The Oval Portrait," in 1842. Here the protagonist finds an illustrated book with a painting of a beautiful woman. He learns that the portrait was made by a painter who had asked his young bride to sit for him. But the painter's demands were so extreme, and his obsession with his art so complete,

---

[6] For examples of these images, see (Rice, n.d.). I first learned of this story through Orhan Pamuk's 1998 novel *My Name is Red*, where it is mentioned numerous times, with the names written in English as Husrev and Shirin.

[7] See (Hoffmann 1967). For more on the Hoffmann story and Freud's (mis)interpretation of it, see (Freeland 2004).

that he ignored and exhausted his wife in the process. So when he finally finished the picture and turned to show her, it was only to find her dead.

Another creepy portrait story is offered by by the Russian writer Gogol (1809-1852). In his "The Portrait" (1835), a struggling young artist is so taken with a lifelike image of a man he finds in an art gallery that he buys it despite his poverty. That night he dreams that the man in the portrait steps out of it and offers him money. The next day he finds a fortune in gold hidden in the picture frame. He becomes a wealthy and successful artist but squanders his talent in making popular but mediocre work. He ultimately realizes his horrible mistake when he discovers a portrait by someone else that reveals genuine artistic creativity.

These Romantic tales of tragic love and dangerous portraits have echoes later in Hollywood movies about mysteriously potent pictures, especially films from the classic period of the 1940s and 1950s. Several films tell stories about characters who fall in love with or become somehow obsessed by portraits. The 1940 classic film noir *Laura* is a fine example. This murder mystery is still engrossing in part because it features wonderful dialogue, fun acting, and a wealth of suspects. The police detective hero (Dana Andrews) is charged with investigating the murder of a beautiful and alluring New York society maven, Laura (Gene Tierney). The cop searches for clues in her fabulous apartment, sniffing her perfumes and touching her dresses. He then meditates while drinking her liquor in a living room dominated by a large portrait of the victim. Eventually this hard-boiled cop becomes so entranced by Laura's portrait that a suspect accuses him of falling in love with his victim. *Laura* is a film whose plot takes many surprising turns, and it even manages to provide an unexpectedly happy ending.

There are more dire outcomes that involve portraits in several Alfred Hitchcock thrillers. In *Rebecca* (1940), based upon the novel by Daphne du Maurier, the newly married Mrs. de Winter (Joan Fontaine) is intimidated by everything she hears about the beautiful and imperious Rebecca, her predecessor as mistress of the grand estate Manderley. Her insecurities are furthered by the chilly housekeeper Mrs. Danvers (Judith Anderson), who remains devoted to her dead mistress. Danvers tricks the young woman into wearing a dress for a ball just like one that Rebecca had worn. The young woman lingers in awe before Rebecca's imposing portrait and has a copy made of her glamourous dress. But this has disastrous consequences: her appalled husband (Laurence Olivier) becomes furious when he sees it and orders her to change clothes immediately. She feels spurned because of some inadequacy, whereas in fact, as we eventually discover, he had detested his deceitful first wife.

Just as Rebecca's portrait preoccupies the second Mrs. de Winter, so too is the heroine of *Vertigo* obsessed with a portrait. The young and beautiful blonde

woman, Madeleine Elster (Kim Novak), makes repeated visits to a museum to stare at the portrait of her ancestor Carlotta Valdes. Along the way, she is tracked by the hero, a sleuth named Scottie (Jimmy Stewart). He finds her sitting in a room of the Legion of Honor in San Francisco, staring at a woman on the canvas whom she distinctly resembles. Madeleine wears similar clothes and the same hairstyle, and she even carries a bouquet like the one Carlotta holds in the picture. (The film-maker actually had a portrait created especially for this scene.[8]) Although we later discover that much of this scenario rests upon a false setup, the young woman as we first meet her appears to be preoccupied, even haunted by, the mysterious Carlotta. Scottie falls in love with the beautiful Madeleine. Later, after he believes her to have died in an unfortunate accident, he becomes obsessed with her, just as she had appeared to be obsessed with Carlotta. He even attempts to recreate her as his ideal woman by refashioning another woman, Judy (also played by Novak), in Madeleine's image. We are shown his often cruel and controlling (Professor Higgins-style) efforts to mold and shape Judy into Madeleine's duplicate. The plot of *Vertigo* is complex, with multiple levels of deception and imitation; but the results are, inevitably, tragic.

The interesting thing about these two Hitchcock films, *Rebecca* and *Vertigo*, is that the power of the portraits they feature is not attributed to the skill of their painters but rather to the strong responses that they elicit in certain viewers (as happens too in *Laura*). Madeleine is (supposedly) being manipulated into suicide by Carlotta's portrait, and the second Mrs. de Winter is cowed by the portrait of Rebecca. In these cases, the magical power of portraits is an evil one, because the beautiful people (here women) in these portraits exert a hypnotic effect on viewers. Their portraits mysteriously entrap victims into layers of subterfuge, lies, and danger.

3 The Uncanny

Portraits that seem too real or too powerful, like those in Hitchcock's movies, can be creepy or eerie. Sigmund Freud wrote about this feeling in discussing Hoff-mann's "*The Sandman*" in his famous essay "The Uncanny" (1919).[9] Freud described the uncanny as an uncomfortable feeling that arises about something familiar that is coupled with strangeness. It often has to do with stories of a double or doppel-ganger; the discovery that one has an unexpected twin can cause feelings of great unease.

---

[8] See (Dowd 2019).
[9] (Freud 2001). See also (Windsor 2019).

Uncanny tales involving doubles, and films with images of twins, often appear in the horror genre. A well-known example is the twin girls in Stanley Kubrick's film version of *The Shining*. They appear in several visions that the young boy Danny sees while riding his tricycle down the long hallways of the mysterious hotel where his family is staying for the winter. We learn that the girls are ghosts, because they were in fact killed by their father years ago. We eventually see their bodies ourselves, along with young Danny, in a very chilling scene of the film that presents a sudden vision of them lying butchered and bloody.[10]

Freud laid out his influential account of the uncanny in 1919, but the term has acquired a new usage more recently that applies to a field unanticipated by him: the creation of robots, androids, and more generally, products of artificial intelligence. The new descriptive term "the Uncanny Valley" was coined and explained in a 1970 essay in a Japanese journal by the Tokyo professor of robotics Masahiro Mori.[11] Mori's basic idea is that we humans are attracted by examples of very humanlike robots, but that as their realism is increased, we eventually test it and become disappointed as we ascend and cross over a kind of slope of similitude– the uncanny valley. Although greater lifelikeness is initially appealing, after such creations reach a certain point of realism, there will be inevitable flaws or cracks, at which point the robot is seen as so eerie or uncanny that it becomes repulsive.

This is very much like what Nathanael in "The Sandman" experiences when he begins to sense that the beautiful Olympia is not actually a living human being—especially when he realizes that her eyes were artificial. Such a discovery would indeed be repulsive. Mori's key premise was that robot designers desire to create a truly lifelike robot. But the robots at the time he wrote were not very lifelike. Robot designers can increase our human affinity for their creatures by adding certain realistic features, such as covering a prosthetic hand with flesh and the right sort of shapes. And yet even if we encounter a very "real-looking" hand in a certain context it will drop us into that uncanny valley and become repulsive. He explains,

> When we realize the hand, which at first sight looked real, is in fact artificial, we experience an eerie sensation. For example, we could be startled during a handshake by its limp boneless grip together with its texture and coldness. When this happens, we lose our sense of affinity, and the hand becomes uncanny. (Mori 1970, 2012)

---

[10] See chapter 7 of my book (Freeland 2000). There are numerous horror movies about the creepy nature of twins; see for example *Sisters* (1973) and *Dead Ringers* (1988).
[11] The original essay can be found in English translation here: (Mori 2012). For some updates and critical assessments, see ("Who's Afraid of the Uncanny Valley?" 2010).

For example, a robot with a face designed to smile in a sequence of motions roughly like those produced by our own face muscles would appear creepy if the speed with which it made these relevant shifts was a bit off. Mori suggests that perhaps robotic replacements for human parts should have a deliberately false design so as to avoid the uncanny valley effect. He gives the example of eye-glasses, which do not try to be replicas of human eyes but can have their own charming designs. Beings like zombies or corpses are also uncanny. Mori wonders about the origin of this feeling in humans–much as Freud did–and speculates that the uncanny evolved as an essential part of human nature as "an integral part of our instinct for self-preservation" (Freud 2001).

Although the concept of the uncanny valley is still much-cited, more recent researchers in robotics and AI have expressed reservations about its validity as an actual scientific hypothesis. Remember that Mori emphasized the importance of how robotic creations like prosthetic hands or facial muscles move. However, some current critics feel that believable, socially familiar kinds of motion in a robotic creation may overcome the fact that the item in question (a hand, a face) doesn't actually look like an actual replica of the human version.

4   Dorian Gray

Probably the most famous tale of an uncanny portrait is Oscar Wilde's 1890 novella *The Picture of Dorian Gray*, the story of a portrait that takes on an unusual life of its own. I will discuss this book in detail here in the final part of this essay. Not only does it offer an interesting variation on the theme of the magical portrait, but it also serves as a meta-reflection on the nature and value of a life devoted to artis-tic pursuits. Wilde describes diverse modes of artistic creation that range from painting to acting to music. Also, in the revised and expanded version published in 1891, the author included a preface that made bold pronouncements about the relations between art and morality.

The Picture of Dorian Gray opens with the unveiling of a new portrait by the artist Basil Hallwell. He knows this new painting is his best one ever, but fears that it will reveal his idolatry of the beautiful young man, Dorian Gray, who is its subject. Dorian has remarkable good looks, with golden hair and ruby lips, and is especially attractive because of his purity. "All the candour of youth was there, as well as youth's passionate purity… one felt he had kept himself unspotted from the world.[12]" Besides being the portrait's subject, Dorian is also in a sense the

---

[12] All quotations from *The Picture of Dorian Gray* are from the version which may be downloaded as a pdf for free here: (Wilde, First published 1890).

painter's product, in that the painting serves to reveal Dorian to himself. Although he has been told that he is both beautiful and charming, only upon seeing himself through the painter's eyes does he realize this truth.

A third man is present at the unveiling of the portrait, Lord Henry Wotton. He praises the painting and he also finds Dorian charming. But instead of admiring Dorian's purity, Lord Henry sees him as naïve and unformed, raw material to be manipulated and directed. The worldly-wise Lord Henry appears to find it amusing to experiment upon such an innocent nature. As the crafty man remarks upon Dorian's beauty, the young man begins to feel jealous of his own portrait. Henry's comments awaken a chord of vanity in him. He realizes that he will age while his picture does not, and fervently wishes that these processes could happen in the reverse way.

Strangely enough, Dorian's wish comes true. He first notices that the portrait has altered after behaving very badly to the young woman he was in love with, Sybil Vane. He had found her, a pure beauty blossoming like a lotus in the gross swampy environment of a cheap theater run by a nefarious man. Sybil captivates Dorian not only by her beauty but by her transcendent acting in many famous roles. He proposes and she accepts. But when Dorian takes his friends Basil and Lord Henry to see Sybil acting as Juliet, they are appalled to discover that she is a dreadful actress. She explains later to Dorian that her ability to act has left her because now she has a real romance, whereas before she had always believed in the fictional romances she was enacting. Sybil's transition from illusion to reality causes a parallel transition in Dorian: he now sees her as common and vulgar, and cruelly breaks off with her. Later that night when he examines the portrait, he notices that there are new nasty lines around the mouth of the boy shown there.

Over the course of the novel, Dorian Gray's portrait continues to change; his innocent visage in it becomes more and more stained and ugly. As a rich and high-ranking young man admired by everyone for his charm and beauty, he enjoys access to society. He travels and entertains, but gradually his reputation becomes sullied as stories and rumors circulate about him ruining others and even causing suicides. Dorian is driven to the wildest extremes of experience and desire. As his depravity increases, so does the hideousness of his picture. Recognizing what the portrait reveals, the horrified Dorian hides it away under lock and key.

There is an uncanny doubling in *The Picture of Dorian Gray*. The person takes on the quality we expect of the picture, and vice versa. In the ordinary case, living means changing and aging, but Dorian never ages. This is part of what I have been calling the "magic" of portraits. Dorian remains beautiful and is a voluptuary of elegance in all forms. The selfish and corrupt activities he pursues leave no mark on his person, only on his portrait. And yet he finds his beauty insufficient for

happiness. He feels shame over his "sins" of murder and worries that he may come to "loathe his own soul." When he finally confronts the flaws revealed in his portrait, he cannot abide the knowledge and attempts to destroy it. What is left behind, mysteriously, is his own aged and decrepit body at the feet of the now restored, unspoiled picture.

The manipulative Lord Henry who sparked Dorian's downfall continues to egg him on with praise of his continued beauty. Lord Henry tells Dorian that beauty is above all else in nature, even above genius.

> It needs no explanation. It is one of the great facts of the world, like sunlight, or spring-time, or the reflection in dark waters of that silver shell we call the moon. It cannot be questioned. It has its divine right of sovereignty. It makes princes of those who have it.

Wilde's ambitious tale addresses important issues about the relations among art, beauty, genius, creativity, morality, and death. Commentators have sometimes remarked that there seems to be a contradiction between the apparently moral "message" of *The Picture of Dorian Gray* and Oscar Wilde's well-known "aestheticist" position, according to which art is something apart from morality and beauty is its own justification.[13] He wrote in the preface to the version published in 1891 that "There is no such thing as a moral or an immoral book. Books are well written, or badly written. That is all."

How does this bold proclamation fit with the strong moral elements to the story, in which Dorian Gray's trajectory appears to be a tragic downfall? Dorian feels he has "blood on his feet" from the murders he has committed. His moral self-assessment is very clearly stated:

> He knew that he had tarnished himself, filled his mind with corruption and given horror to his fancy; that he had been an evil influence to others, and had experienced a terrible joy in being so; and that of the lives that had crossed his own, it had been the fairest and the most full of promise that he had brought to shame.

We must keep in mind that the supposed moral viewpoint of a novel or any other artwork is not necessarily that of a character depicted in it. Also expressed in the book is what appears to be a version of Wilde's own position in the remarks of Lord Henry, who insists that there is no such thing as the soul, because everything in

---

[13] The term "aestheticism" is used differently in literary studies than in philosophical aesthetics, but Wilde is associated with it in both fields. In aesthetics this position is sometimes instead labeled "autonomism." For further discussion of these terms and related issues, see (Gaut 2009).

life depends upon the body and on sensations. He claims that art is independent of action (and hence presumably of moral action) when he tells Dorian,

> Art has no influence upon action. It annihilates the desire to act. It is superbly sterile. The books that the world calls immoral are books that show the world its own shame. That is all.

These remarks resonate with Wilde's point in the preface that "All influence is immoral, because the aim in life is self-development."[14]

Does Lord Henry's remark apply to the book in which he is a character? Perhaps we should regard *The Picture of Dorian Gray* as an inert artwork. It may only seem immoral if it "shows the world its shame"—that is, our own shame. The book is as innocent as Dorian's portrait. And if the true aim in life is self-development, we can judge that although Dorian did devote himself to sensory and other pleasures, these did not leave him satisfied with the self he had created. He used the portrait as an excuse to act immorally because it would spare his beauty from the consequences of his behavior. But as I just noted, the portrait itself bears no responsibility for Dorian's actions. This is why in the end it remains pristine and beautiful, while Dorian is hideous—and dead.

The Picture of Dorian Gray is a complex novel about characters who all practice art: Dorian is a skilled pianist, Sybil Vane an actress, Basil Hallward a painter, and Lord Henry a man molding those around him—a blend of sculptor-cum-dramaturg. As the book's author, Oscar Wilde created a work that is a better as a story—as art—because of its tragic ending. Consider: A beautiful man gets an amazing gift of a portrait that will age in his place, but he cannot live well with this gift. The portrait makes him "look… on evil simply as a mode through which he could realize his conception of the beautiful." Dorian's conception of the beautiful was a weak one that did not satisfy him. So when he attempts to destroy his fateful portrait, he only kills himself, leaving the portrait to live on, eternally beautiful. In the end, the book functions as proof of the superiority of art over life.

5   Conclusion

Portraiture involves a triangular relationship among the artist, the sitter, and the viewer. The magic of portraits can happen at any point in the triangle. Sometimes the artist is the source, with wizardly abilities to render a person lifelike and knowable. This aspect of portrait creation is emphasized in the Harry Potter series. In

---

[14] Wilde also says in the final sentence of the preface that "All art is quite useless."

other examples, what makes a portrait magical rests with the sitter who is some-
one with a mysterious allure, like the stunning women pictured in *Laura* or the
Hitchcock films I discussed. Finally, magical effects sometimes happen because
a viewer is particularly vulnerable to becoming magnetized by the person in the
picture. This was true of Nathanael in "The Sandman," whose passion blinkered
him so that he could not see the truth about the lovely Olympia. *The Picture of Do-
rian Gray* explores all three points of this potent portrait triangle. In the story the
artist is especially inspired, the subject is surpassingly beautiful, and the effect of
the portrait on one viewer of the picture, Dorian Gray himself, is momentous and
deadly.

References

Brilliant, Richard. 1991. *Portraiture.* Cambridge: Harvard University Press.

Dowd, Katie. 2019. "A portrait was hung in the Legion of Honor for 'Vertigo.' No
one's seen it since." https://www.sfgate.com/art/article/hitchcock-vertigo-
portrait-of-carlotta-missing-14434240.php.

Freeland, Cynthia. 2000. *The Naked and the Undead: Evil and the Appeal of Horror.*
Boulder: Westview Press.

———. 2004. "Explaining the Uncanny in The Double Life of Veronique." In *Hor-
ror film and psychoanalysis: Freud's worst nightmare,* edited by Steven Jay
Schneider. Cambridge: Cambridge University Press.

———. 2010. *Portraits and persons.* Oxford: Oxford University Press.

———. 2020. "Moving Picture Portraits." In *Portraits and Philosophy,* edited by
Hans Maes. London: Routledge.

Freud, Sigmund. 2001. "The Uncanny." In *The Complete Psychological Works of Sig-
mund Freud,* translated by James Strachey, vol. 17. London: Vintage.

Gaut, Berys. 2009. *Art, Emotion and Ethics.* Oxford: Oxford University Press.

Hoffmann, ETA. 1967. "The Sandman." In *The Best Tales of Hoffmann, ed,* edited by
Bleiler Everett F. New York: Dover Publications.

Maes, Hans. 2015. "What is a portrait?" *British Journal of Aesthetics* 55 (3).

Meares, Hadley. 2018. "How Unflattering Royal Portraits Could Break a Marriage Contract." https://www.history.com/news/how-portraiture-shaped-royal-marriages.

Mori, Masahiro. 1970. "Bukimi no tani [the uncanny valley]" [in Japanese]. *Energy* 7:33–35.

———. 2012. "The Uncanny Valley: The Original Essay by Masahiro Mori." Translated by Karl F. MacDorman and Norri Kageki. https://spectrum.ieee.org/automat on/robotics/humanoids/the-uncanny-valley.

"Portraits." 2021. https://harrypotter.fandom.com/wiki/Portrait.

Rice, Cory. n.d. "Khusraw Discovers Shirin Bathing." https://www.khanacademy.org/humanities/art-islam/chronological-periods-islamic/islamic-art-late-period/a/khusraw-discovers-shirin-bathing.

West, Shearer. 2004. *Portraiture.* Oxford: Oxford University Press.

"Who's Afraid of the Uncanny Valley?" 2010. https://spectrum.ieee.org/automato n/robotics/humanoids/040210-who-is-afraid-of-the-uncanny-valley.

Wilde, Oscar. First published 1890. *The Picture of Dorian Gray.* https://freeditorial.com/en/books/the-picture-of-dorian-gray/readonline.

Windsor, Mark. 2019. "What is the Uncanny?" *The British Journal of Aesthetics* 59 (1): 51–65.

A Criticism of Two Dogmas in Philosophy in China

Ruipeng Lei
Huazhong University of Science and Technology

Renzong Qiu
Chinese Academy of Social Sciences*

ABSTRACT | The authors of this paper criticize two dogmas which have prevailed in philosophy in China over several decades. These dogmas have had devastating consequences for Chinese philosophy. In the first part of this paper the authors criticize the first dogma: Philosophy is the history of philosophy. They describe the first wave of criticism against this dogma in China, and summarize Hegel's main points in his lectures on the history of philosophy. The authors conclude that Hegel does not endorse this dogma, as some have thought. The authors then put forward their own arguments against this dogma. In the second part of the paper, the authors criticize the second dogma: Philosophers ought to design a universal system of philosophical theory from which the solution of any future philosophical issues can be deduced. In their arguments against the second dogma, the authors employ Popper's arguments that it is impossible to predict future human activity. Popper's arguments imply that it is impossible to design the kind of universal system in the second dogma. Finally, the authors emphasize the importance of weighing the values of different options on a case-by-case analysis, taking into account the actual practices of human activities.

KEYWORDS | Dogma; Chinese Philosophy; the History of Philosophy; Deduction; Beneficence

---

* *Correspondence:* Renzong Qiu – no.2121 Building 9, Caoqiao Xinyuan San Qu, Fengtai District, Beijing 100067, China. Email: qiurenzong@hotmail.com

1 The First Dogma: "Philosophy Is the History of Philosophy"

Perhaps, broadly speaking, philosophers are not aware that there is a dogma which has dominated philosophy in China for more than half a century. The dogma is: Philosophy is the history of philosophy.

The proponents of this dogma have never articulated arguments for the claim in a philosophical journal or academic presentation: Their work simply assumes the truth of the dogma. It is, after all, a dogma. Yet it continues to exert influence. It has been perpetuated by younger scholars. Nobody challenged the dogma until 2012 when Chen Bo, a professor from Department of Philosophy at Peking University published his paper titled as "Facing Philosophical Problems and Participating in Contemporary Philosophy Construction" (Chen 2010).

1.1 The First Wave of the Criticism against the Dogma

In his paper Professor Chen describes the situation of China's philosophy as follows:

> As far as we can see, there are almost all "historical" studies in the philosophy circle in China…Almost the whole Chinese philosophy circle is doing the "history" of philosophy, few are really doing "philosophy"… almost all people focus on the past of philosophy, few people "live" in the present and current time of philosophy, participate in the contemporary construction of philosophy. (Chen 2010, p. 12)

> In contemporary China, there are some historians of philosophy, many professors of philosophy, and many philosophical scholars, but there is no "philosopher" in the real sense, and no one dares to call himself a "philosopher." (Chen 2010, p. 14)

In his critique of this dogma, Chen Bo points out that philosophy is not the history of philosophy, and neither does the study of philosophy amount to the study of the history of philosophy. This is, of course, amounts to a denial of the first dogma. From the viewpoint of the nature of philosophy, the prevalence of this dogma has been harmful. For most students who study in a philosophy department, the purpose of learning the history of philosophy is to accumulate the necessary cognitive resources for independent thinking, not to become a historian of philosophy. By focusing exclusively on the history of philosophy, contemporary Chinese philosophers have not properly foregrounded their own independent lines of inquiry. Instead, contemporary Chinese philosophers should establish their own identity and

win the dignity of philosophers through their own efforts (Chen 2010 p. 14). Chen Bo continued with several proposals: first, that the source of philosophy is always some problem or question; second, that the refinement of philosophical problems leads to the specialization of philosophical research and the emergence of new branches of philosophy; third, that the principles of philosophical study are free inquiry and serious criticism; and finally, that the method of philosophical study is argumentation (Chen 2010, p. 15 ff.).

Immediately following Chen's paper, Jia Yushu, a philosopher working at Institute of Science, Technology and Culture, College of Ordnance Engineering, People's Liberation Army published a paper titled "Departure from the Wrong Area Where Philosophy Is Defined as History of Philosophy" in the same journal (Jia 2010). Jia Yushu first identifies as a dogma the claim that philosophy is the history of philosophy (Jia 2010). He argues that this dogma is "the product of misunderstanding and misinterpretation of Hegel's philosophy by Chinese philosophers. Hegel never said that 'philosophy is the history of philosophy'. Therefore, the advocacy of the dogma 'philosophy is the history of philosophy' by the Chinese philosophy circle is nothing but to dissolve the unfamiliar philosophical research into the familiar historical research with the help of the intersection of philosophy and history of philosophy." The author argues that "Chinese philosophy must say goodbye to history, and instead, it must face the real world" (Jia 2010).

We don't agree with Jia Yushu that the dogma has a goal of promoting a postmodernism ideological program; nor do we agree that Chinese philosophy must say goodbye to history. The fact is that this dogma did not promote a postmodernism ideological program in Chinese philosophy. It did, however, lead to a great number of editorial and translational works on the history of philosophy which are deemed to be futile–for example, the translation of Hegel's complete works into Chinese. We believe that the best and productive approach to Hegelian Philosophy is to read Hegel's works in the original German rather than in Chinese translation. We also claim that studying the history of philosophy is useful in order to study the ideas, thoughts and principles of philosophy. We must not say goodbye to the history of philosophy. However, for the majority of the people who are interested in learning philosophy, the history of philosophy is only a small part of their work. We agree with Professor Jia that Chinese philosophy must face the real world, and the Chinese philosophers must focus on philosophical issues that are raised in social practices during the transformation of Chinese society. For example, in the introduction of emerging technologies into China, a number of philosophical issues have been raised. These philosophical issues include ontological, epistemological and ethical issues, such as: what is the entity of a robot (is it an agent or moral agent?) or a synthetic organism (is it a machine or a form of life?); how to know

the entity of these kinds (e.g., do we know life through creating life?); and ethical issues (ought we to prohibit, permit or promote the research, development and application of emerging technologies?).

1.2   Hegel on the Relationship between Philosophy and the History of Philosophy

There is a lot to say about Hegel's idea about the relationship between philosophy and the history of philosophy; the dogma "philosophy is the history of philosophy" stems partly from a misreading of Hegel's lectures on the history of philosophy, and partly from a dubious source.

In Chinese philosophy, the proposition that "philosophy is the history of philosophy" has been smuggled in without philosophical argument, accepted as the golden principle of philosophy. Some claim that it was Hegel who said it. We cannot find the source for this claim. Hegel has never said that "philosophy is the history of philosophy" in his *Lectures on the History of Philosophy* (Hegel 1892). Interestingly enough, Hegel argued that the concept of the history of philosophy is a self-contradiction in his lectures about the notion of the history of philosophy:

> The thought which may first occur to us in the history of Philosophy, is that the subject itself contains an inner contradiction. For Philosophy aims at understanding what is unchangeable, eternal, in and for itself: its end is Truth. But history tells us of that which has at one time existed, at another time has vanished, having been expelled by something else. Truth is eternal; it does not fall within the sphere of the transient, and has no history. But if it has a history, and as this history is only the representation of a succession of past forms of knowledge, the truth is not to be found in it, for the truth cannot be what has passed away. (Hegel 1892-1896)

After this paragraph Hegel tries to explain his argument through a comparison with Christian religion and "the other sciences". Hegel argues that in order to get a better understanding of this contradiction, we must distinguish between the outward history of a religion or a science with the inward history of the subject itself. As for the history of philosophy, we must take into account that because of the special nature of its subject-matter, it is different from other histories. Now it is evident that the contradiction in question could not refer to the outward history, but merely to the inward history–that of the content itself. Hegel further explains the difference between the outward or external history and inward or internal history. He writes that there is a history of the spread of Christianity and

of the lives of those who have avowed it, and its existence has formed itself into that of a Church. This constitutes an external history of Christianity. In contrast with this external history, the Christian doctrine has its inward history, but it necessarily soon reached its full development and attained to its appointed powers. In Hegel's analysis, that the history of this doctrine in its wider sense includes two elements: first the various additions to and deviations from the truth formerly established, and secondly the combating of these errors, the purification of the principles that remain from such additions, and a consequent return to their first simplicity (Hegel 1892-1896).

Then Hegel returns to Philosophy. He writes that the other sciences, including Philosophy, also have an external history like Religion. Philosophy has a history of its origin, diffusion, maturity, decay, revival; a history of its teachers, promoters, and of its opponents–often, too, of an outward relation to religion and occasionally to the State. However, Hegel argues that the inner content of Religion, Philosophy and Sciences seems to have little historical content. He writes:

> In the comparison between the history of Religion and that of Philosophy as to inner content, it is found that there is not in the latter as there is in Religion a fixed and fundamental truth which, as unchangeable, is apart from history. The content of Christianity, which is Truth, has, however, remained unaltered as such, and has therefore little history or as good as none... The other sciences, indeed, have also according to their content a History, a part of which relates to alterations, and the renunciation of tenets which were formerly current. But a great, perhaps the greater, part of the history relates to what has proved permanent, so that what was new, was not an alteration on earlier acquisitions, but an addition to them. (Hegel 1892-1896)

Hegel takes an example of a science like Mathematics. In Mathematics, history more or less has only the pleasant task of recording further additions. Elementary geometry, insofar as it was created by Euclid, may from his time on be regarded as having no further history.

Hegel's ideas on the history of philosophy might be summarized as follows:

First, for Hegel the subject History of Philosophy as an accumulation of opinions would be a most superfluous and tiresome science:

> History, at the first glance, includes in its aim the narration of the accidental circumstances of times, of races, and of individuals, treated impartially partly as regards their relation in time, and partly as to

> their content. The appearance of contingency in time-succession is
> to be dealt with later on. It is contingency of content which is the
> idea with which we have first to deal-the idea of contingent actions.
> But thoughts and not external actions, or griefs, or joys, form the
> content of Philosophy. Contingent thoughts, however, are nothing
> but opinions, and philosophical opinions are opinions relating to the
> more special content of Philosophy, regarding God, Nature and Spirit.
> (Hegel 1892-1896)

Hegel argues that the view usually taken of the history of Philosophy ascribes to
it the narration of a number of philosophical opinions as they have arisen and
manifested themselves in time. This kind of matter is called opinion; those who
think themselves more capable of judging rightly, call such a history a display of
senseless follies, or at least of errors made by misguided men. Then he concludes
that:

> If the history of Philosophy merely represented various opinions in
> array, whether they be of God or of natural and spiritual things exis-
> tent, it would be a most superfluous and tiresome science, no mat-
> ter what advantage might be brought forward as derived from such
> thought-activity and learning. What can be more useless than to
> learn a string of bald opinions, and what more unimportant? Literary
> works, being histories of Philosophy in the sense that they produce
> and treat the ideas of Philosophy as if they were opinions, need be
> only superficially glanced at to find how dry and destitute of interest
> everything about them is. (Hegel 1892-1896)

Second, Hegel further argues that philosophical knowledge obtained through the
history of philosophy itself is futile. Hegel writes that in view of the many opinions
and philosophical systems throughout the history of philosophy, it is a perplexing
task to know which one ought to be accepted. It is evident that the greatest minds
have erred, because they have been contradicted by others. "Since, this has been
so with minds so great, how then can ego homuncio attempt to form a judgment?"
When it is admitted that Philosophy ought to be a real science, and one Philosophy
must certainly be the true, the question arises as to which Philosophy it is, and
when it can be known. Each one asserts its genuineness, each even gives different
signs and tokens by which the Truth can be discovered; sober reflective thought
must therefore hesitate to give its judgment. Then eventually the whole of the
history of Philosophy becomes a battlefield covered with the bones of the dead;
it is a kingdom not merely formed of dead and lifeless individuals, but of refuted

and spiritually dead systems, since each has killed and buried the other, as Hegel describes (Hegel 1892-1896).

Third, Hegel argues that in the history of Philosophy we have to deal with Philosophy itself. He writes that

> This makes it easy to us to comprehend the aim of Philosophy, which is in thought and in conception to grasp the Truth, and not merely to discover that nothing can be known, or that at least temporal, finite truth, which also is an untruth, can alone be known and not the Truth indeed. Further we find that in the history of Philosophy we have to deal with Philosophy itself. The facts within that history are not adventures and contain no more romance than does the history of the world. They are not a mere collection of chance events, of expeditions of wandering knights, each going about fighting, struggling purposelessly, leaving no results to show for all his efforts. Nor is it so that one thing has been thought out here, another there, at will; in the activity of thinking mind there is real connection, and what there takes place is rational. It is with this belief in the spirit of the world that we must proceed to history, and in particular to the history of Philosophy. (Hegel 1892-1896)

Finally, Hegel defines the study of the history of philosophy is the study of philosophy itself.

Hegel concludes in his lectures that

> The study of the history of Philosophy is the study of Philosophy itself, for, indeed, it can be nothing else. Whoever studies the history of sciences such as Physics and Mathematics makes himself acquainted with Physics and Mathematics themselves. But in order to obtain a knowledge of its progress as the development of the Idea in the empirical, external form in which Philosophy appears in History, a corresponding knowledge of the Idea is absolutely essential, just as in judging of human affairs one must have a conception of that which is right and fitting. Else, indeed, as in so many histories of Philosophy, there is presented to the vision devoid of idea, only a disarranged collection of opinions. To make you acquainted with this Idea, and consequently to explain the manifestations, is the business of the history of Philosophy, and to do this is my object in undertaking to lecture on the subject. Since the observer must bring with him the Notion of the subject in order to see it in its phenomenal aspect and

> in order to expose the object faithfully to view, we need not wonder at there being so many dull histories of Philosophy in which the succession of its systems are represented simply as a number of opinions, errors and freaks of thought. (Hegel 1892-1896)

From all this, it is clear that the Chinese philosophers who favor the dogma misread Hegel's thought on the history of philosophy in his lectures on the history of philosophy, where he focused on what the subject of the history of philosophy should be rather than on what Philosophy should be. Furthermore, they have misinterpreted Hegel's ideas. Hegel said that the study of the history of philosophy is philosophy, but he never said that philosophy is restricted to the history of philosophy. Hegel has never argued for such a claim and at least there is no evidence to prove he has said it.

One more astonishing thing is that in the Chinese version of *Hegel's Lectures on the History of Philosophy* (黑格尔 1959) the translators (Professor He Ling and his colleagues) inserted the following sentence (which is one of Hegel's addition in his other writing) into the Chinese translation of *Hegel's Lectures on the History of Philosophy*:

The Chinese version is:

哲学史的本身就是科学的，因而本质上它就是哲学这门科学

The German original text is:

die Geschichte der Philosophie selbst wissenschaftlich ist und sogar zur Wissenschaft der Philosophie, der Hauptsache nach, wird

The English text is:

The history of Philosophy is itself scientific, and thus essentially becomes the science of Philosophy.

It is quite extraordinary that an authentic Chinese expert in German Philosopher (He Ling) illegitimately inserted such a sentence which was absent from the original of *Hegel's Lectures on the History of Philosophy*.

Many Chinese proponents of the dogma "Philosophy is the history of philosophy" seem to cite this sentence as the basis of their acceptance of the dogma. However, the passage in which the inserted sentence appears is really an argument for what the history of philosophy should be if it becomes a legitimate subject, not for what philosophy as such should be (Han 2015; Yao 2020; "Positivism" 2021).

1.3   Arguments against the First Dogma: "Philosophy is more than the History of Philosophy."

We would like to argue that philosophy is more than the history of philosophy. If philosophy is limited to the history of philosophy, it would not be productive and creative; it would produce only glossaries of previous philosophical works. Let's suppose that there is a first philosopher A, who has written a lot of philosophical papers, then a successor B studies A, and writes a lot of papers on the philosophy of A, and after B there is a C who studies A and B, and produces a lot of papers, and so on. Some people comment that philosophy in China is nothing but an annotation to the six Confucian classics–or, likewise, that all Western philosophy is nothing but footnotes to Plato. Despite being different from science in some aspects, philosophy should share with science the possibility of breakthroughs. If philosophy is nothing but annotations to previous philosophical works, it would tend to be confined in an updated world view. And indeed, Chinese traditional philosophy and science has retained the conceptual framework which consists of *qi*, *yin yang* and *wuxing*, so neither Chinese science nor philosophy have entered a modern time. In fact, current philosophy, both in China and in the West, is much more advanced and sophisticated than the six Classics and Platonic philosophy, precisely because philosophers on both sides have not been confined to the outdated world view formulated in the works of previous philosophers.

We can see that the history of philosophy is filled with creative and wonderful ideas, thoughts and principles which were invented by great philosophers over successive generations. These philosophers include philosophers in the West such as Plato, Aristotle, Democritus, Epicurus, Thomas Aquinas, Francis Bacon, John Locke, John Stuart Mill, Immanuel Kant, Georg Hegel, and Karl Marx. The great contributions that they made are far beyond annotations or glossaries to ancient Greek philosophers who are the first generation of philosophers, such as Thales and Heraclitus. As cited above, Hegel claims that philosophy is not just the sum of opinions. In our opinion philosophy should be a systemic knowledge on what entity exists or is real in the world (ontology), how to know it (epistemology) and what we ought to do to it (ethics), all with rational philosophical argumentation. A good philosophical theory should have greater explanatory, predictive, and performative (changing the world) power than others (Qiu 1996). If a philosophy consists only of glossaries of previous works done by previous philosophers it would not have adequate explanatory, predictive and normative power to deal with newly found entities or facts. Consider: we know that a person who is in an irreversible coma but on the respirator can have normal cardiopulmonary function. How to explain the fact? What is the entity: a living person or a brain dead? Will he/she

exist in this state forever? How should we treat her/him? We cannot find the solution of all these issues if we limit us to the history of philosophy or the glossaries of the works of previous philosophers, such as the six Confucian classics and Platonic philosophy. There would be a great blank space in philosophy except the part of the history of philosophy. To properly solve these issues we have to learn new facts, and invent new theories (such as new concept of death) out of the philosophical ivory tower.

This dogma is institutionalized in China, upheld by senior and younger philosophers for lifetimes. It controls not only the philosophy departments of universities and the philosophy institutes of the academy of social sciences, but also the publishing houses and even the funding and professional title review institutions. If you don't comply with this dogma, your work is not considered philosophical. Your chance of getting published is poor, your opportunity of professional promotion is slim, and your graduate students cannot obtain their degrees. The proper work of both university philosophy departments and of the philosophy institutes of the academy of social sciences is to compile or translate the history of philosophy, or write about historical philosophers. Instead of philosophies developed by Chinese own philosophers, philosophy in China has become a philosophy of talking about the philosophies developed by the ancient philosophers.

This dogma has become a defense to support philosophers who feel at ease in their ivory tower and are not concerned with philosophical issues raised in social practices and actual reality, especially those issues raised by emerging technologies and social reform. A visit of Chinese philosophers to Europe is indicative in a couple of ways. When Chinese philosophers visited Britain in 1980s, they found a paper titled "The doctor, the pill, and the fifteen-years-old girl: a case study in medical ethics and law" written by Ian Kennedy (Kennedy 1985). This paper was of great interest because there is never such an article in China which is so practical and also very creative. Kennedy provides a convincing answer that the doctor ought to prescribe the conceptive pill to the fifteen-years-old girl with rigorous arguments of doctor's obligation to her/his patient and possible outcomes if he/she refuses to prescribe. Kenney's paper, together with other papers written by British philosophers and lawyers, led to revise the legal regulation which limited the age of girls who could get access to contraceptive pills to 16 years-old. This event reveals the fact that philosophy can change the world if philosophers can persuade relevant policy-makers or law/rule-makers to change their decision. Then Chinese philosophers visited Karl Marx's grave and on the grave there is an epitaph which says "The philosophers have only interpreted the world in various way, the point, however, is to change it" (Marx 1976). Kennedy's article is exemplary of philosophy which changes or improves the world, a clear contrast with Chinese philoso-

phers who have been immersed in the history of philosophy without concern for real-world affairs. Because of their complexities, uncertainty and transformational power, emerging technologies (technologies such as gene editing, synthetic biology, artificial intelligence, and robotics) have raised sophisticated philosophical issues–issues which need to be solved for application and further research. For example, in the case of synthetic biology there have been a series of philosophical questions regarding ontological, epistemological and ethical issues such as "Is synthetic genome a machine or a form of life?", "Do we know the life via creating life?", "Is synthetic biology Playing God, so it should be banned?", "Does synthetic biology create an unnatural life so compromise the value of life?", "Does synthetic biology undermine the relationship between the nature and human beings?" and many others. When the People's Republic of China was founded in 1949 an urgent task was to liberate women from gender injustice which is deeply rooted in traditional ideology and institutions, and mobilize them to participate in the construction of new republic, almost no philosophers who immersed into the history of philosophy stood out to criticize the gender prejudice in traditional culture including Confucianism which has notorious gender prejudice in its core theory.

2 The Second Dogma: "All Issues in the World Can Be Solved by a Well-designed Universal System of Philosophical Theory."

Now let us to move to the second dogma, that is, that all issues in the world can be solved from a well-designed universal system of philosophical theory with deduction. Many younger philosophers and Ph.D. students in philosophy departments have been haunted by this dogma, spending their whole lives to build a well-designed universal system of philosophical theory. The hope is that all issues in the world can be solved by deduction from such a theory. So far, however, no such system has been developed. In what follows we will argue that on the contrary, real philosophical issues that arise in practice–in particular, ethical issues–cannot be solved by mere deduction from a universal theoretical system which is designed and constructed only in a certain period of time in human history.

2.1 Arguments against the Second Dogma: "Impossibility of predicting the future events"

The second dogma relies upon several assumptions, such as the following:

> Assumption 1. The past is similar with the present and future, there is nothing new under the sun in the human history;

Assumption 2. We can know everything which has happened in the human history;

Assumption 3. We can predict what will happen in the future of human activities.

It follows from these assumptions that we can solve present and future issues by merely deducing from a universal theoretical system which is designed and constructed only in a certain period of time.

We think that a strong argument against the second dogma is this: when designing a well-designed universal system of philosophical theory to solve all present and future issues, the designer cannot know everything that has happened in the human history; nor can he/she predict what will happen in future human activity. There is no way to know which issues would be raised after the system is designed because the present will be quite different from the past, and the future will be more different from the present, and the new events always happen without precedent and beyond the expectation of all people. In such situation the designer cannot predict what issues will be raised, so he/she cannot contain the solution of these new issues into her/his system. The impossibility to predict the future issue is based on the impossibility of predicting the future of the humankind. As Karl Popper argues in his *The Poverty of Historicism* (Popper 1957), there are four reasons why it is impossible for us to predict the future of the humankind:

First, a description of the whole of history of the whole of society is impossible because the list of characteristics making up such a description would be infinite. If we cannot know the whole of the present state of the mankind it follows that we cannot know the future of the humankind. Popper writes:

> If we wish to study a thing, we are bound to select certain aspects of it. It is not possible for us to observe or to describe a whole piece of the world, or a whole piece of nature; in fact, not even the smallest whole piece may be so described, since all description is necessarily selective. (Popper 1957, p. 77)

Second, human history is a single unique event. Therefore, knowledge of the past does not necessarily help one to know the future (Popper 1957, p. 108).

Third, individual human action or reaction can never be predicted with certainty, therefore neither can the future: "the human factor is the ultimately uncertain and wayward element in social life and in all social institutions. Indeed this is

the element which ultimately cannot be completely controlled by institutions (as Spinoza first saw)" (Popper 1957, p. 158).

Fourth, it is impossible to know the future course of history when that course depends in part on the future growth of scientific knowledge, because we cannot predict the growth of scientific knowledge. So we cannot predict the future course of human history. In last century, nobody expected the emergence of relativity theory and quantum mechanics. In this century, nobody expected the invention of gene editing and synthetic genome; nor did they software of artificial intelligence such as DeepBlue which beat the chess Master Kasparov or AlphaGo which won the Go Master Lee Sedol. Suppose that the designer of this universal system lived before gene editing technology was invented. The designer cannot predict any ethical issues raised by this technology. For example: Should we permit ex vivo fetal research in gene editing? Should we permit to conduct clinical trials on somatic gene editing and treat it an ordinary therapy? Should we conduct clinical trials on inheritable gene editing? Do we have obligations for the wellbeing of future generations the health and life of which will be impacted on by inheritable gene editing? Should we enhance our bodily and mental functions beyond those that our species have? And: should we conduct gene editing on non-human organism, including eliminating harmful species such as *Aedes aegypti* which spreads malaria and Dengue fever, reviving eliminated species such as mammoth, or produce legend species such as unicorn? And so on. If the designer has no idea of these ethical issues, he or she cannot include the solution of these ethical issues into her or his system.

2.2   The Possible Counter-Arguments

One argument may claim that there are many cases in ordinary life in which the issue raised in practice can be readily solved by some existing theory or principle or rule. This is true, but the issue must be a routine issue. Consider the following inference:

(1) All patients who suffer from pneumonia without any hypersensitive or immune reactions and other complications should be prescribed with penicillin;

(2) Patient A suffer from pneumonia without any hypersensitive or immune reactions and other complications; and

(3) Patient A should be prescribed with penicillin.

In this routine case there is no ethical difficulty: we can easily see what the physician ought to be done by deducing from existing rules.

However, consider another case. In a far more significant example, the deduction does not work anymore:

Patient B suffers from pneumonia without any negative reactions or other complications, but he refuses to be treated. An ethical issue is raised: What ought the physician to do? Should the physician prescribe penicillin and force the patient to receive the treatment, or should he withhold treatment and respect the patient's preference? The physician cannot solve this ethical issue merely by deduction from any systematic theory, either from a novel well-designed philosophical theory or from a historical theory. An ethical issue is an issue about what ought to be done. And the designer of any particular philosophical system, contemporary or historical, cannot predict it when he or she design the system as human action is uncertain and wayward both in patient's side and physician's side.

As Popper argues above, individual human action or reaction can never be predicted with certainty. Therefore, the future cannot be predicted. In the case in question, the physician should not look for an answer from any available theory. Instead, he or she ought to ask B why he or she decides to refuse the treatment. The reasons of which the patient B refuses to take penicillin may be various:

The patient B may tell the physician that he or she has been at the late stage of bone cancer with intractable pain–that pneumonia is "the friend of elderly" as it causes death with less suffering in the elderly even if it results in B being dead earlier.

The patient B may tell the physician that one of her or his friend was killed by injecting penicillin, so he or she is hesitant as he or she has no confidence in the quality of penicillin available. On these grounds, B may refuse to take penicillin at least temporarily.

The patient B may tell the physician that he or she has heard that the mycobacterium tuberculosis generates a new variant via gene mutation which has resistant power to penicillin.

The patient B may tell the physician that he or she has a phobia of being injected.

And B may have other reasons to refuse to take penicillin. Also it may be the case that the patient B may be in an oscillating situation and he or she may switch from one option to the other option from time to time.

As an individual human the doctor D who treats B may have various belief-value system and he or she may come from different culture:

D may be a medical paternalist and insist that the clinical decision should be made by the physician rather than the patient or anybody else.

D may be a Confucian from China and think that in clinical context a physician must put the priority on the principle of beneficence and assume the responsibility

for patient's good, that is: treat her or disease, restore health, and prolong her or his life as far possible despite the patient's refusal.

D may be a believer of Kantian philosophy and think that a doctor must put the priority on the principle of respect for patient's autonomy, following the patient's decision even if it is irrational or uninformed.

These scenarios show that the designer of a universal philosophical system cannot predict all of the various human uncertain and wayward ideas and actions, so the system cannot provide the solutions to future ethical issues.

When confronted with such variables from the patient's and physician's preferences, the physician is faced with a moral dilemma: Allow patient B to die from pneumonia without the treatment of penicillin, complying with the ethical principle of respecting the patient's autonomy; or allow the patient to linger with intractable pain, complying with the ethical principle of beneficence (treating disease, restoring health, avoiding premature death). The physician is faced with conflicting obligations. One of his or her obligations is "provide benefit to the patient" (beneficence); the other is "respect the patient as a person" (autonomy). When these two obligations cannot be carried out at same time, the physician is trapped in a moral dilemma.

How does the physician escape from this dilemma? When we deal with this kind of moral dilemma, we find two things. First, we find that these cases present different values which we have to weigh. In the case above, one value is patient B's preference or autonomy, and the other value is the patient B's life being prolonged in a certain time.

We could formulate the moral dilemma above as follows (P=the physician, M and N=options of clinical decision):

(1) P ought to do M;

(2) P ought to do N;

(3) P cannot do M and N both at same time;

(4) If M overrides N, P should make decision of doing M;

(5) If N overrides M, P should make decision of doing N;

(6) If M does not override N, and N does not override M because of uncertainty and ignorance; now what P ought to do?

In the formula above it is commonly accepted by both the medical community and the public that the physician has the obligation to do benefit patients (M) while at

the same time respecting the patient's autonomy (N). In (3) M and N are in conflict; P is cast in a moral dilemma in which he has to weigh which obligation between M and N should be prioritized. In (4), P judges that M overrides N and makes the decision of doing M. In (5) after weighing P judges that N overrides, P makes the decision of doing N. However, in (6) M does not override N, nor N does override M. In such a case it is difficult for P to make a decision following any rule. Instead, the decision must be made on a case-by-case basis. Suppose that P lives in a plural cultural context in which whether M overrides N can only be decided on a case-by-case analysis, and P's own value system will have great impact in clinical decision making. Indeed, this is just the case in China, where both Confucianism (beneficence should override respect for patient's autonomy) and Kantianism (respect for patient's autonomy should override beneficence) are influential in both the medical community and the public. In this context, a doctor such as P ought to weigh the outcomes of different options: Which will bring about lesser harms and more benefits or interests to B, but not the deduction from a philosophical system which is designed by some philosophers who fail to predict all possible variables on the side of the patient (such as B) and on the side of the physician (such as P). Our argument above indicates the futility of pre-designed universal systems of philosophical theory. P's appropriate decision cannot be deduced from the system. Instead, P has to weigh the value in his or her own settings.

There may be a counter argument which says that if we design a Kantian system, the above case can be solved by prioritizing the patient's (B's) autonomy. However, in the Chinese cultural context human life is assigned more weight than human autonomy. Furthermore, the decision made by the patient (or the family) may be made out of irrationality or ignorance. In these cases the administrative authority can order the treatment. After such a successful mandatory treatment, the patient or family always expresses great gratitude to physician for saving their life. In the efforts to fight against COVID-19, China is one of the countries in the world with the most effective control over the epidemic. One reason is that nobody refuses to wear face mask, and nobody adopts the slogan, "Mask no, freedom yes." People observe social distancing receive vaccination. Chinese people believe that freedom only exists for a human when that human is alive. So a Kantian system will not work in China. Yet a Confucian system also does not work, because in the Chinese pluralistic context there are many groups who do not favor Confucianism, instead they prefer individual autonomy or self-determination. Only decisions made with the patient's best interest in mind may be accepted by both the patient and the physician. However, even in such a case, what is the best interest for the patient can be determined only on a case by case basis; both sides need to weigh the specifics and communicate with each other to reach agreement.

The pre-designed system is futile in this endeavor.

We wish to make two final points. First: in the clinic or during an epidemic the situation changes quickly. Under such conditions, philosophers cannot predict all possible issues raised by these changes and develop in advance the rules to solve these issues by a pre-designed system of philosophical theory. Second: when we make a moral judgment, we must know the facts, especially facts which cause a new ethical issue or moral dilemma. Take the example of COVID-19 pandemic. When many people became infected with coronavirus and some of them died, we needed to know what the coronavirus is, what medical conditions it caused, how to treat these conditions, and how to prevent or control the epidemic. The pandemic raised philosophical issues in scientific, technological, medical practices. To solve these issues, we cannot appeal to any philosophical theory, contemporary or historical, designed before the coronavirus pandemic. Instead, we have to study what actually happens in the real world.

3   Conclusion

The conclusion is: Pre-designed systems of philosophical theory which are intended to solve all present and future issues are in vain.

References

Chen, Bo. 2010. "Facing Philosophical Problems and Participating in Contemporary Philosophy Construction." *Academic Journal of Jinyang* 4.

Han, Donghui. 2015. https://www.zhihu.com/question/27695967.

Hegel, G.W.F. 1892. *Lectures on the History of Philosophy.* Translated by H. Sanderson. London: K. Paul.

———. 1892-1896. "Hegel's Lectures on the History of Philosophy." Translated by E S. Haldane. https://www.marxists.org/reference/archive/hegel/works/hp/hpintroa.htm.

Jia, Yushu. 2010. "Out of the Misunderstanding of 'Philosophy Being the History of Philosophy'." *Academic Journal of Jinyang* 6.

Kennedy, Ian. 1985. "The doctor, the pill, and the fifteen-years-old girl: a case study in medical ethics and law." In *Moral Dilemmas in Modern Medicine,* edited by M. Lockwood. Oxford: Oxford University Press.

Marx, Karl. 1976. "Theses on Feuerbach." In *Marx K. Engels F. Collected Works, volume 5, 1845-1847,* vol. 5. New York: International Publishers.

Popper, Karl. 1957. *The Poverty of Historicism.* London: Routledge.

"Positivism." 2021. https://en.wikipedia.org/wiki/Positivism.

Qiu, Renzong. 1996. "How to know what rises up is the Moon? – On the concept of realism and the irrelevancy of quantum mechanics to the debate on realism vs antirealism." In *Realism and Anti-Realism in the Philosophy of Science,* edited by R. Hilpinen Cohen and Renzong Qiu. Dordrecht: Kluwer Academic Publishers.

Yao. 2020. https://www.zhihu.com/question/422560698/answer/1491688105.

黑格尔. 1959. 《哲学史讲演录》, 第一卷. 贺麟, 王太庆译. 北京: 商务印书馆.

Dao De Jing and *Mūlamadhyamakakārikā*
Making Sense of Ineffability

Graham Priest

Departments of Philosophy, CUNY Graduate Center*
University of Melbourne
Ruhr University of Bochum

ABSTRACT | In a number of philosophical traditions, East and West, there are things that are held to be beyond the limits of language, that are ineffable. Philosophers in these traditions also explain *why* there are such things, in the process, describing them. The things would hence appear to be both effable and ineffable. Though one may try to wriggle out of these contradictions, the most straightforward response is simply to accept that we are dealing with contradictory objects. In the first part of the paper, we will see how the phenomenon in question arises in the Daoist and Mahāyāna Buddhist traditions. In the second, we will see how the contradiction concering an object which is both effable and ineffable can be handled using the techniques of paraconsistent logic.

---

* *Correspondence:* Graham Priest – CUNY Graduate Center, 365 5th Ave, New York, NY 10036, USA. Email: priest.graham@gmail.com

1   Introduction

This paper concerns a curious phenomenon: the effable ineffable. One finds this, amongst other places, in certain interpretations of Daoism and in certain kinds of Buddhism – most notably Madhyamaka, but more generally other branches of Mahāyāna. According to both of these views, there is something that is ineffable. Both of these views, however, talk about this ineffable thing: indeed, they explain *why* it is ineffable. That is, of course, a contradiction. Moreover, it is a contradiction which is entirely obvious. What to make of the matter?

We might argue about what, exactly, it means to say that an object is ineffable, though the following discussion is largely independent of the precise details of a definition. Thus, in an attempt to get out of the problem, one might insert an *ad hoc* escape clause: an ineffable object is one about which one can say nothing, *except* that it is ineffable. This won't help here, since, as we will see, a lot *more* is said about the objects we will be concerned with.

One may, of course, try to defuse the contradiction in some more sophisticated way. And certainly some Buddhism thinkers did just this,[1] though I will not go into the matter here.[2] Another strategy – and the one investigated in this paper – is simply to accept the contradiction. This is certainly a possibility if one is a dialetheist and is prepared to deploy a paraconsistent logic. But how, exactly, does one deploy these techniques? This paper provides an answer to this question.

In the first part of the paper, we will look at the Daoist and Buddhist texts which generate the phenomenon. In the second part of the paper, we will look at the logical techniques which may be deployed to handle it. An interlude on paraconsistent logic explains the relevant basics of paraconsistent logic for those who have not met them before.[3]

2   Daoism and Buddhism

2.1   The *Dao De Jing*

First, the *Dao De Jing*, 道德經. This is traditionally supposed to have been written by a character called Laozi, 老子, a rough contemporary of Confucius (5th c. BCE).

---

[1] For example, Gorampa (1429-1489). For a discussion, see (Priest 2018, 6.2).

[2] Some familiar moves in the game are discussed in (Priest 202+).

[3] This paper is a written-up version of a talk with the same title given at the conference *Dialetheism and Related Issues in Analytic Asian Philosophy: an International Workshop*, Kyoto University, June 2017. Parts of it have since appeared in (Priest 2018), esp. (Priest 2019b, ch. 5), (Deguchi, Garfield, Priest, and Sharf 2021). Thanks go to an anonymous referee of this journal for helpful comments on an earlier draft.

However, modern scholarship suggests that it is a text of a much later date, and is simply a collection of sayings of old masters (one meaning of *laozi*). That is certainly the way it reads, since it appears to contain multiple voices.

Exactly how to interpret the text and its passages is a matter of some contention. However, we are going to focus here on just one verse, and the interpretation given to this by one of the most influential Neo-Daoist commentators, Wang Bi (王弼, 226-249 CE).

The *Dao De Jing* opens famously with the lines (Lynn 1999, p. 51):[4]

> The Dao that can be described in language is not the constant Dao; the name that can be given it is not the constant name.

Wang Bi's commentary on this goes as follows (*Ibid*):

> The Dao that can be rendered in language and the name [*ming*] that can be given it point to a thing/matter [*shi*] or reproduce a form [*xing*], neither of which is it in its constancy [*chang*]. This is why it can neither be rendered in language nor given a name.

The *Dao De Jing* then continues (*Ibid*):

> Nameless it is the origin of the myriad things; named it is the mother of the myriad things.

Wang Bi's commentary (*Ibid*):

> Anything that exists originates in nothingness [*wu*], thus, before it has forms and when it is still nameless, it serves as the origin of the myriad things, and once it has forms and is named, it grows them, rears them, ensure them their proper shapes, and matures them as their mother. In other words, Dao, by being itself formless and nameless, originates and brings the myriad things to completion. They are originated and completed in this way yet do not know how it happens. This is the mystery [*xuan*] beyond mystery.

The thought is clear. Behind the flow of phenomenal events, there is a certain principle, Dao, 道, which generates the flux. Dao, though, cannot be named, that is, cannot be described.

Neither the *Dao De Jing* nor Wang Bi, gives an argument for the existence of Dao. However, Wang Bi does give an argument for its ineffability. In fact, he gives two. Here is the first (Lynn 1991, p. 30):

---

[4] Quotations are taken from (Lynn 1999), to which page references refer. Interpolations are the translator's.

> The way things come into existence, and efficacy [*gong*] comes about, is that things arise from the formless [*wuxing*], and that efficacy emanates from the nameless [*wuming*]. The formless and the nameless [the Dao] is the progenitor of the myriad things. It is neither warm nor cool and makes neither the note *gong* nor the note *shang*. If it were warm, it could not be cold; if it were the note *gong*, it could not be the note *shang*. If it had a form, it would necessarily possess the means of being distinguished from other things; if it made a sound, it would necessarily belong among other sounds.

In other words, Dao cannot be characterised in any way. For if it could, it would not be able to be something else. But since it can be all things, it can't be characterised.[5]

Wang Bi's second argument goes as follows (Lynn 1999, p. 32):

> … even the most replete [Dao], as long as it can still be expressed in words, would never have the capacity to govern Heaven and Earth, and the greatest thing that can possibly have form would never be large enough to house the myriad things. … Any name for it would fail to match what it is. Any comparison for it would fail to express all that it is. A name necessarily involves how one thing is distinct from other things, and a comparison necessarily involves how the tenor of one thing depends on the vehicle of another. Making distinctions, any name would fail to be inclusive; being dependent, any comparison would fall short of all that it is. As it cannot be perfectly inclusive, any name for it would deviate greatly from the truth; as it cannot express all that it is, any comparison for it would fail to designate what it really is.

The thought here is that Dao is so great that it transcends any linguistic categorisation.[6]

Now, whether or not these are good arguments, it is clear that Wang Bi is explaining why Dao is ineffable, and, in doing so, describing, that is, characterising, it in certain ways.

Here, then, is our phenomenon in the *Dao De Jing*. A text that talks about the ineffable.

---

[5] It is worth noting that an argument of exactly the same kind is given by Plato in the *Timeaus*, for the ineffability of the stuff which receives the forms, $\chi\omega\rho\alpha$. See (Sorabji 1988, pp. 32ff.).

[6] Again, it is worth noting that a similar argument for the ineffability of God was used by the 15th Century Neo-Platonist, Nicholas of Cusa. See (Priest 2002, pp. 22ff.).

2.2 Buddhism

Now let us turn to Buddhism. All forms of Buddhism distinguish between the conventional reality of the way that the world appears to us (Skt: *saṃvṛti satya*), and the ultimate reality of the way that the world really is (Skt: *paramārtha satya*). Different schools of Buddhism, however, interpret the distinction in different ways. Here we are concerned with Mahāyāna Buddhism, and particularly Madhyamaka.

Around the turn of the Common Era, a new kind of sūtra started to emerge, the *Prajñāpāramitā* (Perfection of Wisdom) *Sūtras*. One thing these do is to claim that to grasp ultimate reality, words are of no use. Thus we have in the *Aṣṭasāhasrikā Prajñāpāramitā Sūtra*:[7]

> All words for things in use in this world must be left behind,
>
> All things produced and made must be transcended –
>
> The deathless, the supreme, incomparable gnosis is then won.
>
> That is the sense in which we speak of perfect wisdom.

And in another *Prajñāpāramitā* sūtra, the *Vajracchedikā* (Diamond) *Sūtra*, we have:[8]

> [The Buddha said]: Subhūti, words cannot explain the real nature of the cosmos. Only common people fettered with desire make use of this arbitrary method.

The point is taken up by Nāgārjuna, who flourished some time in the 2nd c. CE, in his *Mūlamadhyamakakārikā* (MMK), the first philosophical text to try to make sense of the new sūtras.[9] Thus we have (MMK XIII: 8):[10]

> The victorious ones have said
>
> That emptiness is the elimination of all views.
>
> For whomever emptiness is a view
>
> That one will accomplish nothing.

The views to be eliminated are not, of course, those of conventional reality. They are perfectly good, just merely conventional. The views to be given up are those concerning ultimate reality.

---

[7] (Conze 1973, p. 12).

[8] (Price and Wong (trs.) 1990, p. 51).

[9] It must be said that interpreting this text is a contentious matter. The interpretation given here is in line with that to be found in (Garfield 1995), and (Priest 2018, ch. 4).

[10] Translations from the MMK are taken from (Garfield 1995).

But why should ultimate reality be ineffable? Nāgārjuna indicates the reason in the following passage (MMK XVIII: 9):

> Not dependent on another, peaceful and
>
> Not fabricated by mental fabrication,
>
> Not thought, without distinction.
>
> That is the character of reality.

The reality in question is, of course, ultimate reality. The point is that conventional reality is conceptually constructed. Ultimate reality is what one arrives at when one "peels off" the concepts. Concepts (mental fabrications), cannot, then, be applied to ultimate reality. To describe it, one would have to deploy concepts, and so make it merely conventional; so one cannot describe it.

So important is the point, that it is made at the very start of the MMK in its dedicatory verses:

> I prostrate to the Perfect Buddha
>
> The best of teachers, who taught that
>
> Whatever is dependently arisen is
>
> Unceasing, unborn,
>
> Unannihilated, not permanent,
>
> Not coming, not going,
>
> Without distinction, without identity,
>
> And free from conceptual construction.

Again, the ultimate is concept-free.

But as we see, Nāgārjuna characterises the ultimate in many ways. Indeed he does so in saying that it is not conceptually constructed. Here, then, is our Buddhist example of the phenomenon: a text which talks about the ineffable.

2.3   Taking Stock

Let us now take stock. The situation is as follows. Dao/ultimate-reality is ineffable. But this is explained, thereby describing it. So it is effable.

In fact, what is driving this contradiction is something deeper. The effability/ineffability of something is grounded in its very nature: its objecthood/non-objecthood. For note that:

• Something is an object iff one can say something about it.

If something is an object, one can say something of it – for example that it is an object, or that it is self-identical. Conversely, if something is not an object, one can say nothing of it. For to say anything is to predicate. This requires an object to which to apply the predicate.[11]

This thought will guide us into the formal analysis which follows. But before we turn to that, let me point out the following. It might be thought that the phenomenon we are dealing with is something peculiar to the Asian philosophical traditions. It is not. It is to be found in Western philosophical traditions too. Let me give just two examples.

In the *Tractatus*, Wittgenstein provides an analysis of language, reality, and the relationship between them. Language is constituted by propositions. These are composed of names, enformed in a certain way. Reality is constituted by states of affairs. These are composed of objects, enformed in a certain way. A proposition, *p*, describes a state of affairs, *s*, if the names in *p* refer to the objects in *s*, and *p* and *s* have the same form. But form is not an object: it is the way that names or objects are put together. So one can say nothing about it. But the *Tractatus* is full of statements about form: indeed, it explains why one cannot talk about it. Wittgenstein is, of course, well aware of the matter, and it leads to the stunning dénoument of the *Tractatus*.

In *Sein und Zeit* Heidegger asks the *Seinsfrage*: what is being? And he immediately tells us that there is a singular mistake to avoid. Being is not itself *a* being, that is, it is not an object: it is that in virtue of which objects (beings) *are* objects (beings). But of course, in an attempt to answer the *Seinsfrage*, Heidegger's works are full of statements about being. And to say anything of something is exactly to treat it as an object—indeed, even the *Seinsfrage* itself treats it as an object. Heidegger is also well aware of the matter, and struggles in numerous ways to deal with it in his later writings.

What Wittgenstein and Heidegger made of these matters, is not our concern here. I simply want to point out that the phenomenon we are dealing with is not restricted to Eastern philosophy. Indeed, it is one of the most profound metaphysical issues, East and West. A thinker comes to the conclusion that there are limits to what can be said. Some things are ineffable. And the thinker explains exactly why they are so, thus describing the ineffable.[12]

---

[11] Indeed, to say of some thing that it is not an object must be false (i.e., have a true negation). Any non-object will, then, be contradictory—as we shall see.

[12] For more on the matter, see (Priest 2002). On Wittgenstein, see (Priest 2002, ch. 12). On Heidegger, see (Priest 2002, ch. 15).

3   Interlude: Paraconsistent Logic

In the next section, we will turn to the matter of how to analyse the situation in which our thinkers find themselves, by deploying the tools of paraconsistent logic. These tools may well be unfamiliar to some readers, so in this interlude, I will explain the relevant basics of simple paraconsistent logic, LP.[13]

First, classical (propositional) logic. Every situation – or interpretation as logicians call it – divides up sentences into those that are true, $\mathfrak{T}$, and those that are false, $\mathfrak{F}$, these two categories being mutually exclusive and exhaustive. Negation toggles a sentence between these two zones. So if A is true, $\neg A$ is false; and if B is false, $\neg B$ is true. We may depict the situation thus:
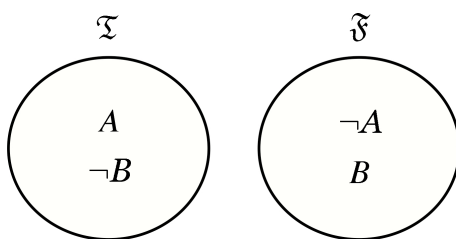
A valid inference is one such that there is no interpretation where the premises are true, and the conclusion is not. But as is clear, a sentence and its negation can never both be in the true zone. It follows that the inference of explosion is valid:

- $A, \neg A \models B$

If there is no interpretation in which A and $\neg A$ are both true, then, *a fortiori*, there is no interpretation in which they are both true *and* B is not.

Now to paraconsistent logic. This is exactly the same, with just one crucial difference: namely, the true and the false zones in an interpretation may overlap.[14] Negation still takes a formula from the true zone to the false zone, and vice versa.[15] But now, suppose that something is both true and false, that is, in the overlap between the two zones. Then its negation is both false and true, that is, it is in the overlap as well. Thus we can have the following interpretation:

---

[13] For full details see, e.g., (Priest 2008, chs. 7 and 21).
[14] There are some paraconsistent logics in which they may underlap as well, but that is not relevant to this story.
[15] Beware: being false is now *not* the same as not being true. (These correspond to different areas in the following diagram.)
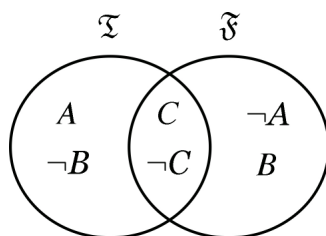
In this, C is in the overlap, so both C and $\neg C$ are true (and false as well, but that does not matter). B is not in the true zone; so we have a counter-example to explosion:

- $C, \neg C \nvDash B$

There is one more thing which we need to note. This concerns not propositional logic, but predication. A predicate has an extension and an anti-extension; these comprise the objects of which the predicate is true and the objects of which it is false, respectively. In classical logic, extension and anti-extension are the complements of each other. That is, the two zones are mutually exclusive and exhaustive. In paraconsistent logic, as one would expect, they, also, may overlap in an interpretation. So if P is monadic predicate, its extension and anti-extension are subsets of the domain. And if 'a' refers to an object in the overlap, Pa and $\neg Pa$ are both true (and false).

One predicate, in particular, will concern us in what follows: the identity predicate, $=$. This is a binary predicate, and so its extension and anti-extension are sets of *pairs* of the domain. As one would expect, the extension of the predicate is:

- $\{\langle d, d \rangle : d \in D\}$

That, in fact, is sufficient to guarantee all the standard properties of identity. The anti-extension of the identity predicate can be any set which *contains* the complement of this. In particular, then, the extension and anti-extension of the identity predicate may overlap. Thus, whatever d is, $\langle d, d \rangle$ is in the extension of $=$; so if 'a' refers to d, then $a = a$ is true. But if $\langle d, d \rangle$ is in the anti-extension as well, then $a = a$ is false; that is $\neg a = a$ is true as well.[16]

---

[16] Quantifiers, I might note, work in exactly the same way in paraconsistent logic as they do in classical logic.

4   Making Sense of the Effable Ineffable

4.1   Objecthood and Identity

Now let us return to the topic of ineffability. First, what is it to be an object? The answer is simple: to be an object is to be something. That is, x is an object iff:[17]

- $\exists y\ y = x$

$x = x$ is a logical truth. It follows that $\exists y\ y = x$, and so that $\forall x \exists y\ y = x$. That is, everything is an object. No surprises there.

   But suppose that x is not an object. That is, $\neg \exists y\ y = x$; then $\forall y \neg y = x$; and so, in particular, $x \neq x$. So if x is an object that is not an object, $x = x$ and $x \neq x$.[18]

4.2   Naming

Next, we turn to the issue of naming. An intuitively correct principle concerning truth is the familiar T-Schema:[19]

- $T \langle A \rangle \leftrightarrow A$

Here, Tx is the truth predicate (x is true), A is any closed sentence, and angle brackets are a name-forming device. Naturally, cognate semantic notions, such as satisfaction and denotation, are governed by similar schemas. The one for denotation is the less-familiar D-Schema:[20]

- $\forall x (D(\langle n \rangle, x) \leftrightarrow n = x)$

$D(y, x)$ (y denotes x) is the denotation predicate, and n is any name. Now, take any object that is not an object. Call this 'μ'. Then, instantiating the D-Schema, we have:

- $D(\langle \mu \rangle, \mu) \leftrightarrow \mu = \mu$

---

[17] The particular quantifier here, $\exists$, note, should not be taken as "existentially loaded". That is, $\exists y$ is simply read 'some y is such that', not as 'there exists a y such that'.

[18] The converse also holds. For suppose that $x \neq x$. Then either $x = y$ or $x \neq y$. In the first case, $x \neq y$, by the substitutivity of identicals. So in either case $x \neq y$. That is, $\forall y \neg x = y$, i.e., $\neg \exists y\ y = x$.

[19] Of course, some people (though not myself) do not accept this in full generality, since, given a few other conditions, it delivers contradictions such as the liar paradox. These concerns are not relevant here, and so I ignore them.

[20] See, e.g., (Priest 2005, ch. 8.2).

Since $\mu = \mu$, it follows that $D(\langle\mu\rangle, \mu)$. That is, '$\mu$' names μ (naturally).

But since 'μ' refers to something that is not object:

- $\neg\exists y \ y = \mu$

That is:

- $\forall y \ y \neq \mu$

So for any name, 'n':

- $n \neq \mu$

By the D-Schema again:

- $D(\langle n\rangle, \mu) \leftrightarrow n = \mu$

So by contraposition:[21]

- $\neg D(\langle n\rangle, \mu)$

That is, μ has no name – not even 'μ'!

4.3   Statements about μ

And now, at last, we can turn to ineffability. Note that if the statement 'Pn' is about an object, x, then 'n' must refer to x. But as we saw in the last subsection, μ has no name, so one can say nothing about it. Hence, it is ineffable.

But again, and as we also saw in the last subsection, 'μ' is a name for μ, and so one *can* say things about it, such as that it is self identical, $\mu = \mu$; or that it is and is not an object, $\exists y \ y = \mu$ and $\neg\exists y \ y = \mu$.

The simple paraconsistent technical machinery delivers exactly what is required.

5   Conclusion

In the first half of this essay, we saw that versions of both Daoism and Buddhism are committed to there being something that is ineffable (Dao, ultimate reality). Both, however, talk about it, and so are committed to its effability as well.

---

[21] There is an issue here about whether one should expect the conditional employed in schemas of this kind contrapose. However, let me set that issue aside here.

In the second part of the essay, we have seen how one can make precise and rigorous sense of the matter, using some simple techniques of paraconsistent logic.

That does not, of course, show that these views are true; that is an entirely different matter. However, it does show that the dialetheism of these views cannot be rejected simply on the grounds of logical incoherence. The situation provides another possible application for a dialetheic application of paraconsistent logic, to boot, adding to those concerning the paradoxes of self-reference, motion, vagueness, and the law.[22] Is this application anachronistic? Literally, yes, of course. Objectionably so? No. No more so than is using 20th Century mathematics to analyse Newtonian physics.

References

Conze, Edward (tr.) 1973. *The Perfection of Wisdom in Eight Thousand Lines and its Verse Summary.* Delhi: Sri Satguru Publications.

Deguchi, Yasuo, Jay L. Garfield, Graham Priest, and Robert H. Sharf. 2021. *What Can't be Said: Paradox and Contradiction in East Asian Thought.* New York: Oxford University Press.

Garfield, Jay L. (tr.) 1995. *The Fundamental Wisdom of the Middle Way.* New York: Oxford University Press.

Lynn, Richard J. 1999. *The Classic of the Way and Virtue: a New Translation of the Tao-te Ching of Laozi as Interpreted by Wang Bi.* New York: Columbia University Press.

Price, A. F., and Mou-lam Wong (trs.) 1990. *The Diamond Sūtra and the Sūtra of Hui-Neng.* Boston: Shambala.

Priest, Graham. 1987. *In Contradiction.* Dordrecht: Martinus Nijhoff; 2nd edn, Oxford: Oxford University Press, 2006.

———. 2002. *Beyond the Limits of Thought.* 2nd ed. Oxford: Oxford University Press.

---

[22] See, e.g., (Priest 1987), (Priest 2019a).

———. 2005. *Towards Non-Being.* Oxford: Oxford University Press; 2nd edn, 2016.

———. 2008. *Introduction to Non-Classical Logic: From If to Is.* Cambridge: Cambridge University Press.

———. 2018. *The Fifth Corner of Four: an Essay on Buddhist Metaphysics and the Catuṣkoṭi.* Oxford: Oxford University Press.

———. 2019a. "Dialetheism and the Sorites Paradox." In *The Sorites Paradox,* edited by S. Oms and E. Zardini. Cambridge: Cambridge University Press.

———. 2019b. "Objects that are not Objects." In *Quo Vadis Metaphysics? Essays in Honor of Peter van Inwagen,* edited by M. Szatkowski. Berlin: De Gruyter.

———. 202+. *God and the Paradox of Ineffability.* to appear.

Sorabji, Richard. 1988. *Matter, Space, and Motion.* London: Duckworth.

Phenomenology of Precarity

Alice Pugliese

University of Palermo*

ABSTRACT | Edmund Husserl's phenomenological method has mostly been applied to epistemological questions. However, it also has great potential for analysing phenomena of social and political relevance. This text outlines a phenomenology of social precariousness, showing how it impacts the experiential, temporal, perceptual and self-perceptual structures of the subject, thereby influencing intergenerational relations.

KEYWORDS | Phenomenological Time; Autonomy; Precariousness; Precarity; Public Ethics

* *Correspondence:* Alice Pugliese – Department of Humanities, University of Palermo, Viale delle Scienze - Edificio 12, 90128 Palermo, Italy. Email: alice.pugliese@unipa.it

Alice Pugliese

1   Between Public and Private:
   the Point of View of Precarity

In what follows, I aim to use some fundamental instruments of the phenomenological method to address the wide-ranging social phenomenon of precarity. Precarity is characterized by insecurity of work and life conditions, pressure to adapt to constantly changing work tasks, times, and place of residence, as well as the reduction of the horizon of predictability of personal and social development.[1] This analysis of the growing precarity of life is of interest beyond only its relevance to our current situation. Rather, it relates to a constellation of problems suitable for showing the close constitutive relationship that exists between the personal, private, psychological, and individual dimensions of human existence and the institutional and collective dimension of associated life. In this sense, Judith Butler's description of precarity as a "politically induced condition"[2] acquires its full meaning. The explicit reference to the political structure of society not only indicates the genesis of precarity as a modern phenomenon, but also reveals its ultimate significance for anthropological and philosophical research.

Accordingly, describing precarity means observing the processes that dynamically bind the public and the private. I believe that phenomenology can provide fruitful methodological tools to explore the subjective meaning of this dynamic, thereby proving itself useful as a method for developing a possible public ethics. In this sense, the analysis of precarity gains an exemplary character within phenomenological scholarship since it questions the potential of phenomenological description to ground a reflection that does not separate individual moral choices from social context, without, however, dissolving the sense of subjectivity into macro-social dynamics.

---

[1] Relevant in this context is the distinction between the precariousness of life, on the one hand, intended as an ontological condition of finitude and contingency, also stressed by religious views and particularly by Christianity, and, on the other hand, "precarity" as a specific social phenomena connected to the current organization of work and social hierarchy. This distinction has been investigated by (Lorey 2015) The term precarity can be traced back to French sociologist Pierre Bourdieu, who intended it primarily as the expanding condition of job insecurity in meaningful connection with personal subordination (Bourdieu 1998, p. 82).

[2] In her approach, Butler not only distinguishes the precariousness of life from precarity, but also underlines how political and social institutions produce precarity by failing at their proper function, which is balancing the natural precariousness of the human condition, thereby *producing* precarity (Butler 2009, p. II).

2   Phenomenology as a Radical Empiricism

The French branch of the philosophy initiated by Edmund Husserl – animated by the influential work of J.-P. Sartre, M. Merleau-Ponty, P. Nizan and later E. Levinas – was forced to take a stand in the politically challenging context of the post-war era and was soon mixed with existentialism. Yet, phenomenology in the Husserlian sense remained for a long time caught up in what appear to be purely theoretical and epistemological questions, thereby keeping it far away from more concrete political issues. Only in recent years has interest in political phenomena grown to the point of inaugurating a new tendency dedicated to "critical phenomenology," which focuses on social and politically relevant phenomena such as race, gender, and justice.[3]

The phenomenological method, nevertheless, is based on an analysis of experience. As a description of phenomena, in fact, phenomenology can be characterised as a 'radical empiricism', i.e. as a philosophy aiming at bringing the peculiarity of experience, its structure, and dynamics to the fore. In the 1910/11 lectures, which represent Husserl's breakthrough from the *Logical Investigations* to a new phase of his reflection, the philosopher offers his audience a succinct statement of his concern: „Wir beschreiben, wie wohl zu beachten ist, nur das, was jedes Ich als solches vorfindet" ("We describe, it should be noted, only that which each I finds as such") (Husserl 2006, p. 3). Accordingly, phenomenology starts from what can be found directly by everyone. Each of us is considered a legitimate source of meaningful experience. Yet, the methodologically crucial demand for accuracy and faithfulness to phenomena harbours the hidden danger of dispersion. This is because beginning with what is found means "something different for each of us" (Husserl 2006, p. 2). The problem of phenomenology as a radical empiricism is, therefore, how to avoid getting lost in the multiplicity of objects and relative descriptions. This problem should not lead us to a relativistic or even sceptical solution. It cannot be solved by simply giving up the goal of a secured and certain knowledge of determined objects. Dealing with the problem of relativism, Husserl rather grasps the concept of "evidence" intended as an intermediate form of knowledge which is neither objective in the sense of the third-person-perspective, nor subjective in a psychological and solipsistic sense. The phenomenological description aims to unfold "evidence," intended as a reliable, meaningful and fulfilled relationship between the subject and the object.

Here, the role of the subject-as-observer comes to the foreground. Nevertheless, the subject itself is neither considered the final authority, which would lead to

---

[3] See the debate raised by the recent publication of (Weiss, Salamon, and Murphy 2019).

relativism, nor an interchangeable abstract hypothesis. Rather, it is characterised by its ability to transform object-related certainty into meaningful *evidence*. The ego functions neither as a creator of the world nor as a mere recipient of stimuli. Phenomenology rather emphasises the 'transformative' power of subjective experience and interprets it as a 'performance,' i.e. as an experience 'of someone' and not just 'of something.' Experience is not exhausted in the external relation between two pre-existing poles, but proves to be a phenomenon of transformation in which subjective and objective elements merge together.

Husserl refers to such a connection with the concept of "constitution,"[4] which points to the meaning-bestowing capacity of the subject. Unfortunately, the notion of constitution caused many idealistic misunderstandings as early as the publication of the *Ideas* (1913). Yet, it is precisely in the lectures of 1910/11 that a simple and very concrete description of the effective peculiarity of the ego can be found. Here, Husserl describes the ego as "'the thing' around which is grouped a material environment that continues into the infinite (das ‚Ding', um welches sich eine ins Unbegrenzte fortgehende dingliche Umgebung gruppiert)" (Husserl 1973, p. 113). The I does not constitute the world in the sense that it simply projects its own thoughts, representations, or wishes on the world, which would be taken as a mere surface for these items to be projected upon. Rather, in the process of constitution, the I concentrates and polarizes its perceptions, experiences, and expectations around the intended thing, thereby conveying it with a certain orientation and meaning. The permanent performance of subjectivity consists in imprinting directions and relations on an environment that would otherwise be a mere collection of things. Egoic life transforms mere environment into a meaningful horizon and – epistemologically – mere *subjective certainty* into *constitutive evidence*.

This means, however, that the full unfolding of subjective life needs a horizon that "continues into infinity"[5] and, at the same time, presupposes the possibility of standing in such a horizon and operating as an actively grouping centre. From this perspective, the constitutional potential of the ego represents, in my view, the lowest and unnoticed level of subjective autonomy that is indispensable for subjective life.

The notion of autonomy, which is traditionally found at the centre of political and moral debates, is thus observed to be originally rooted in the perceptive and self-perceptive performances of the subject, in its capacity to bestow sense to the surrounding world. Therefore, in order to approach the description of the effects of precarity on the subject's life, it will be necessary to briefly explore at least two

---

[4] See the important section dedicated to constitution in (Husserl 1989).
[5] For a phenomenologico-Husserlian analysis of the concept of infinity, see (Altobrando 2013).

elements that characterise subjective life by means of phenomenological analysis: autonomy and the internal temporality of the subject. By clarifying how these essential aspects of subjective life function it will become easier to shed light on the social processes that profoundly affect it, inducing a radical transformation deep into its proper foundations.

3 The Basic Form of Autonomy

As radical empiricism, phenomenology starts from the assumption that subjective life is only possible in and from experience. The syntheses that take place in our experience make up our entire reality, which is the only reality in which cognition, decision, and action are possible for us. This constitutive achievement represents a fundamental prefiguration and precondition of personal autonomy. The analysis of precarity must start precisely from this minimal idea of autonomy in order to focus on the functionality and needs of the subject so as to observe how they are transformed under current life circumstances. In this sense, autonomy should be understood not as a primarily moral character of the subject, but as its constant *performativity*.

The autonomy of the subject is expressed concretely in every perceptive performance in which the subject shapes its world and – above all – itself according to its viewpoint and needs. In this process, subjectivity delineates an environment of meanings and relevances. This environment is ultimately the only world in which we can live. In this infinite process, the ego also "constitutes" itself. The self-referential side of this dynamic that constitutes the individual life-story is inseparable from its world-constitutive side. The I explores the world by discovering and shaping itself. The relevancies that assert themselves in my perception and orient me within the world both point to and reveal certain structures, tendencies, and needs of my own person. Between the world and the subject, we find that there is not only a mere mirroring taking place,[6] but there is rather a reciprocal shaping process in which the two poles – subjective and objective – are only separable by abstraction.

Autonomy is an essential precondition of this complex achievement. Only secondarily does it refer to the dimension of morality and institutions. Much "earlier" in a genetic perspective, or more profoundly in an existential sense, autonomy denotes the sense-bestowing self-reference that produces and grounds myself as

---

[6] The metaphor of mirroring spread by (Rizzolatti and Sinigaglia 2008) to explain the functioning of a specific class of neurons is in this sense as useful as limiting if it is taken as the all-encompassing model to grasp human interaction.

a subject. My essence as a subject (not as an organic being, of course, but as an actor, as a subject of opinions and rights, as a bearer of emotions and relations) is founded on my ability to exercise my multifaceted capacities by relating to the world. The consistency of my own person is the consistency of my world. My own subjective coherence is grounded in the coherence of my relations and worldviews. My contradictions work themselves out in the tensions and divisions of my world. This process must not be understood as a flat projection or reflection of the ego onto the world, nor as an abstract metaphysical idealism. It is rather a creative and open process in which we become ourselves only by shaping our environment. Our being consists in the complex unfolding of personal autonomy through perception, exploration, and action.

4 The Expanding Horizon of Time

The autonomy of the subject is not therefore a self-isolating dispositive. The autonomous position does not isolate the subject from the surrounding world. It does not coincide with a conscious reflection that follows perception or action and makes judgments about it after the fact. It is not yet the inner dialogue that often accompanies our lives, described by Hanna Arendt as the basis of our moral capacity. It is rather the awareness that we can never only deal directly and immediately with things and circumstances of various natures without them effecting our inner form. Our interactions with things are at the same time always interactions with ourselves.

The reverberation of our sensations and activities presupposes an inner plurality. To enable this everyday refractive effect, consciousness must be understood as neither a monolith nor as a linear chain of acts and facts. On this point, the phenomenological description offers to us a representation of consciousness as a unified and interconnected stream that carries many different currents within it. The stream of consciousness harbours the inner effects and counter-effects of manifold drives, desires, expectations, and perceptions. However, this is not merely an empirical phenomenon. Rather, phenomenological analysis identifies the basic precondition of the intrasubjective self-resonance in the form of an inner temporality.

We live in time: 'When? How long? Since when?' are central questions of our communication as well as our self-dialogue. Such questions are not infrequently characterised by a forceful emotional tone and usually convey more than a need for simple information. We live and plan our lives in calendar days and hours. Timelines are the most requested, dreaded, and stressful documents we have to

deliver on every project. Yet, we do not only live in this 'objectified,' measured, conventional time. Rather, philosophical and existential reflection highlights various forms of experienced, subjective time that are just as vital to us.

Even our body displays its own temporality based on the recurring drives and needs for rest, nourishment, proximity, and contact. This dynamic characterises itself as an empirical temporality, which is, however, not merely material and physiological. Rather, each need is grounded in our consciousness, it triggers individual reactions and makes our singular personality echo differently in each case. Each need, fulfilled or unfulfilled, determines subjectivity in its own way, sedimenting itself in consciousness, bending us, creating habits, and thereby making a significant contribution to our self-resonance and self-constitution.

This deep web of experience consists in a felt temporality that constantly inheres to experience. Every experience presents itself in the form of the present, it occupies us and fills our present, thus that the scope of this present itself is thereby determined. The evidential present-ness of experience, however, cannot exist without an inner reference to the past. Every experience carries with it a horizon of the past that determines the 'whence' of experience. Husserl describes such a temporal echo as a "comet's tail"[7] that follows every experience. Every experience slowly sinks into the past. The experience does not disappear with the past, but rather reverberates like a voice that continues to have an effect, although it becomes fainter and fainter. Our connection with what is experienced necessarily consists at the same time in a letting go and a keeping, which transforms contents and forms. Such a transformation instigates the very possibility of experiencing something and integrating it into our conscious lives.

However, the effective reference to the past represents only one aspect of inner temporality. Every experience also carries within it a future horizon, thanks to which it can extend "forward," thereby anticipating and establishing a dialogue with future experiences. The possibilities of experience that are anticipated in this process are not mere fictions, they are not exhausted in fantasies. Rather, they play a necessary role in the orientation and further course of the experience itself. We would not take a step if it were not implicitly "expected" in our perception that the ground would hold, that the laws of gravity would not be overturned, etc. Hence, every experience carries expectations within it, wherein its meaning often consists even more of expectations than of actual present knowledge.

The outlined structures of expectation (future) and the slow sinking of experience (past) form two non-presences that surround and make possible every actual

---

[7] "Aber diese Jetztauffassung ist gleichsam der Kern zu einem Kometenschweif von Retentionen, auf die frühere Jetzpunkte der Bewegung bezogen" (Husserl 1966, p. 30).

present experience. They designate empty horizons that prepare the necessary resonance for every experiential situation. These temporal structures provide the resonance space for every experience and thus for a meaningful life. The human conscious being needs such a temporal virtual expansion to make sense of life and be able to experience the surrounding environment. In this sense, it is not true that our being temporal inevitably makes us contingent and precarious. Rather, lived temporality is simultaneously the condition of possibility of coherence and subsistence of our psychic life.

The vital inner temporality is characterised by its continuity and fluidity. Nothing remains isolated and unconnected in consciousness. Each sinking experience affects the one that follows and each current one anticipates further moments to come. Experience flows and connects. However, in doing so, it is also constantly transforming. Inner temporality denotes an elemental but essential creative force that bars our inner world from any rigidity. Subjective life is rooted in this primitive creativity and defines itself through it. Only thanks to the inner empty horizons of anticipation and retention can the process of ongoing transformation and reshaping that constitutes a meaningful experiential life take place. Only through this can we meet the challenges of the world without breaking down or resigning. Only through the inner space provided by anticipation and retention this will profoundly different experiences be integrated as stages of personal development and appear as cornerstones of our autonomous self-constitution.

With that said, this inner processuality is the exact opposite of the omnipresent schedules, work plans, timetables in which we divide and objectify time. It is in fact based on the creative potential of the empty inner horizons; it requires the possibility to expand in time, to anticipate future experiences and to draw on past ones. How can such an inner temporal resonance be lived out in times of precarious work, flexible life, and technological simultaneity?

5 Disruption of Time, Simultaneity, and Intergenerational Alienation

One might assume that the essential function of inner temporality just outlined would be stored and guarded in the inner life like an intimate treasure or individual resource, thereby grounding the stream of experience. But, as described in §2, our experiential life primarily represents a resonance space for everything we encounter. Environment and surroundings, intersubjective relationships and interactions, incidents and external caesurae form the inescapable fabric of our lives and influence its rhythm. Precarity can be interpreted in this framework as a

peculiar way of organising objective time that has profound effects on inner temporality.

The objective order of time is not merely parallel to subjective time. Rather, serious interferences between the two can be identified, which point to a stressful conflict. As an appropriation of people's vitality and time for the purposes of production and profit, any organisation of work implies profound effects on the lives of individuals and communities. In this respect, we should avoid the idealisation of past, overcome, or marginalised ways of working, which entails an undue trivialisation of earlier – but no less violent – forms of exploitation. The current flexibilisation of work, however, presents some peculiarities that stand in direct and conflict-laden contrast to the structure of inner temporality and to its function as a space of resonance.

(i) The first striking element of the current way of life is the fragmentation of work, i.e. the demand to radically change not only one's own tasks but the whole field of work again and again. Repeated changes in working conditions and areas as well as places of work are now part of the life of workers from all backgrounds as a more or less passively suffered fact. Life loses continuity and becomes a chain of moments closed in themselves that are unable to communicate with each other. The preceding experiences no longer sink into a sedimented horizon, but disappear because the entire environment has become alien and inappropriate to them in each case. By losing this continuity of context (working environment, place of residence, rhythm of life), past experiences also lose their core significance. They are no longer involved in the co-creation of current experience and therefore become alienated, disconnected, and meaningless elements. We are witnessing here a progressive break between past experiences and the processes of shaping present meaning. What has been experienced, learnt, and exercised in past contexts does not resonate when placed within the present context. The horizon of the past becomes alien and the flow of consciousness is interrupted. The inner horizon of the future, which is based on the structured and fundamental expectation of continuity, is equally threatened, especially by forms of work that structurally consist of individual projects. With Heidegger, we could say that such projects miss and even prevent the characteristic "*Entwurf*" (projection) crucial to authentic existence (Heidegger 1927, ch. 31). Imagination, self-imagination, and self-projection into the future are made impossible by working conditions that focus less on the actual usefulness of work and more on whether or not the task at hand has merely been completed. In the performance of isolated, finite tasks, the ultimate goal of the activity and its purpose fade away. The consequences of the action lie in a distant future and lose their significance in the face of the immediate urge of completing the action itself in a fixed time. This dynamic, however,

not only limits the perspective of the future, but renders it ineffective. Thus, both "forward" and "backward," the temporal resonance of the subject is blocked, prevented, frozen. Work organisation demands an exclusive and blind focus on the current present, which sacrifices the complexity of the inner resonance of the self in its past and future dimensions in favour of the manageability and flexibility of current market conditions.

As we noted above, the very vitality of the stream of consciousness depends on the possibility of extending itself into progressively empty, but always operative, horizons of past and future. In this sense, the difficulty of integrating past experiences and future perspectives into the present occupied by defined tasks has an immediate effect on the vitality of consciousness, its elasticity, and its possibility of constituting meaning in an integrated and effective way. The progressive corrosion of internal time seriously weakens agency, performative autonomy, and the possibility of making personal decisions.

(ii) A second aspect of negated resonance points to the characteristic simultaneity that current working conditions demand. Thanks to technological infrastructures that keep us connected constantly and everywhere, working time and private lifetime easily coincide. This requires everyone to be in different places, to interact with different environments, to play out different patterns of interaction at the same time. Instead of being experienced in the life of consciousness in the form of an ongoing deepening of experience as the constitutive result of inner temporality, i.e. as the sedimentation and the complication of the experiencing moments, imultaneity, is now lived out in the opposite forms of simplification, of non-commitment, of immediate coincidence. We shift from the constitutive simultaneity of lived temporality to the fixation of time moments in repeated tasks.

Also in this case, the consequences for the human capacity to synthetize experience and make sense of their own life are deep and difficult to foresee. Simultaneity is no longer the constant achievement of consciousness, but rather takes the form of a relational short-circuiting. The mediation operated by retention and protention, observed in the inner functioning of time, fades away and is replaced by the immediateness of isolated work-tasks.

In this context, no separation between work and non-work is allowed. What's more, this kind of immediate coincidence actually hides much deeper rifts in one's own existential environment and inner temporality. For the worker who is required to live out his private life in the company and whose private sphere is co-shaped by the company, the abysses that cross the stream of experience remain veiled. Its inner and social resonance as a subject is not merely weakened by this new situation and lifestyle. Indeed, it furthermore shifts into externality and is thus transformed into a form of social dependence.

On the one hand, recognition, as the crucial goal of communication and human interaction, loses its characteristic processuality and mediation, in both of which the subject is at play with its various inner components. On the other hand, recognition is played up as an immediate one-sided and narcissistic demand that hinders, rather than promotes, the process of self-knowledge and self-constituting.

(iii) A final aspect of our brief phenomenology of precarity goes back to the generative significance of temporality. Both psychology and phenomenology emphasise the central importance of the interaction between generations for the self-constitution of a subjectivity that is temporally predisposed. At this point, our observation shifts toward a more inclusive and comprehensive dimension. It no longer focuses on the individual, but instead includes intersubjective life in time. Our focus is no longer inner temporality, but the historical temporal dimension that includes different generations. Additionally, at this level, the precarisation of work – and life conditions affects the constitution of meaning and ultimately the efficacy of personal autonomy.

The relationship between earlier and later generations and one's own position, as well as the mechanisms of recognition and distancing that take place in this context, are not contingent factors in the self-constitution of an individual and human group. Rather, they are indispensable and inexhaustible processes that profoundly determine the life of every individual. In our relationship with the older generation, we experience contents and lifestyles of times that are past and yet continue to have an effect on us. In our imagined relationship with future generations, we find an echo of our current desires and expectations.

The precarisation of living conditions stops and distorts such processes by hindering the traditional succession of generational tasks and functions in various ways. The extension of the working period for the older generation and the slowing down of access to work for the newer ones creates a deceptive simultaneity that is not matched by equivalence in power, resources, and social influence.

At the two extremes of this process, the younger generation is forced to wait in order to achieve work autonomy and the older generation is forced to work longer than expected. Both experience a frustration that prevents mutual recognition. Generations drift apart, mutual needs become opaque, resentment grows and is rooted in mutual alienation. The inter-subjective dimension stretching across personal history breaks down, thereby giving rise not to physiological and individuated conflicts but to radical estrangement.

Apparently, we are all called upon to perform in the same way, regardless of age and condition. Apparently, corresponding opportunities for work and consumption are promised to all. Nevertheless, this only leads to the multiplication of expectations and demands without an effective intergenerational resonance

becoming possible.

As we have seen, for phenomenology, temporality is not an extrinsic dimension or a purely individual intuition. Rather, with its internal and intersubjective dimensions, it delineates the proper space of consciousness and provides the essential condition of possibility for the constitution of meaning and the self-constitution of subjectivity. The precarity of lifestyles has a radical influence primarily on the temporality of the subject. It is, in the full sense, an appropriation and transformation of the time of the subject, which is fragmented and redistributed according to a new and more flexible organisation of work. This, however, has important consequences not only for living conditions, but also for the very capacity of perception and self-perception of subjects, and therefore for their possibility of developing an autonomous vision of the world and of themselves.

6   Conclusion

The fragmentation of time horizons through (i) precarious working conditions, (ii) the abstract simultaneity of technology, and (iii) the distortion of intergenerational interactions constitute three moments of alienation that clearly reveal the impact of precarity on the existential structure of time. We witness the loss of present and future horizons due to constant uprooting and the pressure to consistently transform work and life contexts. We have described the transformation of the living simultaneity of the horizons of consciousness into the lifeless simultaneity of technology-mediated relationships. The permanent being connected takes the place of an intimate interconnectedness of one's experiences, expectations, desires. Life-experience and complex knowledge are substituted by pre-defined competences and skills. Finally, intergenerational conflicts take the form of a profound reciprocal alienation of generations forced to give up their personal and professional fulfilment in order to comply with the distorted times of the labour market.

Such a change is not merely theoretical. Rather, it affects the subjective ability to experience inner and social resonance, to reverberate both within oneself and with others, to multiply and thus to unfold one's creative potential. Subjective autonomy is thus so severely challenged that it becomes questionable as to whether an actual human, rational, and meaningful life can be essentially preserved and continued under such circumstances or whether we are not rather facing a radical, as well as threatening, de-subjectification[8] that carries with it a shattering inabil-

---

[8]   Recently, the de-subjectification process is reflected upon by the trans-humanistic understanding of modern anthropology. See (Hansell and Grassie 2011).

ity to answer for oneself and relate to oneself. The "fluidification" (Bauman 2000) of society has produced the serious side effect of a frozen consciousness that remains in mere identity with itself in order not to lose itself definitively. The theme of identity proliferates, invading not only the public space of politics, but also the private dimension of the struggling individual. The whole energy of existence is concentrated in a search for identity that locks the subject into a defensive, individualistic, and hostile position. Even when such a search seams successful, however, the rigid and anticipated identity that results does not really come into play as a resource. Instead, it is a serious obstacle to the necessary process of self-constitution and self-development of personality. In order to search for ourselves, we stop developing the relational potential that would otherwise enrich our personality.

Personal autonomy then becomes a defensive self-identification and abstract distinction rather than a condition for the possibility of further discoveries and designs. Autonomy, which as we have seen above is also linked to the possibility of unfolding internal temporality, becomes a rigid position, a stubborn self-assertion. Finally, precarity threatens to eat away at its own anthropological preconditions: Flexibility turns into rigidity, creativity into fixity of identity.

References

Altobrando, Andrea. 2013. *Esperienza e infinito: contributo per una fenomenologia dell'idea di infinito a partire da Husserl.* Trento: Verifiche.

Bauman, Zygmunt. 2000. *Liquid Modernity.* Cambridge: Cambridge University Press.

Bourdieu, Pierre. 1998. *Acts of Resistance: Against the New Myths of Our Time.* Cambridge: Polity Press.

Butler, Judith. 2009. "Performativity, Precarity and Sexual Politics." *AIBR. Revista de Antropología Iberoamericana* 4 (3).

Hansell, Gregory R., and William Grassie, eds. 2011. *H+/-: Transhumanism and its Critics.* Philadelphia: Metanexus Institute.

Heidegger, Martin. 1927. *Sein Und Zeit.* Halle: M. Niemeyer.

Husserl, Edmund. 1966. *Zur Phänomenologie des inneren Zeitbewusstseins (1893–1917)*. Edited by Rudolf Boehm. Vol. 10. Husserliana. Den Haag: Nijhoff.

———. 1973. *Zur Phänomenologie der Intersubjektivität: Texte aus dem Nachlass Erster Teil: 1905–1920*. Edited by Iso Kern. Vol. XIII. Husserliana. Den Haag: Martinus Nijhoff.

———. 1989. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: Second Book Studies in the Phenomenology of Constitution*. Translated by R. Rojcewicz and A. Schuwer. Dordrecht: Kluwer Academic Publishers.

———. 2006. *The Basic Problems of Phenomenology: From the Lectures, Winter Semester, 1910-1911*. Translated by Ingo Farin and James G. Hart. Dordrecht: Springer.

Lorey, Isabell. 2015. *State of insecurity: Government of the precarious*. Translated by Aileen Derieg. London: Verso.

Rizzolatti, Giacomo, and Corrado Sinigaglia. 2008. *Mirrors in the brain: How our minds share actions and emotions*. Translated by Frances Anderson. Oxford: Oxford University Press.

Weiss, Gail, Gayle Salamon, and Ann V Murphy. 2019. *50 Concepts for a Critical Phenomenology*. Evanston: Northwestern University Press.

Marx the Fichtean

Tom Rockmore

Peking University*

ABSTRACT | We ignore the history of philosophy at our peril. Engels, who typically conflates Marx and Marxism, points to the relation of Marxism to the tradition while also denying it. In his little book on Feuerbach, Engels incorrectly depicts Feuerbach as leading Marx away from Hegel, away from classical German philosophy and away from philosophy and towards materialism and science. This view suggests that Marx is at best negatively related to classical German philosophy, including Hegel. Yet Engels elsewhere correctly suggests that Marx belongs to the classical German philosophical tradition. In the preface to *Socialism, Utopian and Scientific*, Engels wrote: "We German socialists are proud that we trace our descent not only from Saint Simon, Fourier and Owen, but also from Kant, Fichte and Hegel" (Marx and Engels 2010b, p. 459). I will be focusing on Marx's relation to Fichte, who is rarely mentioned in the Marxist debate, but who, I will argue, was doubly crucial both for the formulation of Marx's position and for assessing his contribution. One result will be to indicate that Marx, in reacting against Hegel, did not, as is often suggested, 'leave' philosophy, but in fact made a crucial philosophical contribution.

KEYWORDS | Fichte; Marx; Engels; Marxism; Hegel

---

* *Correspondence:* Tom Rockmore – Avenue du Maréchal, Foch, 06000 Nice, France. Email: tom.rockmore@icloud.com

Tom Rockmore

1   Introduction

We ignore the history of philosophy at our peril. Engels, who typically conflates Marx and Marxism, points to the relation of Marxism to the tradition while also denying it. In his little book on Feuerbach, Engels depicts Feuerbach as leading Marx away from Hegel, away from classical German philosophy and away from philosophy and towards materialism and science. This view suggests that Marx is at best negatively related to classical German philosophy, including Hegel. Yet Engels elsewhere suggests that Marx belongs to the classical German philosophical tradition. In the preface to *Socialism, Utopian and Scientific*, Engels wrote: "We German socialists are proud that we trace our descent not only from Saint Simon, Fourier and Owen, but also from Kant, Fichte and Hegel" (Marx and Engels 2010b, p. 459).

I will be focusing on Marx's relation to Fichte, who is rarely mentioned in the Marxist debate, but who, I will argue, was doubly crucial both for the formulation of Marx's position and for assessing his contribution. One result will be to indicate that Marx, in reacting against Hegel, did not, as is often suggested, 'leave' philosophy, but in fact made a crucial philosophical contribution.

2   Feuerbach and the Marxist Reading of Marx

Engels depicts the relation of Marx to philosophy mainly through Hegel, whom he allegedly rejects, and Feuerbach, who supposedly makes this rejection possible. If it turns out that Fichte did in fact influence Marx, then it will be necessary to revise the Marxist view of his link to German Marxism.

The argument presented by Engels goes something like this: Marx's position arose in reaction against Hegelian idealism. Idealism of all kinds, hence Hegelian idealism, offers a distorted, therefore false, view, based on an inverted conception of the real world as viewed through the lens of bourgeois thought. Marx later freed himself from idealism mainly through Feuerbach, who enabled Marx to simply throw Hegel aside. Idealism and materialism are incompatible opposites. Idealism in all its forms is false, but at least one type of materialism is true. Feuerbach provides a materialist critique of Hegel, hence of idealism. Feuerbach, whose critique of idealism overcomes Hegel, is in turn later criticized by Marx. Marx follows Feuerbach's lead in 'leaving' idealist claptrap behind for materialism, whose dialectical version provides the only correct approach to contemporary society.

This complex claim can be usefully restated as a series of four discrete propositions:

1. Marx's position arose in reaction to Hegelian idealism.

2. Idealism and materialism are incompatible opposites, of which one is true and the other is false.

3. Marx followed Feuerbach's decisive critique of Hegelian idealism in giving up idealism for materialism.

4. In giving up idealism for materialism, Marx moved beyond philosophy.

These assertions are often regarded as true, but each is in fact false. By 'false,' I mean inaccurate, tendentious, or misleading. There is a widespread tendency to understand Marx in terms of his roots in Hegelian idealism. Now, Marx's relation to Hegel and Hegelianism should not be denied. He was obviously influenced by Hegel, whom he read as a teenager, whom he criticized in his early writings while still in his mid-twenties, and on whom he continued to rely for key categories, arguments and insights in later writings, including *Capital*.

Though himself a Young, or left-, Hegelian, Marx was critical of other Young Hegelians. It is initially plausible – but, on reflection, misleading – to understand Marx's position as arising solely or even mainly in reaction to Hegelian idealism, if that means it can somehow be adequately accounted for or understood simply in terms of its Hegelian roots. It is plausible that, as Engels reports, the origin of Marx's position lies in coming to grips with Hegel's *Elements of the Philosophy of Right*. But this led Marx well beyond the confines of his understanding of Hegel, towards other economic and political horizons and, within philosophy, towards thinkers he regarded as supplementing or even correcting Hegel.

The second proposition concerns the relation between idealism and materialism. Most observers regard materialism (or realism) and idealism as incompatible, and believe that a simultaneous commitment to both would be self-contradictory. The view that no version of idealism and materialism (or realism) can be combined within a single position is common to objections raised against idealism in different ways by its Marxist and analytic critics. Elsewhere, I have examined 'idealism' in detail in the context of an account of Kantian, hence German, idealism (Rockmore 2007). There are different types of idealism and materialism. It is doubtful that there is a single, shared doctrinal commitment for either idealism or materialism, whose subtypes appear to overlap in terms of family resemblances rather than a shared essence. In a famous paper, G.E. Moore influentially suggested that idealists of all stripes deny the existence of the external world (Moore 1903).

Yet this is a clear error. Moore does not point out any idealist guilty of this mistake and none has ever been identified. Further, the supposed incompatibil-

ity between idealism and materialism, though often asserted, is nowhere demonstrated. On a closer look, it appears that, if properly understood and under appropriate conditions, idealism and materialism are compatible.

The relation between these doctrines is long and complex. The philosophical term "idealist" seems to have been invented by Leibniz. In responding to Bayle, he objects to "those who, like Epicurus and Hobbes, believe that the soul is material" in adding that in his own position "whatever of good there is in the hypotheses of Epicurus and Plato, of the great materialists and the great idealists, is combined here" (Leibniz 1875). Leibniz's usage of the term implies idealism and materialism differ, but can be combined, in a single position. He suggests, as Fichte later appears to suggest, a simultaneous commitment to idealism and materialism (or realism) (Fichte 1982, pp. 3-28).

Marx is often regarded as following Feuerbach's lead in giving up idealism, which he supposedly vanquished, for materialism. Feuerbach, who was an opponent of Hegel, criticizes the latter in various texts from the perspective of the so-called *Principles of the Philosophy of the Future* (Feuerbach 1986). But it is not the case, as the name of his position clearly suggests, that he vanquishes idealism for extra-philosophical materialism. It is further exaggerated to claim that Feuerbach, who is better known for his contribution to religion than to philosophy, "overcomes" Hegel, a true philosophical giant. At most, he can be read as pointing beyond Hegel in other directions.

Fourth, even if Marx were a materialist, it would not follow that he had moved beyond philosophy. There are numerous philosophical materialists, beginning with Democritus, Leucippus and Epicurus, the materialists of antiquity, and continuing up to the present. Even if Marx supported Feuerbach against Hegel, it would not follow that he moved beyond philosophy.

Engels, who did not graduate high school, was an autodidact, with no more than a cursory philosophical background. In inventing Marxism, he was influenced by a short period of study with Schelling in 1841. Another student in the same class was Kierkegaard. In the Munich lectures, held shortly after Hegel's death, Schelling sharply criticized Hegel's position as negative, in advancing his own supposedly positive philosophy, which ultimately became his theory of revelation. Engels and Kierkegaard both later formulated different versions of Schelling's complaint that Hegel was unable to grasp concrete existence. In Marxism, this became the difference between theory and practice, or *Praxis*.

Engels developed Schelling's distinction between negative and positive philosophy in substituting the familiar distinction between materialism and idealism. According to Hegel, idealism is concrete and materialism is abstract. But according to Engels, idealism is abstract and cannot grasp the real social context. It is

grasped only by materialism, which, unlike idealism, is concrete. Marxism, which is only distantly related to Marx's own position, is a philosophical amalgam thrown together by borrowing from different sources, constructing a crude but highly misleading view of the Western philosophical tradition on the basis of a simplistic account of German idealism.

The most influential statement of this theory is found in Engels's little book, *Ludwig Feuerbach and the End of Classical German Philosophy* (Engels 1941). Here and elsewhere, Marxism promotes a misleading, simplistic three-fold claim regarding the relation of Hegel to prior philosophy, the relation of philosophy to philosophical problems, and the relation of Marx to philosophy and philosophical problems. According to this view, philosophy came to a peak and to an end in Hegel.

Kant made a version of this claim. He suggests the critical philosophy forever solved the problems of philosophy and, hence, could not be revised. Hegel, who never made any version of this claim, contradicted Kant by suggesting that all positions, including his own, belong to the history of philosophy. No philosophical theory, position, insight or argument can suffice to bring the philosophical tradition to an end. Earlier theories are either ignored or refuted by later theories, which continually take the discussion beyond any given point. Engels generalized Schelling's view of Hegel's supposedly negative philosophy to philosophy in general.

According to Engels, philosophy is inadequate to solve, resolve or otherwise dispose of its problems, concerns, or difficulties. His basic insight that reason must be adapted to – or, in another formulation, made congruent with – its object goes back in the early Greek tradition at least to Parmenides. In distinguishing between the way of error, which is straight, and the way of truth, which is circular, he indicates that the criterion of knowledge is the identity of thought and being. For the cognitive instrument must be adapted to its cognitive object, which it seeks to know. This idea is later restated many times: for instance, in Kant's so-called Copernican revolution, which centers around the claim that one can only know what one in some sense constructs. Engels suggests that philosophy is inadequate to come to grips with its problems, which are, however, real. These problems are resolved by Marxism only, which is situated beyond philosophy.

Engels's unsupported blanket-claim rapidly became an article of Marxist faith, with roughly the same status as religious beliefs. Such beliefs need neither argument nor demonstration in order to be accepted, and cannot be refuted through ordinary forms of argument. Engels, who did not demonstrate any of his claims, made no pretense of arguing for his interpretations, which remain as mere assertions. He did not, for instance, show that philosophy came to an end in Hegel,

that it cannot carry out its self-assigned tasks, that idealism goes from thought to being rather than going from being to thought, nor that Marxism can provide an extra-philosophical solution to philosophical problems. It seems doubtful that philosophical questions can be answered from a position "outside philosophy."

3 Lukács and Hegelian Marxism

In Engels's reading of Marx, as already pointed out, Feuerbach enabled the latter to 'leave' philosophy in favor of science, or at least an extra-philosophical, scientific perspective situated outside it in order to solve, resolve or overcome its difficulties, problems and concerns. This suggestion is doubly problematic. On the one hand, there is no trace of this approach, which Engels illustrates, in Marx's writings. On the other hand, this is tantamount to suggesting that, as Althusser later insisted, Marx turned away from an approach based on the actions of one or more individuals, which is supposedly illustrated in German idealism, to rely on modern science. On the contrary, when we inspect Marx's texts, we see that his position relies on rethinking the conception of the subject that does not leave behind, but rather depends on, German idealism, especially Fichte.

Engels's simplistic, unargued account of the relation of Marx to Hegel and German idealism is literally transformed in Lukács's complex, closely argued account. Simultaneously with Karl Korsch (Korsch 2012), Lukács invented Hegelian Marxism. Unlike Engels and most students of Marx, including Korsch, Lukács had a thorough grasp of classical German philosophy. He did, for instance, early work in Kantian aesthetics before turning to Marxism. His particular form of Hegelian Marxism has two characteristics. First – like Korsch, and like other Hegelian Marxists – he resisted a simplistic, binary reading of the relation between Marx to Hegel, in formulating a richer, multi-dimensional account in the broader context of German idealism, especially Hegel. Second – unlike Korsch as well as other Hegelian Marxists – Lukács, in emphasizing Hegel, also pointed to the importance of Fichte for Marx's position. Lukács's most significant account of Hegelian Marxism occurred in *History and Class Consciousness* (Lukács 1971), which appeared in 1923, the same year as Korsch's important study, *Marxism and Philosophy*. Lukács, who employed a Marxist reading of Marx with Kantian and neo-Kantian elements, comprehends Marx's theory as a form of commodity-analysis. According to Lukács, only Marxist political economy is capable of comprehending the economic structure of advanced industrial society.

His Kantian argument for this claim consists of two points. First, non-Marxist political economy cannot know its object, that is, the real structure of the so-

cial context. So-called bourgeois political economy, which is limited to grasping false appearance, is implicitly irrational and therefore unable to grasp the mind-independent real. This aspect of Lukács's argument depends on his reading of the little known but important German neo-Kantian Lask (see, for discussion, (Lask 1902)).

Second, Marxian political economy grasps true appearance through the Marxian theory of commodity-analysis, and hence is implicitly rational. It is the only approach that can lead to knowledge of social reality. Marx's theory of commodity-analysis, as Lukács asserted in a dazzling example of Marxist faith, can resolve any and all problems (Lukács 1971, p. 83).

Lukács's attitude towards Engels was both positive and negative: positive in that he supplied arguments to buttress the latter's simplistic assertions in restating Marxism on a philosophical basis, but negative in that he sharply criticized Engels's philosophical inadequacies, such as the latter's simplistic treatment of Kant's key conception of the thing-in-itself (Lukács 1971, pp. 131–133). Engels simply claimed that philosophy reaches its peak and end in Hegel without being able to resolve its problems. Lukács, on the contrary, argued for this claim in identifying a specific flaw in classical German philosophy on the level of the subject, a conception which is allegedly corrected by Marx.

Lukács's argument in favor of Marxism extends Kant's analysis of the thing-in-itself throughout classical German philosophy, which is by definition unable to know its object. According to Lukács, Kant advanced an inadequate conception of the subject, whose difficulty culminated in Hegel's appeal to a mythological concept of the absolute, expressing a manifest inability to understand the real historical subject, the proletarian class, or identical subject-object.

Lukács adduced three reasons, all well-known in the Hegel literature, for Hegel's supposed failure to provide an adequate conception of the subject. First, the relation of reason to history is merely contingent, since reason is not actually imminent to history. This is a version of the familiar Marxian view that Hegel begins from an abstract, theoretical perspective, which never grasps the social and historical context. Second, Hegel supposes that history has an end, which lies in the Prussian state. In this context, Lukács restated the frequent claim that Hegel later turned away from the revolutionary ideals of his youth in assuming a reactionary political stance. Third, he complained that in the *Encyclopedia of the Philosophical Sciences*, in an abstract, contemplative discussion, Hegel separated genesis from history in a merely logical analysis of the transition from logic through nature to spirit. The resultant conception of the absolute only seems to make history. This is a form of the well-known assertion – which Lukács never abandoned, and which formed the basis of his critique of Hegel in *On the Ontology of Social Being*

(*Zur Ontologie des gesellschaftlichen Seins*) – that Hegel's philosophy is a panlogism (Lukács 1971, 1984). According to Lukács, the interest of the German idealist tradition consists of pointing through its method towards the way beyond these limits. The correct path lies in a return to the early Marx's discovery of the true historical subject. Through the dialectical method as the true historical method, we identify the real "we" of the historical process in the proletariat, as the identical subject/object of history.

Lukács writes: "The continuation of that course which at least in method started to point the way beyond these limits, namely the dialectical method as the true historical method, was reserved for the class which was able to discover within itself on the basis of its life-experience the identical subject-object, the subject of 'action' the 'we' of the genesis: namely the proletariat" (Lukács 1971, pp. 148–149).

The claim for the proletarian standpoint as the solution to the problem emerging from the thing-in-itself is, in fact, a transparent restatement of the Young-Hegelian view that philosophy comes to an end in Hegel's thought. In other words, the theory of the proletariat discovered by Marx and continued by Marxism provides the solution for the crucial problem left unsolved by classical German philosophy. In the final analysis, philosophy does not end in the Hegelian synthesis, which, since it depends on an incorrect view of the subject, simply fails. Rather, it is completed and, hence, comes to an end in the Marxian transformation of absolute idealism, which is seamlessly prolonged in Marxism.

We can summarize as follows: according to Lukács, Marx's key move, which is unrelated to exposing the anatomy of modern industrial society, lies in rethinking the subject in his early writings. In this context, Lukács turned his attention towards Fichte. Lukács was critical of the Fichtean concept of activity, whose importance lies in a prototypical solution of the relation of theory and praxis, subjectivity and objectivity. He followed others in maintaining that Fichte failed to understand the true nature of human activity, which he incorrectly assimilated to mental activity alone.

Fichte's conception of the subject as wholly active and never passive is significant as a contribution to the Kantian problem. According to Kant, there has never been a theory of metaphysics, which in the critical philosophy presupposes an inference from the a priori to the a posteriori. Kant makes practice dependent on theory, which includes and hence resolves any and all practical concerns. Fichte responds to Kant in inverting the relation between theory and practice. According to Fichte, theory, which is not disconnected from practice, is not irrelevant, but relevant to practical concerns from which it arises and to which it returns.

It is widely known that in the *Transcendental Deduction* (§16) Kant deduces a transcendental subject unrelated to human being. Fichte goes beyond Kant in

correctly locating the unity of subject and object in *activity*. Lukács follows Fichte suggests in rethinking the unity of subject and object as intrinsically active or activity (Lukács 1971, p. 123). According to Lukács, Fichte showed that the given can be understood as the product of the identical subject/object, which derives from this unity. The importance of Fichte's view for Lukács becomes clear in his argument that the unity of subject and object, which Fichte allegedly located in mental activity, is, in fact, brought about through the activity of the proletariat.

4   Fichte and the Marxian Conception of the Subject

Western Marxism is with some exceptions a series of forms of Hegelian Marxism invented independently by Lukács and Korsch in pioneer formulations of Hegelian Marxism in 1923. Lukács's reading of Marx in relation to German idealism is ground-breaking but remains incomplete. As a co-founder of Hegelian Marxism, he provided a richer, better informed, more nuanced account of the relation of Marx to Hegel and classical German idealism than any other observer has before or since.

Two points are relevant here. Though his intention was to show that Marx went beyond German idealism in answering Hegel, Lukács correctly points out Fichte's significance for Marx's conception of the subject. Lukács thereby undercuts any version of the familiar Marxist claim that Marx left philosophy behind. At the same time, he undercuts his own effort to depict Marx in replying to Engels's suggestion that we must understand Marx through Hegel in pointing beyond him to Fichte.

In the wake of Descartes, the problem of the subject recurs throughout modern philosophy, including German idealism. In reacting against Hegel, Lukács suggests that the Hegelian synthesis is finally completed in the Marxian transformation of absolute idealism, a transformation that is seamlessly prolonged in Marxism.

We can summarize the argument as follows: according to Lukács, Marx's key move lay in rethinking the subject, or more precisely in discovering the real historical subject in replacing the absolute by the proletariat in his early writings. In this context, Lukács turned his attention towards Fichte. Lukács was critical of the Fichtean concept of activity, whose importance lies in a prototypical solution of the relation of theory and praxis. He followed others in maintaining that Fichte failed to understand the true nature of human activity, which he assimilated to mental activity alone.

Kant's problem lies in bringing together objectivity and subjectivity, theory and practice in, as Lukács says, solving the enigma of the thing in itself, in other words in finally grasping modern bourgeois society. This problem occurs on different levels, including theory, practice, and the unity of theory and practice. On

the theoretical level it is the problem of how to know modern industrial society that according to Lukács cannot be grasped from the bourgeois perspective and can only be grasped from the perspective of the proletariat. On the practical level it concerns the relationship of theory and practice that Kant resolves in the moral writings in subordinating practice to theory. The difficulty lies in bringing together theory and practice in a unity that as Lukács points out lies in activity. According to Lukács, in this respect Fichte went beyond Kant and "put the practical, action and activity in the center of his unifying philosophical system" (Lukács 1971, p. 123).

Yet, since Lukács is committed to the superiority of Marxism, his view of Fichte is equivocal. He sees but fails to understand the importance of the Fichtean view of the subject as intrinsically active since he is committed to the view that, as he writes, "*only the practical class consciousness of the proletariat* possesses this ability to transform things" (Lukács 1971, p. 205).

In fact, the situation differs from Lukács's depiction of it. Marx does not merely dismiss but rather relies on Fichte's conception of the subject as wholly active and never passive as an indispensable clue to the solution of the real historical subject. According to Lukács, who follows Lask, Fichte shows that the given can be understood as the product of the identical subject/object, which derives from this unity. The importance of Fichte's view for Lukács's Marxist reading of the relation of Marx to classical German philosophy becomes clear in his argument that the unity of subject and object, which Fichte allegedly located in mental activity, is in fact brought about through the activity of the proletariat.

5   Fichte and the Marxian Conception of the Subject

After these remarks on Lukács's view of Fichte we turn immediately to the relationship between Fichte and Marx. In reacting against Hegel, the Young Hegelians, including Feuerbach, turned to Fichte in order to formulate an adequate account of subjectivity. Kant considers the question of what "man" is to be the single most important theme. He worked out his view of the subject in the *Critique of Pure Reason* in the transcendental deduction, in isolating the conditions of knowledge in general from psychological factors. Though he accounted types of experience through types of activity, he was unable to formulate a unified theory of the subject. In Kant's wake, Fichte for the first time in classical German philosophy formulated a unified theory of the subject based on its activity. After Hegel, a number of Young Hegelians, including Feuerbach and Marx, if Marx was in fact a young Hegelian, turned to Fichte for a model of subjectivity that they then proceeded to develop (Cornu 1955–1970).

In the very early "*Critique of Hegel's Philosophy of Right*," Marx argues that Hegel substitutes a mere concept or idea for the real historical subject. Hegel's supposed failure to identify the real subject is an important factor in Marx's turn in the *Paris Manuscripts* from Hegel to Fichte to overcome a perceived deficit in the Hegelian theory of the modern state. The family and civil society in reality produce the state through the unfolding of the concrete social context. Yet, Marx contends, Hegel incorrectly sees them as produced by the idea (Marx and Engels 2010a, pp. 8–9) or, in Hegelian language, by a self-realizing concept. Marx sums up his methodological criticism by accusing Hegel of a Feuerbachian inversion of subject and predicate: "The fact which is taken as a point of departure is not conceived as such, but as a mystical result" (Marx and Engels 2010a, p. 9). In other words, Hegel conflates causes and effects in substituting effects for causes and conversely.

The young Marx, who was familiar with Fichte's position, later maintained an interest in the latter's views. This interest in Fichte is clear in his early writings, especially in the *Paris Manuscripts* of 1844. In the third *Manuscript*, in the section known as the '*Critique of Hegel's Dialectic and General Philosophy*,' Marx objects to the conceptions of the subject in Fichte and Hegel. It is, he maintained, as much a mistake to consider human being through self-consciousness as to reduce the object of consciousness to a purely mental creation. In an important but little noticed passage, Marx utilizes Fichtean terminology against Fichte, in writing:

> When real corporal man … posits [setzt], the positing [das Setzen] is not the subject of this act … An objective being acts objectively … It creates and establishes [setzt] only objects … In the act of establishing it does not descend from its 'pure activity' to the creation of objects [In dem Akt des Setzens fällt es also nicht aus seiner 'reinen Tätigkeit' in ein Schaffen des Gegenstandes]; its objective product simply confirms its objective activity, with its activity as an objective, natural being.[1]

---

[1] (Marx and Engels 2010a, p. 336). See also (Marx and Engels 1982, p. 295): 'Wenn der wirkliche, leibliche, auf der festen wohlgerundeten Erde stehende, alle Naturkräfte aus- und einatmende Mensch seine wirklichen, gegenständlichen Wesenskräfte durch seine Entäußerung als fremde Gegenstände setzt, so ist nicht das Setzen Subjekt; [A*] es ist die Subjektivität gegenständlicher Wesenskräfte, deren Aktion daher auch eine gegenständliche sein muß. Das gegenständliche Wesen wirkt gegenständlich, und es wurde nicht gegenständlich wirken, wenn nicht das Gegenständliche in seinen Wesensbestimmung läge. Es schafft, setzt nun Gegenstände, weil es durch Gegenstände gesetzt ist, weil es von Haus aus Natur ist. In dem Akt des Setzens fällt es also nicht aus seiner "reinen Tätigkeit" in ein Schaffen des Gegenstandes, sondern sein gegenständliches Produkt bestätigt nur seine gegenständliche Tätigkeit, seine Tätigkeit als die Tätigkeit eines gegenständlichen naturlichen Wesens.'

Here, we see Marx insisting on the objectivity of the external world, in opposition to Fichte, who was widely but incorrectly understood as believing that reality is wholly a product of thought. Marx further insists that if human individuals are not solely created through mental activity, they also cannot be understood through their mental capacities. It is remarkable how far the view that Marx here insists on – presumably for the most part against Hegel, and perhaps against Fichte as well – resembles Fichte's own conception of the human subject. In order to bring out this point, it is useful to quote the relevant passage at some length.

"Man," Marx writes,

> is directly a natural being. As a natural being, and as a living natural being he is, on the one hand, endowed with natural powers and faculties, which exist in him as tendencies and abilities, as drives. On the other hand, as a natural, embodied, sentient, objective being, his is a suffering, conditioned and limited being, like animals and plants. The objects of his drives exist outside himself as objective independent of him, yet they are objects of his needs, essential objects, which are indispensable to the exercise and confirmation of his faculties. The fact that man is an embodied, living, real, sentient objective being with natural powers, means that he has real, sensuous objects as the objects of his being, or that he can only express his being in real sensuous objects … Man as an objective sentient being is a suffering being, and since he feels his suffering, a passionate being. Passion is man's faculties striving [strebende] to attain their object.[2]

---

[2] (Marx and Engels 2010a p. 336-337); see also (Marx and Engels 1982, p. 408): 'Der Mensch ist unmittelbar Naturwesen. Als Naturwesen und als lebendiges Naturwesen ist er teils mit naturlichen Kräften, mit Lebenskräften ausgerustet, ein tätiges Naturwesen; diese Kräfte existieren in ihm als Anlagen und Fähigkeiten, als Triebe; teils ist er als naturliches, leibliches, sinnliches, gegenständliches Wesen ein leidendes, bedingtes und beschränktes Wesen, wie es auch das Tier und die Pflanze ist, d.h. die Gegenstände seiner Triebe existieren außer ihm, als von ihm unabhängige Gegenstände; aber diese Gegenstände sind Gegenstände seines Bedurfnisses, zur Betätigung und Bestätigung seiner Wesenskräfte unentbehrliche, wesentliche Gegenstände. Daß der Mensch ein leibliches, naturkräftiges, lebendiges, wirkliches, sinnliches, gegenständliches Wesen ist, heißt, daß er wirkliche, sinnliche Gegenstände zum Gegenstand seines Wesens, seinen Lebensäußerung hat oder daß er nun an wirklichen, sinnlichen Gegenständen sein Leben äußern kann. Gegenständlich, naturlich, sinnlich sein und sowohl Gegenstand, Natur, Sinn außer sich haben oder selbst Gegenstand, Natur, Sinn fur ein drittes sein ist identisch.> Der Hunger ist ein naturliches Bedurfnis; er bedarf also einer Natur außer sich, eines Gegenstandes außer sich, um sich zu befriedigen, um sich zu stillen. Der Hunger ist das gestandne Bedurfnis meines Leibes nach einem außer ihm seienden, zu seinen Integrierung und Wesensäußerung unentbehrlichen Gegenstande. Die Sonne ist der Gegenstand der Pflanze, ein ihr unentbehrlicher, ihr Leben bestätigender Gegenstand, wie die Pflanze Gegenstand der Sonne ist, als Äußerung von der lebenser-

To the best of my knowledge, no other single passage anywhere in Marx's voluminous writings offers a more detailed statement of his understanding of the human individual. This passage is, furthermore, fascinating for the remarkable resemblance between Marx's comprehension of finite human being and Fichte's view. Man is described, in Fichtean language, as 'natural,' possessed of 'drives,' as 'suffering' because limited, and as 'passionate' due to his awareness of his limitations: this reflects Marx's awareness of Fichte's theory as well as his specific conception of finite human being. Though Marx's overall position differed from Fichte's, he clearly accepts the main lines of Fichte's conception of the human individual as a natural being, hence obliged to meet his needs outside himself, limited by and only able to realize himself in relation to others through transforming the surrounding social context.

6   Marxian Man and Marxian Economics

Above it was pointed out that Marx offers a solution to the modern form of the ancient problem of human flourishing in its modern reformulation by Rousseau and later thinkers including Kant, Hegel and others. Marx suggests that in and through its activity human being meets two kinds of needs. They include basic reproductive needs, such as the proverbial food, clothing and shelter, which must be met in order for workers to continue to work in meeting their basic needs and capitalism to continue to function through realizing and appropriating surplus value. They include as well human needs, or the need to realize one's specific capacities as a finite human being, or the need to surpass mere reproductive needs in taking one's place as a fully individual member of society.

   The different kinds of human needs are met in practice through different forms of human activity. In skeletal form Marx's position includes a theory of the main forms of activity through which finite human beings meet their basic or species needs, an account of surplus value is created and appropriated by the owners of the means of production, as well as the revolutionary activity through which private property is abolished in the transition from capitalism to communism, and finally the activity through human beings finally become fully human individuals all depend on his reworking of the central Fichtean insight into human being as basically active and never passive.

   If this is correct, then it follows that the Fichtean view of human being as active is not, as Lukács suggests, a bourgeois, hence a non-proletarian and incorrect supplement to Marx's effort to rethink the subject in Hegel's wake. It is rather key

---

weckenden Kraft der Sonne, von der gegenständlichen Wesenskraft der Sonne.'

to Marx's effort, in relying on Fichtean insights to construct a replacement for what he regarded as Hegel's supposedly inadequate conception of the human subject. Marx's effort to respond to Hegel, which Marx rethinks on a Fichtean basis, is, since the conception of the subject is central to all the classical German thinkers, central to Marx's theory as well. For at the end of the day, if we leave aside the Marxist rhetoric to concentrate on the practical problems of modern industrial society, we see that Marx's conception of the human being as active is key to the Marxian conception of the solution to the ancient theme of human flourishing in modern industrial society.

7   Conclusion

I began by a concern about the relation of Marx to Fichte. The answer depends on what it means to be a Fichtean. This question was already controversial in Fichte's time. In the period between Kant and Hegel Fichte was enormously influential. He was still very influential when Marx was a philosophical graduate student. The young Schelling and the young Hegel were Fichteans for a time, though Fichte, who thought he was misunderstood, rapidly rejected Schelling as a disciple and Hegel just as rapidly moved beyond this early phase. If to be a Fichtean means to accept the main lines of Fichte's position, then neither Marx nor, arguably, anyone else was ever a Fichtean. Even Fichte was, arguably, never a Fichtean, since he continually altered his position, a position that he was never able to state satisfactorily, in some 16 versions of the *Wissenschaftslehre.* If, on the contrary, to be a Fichtean means to accept one or more central Fichtean ideas, then it seems clear that in an important sense, Marx is a Fichtean, above all with respect to the conception of finite human being as essentially active. At the dawn of modern philosophy, Descartes invented two views of the human subject, the widely-known, 'official' spectator theory, and the little-known, but perhaps more interesting, so-called actor theory, implicit in the famous remark in the *Discourse*, but never worked out, about 'trying to be a spectator rather than an actor in all the comedies the world displays' (Descartes 1970, p. 99). The so-called spectator theory of subjectivity has long been popular. But in general, the more interesting views of the subject in the wake of Descartes are different forms of the largely undefined actor view, through which various thinkers strived to understand knowledge, morality and the social surroundings through the prism of the activity of finite human beings.

Marx belongs to this ongoing effort, arising in the first instance in the reaction to the wholly theoretical subject in Kant and, at least from his perspective, the supposed lack of a historical subject in Hegel, a lack that is initially overcome

through the Fichtean conception of the subject as always active and never passive, but whose activity is constrained by its self-constructed social surroundings. Marx is certainly not Fichte, though in some respects the resemblance runs very deep. In Fichte's wake, Marx participated in a Fichtean effort to rethink the subject as defined by work – or, indeed, labor [*Arbeit*] – in modern industrial society. He understood capitalism, communism and, if there is a distinction, socialism through the self-production of finite human beings – within capitalism, in the form of work through which one meets one's basic reproductive or subsistence needs; and in a future form of society, perhaps unrealistically, situated beyond the limits of human needs anywhere one develops one's human potentials through what we can call free human activity.

Marx's entire position turns on working out an understanding and account of the real conditions of human freedom in modern industrial society on the basis of a theory of human activity. This approach to human beings through human activity was formulated by Fichte in reaction against Kant and then appropriated and transformed by Marx in reaction against Hegel.

Marx's position, which cannot be reduced to Fichte's, obviously resembles and depends on it as it works out an approach to finite human being and all the many forms of society through a conception of human activity. This approach goes back in the Western tradition at least as far as Aristotle, who advanced a theory of life as activity in the *Nicomachean Ethics*. But its proximal version, which influenced the Young Hegelians in the mid-nineteenth century as they rebelled against Hegel, was in Fichte's position. Fichte is, in this respect, the origin of Marx's conception of human being, which is the centerpiece of his entire position. I conclude that, in this sense, and perhaps others as well, Marx is, indeed, a Fichtean.

References

Cornu, Auguste. 1955–1970. *Karl Marx et Friedrich Engels*. Paris: Presses universitaires de France.

Descartes, René. 1970. *The Philosophical Works of Descartes*. Translated by E. Haldane and G.R.T. Ross. Vol. 1. New York: Cambridge University Press.

Engels, Friedrich. 1941. *Ludwig Feuerbach and the End of Classical German Philosophy (1886)*. Edited by C.P. Dutt. New York: International Publishers.

Feuerbach, Ludwig. 1986. *Principles of the Philosophy of the Future.* Indianapolis: Hackett.

Fichte, Johann Gottlieb. 1982. *The Science of Knowledge: With the First and Second Introductions.* Edited by P. Heath and J. Lachs. Cambridge: Cambridge University Press.

Korsch, Karl. 2012. *Marxim and Philosophy.* Translated by Fred Halliday. London: Verso.

Lask, Emil. 1902. *Fichtes Idealismus und die Geschichte.* Berlin: Imberg & Lefson.

Leibniz, Gottfried Wilhelm. 1875. *Die philosophischen Schriften: Leibniz 1875–1900.* Edited by C.I. Gerhardt. Vol. 4. Berlin: Weidmann.

Lukács, Georg. 1971. *History and Class Consciousness: Studies in Marxist Dialectics.* Translated by Rodney Livingstone. Cambridge: MIT Press.

———. 1971, 1984. *Georg Lukács, Zur Ontologie des gesellschaftlichen Seins.* Edited by Frank Benseler. Vol. 14. Neuwied: Luchterhand Verlag.

Marx, Karl, and Friedrich Engels. 1982. *Marx Engels Gesamtausgabe.* Vol. I.2. Berlin: Dietz Verlag.

———. 2010a. *Marx Engels Collected Works.* Vol. III. London: Lawrence & Wishart.

———. 2010b. *Marx Engels Collected Works.* Vol. XXIV. London: Lawrence & Wishart.

Moore, George Edward. 1903. "The Refutation of Idealism." Reprinted in G. W. Moore, *Philosophical Studies*, London: Routledge& Kegan Paul, 1958.

Rockmore, Tom. 2007. *Kant and Idealism.* New Haven: Yale University Press.

The Necessary Uniformity of Nature

Michael Slote

University of Miami*

ABSTRACT | In this essay, I shall defend the uniformity of nature in a new way, and that defense, in turn, will require us first to consider some a priori/conceptual ideas about what is psychologically impossible. Those ideas depend on an insight one can find in G. E. M. (Elizabeth) Anscombe's famous book *Intention.* But I will attempt to generalize that insight and show how it allows us to justify a wide range of conclusions about what is, on conceptual grounds, impossible for the mind or xin. Then I will argue that there are parallel conceptual impossibilities in nature outside the mind or xin, and that discussion will ultimately lead to the conclusion that nature both inside and outside the mind has to be uniform.

---

  * *Correspondence:* Michael Slote – University of Miami, Coral Gables, Florida 33124, USA. Email: mslote@miami.edu

Michael Slote

1 Psychological Impossibilities

I believe that the contents of the mind, and especially beliefs and desires, are necessarily more restricted than previous philosophical thinking East or West has ever suggested. There are things it is metaphysically impossible to believe or desire, and Elizabeth (G. E. M.) Anscombe has pointed out one very plausible and striking example of such a desire that no one could have. If we go with that example, we can easily mention others, and our discussion will then spill over into ideas about how certain sorts of beliefs (and inferences) are also metaphysically impossible. Anscombe argues that it is impossible in the strictest sense for anyone to simply want a saucer of mud, to want it for no reason beyond or ulterior to that simple desire. And I find this idea pretty compelling. Yes, you could want a saucer of mud because the saucer shaped mud in a good way in preparation for a facial, but that is not simply wanting a saucer of mud. Yes, you or a robot could be built with a mechanism and software that caused you to choose or pick saucers of mud whenever you were asked to choose decoration for your home or office, but this wouldn't amount to desiring or wanting a saucer of mud unless obtaining it made one feel pleased or delighted; and it is dubious that such an intentionally-loaded and vivid state of mind/xin could mechanically fall out of a software program or, for that matter, out of pure instinctual reflexes or automatisms. So I am persuaded by Anscombe's example.[1]

However, many philosophers have at least implicitly rejected Anscombe's example. R. M. Hare's prescriptivism and many versions of emotivism seem committed to the idea that emotion and reason/cognition are separate, so that any fact or reality could be the object of any emotion. Such views treat emotion and desire as in some sense ultimately arbitrary: anything can in principle be desired or hated, and evaluative judgments simply express desire or emotional/conative attitudes toward various entities without being constrained by any other factors. But Anscombe's example and others like it tell a different story.

Here is another example inspired by some work of Philippa Foot. Could anyone have a basic desire, could anyone simply want, to (always) turn northeast after turning northwest? Again, I think this makes no sense, and in fact Foot used her original version of this example to argue against Hare's prescriptivist view that our ultimate values (or desires) can be *absolutely arbitrary*.[2]

However, it is not just desires and emotional attitudes that are metaphysically limited in this way. Similar points apply to beliefs, though the examples I want to

---

[1] Anscombe's example comes from her book (Anscombe 1957, p. 70 and passim).

[2] The example, which I have somewhat modified for present purposes, comes from Foot's "Moral Beliefs," reprinted in her (Foot 1978).

rely on are different from any I know of that have been mentioned in the literature surrounding belief. Let me give you one plausible instance taken from the literature relating to Nelson Goodman's New Riddle of Induction.[3] Goodman defines a predicate "grue" as applying to green things examined before now and to blue things not so examined (examined only after now or never examined at all), and since all emeralds examined up till now have been both green and grue, the question Goodman raised was whether or why we have reason to believe other emeralds to be green rather than grue and therefore blue. No one who has previously discussed this issue (including my earlier self) has ever disputed the possibility that some creatures other than ourselves might find it natural to generalize with "grue." We have just wanted to know why it is reasonable at least for us to reject such generalizations.

Recently, though, I started to think more generally about the implications of what Anscombe had said about desiring a saucer of mud, and in particular I started looking at the New Riddle of Induction in the light of Anscombe's example. This made me reconsider my assumption, shared with everyone else who had discussed the matter, that people who have seen only green emeralds could inferentially believe that the next emerald to be viewed would be grue and blue, rather than green. It began to occur to me that such an inference is really not intelligible to us and that there is reason, based on the analogy with Anscombe's example, to hold that such an inference simply couldn't occur – if you will, that such an inferential belief is metaphysically impossible. After I had arrived at this conclusion, my then research assistant David DiDomenico put me in touch with a passage in which David Lewis analogized between the desire for a saucer of mud and the inferential belief that emeralds not yet examined are, or are likely to be, grue/blue.[4] Lewis very circumspectly stopped short of pushing the analogy fully home and never quite claimed that inferring with "grue" is just as impossible as simply wanting a saucer of mud. But that is precisely what I want to do here and have done at greater length elsewhere. (See the second appendix of my book (Slote 2018), with side-by-side English language and Chinese language texts.) Thus it is not just desires and attitudes, but also beliefs and inferences, that have limits as to their metaphysical possibilities.

---

[3] See (Goodman 1965).

[4] See (Lewis 1996, p. 306).

2   Physical Impossibilities

Some psychological items turn out to be metaphysically impossible, even though many of us have confusedly thought in many instances that they were possible. Corresponding to this and similarly unfamiliar are physical events, facts, or things that turn out to be metaphysically impossible despite our long having thought that they were possible. I have not yet said or said very much about what kinds of physical things fall into this category, but before saying any more, I would like to anticipate a further parallel between the mental and the physical as regards metaphysical/conceptual impossibility .

As we shall see in a moment and may have been obvious to some readers all along, there are two basic ways in which contents of xin turn out to be metaphysically impossible. One way, as with desiring a saucer of mud, doesn't rest on any impossibility having to do with particulars or particularity; the other way does, as when the present moment is believed to make all the difference to the color of emeralds according to the grue hypothesis. And there is a parallel within the category of (surprising or unanticipated) physical impossibilities. In some cases (as I hope to show you) the impossibility has something to do with particulars or particularity; in other cases it doesn't. But in order to introduce this distinction within physical impossibility, I think I should first say a bit more about the two ways in which mental (i.e., psychological) items or facts can be metaphysically impossible.

I mentioned the grue example as illustrating a psychological impossibility (of belief or belief formation) that seems to arise from a relation to some particular (the present moment), and that seems, therefore, related to particularity. However, even Anscombe's example can be changed into something involving a particular. If no one could simply desire a saucer of mud, no one could simply desire this saucer of mud rather than some other. The impossibility derives at least in part, of course, from what general kind of thing is said to be wanted. But particularity seems to play a role too. One can't just want this delicious ice cream, want it for no reason other than the fact that it is this ice cream rather than that ice cream over there. (If it is harder to reach the ice cream over there, then the desire for this ice cream isn't is just based on the fact that it is this ice cream.) And given the similarity to this ice cream case, the desire for a particular saucer of mud can be said to be impossible at least partly because of the particularity it involves. (Impossibility may be conceptually or metaphysically overdetermined in such cases.)

So some cases of impossible desire or belief depend on particularity, but, of course, Anscombe's famous original example does not. Moreover, once we recognize the impossibility of that sort of psychological item, other examples depending not on particularity but on the general kind of thing that is supposed to be

wanted or believed can easily be thought up. No one could simply want (to possess) a/some glass or container full of water: there has to be some reason why one wants it. And no one could simply believe that the perennial popularity of chocolate ice cream is or would be evidence that Pluto has no atmosphere. So here in the area of (inferential) belief we have an example of impossibility that, unlike the grue case, makes no essential reference to any particular time.

So there are two patterns to what is metaphysically impossible for or in the mind or xin, one having to do with particularity, the other not; and I want now to show you that analogous metaphysical impossibilities in or of the physical realm likewise break down in two (sometimes overlapping) classes or categories: those where the impossibility derives from particularity and those where other reasons make some physical fact or entity metaphysically impossible. With regard to the former, I shall be banking on an assumption that I think most philosophers or philosophers of science make or would be willing to make. I am going to presume at the outset that particularity cannot operate at the most fundamental level as causally/physically explanatory (though I think that what I have to say using this assumption will actually tend to support it).[5] Given this assumption, philosophers by and large (myself included) have tended to be rather superficial and confused with regard to entities or things that we have believed to be metaphysically possible but merely ruled out by the contingent laws of nature of the actual universe. These entities, these facts, these fundamental explanations are arguably (and I will give the arguments) metaphysically impossible, impossible in principle, and not just impossible in relation to contingent empirically known facts about the universe.

What do I have in mind? Well, none of us or almost none of us believes in magic (I am not here speaking of magic *tricks*). But I have never heard or seen any philosopher or, for that matter, anyone else say that magic as traditionally conceived is a metaphysical impossibility. Yet if we buy into the assumption about particularity I said I would rely on, then many kinds of magic are ruled out a priori and not (merely) as a result of accumulating empirical knowledge, because many kinds of magic implicitly assume the efficacy of the particular or the particularity of some entity.

There are many legends, stories, or myths according to which a particular in-

---

[5] This assumption doesn't assume that there can't be indeterministic elements in causation. It might be the case that the movement of a given electron is undetermined and that things have been set up in such a way that if it moves in one direction, it will cause the death of a cat and if it moves in another direction, the death of a dog. (Shades of Schroedinger's cat!) But this doesn't entail that there is a *basic physical/causal explanation* of why the electron came to cause the death of a cat, if that turns out to be what happens.

dividual (and what I am saying about fictional individuals can also be said about various fictional *things*) uniquely has the power to accomplish some task but nothing is *said* about how they and they alone got that power. You might argue that in such cases it is always *assumed* in the (telling of the) legend or story that there is a causally understandable source of that power that separates the individual with the special power from all the others who lack it, but I don't think that that would be accurate. In some stories no such assumption seems to be explicitly or implicitly made, and the idea is conveyed or insinuated that the given individual has their amazing power just because of who they are and without any further explanation being needed. This is an instance of fictional magic power that is at least implicitly understood as grounded in particularity, and I am saying that such particularistic explanation cannot in principle apply to or in non-fictional reality. (I am indebted here to discussion with Nancy Snow.)

Let me give you another and more specific example of where particularity is assumed to have explanatory force in regard to (non-fictional) physical events: unlucky numbers like thirteen in the West and in China the number four. It has been said that belief in the unluckiness of thirteen can be traced to the recognition, among the ancient Hebrews, that the thirteen letter of their alphabet is also the first letter of the Hebrew word for death. But even if that, fancifully, might be a reason to believe the number unlucky, I think or hope no one believes the number is unlucky *because* of its relationship with the Hebrew alphabet. I don't think the fact about the alphabet and the Hebrew language is supposed to be causally explanatory. The Last Supper offers us what may seem to be a better argument for believing in the bad luck supposedly inherent in the number thirteen. The fact that there were thirteen at the Last Supper may be thought to be the best causal explanation of *why* thing went so badly (from a limited perspective) right after the Last Supper. The idea here (as I am for our purposes conceiving it) is not that the Last Supper caused the number thirteen to be unlucky, but rather that the inherent unluckiness of that particular number might explain why things went so wrong immediately afterward.

But I wonder whether this is really intelligible. If the unluckiness of the number is not due to events in the history of the Jews, it has to be inherent in that number as that number, and that doesn't make sense. Remember too that the unluckiness of thirteen isn't just a matter of cardinality, where groups containing thirteen members are somehow jinxed because of the number of their members. Thirteen is also thought of as ordinally unlucky, as when people don't want to live on the thirteenth floor of a building and floors are numbered accordingly. (The same thing happens in Taiwan with the number four.) So the unluckiness of the number in the popular imagination is more general and in some sense more abstract than

anything to do with the cardinal number thirteen or its ordinal counterpart, and is something that abstract really capable, even in principle, of entering into physical explanations (e. g., of why the thirteenth house on the street was destroyed by a hurricane and the other houses were not)? I think not. To hold that the particular number in itself is unlucky and capable of affecting and explaining physical events is very much like saying that the Platonic Forms might all on their own affect the material world, and, of course, Plato himself precisely didn't say this, but rather in the *Timaeus* invoked the Demiurge to explain how and why things happen in the material world. We have reason to believe that numbers are too abstract to be able to causally explain things in the physical world, and so in addition to the fact that the fact that thirteen is the particular number it is cannot explain physical events, it cannot explain such events *because it is just a number*.

The present example thereby helps us to see that there are at least two ways in which (the truth of) physical explanations may be impossible. A certain kind of ultimate physical explanation may be metaphysically impossible on a priori grounds because it relies on a certain kind of particularity, but other sorts of explanatory physical fact are impossible because of the general concepts that would apply to them, the general kinds of factors they would have to be based on or in. Numbers cannot explain because they are numbers, and this takes us beyond particularity.[6] But just in case you aren't entirely convinced, let me mention another

---

[6] I have not mentioned psychokinesis, but what I have been saying pretty clearly entails that "pure" psychokinesis is metaphysically impossible. Is that plausible? Well, it is interesting that defenders of psychokinesis have typically gone in to offer further explanations of what happens when mind influences distant bodies. For example, in their famous/infamous paper (Mattuck and Walker 1979), R. Mattuck and E. Walker argue that psychokinesis is underlain by quantum phenomena. This paper has been roundly criticized for the specific theory it offers, but the main point for my purpose is the fact that they were not satisfied to treat psychokinesis as a pure phenomenon subject to no further explanation. Other theorists have offered different possible explanations of psychokinesis: e. g., via a fifth force additional to those physics already acknowledges, but this too fits well with the idea that the pure thing is impossible. However, in (Broad 1949), C. D. Broad offers an account of psychokinesis in terms of our having extended intangible and invisible bodies as a possible alternative to the pure theory, but also says that the pure theory might be true. Others are somewhat inclined to disagree with this. In (Candlish 2007, pp. 67, 307 and passim), Stewart Candlish suggests that pure psychokinesis may very well be impossible in principle, but doesn't state this definitively. He also compares psychokinesis to magic. (He is not the only author who does this.) My own take on this literature is that it should be considered together with the whole supposed realm of magic. When we do that, I think we are led toward the conclusions reached in the main text, and this implies or entails that pure psychokinesis is impossible. I think similar points can perhaps be made about astrology and the idea that heavenly bodies affect (at a distance?) human personalities. This sort of thing may *au fond* constitute magic in the sense we are arguing is impossible. Finally, let me mention David Robb and John Heil's insightful article on (Robb and Heil 2008). At the end of section 2.2 they consider the supposed possibility (defended by some dualists they mention) that particular individual souls might have the power to influence only one particular body and they compare that idea to holding that a given key

example where certain general kinds of numbers simply cannot explain physical facts in any ground-floor way.

This example comes from a private discussion I had with Robert Nozick more than forty years ago. Nozick was interested in questions about the nature and existence of God, and noting that the empirical evidence for God's existence is unclear at best, he mentioned to me a possible case, a possible universe, where intelligent enough beings would definitely have evidence that there is a God. He suggested that astronomers might discover or might have discovered that every solar system they investigated had a prime number of planets in it (including our own, as was imagined in the days when there were thought to be seven planets). This, he said, and I agreed with him, would be evidence that there was a God and evidence, moreover, that God wished to reveal his (or her or its) existence to sufficiently intelligent beings.

However, in agreeing about all of this, *neither of us ever considered the possibility that the fact that every solar system had a prime number of planets might be a basic physical fact about the universe*; and I think that is because both of us were implicitly assuming such a physical explanation was impossible, made no coherent sense. The hypothesis of God had to be the explanation (if any explanation was possible) of the physical fact that there were only prime numbers of planets in every given solar system. (We should probably imagine further that there are not only different prime numbers of planets for different solar systems, but also that in some cases the number was relatively high: like 101 planets in a given solar system and 103 in some other, but never 100, 102, or 104 planets in any given solar system. We could make the thing even more evident if we talked of paired prime numbers of planets or of perfect numbers of planets, but enough is enough here.)

I think this example makes it even clearer why numbers or general facts about numbers cannot be ultimate physical explainers. Again, the supposition that the primeness of numbers ultimately explains certain physical facts seems to smack of the confusion one could have accused Plato of if he had treated the Forms as capable of shaping the physical/material world. This sort of explanation doesn't invoke magic, but it somehow confuses different levels or realms of reality. Gottlob Frege in *The Thought* (1918 – available on-line) spoke of a third realm of entities to which thoughts, or Gedanken, belong – the other two realms being the mental and the physical. And the fact that numbers seem more to be associated with the

---

might have the power to unlock one particular lock and no other exactly like it, claiming that "…it is by no means obvious that powers *could* be individualistic in this sense." The parallelism they draw between individualistic mental causation and individualistic purely physical causation resembles what I have been saying here about particularity, but Robb and Heil are not as committed as I am to denying such possibilities.

third realm presumably has something to do with the fact that particular numbers and general facts about or properties of numbers cannot explain physical events or realities.

The idea that the primeness of numbers could explain physical facts about solar systems doesn't seem exactly to be an instance of *magical* thinking, but in any event what we have said about that kind of example and about others where magic is definitely supposed to be involved shows us that the metaphysical impossibilities that exist with respect to the physical really do run parallel to those we have unearthed regarding what xin can contain. Among both the physical and the psychological cases of metaphysical impossibility particularity can be and often is a (philosophically) explanatory factor, but considerations other than particularity are relevant to metaphysical impossibility in other instances.

At this point, though, rather than end our discussion of this parallelism between the psychological and the physical, I want to illustrate and enrich the general points being made here by speaking of another kind of example of putative physical explanation where what is often thought of as merely contingently nonexistent turns out (on a priori grounds) to be metaphysically impossible. This further example concerns magic spells and incantations and magic words or phrases (like "Open sesame!"). Does it really make sense to think such spells can work in the way we are sometimes asked to think of them: as ultimate explainers of physical events. For Ali Baba, saying "open sesame" caused a cave door to open, and in other cases magic spells or incantations are thought of as having immediate effects in or on the world. (These words are used in "Ali Baba and the Forty Thieves," a story that comes from later versions of *A Thousand and One Nights*, a collection ultimately of Islamic origin.) Of course, many spells have counter-spells, but there doesn't seem to have been any counter-magic to block the efficacy of "open sesame," and even where there is a counter-spell, the point is that the spell is efficacious if the counter-spell *isn't* invoked. Could the saying of certain particular words in and of themselves and without any deeper or further explanation be efficacious in the way magical thinking supposes?

It is worth noting that in the case of "open sesame" and in many other cases of supposed magic as well, the incantation bears a semantic relation to what is supposed to happen: the cave is supposed to open. Perhaps this is no accident and can give us insight into what is supposed to happen or be possible when a magic spell or magic words are uttered. Why should a cave door open to words telling it to open rather than to words telling it something else? Couldn't it be basic that the door would open to the words "close sesame"? Well, no believer in magic and no tale about magic would ever suppose such a thing, and I think the reason, the deep reason, why that is so is that the magic is supposed to work at a

Michael Slote

linguistic/semantic level, as if the door understood the words and only responded because of what they mean. But this is truly incoherent, and I am inclined to say that similar incoherence attaches to what is being supposed in other cases of magic spells or phrases. Physical things don't and can't understand language unless, possibly, they are governed by software and hardware that allows them to do that. But in the case of Ali Baba's cave no such thing was being supposed. There is something really quite absurd in the idea that spoken words can govern physical realities without their being any further or deeper physical mechanism by which they work. For example, the doors of a household can be set up to respond to voice commands like "open" and "close" but this doesn't just happen because those words are said. There have to be underlying physical mechanisms and software that ensure that the system works, and, again, we are precisely assuming that there is no need for such things when we speak of phrases like "open sesame." The whole idea of the magical efficacy of certain words just doesn't make sense and represents something that metaphysically cannot be.

Let me say just a bit more deriving from our example of "open sesame." Could there be a world in which every time anyone wished for something, the very opposite happened or remained the case and where this was an unexplainable basic fact about events in physical reality? Clearly, such a supposition is absurd, but its absurdity shows there to be something similarly absurd about the idea of a world where willing something to happen in the world automatically leads to its happening. It seems absurd to suppose that the world might automatically counter our wishes because it is absurd to think the world has thoughts and desires about us. What other than a desire to counter what it understands our wishes to be could explain why the universe always produced the opposite (in some sense) of what we wanted or willed to happen? So the idea that it could be basic to the universe that things happen contrary to what people wish for makes no sense, describes what is metaphysically impossible, and what we say about this kind of case transfers readily to the idea of the universe's always producing what we do will or want. This too seems to suppose that the universe understands us, and it supposes in addition that the universe is favorably disposed toward us, and none of this makes sense or is possible. In a way, the idea that there could be some sort of connection between willing and what happens in the universe anthropomorphizes a universe that is at the same time regarded as not being ensouled, and this is a logically lethal combination: we could call it "logical animism." In a nutshell, then, and to return to our specific example, if it is impossible for "close sesame" to have magical efficacy, then neither, in the end, is such a thing possible for the words "open sesame."

So the category of what is physically impossible appears to be just as rich with

impossibilities as what we saw in connection with the psychological category that includes beliefs, desires, attitudes, etc., that metaphysically cannot be.[7] And given our previous arguments, it follows that, with respect to the physical realm at least, philosophers (not to mention the general public) have held very mistaken views. They have assumed that magic is metaphysically possible at the same time they supposed there to be no such actual thing or phenomenon, whereas the picture I have been offering here entails that they or we have all been mistaken on the issue of metaphysical possibility.

Perhaps we should shy away from attributing such widespread error, but is the case really very different from what Saul Kripke says about fictional creatures? We all tend to think that Pegasus and unicorns are metaphysically possible but just not actually existent, but Kripke's reference-fixing semantics entails (though this is a very informal way of putting it) that Pegasus and unicorns not only *don't* and *didn't* exist but never could have existed. This runs parallel to what I am saying about magic, and perhaps there is safety in numbers here, not to mention the possible strength of the arguments I have offered for my conclusions.[8]

Overall so far, this paper has explored a rather extensive parallelism that exists between the psychological and the physical and has attempted to show that magic is a metaphysical impossibility. But let me now indicate what I take to be the most important implication or set of implications of what we have been arguing for so far.

3   From the Impossibility of Magic to the Uniformity of Nature

What we have been saying here about magic can lead us toward questioning a fundamental assumption about reality that both Hume and Kant made and that most (analytic) philosophers today also make. That assumption is that the uniformity of nature is not a given, is not guaranteed a priori or with metaphysical necessity. We all know that Hume made this assumption, but Hume's enormous influence on Kant largely depends on the fact that Kant appreciated Hume's point and thought the only way around it was to distinguish phenomena or appearances

---

[7] I also hope it is clear that the parallelism between the psychological and the physical that I have been describing has nothing to do with psychophysical parallelism, the view that the physical and mental/psychological run parallel but never causally affect each other.

[8] See (Kripke 1980), especially Lecture 3. Incidentally, people tend to think that there are no such things as monsters but that monsters could have existed. But any living creature has to have a determinate biology and wouldn't count as an "unnatural creature" of the sort we conceive monsters to be – only monsters in fiction can be thought of that way. So in parallel with what Kripke says about unicorns and I am saying about magic, it seems a mistake to regard monsters as only contingently non-existent. On this see (Slote 2010).

from noumena or things in themselves. Philosophers today seem largely to acquiesce in Hume's view of this matter and mainly look for non-Kantian ways of dealing with the issues about induction Hume raised.[9] But if we say that nature has to be uniform in Hume's sense, that the future has to resemble the past, we are advocating a further metaphysical limit or limitation on reality beyond those discussed above, one that, I think you will agree, is or would be far more philosophically significant than any of those mentioned and argued for earlier in this paper. But at the same time that further conclusion seems to me to be supported by what was said earlier about the metaphysical limits of what reality can accommodate, and I shall shortly explain how this works. For the moment, however, let me give a bit of background for what I shall be saying.

Those who even today – and that includes most of us and myself too up till very recently – regard the uniformity of nature as lacking any necessary metaphysical backing do so in relation to normal predicates and scientific categories. If you bring in Goodman-like predicates like "grue" ("green if examined before now and otherwise blue"), it becomes in one sense easier to maintain the necessary uniformity of nature: whatever happens, some universal law making use of a Goodman-like or of a normal or familiar predicate will inevitably hold, and nature will in that sense and to that extent be uniform. But of course if nature is uniform in relation to Goodmanian predicates, it won't be uniform with respect to ordinary predicates, so one might conclude on that basis that it is impossible for nature to be uniform in every way, or that the whole idea of uniformity is too predicate-relative to be a clear or useful notion. Therefore, when typical philosophers assume that nature needn't be uniform, they do so independently of what they think about the implications of Goodman's thought; and I am going to argue that if we stick to the kind of (Humean or pre-Goodmanian) thinking such philosophers engage in with respect to the uniformity of nature, it turns out that, contrary to common philosophical belief, nature has to be uniform.

Now above I defended the view that no intelligent being could, metaphysically could, think inductively in terms of grueness rather than greenness, but even if that is correct, one might still hold that it is metaphysically possible that emeralds observed (if at all) only after now will all be grue and therefore blue. Even if, given present evidence, one has to *think* or *infer* that future emeralds will be green, it still seems possible that those emeralds not be green but rather blue. More

---

[9] On Hume's present-day influence regarding this point, see (Schwartz 2009). Hume's discussion of induction and the uniformity of nature occurs both in the *Treatise of Human Nature* and in the *Enquiry Concerning Human Understanding*. Kant's discussion of Hume's views on this topic (which is often mixed in with discussion of Hume's ideas about causation) can be found both in the *Prolegomena to Any Future Metaphysics* and in the second edition of the *Critique of Pure Reason*.

generally, even if one holds that no one can *think* the future will be totally unlike the past, one may still hold that the future *might/may* be totally unlike the past (our past). This last belief seems common to philosophers of the Western tradition, but I want to try to persuade you now that there are reasons for doubting it based on what was said here earlier.

Our discussion above questioned, among other things, whether magic causal particularity makes any sense. If someone tells me of a door that opens just because the magic words "open sesame" were said in front of it (and not because there is some physical device governing the door that is sensitive to the particular sound waves that exist when one says these words), I am not going to believe them because, for all the reasons mentioned above, what they are saying doesn't make any metaphysical sense. By the same token, it makes no sense to think of "Change to blue!" as special words whose utterance magically can change a green emerald into a blue one. Now Goodman's riddle doesn't depend on any emerald changing from green to blue, but many writers on the riddle make the mistake of assuming that it does.[10] And the mistake, I think, is a natural one because what the actual riddle requires us to imagine, that someone infers from past green/grue emeralds that future ones will be blue/grue, seems to require, for ordinary ways of thinking, an inexplicable epistemic shift in someone's beliefs about emeralds: from seeing past ones to be green to thinking future ones (I am speaking roughly) will be blue.

Well, those who have imagined that Goodman's riddle is about or can be about emeralds inexplicably changing from green to blue have substituted an inexplicable metaphysical event for an inexplicable epistemic shift in someone's thinking, but their misinterpretation constitutes a kind of appreciation of just how inexplicable any green/blue epistemic or metaphysical shift would be. *And this is entirely consonant with received modern and recent views about the uniformity of nature.* If (as those who misunderstand the riddle imagine) it is possible for an emerald to shift unaccountably from green to blue and if the epistemological problem raised by the riddle concerns the justifiability or unjustifiability of the possible belief that such a shift will occur, then, given the truth of that belief, the future of emeralds will not fully resemble their past, but there will be no specific causal reason why this possible change should occur. And the emeralds can be a microcosm of the universe as a whole. If it is possible for an emerald to shift unaccountably from green to blue, why shouldn't this happen mutatis mutandis and with a larger complement of properties and things to the universe as a whole? So this thinking about what is metaphysically possible for emeralds is just an instance of the more

---

[10] D. M. Armstrong in (Armstrong 1983, p. 58) treats the riddle as involving emeralds' changing from green to blue and others have made the same mistake. The fact that it is a mistake has been frequently pointed out. See, for example, (Schwartz 2009).

general and widely held belief that nothing metaphysically ensures that nature is uniform, that the future will resemble the past.

But does it actually make sense to assume that an emerald might change color for no reason at all? If it makes no sense to assume that *particularistic magic* could bring about such a change, do we really have reason to hold that it makes sense to assume that such a change could occur without there being *any causal-explanatory reason* (embedded in all emeralds, say) for it to occur? Yet that is precisely the sort of thing that belief in the possibility that nature isn't uniform involves, a shift in nature or the world that isn't due to factors operating previously in the world. If it is (seen as) due to such factors, then nature is (being conceived as) uniform and regular, though at a higher or more abstract level than was previously realized. And (once again) I am not talking here about Goodmanian predicates. If one cannot simply believe that previously unobserved emeralds will be unlike those one has already observed, then neither could anyone think that unobserved grue emeralds resemble observed ones more than unobserved green emeralds do, and in that case the uniformity of nature also cannot be conceived in terms of Goodmanian predicates. The "typical" philosophers I mentioned earlier are actually conceiving the uniformity of nature in the only way in which it is possible or makes sense to conceive it.

Now someone might object at this point that since electrons can be undetermined in their movements, it makes sense to suppose that emeralds might be undetermined in their color and so for no deterministic reason sometimes shift from being green to being blue. But it is part of the nature of electrons that their movements are or can be indeterministic. That hardly means that everything about electrons is indeterministic and inexplicable, and when an electron moves indeterministically, that is in accordance with the laws and regularities that govern electrons. A higher-level explanatory theory makes room for the indeterministic movement of electrons, and a parallel in the case of emeralds might be that they contain electrons whose indeterministic movements cause them to change color (let's ignore the implausibility of this assumption relative to present-day physical theory). But this would not be an uncaused change of color of the sort imagined by those who misunderstand Goodman's riddle; rather and to the contrary, the change would causally depend on uncaused smaller motions. This is far from magic and does nothing to undercut the general uniformity of nature.[11]

However, someone might then speculate that emeralds might change color indeterministically or without cause and do so *independently* of the indetermin-

---

[11] By the same token, if there is a bomb that for lawlike physical reasons can destroy the universe as we know it and the bomb goes off with precisely that effect, we haven't got an instance of the non-uniformity of nature.

istic movements made by the particles they contain; but this is dubious because it suggests that what happens to a thing can be independent of what happens to the entities it is constituted of. In any event (and independently of the issue just raised about emeralds, which is applicable to the universe as a whole), if the universe changes deterministically over time, it may well end up being very different from anything we have expected, but there is no reason to hold that such change would be incompatible with our offering a true general explanatory scheme or theory about how the universe operates and has operated, and such change will therefore do nothing to undermine the idea that nature as a whole is uniform. The resemblance and uniformity between future and past will just exist at a higher level or be more complex than what we ordinarily imagine it to be, and the most that would then be true is that we didn't know the (causal potentialities of the) past world all that well and so didn't (fully) know the *nature of the uniformity* that governs the universe.

Here is another way to reach the same conclusion. For nature to be or become non-uniform, it has to change (relatively) suddenly and as if by magic lose the power to propel/perpetuate itself into the future. If, as I argued earlier, the existence or (therefore) the onset of positive or active magic power (or regularity) is impossible at any given time, then there seems just as much reason to say that that a sudden magical *lapse*, *loss*, or *deflation* of power (or regularity) is metaphysically impossible too. But that is just what denying the necessity of uniformity involves, so the universe turns out on philosophical/conceptual grounds to be necessarily uniform in the sense relevant to Hume's skeptical worries and Kant's first *Critique*. It is harder to coherently conceive nature as non-uniform than we have realized, and if the arguments I have been offering have any force, then they obviate the need for the distinction between noumena and phenomena that Kant thought was necessary to combatting Hume's views about the a posteriori character of nature's uniformity[12]. Hume may simply have been wrong about the possibility of nature's lack of uniformity, though it may not be easy to see that until one sees the relevance of magic and of its metaphysical impossibility to this whole issue. That is the main upshot of what I have been saying in this essay.

---

[12] In a way Kant's approach to the uniformity of nature and the one taken here are diametrically opposed. Kant sought to demonstrate the necessary uniformity of nature by arguing that a certain kind of causality (based on Newtonian assumptions) is *necessary* to the *appearances of things*. I have sought to show the necessity of such uniformity on the basis of arguing that a certain kind of causality (magic causality) is *impossible* with respect to *things simpliciter*. In addition, there is general disagreement as to how well Kant demonstrates universal causation in the Second Analogy of Experience and elsewhere. I personally have always found his argument quite unpersuasive, beginning when I was Charles Parsons's assistant many many years ago in the Kant course at Harvard. That is at least part of what has motivated the approach I have taken here.

But before we close up shop on these issues, let me briefly connect what has been said just now with what Hume famously said about induction (though he never uses that specific term). Among other things Hume held that when we make any inductive inference from past to future (e. g., from the color of past emeralds to that of any future emerald), we have to assume that nature is uniform; and Hume thinks this assumption cannot be metaphysically guaranteed or assured a priori and can at best only be justified in terms of previous or other inductive arguments (which leads to a justificatory regress). Most philosophers think this gives us at least some philosophical reason to doubt the validity of inductive inferences, but the doubt rests on the assumption that the (principle of the) uniformity of nature is purely contingent, *and that is precisely what our previous arguments call into question.* I think we can conclude, then, that Hume's original doubts about the uniformity of nature can be answered and that the inductive skepticism he (is often thought to have) defended based on what he said about the uniformity of nature can also be answered.[13]

References

Anscombe, G. E. M. 1957. *Intention.* Oxford: Basil Blackwell.

Armstrong, D.M. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.

---

[13] In an article published many years ago that deserves more present-day attention, Wesley Salmon convincingly argues that even if the principle of the uniformity of nature is true and known to be true, that won't by itself help us to justify particular inductions: see (Salmon 1953, p. 42). Salmon's reasoning can be stated, roughly, as follows: If I have seen nothing but black crows, this may nonetheless be part of a pattern of variability with respect to color in crows that would be consistent with the uniformity of nature. So the uniformity of nature doesn't in itself tell us that we have reason (even other things being equal) to make the usual and supposedly justified inference that the next observed crow will probably be black. Knowing that some natural seeming regularities have to be true might not help us, on the basis of our empirical sampling of events and objects in the world, to in any way identify which ones they are. I say that Salmon's argument convinces me, and I mention it because the argument of the present paper doesn't actually assume that the uniformity of nature justifies induction. If Salmon is right, then Hume is simply mistaken to think that we could argue for specific inductive conclusions if we were somehow allowed to assume nature's uniformity. But the reason why the assumption of the uniformity of nature doesn't help us with specific inductive inferences is that it is too general to do that: nature can be uniform independently of any specific such inference or its reasonableness.

Broad, Charlie Dunbar. 1949. "The relevance of psychical research to philosophy." *Philosophy* 24 (91): 291–309.

Candlish, Stewart. 2007. *The Russell/Bradley Dispute and its Significance for Twentieth Century Philosophy.* New York: Palgrave Macmillan.

Foot, Philippa. 1978. *Virtues and Vices and Other Essays in Moral Philosophy.* Oxford: Blackwell.

Goodman, Nelson. 1965. *Fact, Fiction, and Forecast.* 3rd ed. Indianapolis: Bobbs-Merrill.

Kripke, Saul A. 1980. *Naming and Necessity.* Oxford: Blackwell.

Lewis, David. 1996. "Desire as Belief II." *Mind* 105 (418): 303–313.

Mattuck, Richard D, and Evan H Walker. 1979. "The Action of Consciousness on Matter: A Quantum Mechanical Theory of Psychokinesis." In *The Iceland Papers,* edited by Andrija Puharich, 111–160. Amherst, WI: Essentia Research Associates.

Robb, David, and John Heil. 2008. "Mental Causation." In *The Stanford Encyclopedia of Philosophy,* Winter 2008, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2008/entries/mental-causation/.

Salmon, Wesley C. 1953. "The uniformity of nature." *Philosophy and Phenomenological Research* 14 (1): 39–48.

Schwartz, Robert. 2009. "Goodman and the Demise of Syntactic and Semantic Models." In *Handbook of the History of Logic,* edited by Dov M Gabbay and John Hayden Woods, 10:391–413. Elsevier.

Slote, Michael. 2010. "Inapplicable Concepts." In *Selected Essays.* New York: Oxford University Press.

———. 2018. *The Philosophy of Yin and Yang: A Contemporary Approach.* Beijing: Commercial Press.

EA
JP

This issue launches the new academic journal, the East Asian Journal of Philosophy (EAJP). It consists of invited articles written by a diverse group of prominent global scholars working in different areas of philosophy.

EAJP is an open access, peer-reviewed journal with no publication charge for authors (APC) whose aim is to provide a platform for theoretical proposals and discussions both within specific philosophical traditions, and across different traditions and cultures.