HUNGARIAN PHILOSOPHICAL REVIEW

■

MAGYAR FILOZÓFIAI SZEMLE

Current Issues in Metaphysics

Contents

Editorial foreword

The *Hungarian Philosophical Review's* current volume contains six essays. The essays, besides the fact that they all deal with current issues of metaphysics, are quite disparate.

As editors we would like to thank our authors for their contribution to our volume. We do hope that this supplementary volume of the *Hungarian Philosophical Review* would further the international recognition of philosophy done in Hungary.

Gábor Bács, Dávid Márk Kovács, János Tőzsér

GÁBOR BÁCS – JÁNOS TŐZSÉR

Works of Art from the Philosophically Innocent Point of View

INTRODUCTION

The *Mona Lisa*, the *Mondscheinsonate*, the *Chanson d'automne* are works of art, the salt shaker on your table, the car in your garage, or the pijamas on your bed are not. The basic question of the metaphysics of works of art is this: what makes a thing a work of art? That is: what sort of property do works of art have in virtue of which they are works of art? Or more simply: what sort of property *being a work of art* is?

In this paper we argue that things are works of art in virtue of *what they are like*, their intrinsic features, that is, in virtue of the fact that they have the perceptual (auditory, visual, etc.) properties they have. In other words: *being a work of art* supervenes on *perceptual-intrinsic features*. Currently, this metaphysical view is extremely unpopular within the philosophy of art. It is unpopular because there allegedly exists a knock-down objection to it, the well-known argument from indiscernible counterparts. Our thesis implies, among other things, that every perceptual duplicate of a work of art is also a work of art. According to the argument from indiscernible counterparts, however, there could be (or even: there are) indiscernible counterparts such that one of them is a work of art while the other is not. Hence things cannot be works of art solely in virtue of what they are like.

Our paper divides into three parts. In the first part we state our views. In the second part we defend it against various versions of the argument from indiscernible counterparts. (In doing so our position will become more plausible, we hope). In the final part we provide some meta-reflections on the matter.

1. THE NATURE OF ARTWORKS

1.1. What is our view exactly?

Let's begin with the notion of perceptual property. The perceptual properties of the *Mona Lisa* are those which are visually presented to us when we look at the painting. The perceptual properties of the *Mondscheinsonate* are those which

are auditively presented to us when we listen to the music. It is crucial that what we call the 'perceptual properties' of the *Mona Lisa* or the *Mondscheinsonate* are *not* the properties of our experiences of them (they are not qualia, or some such mental, phenomenal stuff), but the perceptually accessible intrinsic properties of the works of art themselves what we can see or hear. By definition: the perceptual properties of work of art *O* are those intrinsic properties of *O* which *are manifested to the perceiver during the perception of O*.

Someone might say that notions like 'perception of a work of art' and 'perceptual properties of a work of art' are plausible in the case of fine arts and musical arts (we do perceive such works of art during which their perceptual properties are indeed manifested to us), but they are not plausible in the case of literature.

This isn't so. To begin with, the perceptual properties of some literary artworks are not the perceptual properties of its (printed) textual image. We obviously do not say that the *Chanson d'automne* is a work of art in virtue of the fact that the image of its printed text has the perceptual properties that it has. Instead we say that when you read the *Chanson d'automne*, the work of art is *presented* to you. Not visually, of course, and not auditively, but – to put it this way – *imaginatively*. That is, when you read the *Chanson d'automne* you *perceive* it imaginatively (with your 'mind's eye', so to speak), and the perceptual properties of the *Chanson d'automne* are those which are imaginatively presented to you.

The matter becomes clearer if we bring out a parallel between literary artworks and musical artworks. Suppose that you are a thoroughbred musician and before you lies a copy of the musical score sheet of the *Mondscheinsonate*. When you read the musical score sheet, to you, the qualified musician, the *Mondscheinsonate* becomes imaginatively presented. You do hear the music, imaginatively. Its perceptual properties are manifested to you, imaginatively. We claim that something *similar* is the case when someone reads the *Chanson d'automne*. You do see what the poem says, imaginatively. Its perceptual properties are manifested to you, imaginatively. So the perceptual properties of the *Chanson d'automne* are like the imaginatively given perceptual properties of the *Mondscheinsonate* when a qualified musician is reading its score.

Furthermore, just as you can perceive the *Mondscheinsonate* not only imaginatively, by reading its musical score sheet, but auditively by hearing it, you can perceive the *Chanson d'automne* not only imaginatively, by reading its printed text, but auditively—by hearing its recitation. Just as one can play the *Mondscheinsonate* on a piano (it could be you), one can recite the *Chanson d'automne* (it could be you). Now, it is plausible to think that we gain access to the very *same Mondscheinsonate* when we read its musical score sheet and when we hear it, and we gain access to the very *same Chanson d'automne* when we read its printed text and when we hear it. It is also plausible to think that when you hear the *Mond-*

scheinsonate, and when you hear the *Chanson d'automne*, what you hear is their perceptual properties. Hence we can rightly speak of the perceptual properties of the *Chanson d'automne*, or of any other literary work of art for that matter.

It's by sheer historical coincidence that fewer people can perceive musical works of art by reading musical score sheets than can perceive literary works or art by reading printed texts. We can easily imagine a world at which the only mode of encounter with the *Mondscheinsonate* is by reading its score. It would be odd to say that in that world, the *Mondscheinsonate* has no perceptual properties.

So what we are saying is this: *O* is a work of art in virtue of what it is like, that is, in virtue of the fact that *O* has the perceptually accessible intrinsic features that it has, features that are manifested to us when we perceive *O* (visually, auditively, or imaginatively). The properties manifested to us during perception are the perceptual properties which *alone* can make something a work of art.

Notice that this is not a definition of the *concept* of work of art. The same way in which a physicalist theory of mind is not a definition of the concept of mind. We might ask: among the many perceptual properties a work of art has, which ones are those in virtue of which it is a work of art, and which ones can be used to define the notion of a work of art. Answering such questions lie outside the metaphysician's competence. It's the art historian's, the aesthete's and the art critic' job to provide such answers. Just as it's not the physicalist's task, but the neurophysiologist's to spell out which neurophysiological state is responsible for which mental state. We only say that perceptual properties alone determine works of art, but we are silent about which perceptual properties we're talking about, the same way in which the physicalist only says that physical properties alone determine mental states, but she remains silent about which physical properties she is talking about.

A further aspect of similarity. Just as neurophysiologists can be wrong about which neurophysiological state is responsible for which mental state, art historians, aesthetes and art critics can be wrong about which perceptual properties make something a work of art. So it can all too easily happen (as it did) that during some period of time, a work of art is mistakenly taken for something else before realizing that it is a work of art after all. Mind you, it was a work of art all along, it's just that it wasn't recognized as such.

1.2. What is our main motivation?

Our main motivation is that only this view is in line with our most commonsensical, philosophically innocent beliefs about works of art. From the philosophically innocent point of view, works of art are what they are because of what they are like. The *Mondscheinsonate*, for instance, is a work of art because of what it sounds like.

To us this much is clear. Take David Lewis' simple maxim of honesty: "never put forward a philosophical theory that you yourself cannot believe in your least philosophical and most commonsensical moments" (Lewis 1986, 135). If we heed the maxim, we *cannot* put forward a theory according to which the perception of the *Mondscheinsonate* gives us little clue as to whether it is a work of art or not, that what matters is *not* its perceptual likeness, but something else. We would surely dismiss any such view in our least philosophical and most commonsensical moments.

Why? Well, we usually buy tickets for music concerts to perceive the properties of musical works like the *Mondscheinsonate*, in order to be able to hear certain sounds following one another in a certain rythm, in a certain tempo, and with a certain dynamics. If you were trying to feed us with the idea that the *Mondscheinsonate* is a work of art not because of its perceptual features that you can hear in the concert hall, but because of something else (not perceptual, hence not manifested there and then), we would not buy into it: *everything* that counts is in the concert halls.

Also, if the *Mondscheinsonate* were a work of art because of some non-perceptual feature (like causal history, for instance) which is not manifest when listening to the music, then someone who knows the causal history of the *Mondscheinsonate*, but has never heard the musical work itself is in a better position to judge for himself whether it is a work of art, than someone who knows little about the origin but is thoroughly acquainted with the music. This is counterintuitive. Anyway, whichever non-perceptualist account we end up with, none of them attributes importance to the perception of works of art, at least for judging their artwork status. And this we find extremely implausible—in our most commonsensical moments.

1.3. What views are we against?

So far we have stated our view which was: something is a work of art in virtue of what it's like; and our primary motivation for it was: it accords with our most commonsensical beliefs about works of art. There are many alternative views, however, which are all non-perceptualist.

According to the institutional theory of art, something is a work of art in virtue of the fact that some professional jury, or social institution has conferred the status of artwork upon it (for example, Dickie 1974, Fish 1980). This view is non-perceptualist: on this account, something is a work of art because of the way it is related to the decision of certain people, to some collective intentionality, and such things are obviously not manifest to us when we perceive the artwork itself. According to the mimetic theory of art, something is work of art in virtue of the fact that it mimics a portion of reality (for example, Plato in *The Republic*).

This view is non-perceptualist: a work of art is that which stands in a similarity relation to a portion of reality, which is, again, not manifest to us when we perceive the artwork itself. According to a third theory, something is a work of art in virtue of the fact that we relate to them in a special (aesthetic) way. To use Kant's *bon mot*, with disinterested contemplation (Kant 1790/1997). This view is non-perceptualist: a work of art is that which stands in some relation to a contemplative mode, which is not manifest to us when we perceive the artwork itself. Finally, according to a fourth theory, something is a work of art in virtue of the fact that it has resulted from some deliberate creative artistic intention (for example, Danto 1981). This view is non-perceptualist: a work of art is that which stands in some causal or ancestral relation to a certain artistic intention, not manifest to us when we perceive the artwork itself.

It is quite clear that our view is the opposite of all these *relationist* accounts which claim that works of art are what they are because of the way they are related to other things, where the relation in question (any one of the four) *cannot* be manifested to us in the perception of the artworks themselves. We hold that works of art are what they are because of what they themselves are like, because of their *intrinsic nature* so to speak, and which *can* be manifested to us in the perception of the artworks.

We have already remarked in the introduction that a view like ours is extremely unpopular these days in the philosophy of art, and not only because of the indiscernible counterparts objection.

This unpopularity is also due in part to considerations like the following. It is not uncommon that certain works of art *refer* to other works of art and as such are interrelated. For instance, in the *Don Giovanni*, Mozart makes explicit reference to a well-known part of *The Marriage of Figaro*. Or to use a literary example, in the *Cantos*, Ezra Pound quotes a part from the *Iliad*. Since such references are constituent parts of the works of art in question, their presence indicates that factors like the interrelation between works of art also matters in their artwork status, yet this interrelatedness is beyond the perceptually accessible intrinsic features of any given artwork. This might give a further incentive to the idea that any perceptual–intrinsic account of artworks is doomed to failure.

We think that this consideration concerns the *interpretation* of works of art. We do not deny that the interpretation of a work of art requires much more than taking into account what it is like. In the course of interpreting the *Don Giovanni*, we must take into consideration, among other things, its reference to *The Marriage of Figaro*, and in interpreting the *Cantos*, its quotation from the *Iliad*. No quarrel there. What we do deny, however, is that the ontology of artworks depends in any way on their interpretation, on what we take them to be about.

2. ARGUMENTS FROM INDISCERNIBLE COUNTERPARTS

2.1. The perfect forgery

The argument from perfect forgery runs as follows. It could have been the case that a perfect forgery of the *Mona Lisa* was created. In this counterfactual scenario, the original and the fake are *indiscernible counterparts*, they have the exact same perceptual properties. But alas, while the *Mona Lisa* is a work of art, the perfect replica is *not*. For the *Mona Lisa* is *the original* made by Leonardo, the perfect replica is just a *forgery*. Therefore, it is not true that every (possible) duplicate of a work of art is also a work of art. Consequently, it is not true that a work of art is what it is in virtue of its perceptually accessible intrinsic features.

The argument from perfect forgery is certainly not conclusive. It is no more plausible to maintain that while the original painting is a work of art the perfect forgery is not, than to say that the perfect forgery is *also* a work of art, what's more, the *same* work of art as the original. It is no more plausible, for us at any rate, because we hold that the *Mona Lisa* is a *universal*. Not Platonic, not some abstract stuff lying outside the spatiotemporal realm, but an Aristotelian *immanent* universal which can be perceived. Such an entity, which admits of multiple instances, would be a multi-located entity wholly present at numerous non-overlapping places at the same time. And it would not be a simple immanent universal like whiteness or roundness, but a *structural* universal, an entity that is constituted by all the perceptual properties of all the parts of the particular instantiating it.

Taking the *Mona Lisa* and other works of fine arts as universals offers us the chance to give a simple and *uniform* answer to the ontological question: what kind of things are works of art. In the case of musical works and literary works, universals seem the natural choice. The *same Mondscheinsonate* is played by Evgeny Kissin in Budapest at the Palace of Arts, and played by Yundi Li in London at the Royal Albert Hall. The *same Chanson d'automne* is read by kids in the schools from Paris to Marseille. This sameness can simply be accounted for in terms of the *Mondscheinsonate* and the *Chanson d'automne* as two universals each wholly present at several places (potentially) at the same time. We do not shun a similar explanation of the sameness of the *Mona Lisa* and its perfect forgery. When you stand before the *Mona Lisa* in the Louvre, and when someone else stands before its perfect forgery some place else, the two of you are standing before the *very same work of art*, because the two paintings are one and the same universal wholly present in the Louvre and some place else at once.

There are, of course, many alternative ontologies of art (see Thomasson 2004), but none of them is *prima facie* better than our universalism. To mention but a few. If works of art were mental entities, as Collingwood had suggested (Collingwood 1938), then the *Mona Lisa* would be a shadowy private picture,

everyone having her own in her head, each private picture closely matching the distribution of color patches on the public canvass without being identical to the public picture. This is no less counterintuitive than what we say. Nor are we better off with the suggestion made by Currie (1989) that works of art are event types or action types, because in our commonsensical moments we would surely resist the thought that the *Mona Lisa* is the sort of thing like getting on a bus or taking a sip from a cup of coffee. And finally, we find no less odd the widely shared view that works of art are abstract entities, which implies, among other things, that the *Mona Lisa* is like the number ϖ or the square root of 2.

Anyway, the argument from perfect forgery is based on the contentious assumption that a thing cannot be a work of art if it is not original but fake. This is also reflected in the near zero market value of fakes (once the forgery is discovered) and the big money market of originals. But it is just a contingent fact that our culture is so obsessed with originality and assumes that what is not original is not art. Things could have obviously gone differently. What is more to the point, however, is that originality is not an ontological category, but a historical one. It has nothing to do with the metaphysics of artworks. Originality is the business of art historians, gallery owners and art dealers, not metaphysicians.

2.2. The gorilla's painting

The argument from the gorilla's painting runs as follows. It is possible, no matter how improbable, that a gorilla in the zoo has inadvertently made an exact replica of the *Mona Lisa*. In this counterfactual scenario, the Mona Lisa and the gorilla's painting are *indiscernible counterparts*, they have the exact same perceptual properties. But while the *Mona Lisa* is a work of art, the gorilla's painting is *not*. For the *Mona Lisa* has resulted from a *deliberate creative artistic intention* whereas the gorilla's painting surely did not. Therefore, it is not true that every (possible) duplicate of an artwork is also an artwork. Consequently, it is not true that a work of art is what it is in virtue of its perceptually accessible intrinsic features.

The argument is based on the assumption that only such things can be works of art that have resulted from some deliberate creative artistic intention. That is, that the extrinsic property of being created by deliberate creative artistic intention is a *necessary condition* for being a work of art. We think that this assumption is false or at least questionable.

Let's use intuition pumps. What would you say if it turned out that Leonardo made the *Mona Lisa* while he was dreaming? Or in some other incontrollable state, under hypnosis or under the influence of some drugs? It is uncertain whether we can still speak about a deliberate creative artistic intention in these cases. Yet would you hesitate to say that the *Mona Lisa* is a work of art? What if it turned out that Leonardo made the *Mona Lisa* in pitch dark or in a temporally

blind state? In that case, the Mona Lisa would have been created randomly, much the same way as if the brushes fell on the canvass randomly during an earthquake, miraculously creating a painting just like the *Mona Lisa*.

Take now a case where there is definitely no deliberate creative artistic intention. Suppose that it turns out that one of your favorite readings (say, Mayakovsky's poem, the *Ленин – жил, Ленин – жив, Ленин – будет жить*) was not written by Mayakovsky, but was typed accidentally by a young chimp who sneaked into his room. Would you now say that the *Ленин – жил, Ленин – жив, Ленин – будет жить* is not a work of art, and has no place in the anthologies of Soviet literature?

Forget the monkeys. Imagine a Russell-world, indiscernible from our own, in which everything has come into existence five minutes ago (Russell 1921: 159-160), whereby none of the works of art thought to be older than five minutes have resulted from deliberate creative artistic intention. Would you dispute that in the Russell-world the paintings and sculptures in the art museums, the musical works performed in the concert halls are works of art?

Our intuition tells us that we are dealing with genuine works of art in these counterfactual scenarios despite the fact that the deliberate creative artistic intention is clearly missing. But even if your intuitions about these cases were different, you cannot deny the following. The rain dance of the Hopi Indians, the paintings of Altamira, the diary of István Széchenyi, or the letters of St. Paul are these days seen as works of art. They are the results of deliberate creative intentions, to be sure. But they are certainly *not* the results of deliberate creative *artistic* intentions. So either art historians are wrong, or else a deliberate creative artistic intention is not a necessary condition for artwork status.

2.3 Duchamp's readymade

Duchamp's readymade, the *Fountain* is an *indiscernible counterpart* of an ordinary porcelain urinal in some public toilet, they have the exact same perceptual properties. Agnes Martin's painting the *The Desert* is an *indiscernible counterpart* of a sand-colored plain canvass, they have the exact same perceptual properties (a few years back, *The Desert* cost around 4-6 million dollars, no kidding). John Cage's musical work, the *4'33''* is an *indiscernible counterpart* of a pianist's elongated preparation prior to playing her instrument—they have the exact same perceptual properties. The art world deems the *Fountain*, the *The Desert* and the *4'33''* works of art, but their indiscernible counterparts are clearly *not* works of art. Therefore, it is untrue that no (possible) duplicate of a non-artwork is an artwork. Consequently, it is untrue that a work of art is what it is in virtue of its perceptually accessible intrinsic features.

The argument from Duchamp's readymade (and similar artworks) is the inverse of the previous ones. Those arguments took some *par excellence* work of art and have assumed that it could have had an indiscernible counterpart which is not an artwork. By contrast, the present argument takes some *par excellence* non-artworks (an ordinary urinal, a plain canvass, silence for about 4 minutes and 33 seconds) and points out that they have indiscernible counterparts which in fact are works of art. Furthermore, this argument relies on hard facts, not on farfetched possibilities.

Now in response to the previous arguments, we have claimed that *both* the *Mona Lisa* and its possible indiscernible counterpart are works of art. In the present case, we say that *neither* the *Fountain* installed in Alfred Stieglitz's studio, nor its indiscernible counterpart, the ordinary urinal in a public toilet is a work of art (the same goes for the *The Desert* and the *4'33''*—none of them is a work of art). We claim that the *Fountain* and the likes are not genuine works of art. We don't deny the obvious, of course, that the art world *treats them as* works of art, but we think that all these people are wrong.

Our main reason is the following. If you hold that some works of art are indiscernible counterparts of things that are not works of art, then you must also hold that the object's *likeness* plays no role at all in whether something is a work of art or not. That is, the *object* itself plays no role at all. We do not even have to perceive it to recognize that it is a work of art! Now while this may be true of Duchamp's readymade, it is certainly not true of the *Mona Lisa*.

Imagine the following 'work of art' called the *Knight*. Take an orchestra consisting of an oboe player, a trombone player, a clarinet player, a violinist, a bassoon player, a celesta player and an organist. In front of each musician there is a board (similar to a chess board) of eight by eight squares, and each square contains the name of a tone: one-lined *C*, contra *F*, great *G*, four-lined Gis, small *Ces etc.*. The orchestra has a conductor who stands blindfolded before the musicians and points randomly to one or another musician. The musician then has to play the note in the square which is in 'L-shape' distance from the square whose note was last played. Suppose, the concert goes well (the audience endures it in silence) and the music is played for 60 minutes.

Let us ask of you. Did you have to be at that concert? Did you have to hear what the music sounded like? Or is it merely enough that we told you the punch line of this 'work of art'? We think that the essential feature of such 'works of art' is that knowing their punch line *substitutes* for their perception. Just as we do not need to hear the *Knight*, we do not need to see Duchamp' readymade. It is quite enough to know their punch line.

But this obviously isn't true of genuine works of art. No narrative can substitute the perception of the *Mona Lisa*, the *Mondscheinsonate*, or the *Chanson d'automne*. And the simple reason is that they are works of art in their own right, that is, they are works of art in virtue of what they are like.

We could go even further, and say that 'works of art' like Duchamp's ready-made, not only do not have to be perceived, but do not even have to *exist*! Fictional paintings in fictional art galleries, fictional musical plays performed in fictional concert halls, or fictional literary works read in fictional literary saloons. They could be subjected to interpretation, they could be talked about, they could be analyzed, in general they could function *as if* they were real.

Imagine the following non-existent 'work of art'. In an exhibition room there is a table (which resembles to a kitchen table) and on it there is a meat grinder which grinds little yellow rubber duckies one after the other. Its title is the *Fukuyama's Mistake*.

An art critic, for instance, could write the following upon hearing of it: „According to Fukuyama history has ended, because everywhere in the world liberal democracy has prevailed. Of course this claim is contestable, but the *Fukuyama's Mistake* clearly shows that history is not over yet, because animals are massacred scot-free by the human race. The *Fukuyama's Mistake* draws attention to the fact that history will end only when animal rights are fully acknowledged and respected world-wide."

3. SOME META-REFLECTIONS

It is usually said that what makes it so difficult to give a philosophical account of artworks is the fact that certain artists (indeed the greatest ones) create works of art with the express intent to blow up the actual conceptual framework within which we think of artworks. We are not denying that. Impressionism blew up the confines of realism, dodecaphony blew up the confines of classical harmony theory, and so on. Nonetheless what they have created were such that no amount of punch line-knowing could substitute their perception. With Duchamp's readymade, however, it is not only the case that it blows up the way we traditionally think about artworks, but it is passed *as* a work of art which is *sui generis not*.

In our paper we have expounded and defended a theory of artworks which takes *at face value* our philosophically innocent, commonsensical beliefs about artworks. Namely, that works of art are what they are because of what they are like. In this respect what we do is very similar to what the disjunctive theorist does who defends our naive convictions about perception (see Martin 2004). Instead of saying that there must be some *common factor* between Duchamp' readymade and the *Mona Lisa* in virtue of which they *both* are works of art, which has nothing to do with perceptual-intrinsic features, we said that they belong to *different* ontological types (as according to the disjunctive theory hallucinations are a different type of mental event than the appropriate veridic perceptions which are indiscernible from them). We think that the acceptance of any view

opposite to our's means the renounciation of our commonsensical beliefs about artworks, and so all such views are *error*-theories.

To wit: we have tried to do justice to the layman's intuition, who when confronted by a readymade in a musem, groaned so – this ain't no work of art.

ACKNOWLEDGEMENT

REFERENCES

Collingwood, Robin George 1938. *The principles of Art*. London, Oxford University Press.
Currie, Gregory 1989. *An Ontology of Art*. New York, St.Martins Press.
Danto, Arthur C. 1981. *The Transfiguration of the Commonplace. A Philosopy of Art*. Cambridge, Mass., Harvard University Press.
Dickie, George 1974. *Art and Aesthetics. An Institutional Analysis*. Ithaka, Cornell University Press.
Fish, Stanley 1980. *Is There A Text in This Class? The Authority of Interpretative Communities*. Cambidge, Mass., Harvard University Press.
Kant, Immanul 1790/1987. *Critique of Jugdement* (trans. Werner Pluhar). Indianapolis, Hackett.
Lewis, David 1986. *On the Plurality of Worlds*. Oxford, Basil Blackwell.
Martin, Michael 2004. The Limits of Self-Awareness. *Philosophical Studies* 120, 37-89.
Plato 2000. *The Republic* (trans. Tom Griffith). Cambridge, Cambridge University Press.
Russell, Bertrand 1921. *Analysis of Mind*. London, George Allen and Unwin.
Thomasson, Amie L. 2004. The Ontology of Art. In: Peter Kivy (ed.): *The Blackwell Guide to Aesthetics*, Oxford, Blackwell, 78-92.

Thomas M. Crisp

Temporal Passage: A Shape-Dynamic Account

1. INTRODUCTION

Time passes: one of the many banal facts about time which turns out to be extremely unfriendly to philosophical analysis. Time passes, but what that comes to, no one can say. And why time passes, that's even harder.

I'll do three things in this essay. First, I'll try to shed some light on what the passage of time *is* (which should not be confused with the project of spelling out what the expression 'time passes' *means*; I don't know what the expression means, but I have a favorite theory about what underlying reality makes true our talk about temporal passage). Second, I want to say something about *why* time passes: how the passage of time works. Here I'll draw on recent work by Julian Barbour on shape-dynamic approaches to general relativity. And third, I want to briefly examine J.J.C. Smart's well-known objection to the idea of temporal passage from the perspective afforded by my theory.

A disclaimer: these are tall aims for a short essay. What temporal passage is, why it happens, etc.: these are topics for a book or two. How to say something substantive about them in the space of an essay? By painting in broad strokes, being somewhat impressionistic. That's what I propose to do here: to sketch out in a broad, impressionistic way how to think about these things, leaving the task of filling in the details for another time.

2. TEMPORAL PASSAGE: WHAT IT IS

Everyone believes in temporal passage: it's a datum; one of those features of experience every temporal ontology has to account for. There's a divide among temporal ontologies as to how to characterize it. Those committed to the so-called A-theory—sometimes also called the dynamic view of time—hold that temporal passage is in some sense fundamental, incapable of explanation in more basic terms. Those committed to the so-called B-theory—sometimes called static views of time—hold that temporal passage is not fundamental, that it is explainable in more basic terms that make no mention of any sort of temporal passage.

That's a rough way of characterizing the divide, anyway. Characterizing the divide less roughly turns out to be extremely difficult, as Dean Zimmerman showed in his, "The A-Theory of Time, the B-Theory of Time, and 'Taking Tense Seriously'" (2005). You might have thought you could characterize the difference between the views in terms of tense: the A-theorists are those who "take tense seriously" and the B-theorists don't. But if by that you mean that the A-theorists are those who think that the objects of propositional attitudes—the propositions—are things that change their truth value over time and are governed by the logic of a Prior-style tense logic, and the B-theorists are those who deny this, holding that all propositions have their truth values eternally, you've this problem: David Lewis held the former view, but he most certainly wouldn't be one you would associate with an A-theory of time.

Some try to characterize the difference in terms of a special property of presentness, a property which is both fundamental—perfectly natural, "carving at the joints"—and also moves successively through the B-series, where its passage across the series comprises temporal passage. A-theorists are those who think there is such a property as this fundamental presentness; B-theorists think not. The trouble with this way of carving the territory is two-fold.

First, as Zimmerman shows, B-theorists à la Lewis can accept the existence of a primitive property that is very difficult to distinguish from the A-theorist's presentness. The property *being self-simultaneous*—the property had by something x iff x is simultaneous with itself—the B-theorist might argue, is both fundamental, and such that only one time has, in the present tensed sense of 'has', the property. A-theorists fond of characterizing their theory in terms of fundamental presentness will likely object that presentness is *not* the same thing as self-simultaneity, but saying exactly what the difference comes to turns out to be extremely difficult.

Second, the view I'll endorse below eschews any appeal to fundamental presentness; I don't believe in such a property. But, since, so I think, my view deserves the label 'A-theoretic' as well as any, characterizing the A-/B-theory divide in terms of fundamental presentness is an infelicitous way of characterizing the divide.

Zimmerman thinks the best way of characterizing the A-/B-theory distinction is in terms of the notion of truth *simpliciter*. A-theorists are those who think that what is true *simpliciter* changes over time; B-theorists deny this: though some—the serious tensing B-theorists—hold that the objects of our propositional attitudes change truth value over time, they'll analyze this in terms of dyadic truth-at-a-time, a two-term connection between proposition and time. My belief that I am standing, they'll say, expresses a proposition that is true relative to some times but not to others, where this is a matter of bearing the truth-at relation to some times but not others. Not so, says the A-theorist: my belief that I am standing expresses a proposition that is true *simpliciter*—*that I am standing* is true

relative to various times, to be sure, but is also true *simpliciter*, true period, true full stop. Since it was false *simpliciter*, it's an example of a proposition which is true *simpliciter* but not immutably so. Such, then, is the difference between A- and B-theorists: the former accept and the latter deny that what is true *simpliciter* varies over time.

So far, we have seen Zimmerman's preferred way of characterizing the A-/B-theory divide. Two potential costs for this way of characterizing it, the first of which Zimmerman mentions: First, if you're attracted to a deflationary view of truth, you won't much like this way of characterizing the divide. If there is no such property as truth *simpliciter*, this way of spelling it out is a non-starter. Second, if you're skeptical, as I am, about the existence of propositions—abstract pieces of information encoded or expressed by beliefs, sentences, etc.—then, here again, you won't much like the proposal.

A related suggestion which gets around these two worries postulates the existence of events or states of affairs of the Armstrong/later-Chisholm sort. On this view, necessarily, for any *x* and *y* and relation *R*, *x* bears *R* to *y* iff there is the event (or in Armstrong's language, "state affairs", but I'll stick with event talk for concision) *x*-bearing-*R*-to-*y*, a non-mereological fusion of *x*, *y* and *R*, whose existence depends on the existence of *x* and *y* and the instantiation of *R* by *x* and *y*.

That there are such things as these events is contentious. A reason for thinking there are, alongside the usual ones put forward by Armstrong, Chisholm *et al.* is their usefulness in characterizing the dispute between A- and B-theorists of time, a dispute which intuitively makes sense but is extremely difficult to characterize without adverting to events.

With events in hand, though, the dispute is easily described. Above we saw the A-theorist holding that what is true *simpliciter* changes over time. Here is a variation on that theme: what exists changes over time. More exactly: which events exist changes over time. Or, in terms of quantifiers and tense operators— reading the quantifiers here as unrestricted, ranging over everything whatsoever, ignoring nothing—we may understand A-theorists as those committed to this:

> *Thesis of Temporal Passage*: It is always the case that, for some event *x*, either WAS(for every *y*, *y* is, was and will be numerically distinct from *x*) or WILL(for every event *y*, *y* is, was and will be numerically distinct from *x*).

B-theorists deny the Thesis of Temporal Passage. They'll think it always the case that, for any event *x* you pick, WAS(something will be identical with *x*), or WILL(something was identical with *x*).

So the distinction between A- and B-theories of time may be thought of thus: the A-theorist accepts whereas the B-theorist denies the thesis of temporal passage. Since commitment one way or the other on the Thesis of Temporal Pas-

sage is compatible with any number of views about truth and propositions, we have a way, then, of characterizing the A-/B-theory divide that swings free of deflationism about truth and propositions.

All to the good. We've also a way, then, of answering our question: what temporal passage is. At least, we have a way of answering that question from the perspective of the A-theory, which is the perspective I shall presuppose in this essay.

What is temporal passage, then? It is the coming into being or ceasing to be of events. (It's something like C.D. Broad's *becoming* (e.g., 1923: 66-67), though not exactly that since he thought of the passage of time as involving the coming into being of events, not the ceasing to exist of any events.) In terms of quantifiers and tense operators, taking our quantifiers as unrestricted, temporal passage is a matter of there being some event *x* such that WAS(for no *y* is it the case that *y* is, was or will be identical with *x*) or WILL(for no *y* is it the case that *y* is, was or will be identical with *x*).

3. TEMPORAL PASSAGE: HOW IT WORKS

Next I want to consider the question how temporal passage works. Why does it happen? Why do events come into being and cease to be?

Here is an attractive, if uninformative, picture.

Things have powers. A bit of copper has the power to expand when heated, a stick of dynamite has the power to explode when put into the right conditions, and so forth.

It is an attractive thesis, so I think, that to have a power is to stand in a fundamental, multigrade relation of things to universals. So suppose we have some object *o* with the power to instantiate a property *A* on the condition that it instantiate the property *B*. This, I propose, is a matter of *o*'s bearing a fundamental power relation to the universals *A* and *B*. Using a property abstraction operator similar to those deployed variously by George Bealer (e.g. 1982), we can put it like this: *o*'s having the power to instantiate *A* on the condition that it instantiate *B* is a matter of its being the case that $[Bx \Rightarrow Ax]o$, where we read this as '*o* is an *x* such that *x* has the power to instantiate *A* on the condition that it instantiates *B*, and '$[\_\_ \Rightarrow \_\_]\_\_$' expresses our fundamental power relation. (Why complicate things thus with property abstraction? It'll aid concision below.)

Or: Suppose *o* has the power to instantiate *A* on the condition that it instantiate *B* and *C*. This iff $[Bx \& Cx \Rightarrow Ax]o$.

Or: Suppose o_1 and o_2 jointly have the power to instantiate a relation *A* on the condition that they instantiate the relation *B*. This iff $[Bxy \Rightarrow Axy]o_1,o_2$. And so forth.

This, I say, is an attractive if uninformative picture. It's attractive because it construes power talk as fundamental, as carving at nature's joints. It's uninformative for the same reason: explanation of the fundamental is perforce limited.

Grant me the picture and let us see what it suggests about temporal passage. Suppose there is an x such that (a) x lacks the property A, (b) x has the power to instantiate A on the condition that it instantiate B, and (c) x instantiates B. Then by dint of its having this power and being in what we might call the "triggering condition" of the power, x will exercise its power and jump to a state in which it instantiates A. Likewise with relational powers: Suppose there is an x and y such that (a) x does not bear A to y, (b) $[Bxy \Rightarrow Axy]x,y$, and (c) x bears B to y. Then by dint of x and y jointly possessing this power and being in the power's triggering condition, x and y will exercise their power and jump to a state in which x bears A to y.

When x exercises a power and jumps to a state in which it comes to possess a property A it didn't previously possess, a new event, x-being-A, comes into existence. When that happens, there is an event x-being-A such that, it was the case that, quantifying unrestrictedly, nothing is, was or will be identical with x-being-A. When that happens, time passes.

When x and y jointly exercise their power to jump to a state in which x bears A to y, a relation they didn't previously bear to one another, a new event, x-bearing-A-to-y, comes into existence. When that happens, there is an event x-bearing-A-to-y such that, it was the case that, quantifying unrestrictedly, for no y is it the case that y is, was or will be identical with x-bearing-A-to-y. When that happens, time passes.

So we get this picture of time's passage: at present, reality comprises a large number of particulars, universals, and the events they constitute. Many, many of the particulars comprising reality have powers to jump to new states and are in states sufficient to trigger those powers. As those powers are triggered, the particulars possessing those powers jump to new states, thereby bringing into being new events. At this point, new powers are triggered, resulting in a jump to new states yet, which give rise to further triggering of powers, and so on. Such jumping of things from state to state, thereby successively bringing into being new events, followed by further new events, and so forth, all driven by the exercise of the powers of things, we call the flow or passage of time.

It's a helpful bit of picture thinking to conceive of the phenomenon of temporal passage as something akin to the popping of popcorn. The popcorn kernels each have the power to pop at thus-and-such triggering temperature. A kernel hits its triggering temperature, exercises its power to pop, and jumps to a new state. Another hits its triggering temperature, exercises its power, and jumps to a new state. Another, and another, and another, …. Such is the flow of time: the constant popping of things (points of space, perhaps, or their point-sized

matter/energy constituents) from one state to another as power after power is triggered.

So far, then, a powers account of temporal passage, painting in very broad strokes. Let me next try to fill the picture in some by describing a way of thinking about relativistic physics that fits nicely with the picture.

4. TEMPORAL PASSAGE AND "SHAPE DYNAMICS"

Julian Barbour and collaborators have been working on a Machian view of particle and geometry dynamics for the last decade or so. Ernst Mach was famously suspicious of Newton's absolute space and time on epistemological grounds: they can't be observed, so we have no good reason to postulate them. A major difficulty in realizing Mach's empiricist scruples in our physical theorizing was that it turned out to be extremely difficult to recast Newtonian particle dynamics in a form that makes no appeal to absolute space and time. Newtonian dynamics was eventually superseded by general relativistic dynamics, much of it inspired by Machian ideals, but even here, Einstein ends up postulating an unobservable background space (spacetime) and a fundamental temporal metric (proper time along timelike trajectories), neither of which fits well with Machian scruples.

Barbour (together with his collaborator, Bruno Bertotti) has the distinction of being the first in the history of physics to show how to reconstrue Newtonian particle dynamics in fully Machian form: no appeal to an invisible Newtonian container space (to define inertial motion) or invisibly flowing Newtonian time (to define a temporal metric) (1982). The temporal metric postulated as primitive in Newton's dynamics turns out on Barbour and Bertotti's dynamics to be definable from more fundamental quantities in the theory, as a measure of change in those quantities over time (a useful measure of that change, it turns out, because it yields the simplest mathematical description of the dynamics of that change). Time, in Aristotle's famous phrase, is just a measure of change.

In more recent papers (e.g., 2010), Barbour and collaborators show how to extend the approach to General Relativity (GR). On the standard formulation of GR, the basic equations of the theory are Einstein's field equations, which describe the distribution of metric and matter fields across a four-dimensional spacetime. A non-standard way of thinking of GR is "geometrodynamics", where this is a matter of reconstruing GR as a dynamical theory describing the evolution of three-dimensional geometry over time.

Barbour *et al.* develop a version of geometrodynamics on which the fundamental law governing the evolution of 3-space over time is an action principle determining geodesics through a configuration space, each point of which corresponds to a possible conformal 3-geometry of a closed 3-space, and geodesics through the configuration space corresponds to dynamically possible histories of

an evolving 3-space. The resulting theory corresponds closely to General Relativity (but isn't General Relativity: possible histories in Barbour's theory correspond to only the CMC-foliable models of GR, a subset of the full set of general relativistic spacetime models). So far anyway, it corresponds closely enough to General Relativity as to satisfy all current experimental verifications of GR.

It's an extremely interesting theory of gravity, for several reasons. First, local Lorentz invariance of non-gravitational interactions turns out to be a consequence of the action principle at the heart of the theory (as opposed to the usual approach to GR, according to which the validity of special relativity in local inertial frames is an independent assumption). Second, as with Barbour's Machian Newtonian dynamics, temporal metric (in this case, infinitely many local temporal metrics along timelike trajectories) turns out to be non-fundamental, definable from more fundamental quantities in the theory. And third, spatial metric—the measure of spatial distance—also turns out to be non-fundamental, definable from the dynamics of the theory. (The fundamental geometrical facts, on the theory, are conformal: facts having to do with angles between trajectories in space. It's a theory, then, about the evolution of conformal 3-geometry over time. Hence the name he gives it: shape dynamics.)

There's much to like about the theory in terms of unity and economy: for minimal cost in ontology (no primitive temporal or spatial metric), you get local Lorentz invariance and all known experimental consequences of general relativity. It's a neat theory.

Assume for discussion that it's true and let us consider what it suggests about the picture of temporal passage I have been sketching. The picture, again: The many, many particulars comprising reality have powers to jump to new states and are in states sufficient to trigger those powers. As those powers are triggered, the particulars possessing those powers jump to new states. At this point, new powers are triggered, resulting in a jump to new states yet, which gives rise to further triggering of powers, and on and on. Such constant change—such constant jumping from state to state owing to the exercises of the powers of things—we call the flow or passage of time.

Reflection on Barbour's shape dynamics suggests an interesting development of the account. At the heart of Barbour's shape dynamics is an action principle describing dynamically possible histories of an evolving 3-space. Now, proponents of powers theories of causation will sometimes say that laws of nature should be thought of as descriptions of powers. Let me suggest that that's how we think of the action principle at the heart of Barbour's shape dynamics: as a compendious mathematical description of the powers possessed by Space and its constituent points. (Space: the three-dimensional, enduring container of all mass/energy, whose constituent points are linked by spatial distance relations, whose geometry is variably curved depending on the distribution of mass/energy, and whose geometry changes over time as it lapses through successive

jumps in state. I am assuming, notice, a three-dimensional, as opposed to a four-dimensional, view of the spatiotemporal world. I assume that *all* mass/energy, quantifying unrestrictedly, is housed in an enduring 3-manifold structured by relations of spatial distance as opposed to the usual 4-manifold structured by the spacetime interval. In a word, I assume presentism. See, e.g., my 2003 for further explanation and defense of presentism.)

Perhaps it works like this. Let the *x*s be all and only the points of Space. Then perhaps the *x*s jointly instantiate various power relations: $[R_1xs \Rightarrow R_2xs]xs$, $[R_2xs \Rightarrow R_3xs]xs$, $[R_3xs \Rightarrow R_4xs]xs$, and so forth, where R_1, R_2, ..., we may suppose, are conformal geometrical properties like those described by Barbour's theory. These powers specify that when the *x*s comprising Space are such that R_1xs, they'll jump to a state in which R_2xs, and that when in that state, they'll jump to a state in which R_3xs, and so forth, with the upshot that these powers specify that Space will traverse some one of the extremal curves through a configuration space of Shape Dynamics. Barbour's action principle is just a handy mathematical description of how Space behaves under the action of these powers.

Suppose so. Then two interesting consequences. First, as Space jumps from state to state under the guiding influence of these powers, it successively lapses along a geodesic of a configuration space of Shape Dynamics, and thus successively lapses along a CMC-slicing of some general relativistic spacetime model. All that to say: as Space and its contents lapse from state to state, they behave just as General Relativity predicts: clocks move slower near massive objects; clocks move at different rates in relative motion; massive objects curve the Space around them; gravity waves propagate through Space; and so forth.

Secondly, as Space lapses from state to state under the guiding influence of these powers, there is no fundamental temporal metric measuring its lapse. General relativity's local proper time emerges from more fundamental quantities in the theory as a useful measure of changes in the conformal geometry of Space over time, useful because it yields the simplest mathematical description of that change. But the trajectory-relative temporal metrics of General Relativity aren't fundamental: they don't carve nature at its joints. They're one among infinitely ways to parameterize changes in Space over time, distinguished only in that they enable us to formulate the laws governing the evolution of Space over time more simply than alternatives.

Other ways of parameterizing are, from the standpoint of metaphysics and its attempt to describe fundamental structure, just as correct—and more useful in everyday life to boot. So there's measuring change by solar time, assigning a measure of one solar day to the quantity of change transpiring in some system per rotation of the earth around its axis. There's measuring change over time by ephemeris time, where this is the timescale such that the laws of motion describing the sun and planets in our solar system are approximately those given by Newton. There's reference-frame dependent timescales: time as measured

by a cesium-133 clock in thus-and-such state of motion. All are equally correct ways of measuring cosmic and local change over time. None is fundamental; none carves at the joints. Some make for simpler mathematical description of the change of physical quantities over time, but all are simply measures of change.

5. SUMMARIZING

Taking Barbour's shape dynamics on board, then, here, I want to suggest, is an attractive account of temporal passage:

(1) Individual objects have powers to jump to new states when in certain monadic triggering conditions. Multiple objects jointly have powers to jump to new states in certain polyadic triggering conditions.

(2) Having such monadic and polyadic powers is a matter of entering into a fundamental, multigrade power relation between things and universals.

(3) Things exercise their powers in their triggering conditions, thereby jumping to new states and bringing into being new events: events such that, quantifying unrestrictedly, it was the case that nothing was identical with that event. These new events put their subjects into new triggering conditions, which gives rise to further exercise of powers, which brings into being further new events, which puts their subjects into new triggering conditions, yields further exercise of powers, further new events, and so forth. We call this successive coming into being of events, resulting from successive exercise of powers, the flow or passage of time.

(4) There is such a thing as Space—the manifold of points-at-a-time in which all matter/energy is housed. It is an enduring, three-dimensional space whose curvature varies with distribution of mass/energy and over time. The totality of its constituent points are linked by polyadic power relations which specify certain conformal geometric properties as triggering and manifestation conditions. These power relations are elegantly described by the action principle at the heart of Barbour's shape dynamics.

(5) As the points of Space jointly exercise these powers, they lapse successively through the configuration space of shape dynamics, which corresponds to a CMC-slicing of some general relativistic spacetime model. Wherefore matter and energy behave in accord with the dynamics predicted by general relativity: local Lorentz invariance, gravitational time and length dilation, bending of light, etc.

(6) There is no fundamental temporal measure of this lapse. The usual, trajectory-dependent, relativistic temporal metric emerges from more fundamental quantities in shape dynamics as the simplest way of representing its laws.

Such, in short, is my shape-dynamical account of temporal passage. I close with a brief discussion of its bearing on a classic objection to A-theoretic accounts of temporal passage.

6. SMART ON THE RIVER OF TIME

A classic objection to A-theoretic ways of thinking about temporal passage was first introduced into the philosophical literature by C.D. broad, and was famously defended by J.J.C. smart in his 1949 paper, "The River of Time." The heart of the argument is contained in this passage:

> If time is a flowing river we must think of events taking time to float down this stream, and if we say 'time has flowed faster to-day than yesterday' we are saying that the stream flowed a greater distance to-day than it did in the same time yesterday. That is, we are postulating a second time-scale with respect to which the flow of events along the first time-dimension is measured. 'To-day', 'to-morrow', 'yesterday', become systematically ambiguous. They may represent positions in the first time-dimension, as in 'to-day I played cricket and to-morrow I shall do so again', or they may represent positions in the second time-dimension, as in 'to-day time flowed faster than it did yesterday'. Nor will it help matters to say that time always flows at the same rate. Furthermore, just as we thought of the first time-dimension as a stream, so will we want to think of the second time-dimension as a stream also; now the speed of flow of the second stream is a rate of change with respect to a third time-dimension, and so we can go on indefinitely postulating fresh streams without being any better satisfied. (1949: 484)

How shall we think about this argument? how does it go exactly?
I think the idea is something like this. a river flows through a given area, you might think, only if there is some rate at which the water of the river is passing. Likewise, Smart seems to be thinking, with time: if it flows or passes, there must be some rate at which it flows or passes. As a first premise for the argument, then, we have something like

(1) Time flows or passes only if there is some rate at which it flows or passes.

Suppose this is so; suppose time passes only if there is some rate at which it passes. What rate might that be? At what rate would time pass? Ordinarily, we think of rate as the ratio between some bit of change and a period of time over which that change occurs. So there is the ratio between the change in someone's heart over a period of time (measured in number of beats, say) to some period of time over which that change occurs (measured in minutes, say), arriving at

a heart rate of *x* beats per minute. Or, there is the ratio between the change in someone's position over a period of time (measured in meters, say) to the period of time over which that change occurs (measured in seconds, say), arriving at a rate of *x* meters per second.

The rate at which time passes, then, would be a ratio between the change in time over some period of time to the period of time over which that change occurs. Here one begins to see the problem, for the change in time over a period of one second, say, is one second, and the period over which that change takes place is, well, one second, arriving at a rate of change for time of one second per second. There is something odd about that rate, to be sure. At the very least, it's uninformative to be told that time advances at a rate of one second per second. Smart seems to be thinking there is something nonsensical about it, something incoherent:

> A connected point is this: with respect to motion in space it is always possible to ask 'how fast is it?' An express train, for example, may be moving at 88 feet per second. The question, 'How fast is it moving?' is a sensible question with a definite answer: '88 feet per second'. We may not in fact know the answer, but we do at any rate know what sort of answer is required. Contrast the pseudo-question 'How fast am I advancing through time?' or 'How fast did time flow yesterday?' We do not know how we ought to set about answering it. What sort of measurements ought we to make? We do not even know the sort of units in which our answer should be expressed. 'I am advancing through time at how many seconds per ____?' we might begin, and then we should have to stop. What could possibly fill in the blank? Not 'seconds' surely. In that case the most we could hope for would be the not very illuminating remark that there is just one second in every second. (1949: 485)

Smart is thinking, then, that there is something unhappy about the suggestion that time passes at a rate of one second per second: it's a "not very illuminating" answer to the "pseudo-question" 'How fast does time pass?'.

Well, to be sure, it's not a terribly illuminating answer to this question, but why is that a strike against it? What's wrong with the answer (and with the question)? I'm not sure. Peter van Inwagen (2009) and Eric Olson (2009) have argued that what's wrong with it is that one second per second is not a genuine rate, since one second divided by one second is just one, which isn't a rate. I'm not sure this argument works, but grant it for now; as we'll see, not much hangs on it.

Supposing, then, that it is somehow objectionable to think of time as passing at a rate of a second per second, if you're committed to the idea that time passes and to (1), the idea that it passes only if there is some rate at which it passes, you might think the only coherent way to talk about the rate at which time passes is to postulate some second time scale in terms of which one can describe change in time as featured in the original scale. Were there some such second time

scale, one could say that the rate at which time passes is the ratio between the amount of first-time-scale time over some period to the amount of second-time-scale time over that same period. If you thought all this made sense, you might accept something like

> (2) There is a rate at which time flows or passes only if there is a second time scale (distinct from the first time scale we normally use), in terms of which the rate of time measured by the first scale can be described.

Next, Smart suggests that if we think of some period of temporal passage as measured by our newly introduced second time-scale, the question arises afresh: how fast did that period of second-time-scale time pass? There'll be some rate at which it passed. So:

> (3) If there is a second time scale measuring temporal passage, then there is some rate at which periods of time as measured by our second time scale pass.

But if so, then, says Smart, we'll need some third time scale (distinct from the first- and second- time scales) in terms of which to describe that rate, which sets up a regress:

> (4) If there is a rate at which periods of time as measured by our second time scale pass, there is a third time scale in terms of which the rate of time measured by the second scale can be described, and a fourth in terms of which the rate of time measured by the third can be described, and so forth.

And finally, Smart thinks this regress vicious ("and so we can go on indefinitely postulating fresh streams without being any better satisfied"), suggesting as the final step in the argument something like

> (5) The regress in (4) is impossible,

and the conclusion that

> (6) It is not the case that time flows or passes.

Such, I think, is the gist of Smart's argument that time does not flow or pass. If it's right, my above-adumbrated account of temporal passage is wrong, since, says my account, time flows or passes. What's to say?

6.1 Smart's Argument Examined

To start with, as we have seen, Smart presupposes it nonsensical or incoherent to suppose that time flows at a rate of one second per second. But it's exceedingly difficult to see why that would be. Other than that it's trivial or uninformative, there is nothing incoherent or nonsensical about the suggestion that time flows at a rate of one second per second, as has been nicely argued in a recent paper on this point by Hud Hudson, Ned Markosian, Ryan Wasserman and Dennis Whitcomb (2009). That being so, one could resist Smart's argument by rejecting premise (2) on grounds that we needn't appeal to multiple time scales to make sense of the idea that there is rate at which time flows.

But I'm inclined to accept premise (2), much as I agree with Hudson *et al.* that there is nothing problematic about time's flowing at a rate of one second per second. According to my above-adumbrated account, recall, there is no fundamental temporal metric, no fundamental measure of temporal duration. Say I: there are as many measures of temporal duration as there are measures of change. There are measures of temporal duration associated with sidereal timekeeping; measures associated with ephemeris timekeeping; general relativistic, trajectory-dependent measures of duration; and infinitely more besides. No one way of measuring temporal duration is fundamental; all are equally correct.

That being so, premise (2)'s claim that there is a rate at which time passes only if there is a second time scale (distinct from the first-order time scale) in terms of which the rate of time's flow on the first scale may be measured is perfectly correct. Suppose Space evolves through a sequence of changes during which the earth turns on its axis by exactly 15°. Measured in terms of a sidereal timescale, we'd say that period of time lasted an hour. In terms of Newtonian time, that same sequence of changes will have lasted an ephemeris hour, something just shy of a sidereal hour (or just over; I'm not sure which). Temporal flow, during that period, would be moving, then, at a rate of 1 sidereal hour per 1.01 (let us suppose) ephemeris hours.

This will hold for any period of temporal evolution you pick: for any timescale t_1 you pick, you'll be able to find some distinct timescale t_2 such that temporal passage flows during that period at a rate of t_1 units of time per t_2 units of time. Premise (2) is correct.

Premise (3) looks good, too: True enough, if there is a second time scale measuring temporal passage, then there is some rate at which periods of time as measured by our second time scale pass.

Premise (4) likewise seems fine. Again:

(4) If there is a rate at which periods of time as measured by our second time scale pass, there is a third time scale in terms of which the rate of time measured by the second scale can be described, and a fourth in terms of which the rate of time measured by the third can be described, and so forth.

This is plausible. A temporal metric, from our present perspective, is a conventional measure of change. Since there are infinitely many ways of measuring change, it's plausible that there'll be a well-ordering of temporal metrics of the sort envisaged by Smart. Take, for example, the metric you'd measure with a cesium-133 clock in orbit around the sun at a distance of 1 million miles, the metric you'd measure with a cesium-133 clock in orbit around the sun at a distance of 1 half million miles, the metric you'd measure with a cesium-133 clock in orbit around the sun at one-half that distance, and so forth. Given General Relativity (and shape dynamics too), these yield Smart's well-ordering.

So (1) through (4) of Smart's argument look pretty good. Not so with (5). (5) says that the regress suggested by (4) is vicious. But, manifestly, it isn't. The regress follows from the fact that duration is a conventional measure of change, of which there are infinitely many. Nothing vicious about that. And if so, Smart's argument, interesting and important as it is, makes no trouble for my suggested account of temporal passage.[1]

REFERENCES

Bealer, George. 1982. *Concept and Quality*. Oxford: Clarendon Press.

Barbour, Julian and Bruno Bertotti. 1982. "Mach's principle and the structure of dynamical theories." *Proceedings of the Royal Society* A 382: 295.

Barbour, Julian and Niall O'Murchada. 2010. "Conformal superspace: the configuration space of general relativity." arXiv:1009.3559.

Broad, C.D. 1923. *Scientific Thought*. London: Kegan Paul.

Crisp, Thomas M. 2003. "Presentism." In Michael J. Loux and Dean Zimmerman, eds., *The Oxford Handbook of Metaphysics*. Oxford: Oxford University Press, pp. 211-245.

Hudson, Hud and Ned Markosian, Ryan Wasserman, and Dennis Whitcomb. 2009. "The Rate of Passage: Reply to van Inwagen and Olson." So far unpublished, I think.

Olson, Eric. 2009. "The Rate of Time's Passage". *Analysis* 61:3-9.

Smart, J.J.C. 1949. "The River of Time." *Mind*, 58: 483-494.

Van Inwagen, Peter. 2009. *Metaphysics*, 3rd edition. Colorado: Westview Press.

Zimmerman, Dean. 2005. "The A-Theory of Time, the B-Theory of Time, and 'Taking Tense Seriously'". *Dialectica* 59(4): 401-457.

Ferenc Huoranszki

Powers, Dispositions, and Counterfactual Conditionals

We often say that persons had, have, or will have the power to do certain things. But do we have reasons to ascribe powers to inanimate objects as well? And if we do, is there any difference between ascribing a power and understanding what an object is disposed to do? Are objects' powers dispositions in this sense? In this paper I shall argue that we need to distinguish powers from dispositions for certain theoretical purposes. Most 'disposition terms' in ordinary language do not express causal powers; and many powers cannot be expressed by a conventional disposition term. It is true that when we say that objects are disposed to do this or that, powers are involved. But the converse does not hold because having a power does not entail that objects are disposed to do or to act upon others in certain ways.

The expression 'disposition term' as it is used in contemporary philosophy was introduced by Carnap in the early 20th century. Statements that contain such terms do not describe what is directly observable and hence, according to Carnap, they can have meaning only if they are logically equivalent to other statements which report observable episodes (Carnap 1936, 440). Subsequently, Nelson Goodman went on to argue that, from a logical empiricist point of view, every predicate which refers to enduring properties of objects must be dispositional. Thus all statements that ascribe non-occurrent properties to objects must contain dispositional terms like being fragile, soluble, poisonous, inflammable or soporific. And the meaning of these sentences can be understood only with reference to observable episodes which are the manifestations of these dispositions like breaking, dissolving, poisoning, burning or falling asleep (Goodman 1954, 40).

Most contemporary accounts of powers understand them as dispositions. Otherwise put, they take it that what Carnap and Goodman called disposition terms express what objects are disposed to do; and if objects have powers at all, they are none other than their dispositions. However, I shall argue that the ascription of powers and the ascription of dispositions have distinct theoretical roles. Both dispositions and powers are modal concepts. But whereas the ascription of dis-

positions presupposes some modal truth, the main function of ascribing powers is to ground such truth. Following Hume, many philosophers think that if powers do indeed have this grounding role then we ascribe them in order to ground natural necessities.[1] I shall argue, however, that the main role of properties that are powers is to identify natural possibilities.[2] We ascribe a power in order to identify what a person or an object *can* do; by contrast, we ascribe a disposition in order to express what objects or persons *tend to* do.

Further, I shall also argue that the ascription of powers grounds such possibilities through entailing the truth of certain counterfactuals; for it is with help of such counterfactuals that we can specify powers and hence identify natural possibilities. Propositions that ascribe powers or dispositions to objects are often analyzed with the help of counterfactual conditionals; I shall argue that the main theoretical function of the so-called conditional analysis is different in its application to context in which we aim to understand what objects are disposed to do and in context's in which we ascribe powers.

Traditionally, the conditional analysis is understood as an attempt to provide a semantic analysis of the meaning of disposition terms. And often, the aim of such a semantic analysis is to explain the use of disposition predicates without assuming that objects in fact have non-reducible dispositional properties. As far as the analysis of powers is concerned, however, we rely on counterfactual conditionals in order to explain the link between ascriptions of specific powers and our commitments about what can happen instead of what is going to happen. I shall show that if we understand the aim of the analysis in this way, we can defend a revised, non-reductive, but still relatively simple conditional analysis of powers. Relatedly, I shall also argue that powers are abundant in the sense that objects can have not only intrinsic and generic powers, but also extrinsic and maximally specific ones.

In one respect I propose to follow Carnap's project, even though I do not share his (anti-)metaphysical convictions, and I'm skeptical about his account of the meaning of 'theoretical terms'. When Carnap discusses the problem of disposition terms he does indeed rely on examples borrowed from ordinary language. But it is clear that he is not interested in the semantics of ordinary language. Rather, he wants to explain how we can introduce certain terms into our language in order to provide scientific explanation of certain observable

---

[1] The subtitle of Harré's and Madden's classic work on causal powers is 'A Theory of Natural Necessity'. Later dispositional essentialists like Ellis 2001 and Bird 2007 also argue that powers can make the occurrence of certain events metaphysically necessary.

[2] Although he does not use the concept of power, Goodman seems to hold a similar view (Goodman 1954, 50–54). Closer to the present paper's claim, Rom Harré writes that 'to say that a thing has a power is to say what is possible for it, for that is what it is talk of its dispositions' (Harré 1970, 101). I shall challenge the second part of this claim while fully grant the first.

phenomena. Analogously, the analysis of powers need not be understood as an exercise in the semantics of ordinary disposition terms. Like Carnap, I believe that we use power terms for theoretical purposes. Since my main interest here is metaphysics and not philosophy of science, I shall argue that the theoretical role of powers as properties is to ground natural possibilities. But metaphysics is complementary, and not contradictory, to science. Powers are needed for science as much as they are important for metaphysics.

1. DISPOSITIONS, POWERS, AND TENDENCIES

In this section I aim to show why ascribing powers and ascribing dispositions might play distinct theoretical roles. I shall argue that what philosophers since Carnap call 'disposition terms' can, but needn't, express objects' powers. Afterwards, I shall also argue that there are far more properties which are powers than what conventional disposition terms can express. The problem of powers has often been discussed by using conventional disposition predicates as examples. This strategy can easily blur the distinction between the role of power-like properties and the role of dispositions in our ontology. For even if some conventional terms can express both dispositions and powers, they do so in different contexts. Thus, as far as the metaphysics of properties is concerned, the distinction between powers and dispositions can become crucial.

According to the now standard approach to the use conventional disposition terms, ascribing a disposition is to say something about what objects having the disposition are disposed to do in certain circumstances. Consequently, as David Lewis suggests (and as most subsequent analyses agree) in order to understand ordinary disposition predicates like being fragile or being poisonous, we first need to specify their meaning by a paraphrase. For instance, to say that arsenic is poisonous is to claim that that it is disposed to poison upon ingestion and when no antidote is taken (Lewis 1997, 153). Let us call these paraphrases of conventional disposition terms explicitly dispositional locutions.[3] An explicitly dispositional locution specifies the meaning of a statement ascribing a conventional disposition term D in the sense that an object has D iff it is disposed to M in response to some stimulus S. I shall argue that to the extent that statements containing conventional disposition terms can indeed be so paraphrased, such statements follow a logic distinct from that behind statements that ascribe powers.

This is, I shall argue, the consequence of the fact that having a power to M is not the same property as being disposed to display behavior M in response to stimulus S in circumstances C. Properties as powers and properties as objects'

---

[3] Following Choi 2008.

dispositions to do certain things in various circumstances fulfill distinct theoretical roles. Tautologically, if an object or a person is disposed to *M*, it has the disposition to *M*. But it is exactly in this context in which the assumption that powers and dispositions fulfill distinct theoretical roles can become significant. For having a power to *M* does not invariably imply that something or someone is disposed to *M*. There is an important theoretical difference between the ascription of powers and claims about what things are disposed to do.

Consider Amy, who learned how to swim, and then visits the swimming pool regularly. She has also experienced and enjoyed smoking tobacco. But she has never smoked regularly. Her friend, Bob is a different character. Not only has he tried cigarettes, but he has also become a smoker. As a child, he learned how to swim and to this day, he can swim fairly well. But he cannot recall the last time he swam. The two people's behavior is apparently different, and the difference is very well captured by the fact that Amy, being a swimmer, *is disposed* to swim, but not being a smoker, she is *not disposed to* smoke. Bob, on the other hand, being a smoker, is disposed to smoke, but he is not at all disposed to swim. However, both of them have the same powers: both of them *can* swim and *can* smoke; and hence both of them have the power to swim and the power to smoke. The difference between them is that Amy is not disposed to smoke, while Bob is not disposed to swim.

Why is a difference? As Ryle has already observed, we can use dispositional predicates for many different purposes. Sentences containing such predicates can ascribe certain abilities, capacities or liabilities to certain objects or to certain kinds of objects. Thus, we often use such terms in order to ascribe active or 'passive' (i.e. reactive) powers to objects. But disposition terms can also express tendencies, habits or proneness to do certain kind of things (Ryle 1949, 131). There is an important logical relation between the ascription of powers and behavioral tendencies. A behavioral tendency presupposes some relevant power or ability. No one who is averse to cigarettes because, for instance, she is allergic to cigarette smoke can be a smoker. And no one who is unable to swim can be a swimmer. But many people who do not swim regularly can swim, and at least as many who can smoke or are capable of enjoying smoking, do not smoke. Thus, the ascription of powers does not involve any truth about habits, tendencies or behavioral regularities (Huoranszki 2011, 59).

One might object that the difference disappears if we take the proper specification of dispositions into account, according to which 'has the power to *M* at *t*' is to be specified as 'disposed to *M* at *t* in response to *S*'. When we say that Amy is disposed to swim whereas Bob is not, and this is understood as a tendency or habit, we do not mention a particular type of stimulus to which the manifestation is a response. But the dispositional property that we want to express is more specific than what the conventional use of the term suggests. This might be so, but it does not show that such a paraphrase can express the object's power. If being disposed to *M* in response to *S* is a specification of a power of the person

or the object, then it must entail the more generic power of being disposed to *M*. However, it can be true of Bob that he is disposed to swim when he accidentally falls in deep water, but false nevertheless that he is disposed to swim.

Thus, the power to *M* at *t* does not entail being disposed to *M* at *t* even as a response to *S*. Powers might be called dispositions; but if we do so regard them as such, then we must keep in mind that having a disposition, in the sense of having a power to *M* at *t*, does not entail being disposed to *M* at *t*. This can be seen clearly if we consider that even if it seems natural to specify a disposition *D* as being disposed to *M* in response to *S*, it is not true that having a power *P* can also be specified as being disposed to *M* in response to *S*. If an object has the power to *M* in response to *S* then it must also have the power to *M* because the possession of more specific, i.e. more determinate properties entail the less determinate ones. But as we have seen, if an object is disposed to *M* in response to *S*, it is still not necessarily disposed to *M*. Consequently, it is certainly false that having a power to *M* is the same as being disposed to *M*.

More importantly, however, the distinction between being disposed to *M* and having the power to *M* shall not disappear even if we include 'stimulus condition' in the specifications of habits and tendencies. Tendencies and habits can be conditioned just as powers can be. Bob might be disposed to smoke only if he is in a stressful situation, or when he drinks too much, or to whichever specific kind of 'stimulus' he is exposed. Some habitual behavior might be conditionless, but some others might not be. Conversely, Amy can—or has the power to—smoke in the very same sort of circumstances as Bob can, but she does not precisely because she is not disposed to.

My examples above about drawing a distinction between having a power to *M* and being disposed to *M* involved intentional human behavior. However, I used those examples only to make the distinction as vivid as I could in order to show that there is no natural move from 'having the power to *M*', and hence from 'can *M*', to 'being disposed to *M*'. But the problem is more general: if an object has a causal power (or liability) to *M* which is manifested only in exceptional circumstances, it is misleading to claim that it is disposed to *M*. Medications do have the power to kill someone in certain circumstances; for instance, if they are overdosed or if they are taken by someone who is allergic to them. But medications are not 'disposed to kill people as a response to being ingested'; just as some poisonous material that can cure people in certain special circumstances is not 'disposed to heal'. They are medications, or poisons, after all. Thus, as Ryle indicated, conventional disposition terms can be used for different theoretical purposes. For him, it might be only a question of meaning. But for those who take metaphysics seriously this difference in meaning indicates an important theoretical difference. It shows that properties that are powers and properties that can be expressed with the help of explicitly dispositional locutions play different theoretical roles in our ontology.

2. THE SIGNIFICANCE OF SPECIFIC POWERS

Alexander Bird has argued that the explicitly dispositional locutions which Lewis introduces in order to specify the meaning of ordinary power terms are ambiguous. Such specifications must exclude the presence of counteracting factors like antidotes, but it seems that any interpretation of 'disposed to *M* in the absence of antidotes' is either incompatible with what we ordinarily mean by a disposition, or takes dispositions to be extrinsic properties (Bird 1998, 231). According to one interpretation, we should not ascribe, for instance, the property 'being poisonous' to arsenic if it is taken in conjunction with an antidote. This makes the disposition extrinsic since whether or not we can correctly ascribe it depends on features of the environment in which arsenic is taken as well as on arsenic's intrinsic property. According to the other interpretation, arsenic itself is poisonous even when it is taken together with antidotes. But then, the absence of antidotes must be included in the circumstances of manifestation, and this seems to be in conflict with how we ordinarily understand disposition terms. We understand such terms with reference to the typical 'stimulus conditions', like ingesting arsenic, and not as 'ingesting arsenic in the absence of counteracting factors'.

    If Lewis's 'explicitly dispositional locutions' are indeed unavoidably ambiguous, then I take this as an additional reason for not using them for specifying powers. However, Bird's objection might be interpreted as showing that disposition terms as used in ordinary language can express properties with radically different theoretical roles. So interpreted, the objection shows that by specifying the conditions under which a power is manifested we can identify *different, but not distinct* powers. Further, it also shows that the difference cannot be understood by the paraphrase 'being disposed to *M* as a response to *S*' precisely because the ascription of power-like properties plays a different role in our ontology than the ascription of dispositions does. By specifying the conditions of manifestation, we can capture a more determinate property, which is different, even if not distinct, from the corresponding determinable one.

    Powers, like any other properties, can be more or less generic. When we ascribe a *specific* power, we can simply deny that in the presence of counteracting powers the object has *that specific property* to *M*. As I shall argue below, in certain contexts it is natural to ascribe such powers to objects and persons. Our ordinary disposition terms do not express specific powers, and they are probably not fine-grained enough to identify generic ones. But if we are interested in the nature of properties rather than in the semantics of conventional disposition terms, then this fact should not particularly disturb us. We can refer to any of these powers by means of demonstrative expressions. For instance, we can say *this bit* of arsenic material does not have *the specific property* to poison *these* people (because they have taken antidotes). Or, perhaps closer to standard ordinary usage, we

can say that this bit of arsenic has the *generic property* of being poisonous which entails that someone would be poisoned by ingesting it if it were taken *and* antidotes were not taken, *and* several other unmentioned conditions were met.

Thus the problem of the explicit specification of all relevant factors that are necessary for the power to become manifest arises only when we want to identify generic powers. Consider a particular glass that had just fallen and then broke. Since it did break, there must be a sense in which it could break; hence, since it broke it must have had the property being such that it would break if it were stuck by a hard object at that particular time in those particular circumstances.[4] Thus when we ascribe a power that is *maximally specific*, the problem of how to identify every relevant condition of the power's manifestation does not arise. We could say then that in any situation which differs from the one in which the object actually broke when it was dropped only in that it was not dropped, the object must have the specific power that it would break if it were dropped. Hence we can identify a natural possibility, i.e. what can happen in particular circumstances by ascribing a maximally specific power-like property to an object.

Why should we deny that objects have such properties? It seems that some may want to deny this for three reasons. First, it may be said that our ordinary disposition terms never ascribe such powers to objects. Second, such powers are obviously extrinsic: change the environment and you might change the object's power as well. And third, such properties are extremely abundant. But some would say that if powers are properties at all, they should be sparse rather than abundant.

As to the first reason: what it shows (yet again) is only that the study of disposition terms of ordinary language may not be a good guide to understanding the theoretical role of power-ascriptions. Bird mentions the case in which a sneeze, through 'butterfly effect', can cause a glass breaking (Bird 1998, 231). But he says we would not say that we have the disposition to break windows by a sneeze. This is certainly right; but only because the circumstances in which sneezes can cause the breaking of glasses are 'abnormal' or exceptional. The point about 'butterfly effect' is, however, exactly that there *could* be specific circumstances in which one does have the power to break windows by sneezing. That power might be 'strange' because it seems to be no more than a theoretical possibility that such circumstances actually arise. And that's why it is certainly false that we

---

[4] If we like reasoning with the metaphor of 'closeness of worlds', we can say that, from the perspective of the world in which the glass is not dropped and does not break, (one of) the closest one(s) is ours, i.e. the world in which it is dropped and then breaks. This follows from the fact that even if similarity is not transitive, it is certainly symmetric. Thus if we alter the circumstances *minimally*, so that the only difference between the actual and the counterfactual situation is that in the latter, luckily, the vase does not fall down, then the vase must, in both cases, have the power to break when dropped.

are disposed to break windows in this way if the specific circumstances rarely or never actually arise.

As to the second reason: specific powers are indeed extrinsic, but this does not seem to be a serious problem. Why should we insist that every power is an intrinsic property of the object that has it? Certainly, in some cases whether or not we can correctly ascribe some powers can depend on the *actual* presence or absence of some factors which are, according to some standard account of intrinsicness, extrinsic to objects (McKitrick 2003). But further, it is unclear how these standard accounts are applicable to properties that are powers. Some would say, for instance, that a power is intrinsic if any nomic duplicate of the object has it.[5] But this helps only if we can decide whether or not two objects are nomic duplicates independently of what powers they have. What counts as a nomic property depends on which laws there are. But, as Humeans would have it, what laws there are depends on world-wide regularities that cannot be intrinsic to a particular object. More importantly, it is arguable that nomic relations presuppose powers, and so we cannot characterize powers' features by reference to laws.[6] Hence the notion of extrinsicness as applied to powers is more of a problem than an independent argument for anything.

As to the third reason: one might worry that maximally specific powers cannot be 'real properties' of objects. Real properties are 'sparse'; but maximally specific powers are obviously abundant.[7] But what is exactly the reason to think that properties must be sparse? For the purpose of explanations that aim at unification, the ascription of abundant properties seems idle indeed. But it is one thing to say that for *certain* theoretical purposes it would be wrong to invoke a certain kind of property, and it is quite another to deny that objects can have such a property. The ascription of specific powers can play an important theoretical role in many contexts. Most importantly, if we want to understand natural possibilities as a consequence of objects possessing certain properties, we have good reasons to assume that properties are also abundant. For a certain type of event can occur or cannot occur in one or another specific circumstances, and not just 'in general', without further qualifications.

This does not mean that specific powers cannot play any role in the explanation of what has actually happened. They are often presupposed, for instance, in contrastive explanations of singular events. Such explanations mark out an event *C* in the causal history of *explanandum* event *E*, the absence of which would be

---

[5] Choi 2008.

[6] See also McKitrick 2009, though McKitrick's main concern is causation and not laws.

[7] About the concept and issue of sparseness and abundance of properties see Lewis 1983. As I understand him, Lewis does not deny the existence of abundant properties; he only says that they are not 'natural'. In my view, however, naturalness itself has little to do with parsimony. For an argument that dispositional properties are not sparse (one that is different from mine), see McKitrick 2003a.

sufficient for E's failing to occur. But such explanations work only if we assume that the presence or absence of C does not change the powers of objects that participate in the causal interaction within those particular circumstances. The contrast between the two cases is brought out precisely by the fact that the powers in the specific situations are the same so that the only distinction relevant to the occurrence of E is the presence or absence of C.

We also ascribe powers in order to identify what *things can do* and, pragmatically, what *we can do with them*. Thus, maximally specific powers find a natural place in practical reasoning. When we need to decide what to do, first we want to know what our real options are, i.e. what we can do or *what is in our power* to do *in the specific circumstances* in which we must make a choice. To use the old Lockean example, when I sit in a firmly locked room without keys and other exit etc., I cannot, i.e. I'm not able to, leave it. This means that I do not have the specific power to leave *that room at that time* even if I do have some generic power to leave rooms in the sense that I can move my limbs, see the exit etc. Thus when the question is whether or not I can do something in a particular situation, and I think that I can—I have the power to—do this or that, what I am assuming is that *at that moment and in those circumstances* no intervening factor is active. Even if I cannot explicitly specify all the conditions, I assume that, for all I know, in the specific situation, I have the power to act in the way I choose.[8] And if I'm lucky, I do indeed have that property there and then.

Consequently, some of the powers are objects' or persons' specific properties that ground certain possibilities even if the possibilities are never actualized. And objects can have determinate powers in specific situations even if we do not actually use them with the purpose to make their powers manifest. That objects, or rather certain kinds of objects, may not be disposed to M, not even in response to some type of stimulus S, unless they do display M with some statistically relevant regularity, does not show that they cannot have the power to M in certain specific circumstances.

3. POWERS AND REDUCTIVE ANALYSES

As I have mentioned at the beginning, according to the Humean tradition, the ascription of powers entails modalities in the sense that if objects could have powers, having such powers would entail necessary connections between distinct events. Traditionally, it is the rejection of necessary connections which is the main reason why Humeans have denied that power concepts can refer to genuine properties of objects. However, if I'm right, the theoretical role of

[8] I say more on this and on its implications for the ascription of responsibility in Huoranszki 2011.

ascribing powers is that they ground natural possibilities, not necessities. Thus, someone who holds that many properties are powers can—though needn't—agree with Hume that no event is such that its occurrence can make the occurrence of another distinct event necessary.[9]

Above I argued that we often ascribe a power to persons or objects in order to identify what they can do or what we can do with them. Thus the ascription of powers entails some potentialities or possibilities. This does not mean, as some have complained, that the ascription of powers is *only* talk about objects' relation to possible events. Powers are not 'occult relations to possible events'; they are actual properties of objects that explain why there are certain possibilities. Even if there might be unrealized powers, many powers are more or less specific properties that objects *actually* have.

But Hume's followers insist that power concepts do not refer to properties, or they refer to properties only because they can be reductively analyzed.[10] Reduction usually requires that the instantiation of one (sort of) property depend asymmetrically on the instantiation of some other (sort of) property. In the case of powers, reductive analyses usually assume that statements that ascribe powers express nomologically grounded actual and possible causal connections. The first step of the reductive analyses is to paraphrase statements ascribing powers in terms of what objects are disposed to do. Then it is claimed that such statements are conceptually equivalent to some counterfactuals that express contingent connections between distinct events. And finally, the truth-conditions of counterfactuals should be given purely in non-dispositional terms relying explicitly or implicitly on the concept of laws.

Critics of the Humean approach deny the possibility of such reduction. Interestingly, they do not seem to object to the very first step of the analysis, which I challenged in the earlier section; perhaps because they do not distinguish properties that are powers from those of dispositions. But even setting the issue of explicitly dispositional locutions aside, the possibility of reduction can be denied in two different ways. One way is to deny that the truth-conditions of the counterfactual conditional entailed by the ascription of a power can or should be understood purely in terms of causation, laws and non-dispositional properties. The other is to reject the earlier step of the reductive project, by arguing that the ascription of powers is not logically connected to the truth of any counterfactual conditionals. Many realists about powers, i.e. philosophers who reject reduction and take powers to be genuine properties of objects, believe that there is no

---

[9] In fact, so formulated, the claim is trivial. The question is whether causally connected events are ontologically distinct.

[10] This view has been endorsed in different forms, for instance, by Armstrong 1997, Mackie 1973 and 1977, and Lewis 1997.

interesting logical connection between the ascription of powers and the truth of certain counterfactual conditionals.[11]

Some of them suggest that to individuate a power it is enough to single out the type of events that is its typical manifestation.[12] George Molnar, for instance, argues that having some powers is like being in an intentional state in the sense that powers are directed at their manifestations even when the manifestation does not occur (2003).[13] But even if the parallel with intentional states is adequate, this does not prove that identifying a type of manifestation is sufficient for identifying powers. For intentional states can differ from each other even if they are directed at the same type of intentional object.[14] In fact, the attempt to understand causal powers without using conditionals must face a similar difficulty. The types of manifestations are not fine grained enough to identify distinct powers, unless we already understand them as manifestations of those powers.

Thus I'm going to argue for a third possibility. I accept, along with Humeans, that there must be a conceptual connection between the ascription of powers and the truth of certain counterfactual conditionals. However, the conditional analysis of dispositions can serve different purposes. It may be the case that originally it was introduced as a step towards reduction. The idea was that propositions that ascribe powers, if true at all, can be replaced by talk about actual or potential causal relations between events. But an analysis, even a conceptual analysis, can serve purposes other than conceptual reduction. In fact, very few analyses, if any, have ever provided conceptual reduction of one kind of entity to some other. The conceptual link between powers and counterfactuals can elucidate the relation between properties and natural possibilities without 'reducing' powers or denying that they are genuine properties. So understood, the conditional analysis of powers can play a significant role in our ontology: it is through such an analysis that we can specify powers and hence explain, with reference to objects' properties, which natural possibilities exist.

The non-reductivists' rejection of the conceptual connection between the ascriptions of powers and the truth of certain counterfactuals is based on counterexamples to what is called the 'simple conditional analysis of powers'. According to such an analysis, the ascription of powers entails the truth of some counterfactual conditional. Now, I agree that the instantiation of a property cannot depend on whether or not we happen to hold a counterfactual true. But the

[11] See Martin 1994, Molnar 2003, and Bird 2007, Chapter 2.

[12] See for instance Lowe 2010. Interestingly, most realists say surprisingly little about the way we should individuate powers.

[13] In fact, Molnar—assuming that there are further similarities between powers and intentional states—argues that to have a power is to be in an intentional state. This is a more contentious claim, which has been severely criticized by other realists. See for instance Bird 2007, 114–129.

[14] Thus, in order to identify an intentional state, we also need to identify the intentional mode. See Crane 2001, 32.

power that may or may not be instantiated by an object is nevertheless identifiable only by linking its ascription to the truth of some such counterfactual conditionals. For it is with the help of such counterfactuals that we can individuate the relevant possibilities.

4. FINKS, REDUCTION, AND CONDITIONALS

Alvin Goldman considers—though immediately rejects—an objection to the conditional analysis in one of the footnotes to his book on intentional action. According to the conditional analysis something is soluble iff it would dissolve if it were immersed in water. But imagine that someone has a magical power to make an object soluble whenever it is about to be immersed in water. Then it is true that it would dissolve if it were immersed in water. Nevertheless, the object is insoluble (Goldman 1970, 199–200). Many philosophers thought—perhaps Goldman included—that such examples speak more against the possibility of magic than against the conditional analysis of dispositions. Later, however, in a highly influential article, Charles Martin argued that there are cases logically analogous to the one described by Goldman that do not involve the use of supernatural capacities (Martin 1994).

Martin presents a scenario in which an electro-fink is attached to a wire. An electro-fink is a device that can make a wire live when it is touched by a conductor. Suppose we hold that a wire is live at t iff an electric current would flow through it, if it were touched by a conductor at t. But in the presence of an electro-fink, the conditional is true even if the wire is dead. Finks can work in the opposite way as well. It is possible that a wire is live, but, thanks to the presence of a 'reverse-fink', no electric current would run through it if it were touched by a conductor. In this case, we can correctly ascribe a power to an object even if the conditional is false. Thus, the possibility of finks and 'reverse-finks' raises an obvious difficulty for the simple conditional analysis.

As I noted earlier, in view of such counterexamples, some philosophers want to conclude that there is no interesting conceptual connection between the ascription of powers and the truth of the corresponding counterfactual conditional. Martin himself concludes that counterfactual conditionals in general are 'only clumsy and inexact linguistic gestures to dispositions' (Martin 1994, 8). Perhaps we can use such conditionals to characterize vaguely some causal powers; but the ascription of powers is logically independent of the truth of any counterfactual conditionals.

What Martin and many other realists about powers are supposing is that the simple conditional analysis' failure to accommodate the possibility of finks proves certain metaphysical consequences. They assume that if our ordinary disposition terms cannot be analyzed by means of counterfactual conditionals,

then powers cannot be reduced to non-powers, and hence they cannot be eliminated from our ontology.[15] In fact, I doubt that any such consequence follows. More importantly, however, one can draw such conclusions only if one assumes that the purpose of the conditional analysis *must* be reduction. Accordingly, if the analysis fails, reduction is rejected. But I shall argue for the contrary view: if reduction is rejected, the analysis can be saved and can play an important role in our understanding of powers.

Having said this, it remains true that many philosophers who aim at some reductive account of powers do rely on the conditional analysis. And it is generally agreed that for the purpose of reduction, propositions which ascribe powers must be logically or conceptually equivalent to certain counterfactual conditionals. The possibility of finks does indeed raise a difficulty for the reductive project. Therefore the reductionist needs a more complicated conditional analysis which is immune to such counterexamples.

David Lewis's reformed conditional analysis—perhaps the most influential attempt to amend the original analysis—was conceived exactly in this spirit (Lewis 1997). As Lewis observes, the examples of finks and reverse finks are based on the possibility that the bearer of a power can change during the process of manifestation, and that such change can result in the acquisition or loss of the power to be analyzed. Thus in order to answer the difficulty raised by the possibility of finks, we need to include among the conditions of manifestation that *some property* is retained until the power's manifestation occurs. But what is that property exactly?

Since Lewis rejects what he—following Martin—calls 'irreducible dispositionality', his purpose is to offer a reductive analysis of powers (Lewis 1997, 148). Thus, taking his cue from earlier reductive analyses, he assumes that every disposition must have a *causal basis*—'some intrinsic property B'—the presence of which together with the stimulus event would be causally sufficient for the occurrence of the manifestation event.[16] He then suggests completing the antecedent of the conditional with the condition that in the circumstances of manifestation, the object's relevant intrinsic property B would be retained.

However, the postulation of such intrinsic property gives rise to a number of difficulties. Reduction seems to require that the relevant intrinsic property be non-dispositional; and that it be distinct from the power itself. But it is unclear why every power needs to have a distinct ground or causal base. It has been argued persuasively that many fundamental physical powers do not have such ground (Ellis and Lierse 1994; McKitrick 2003b; Molnar 2003). Further, even

---

[15] This is, for instance, Molnar's view; see Molnar 2003, 82–98. This view is often identified with realism about powers. For important exceptions see Mellor 1974, Mumford 1998 and Mellor 2000.

[16] For earlier analyses relying on the same assumption, see Armstrong 1973, Mackie 1977, Prior et al. 1982.

if we assume that they do, we must explain how dispositions, understood as objects' properties, are related to their ground. If the ground is some non-dispositional property or property-complex, dispositions cannot be identical with it. But if they are distinct, why is the causal basis not sufficient for the occurrence of the manifestation events? Why do we need dispositional properties at all?

Lewis himself, seeing the difficulties with cashing out the nature of the relationship between dispositional properties and the non-dispositional base, tries to be as non-committal as possible about this question (Lewis 1997, 151–152). He insists only that the relevant property must be intrinsic. But even this is contentious. According to Lewis, the ground or basis of dispositions must be intrinsic because dispositions themselves are always intrinsic to their objects. However, our earlier considerations have shown that we have good reason to ascribe specific powers to objects that are extrinsic. And it would be hard to make sense of the claim that such power's ground is intrinsic. Further, some finks may be intrinsic to the object.[17] And if such finks can be removed only together with the intrinsic causal base, Lewis's analysis fails. Hence a large number of philosophers are in no position to agree with Lewis's analysis: on the one hand, philosophers who think that objects can have powers without having distinct non-powers as their grounds; on the other hand, philosophers who think that objects' powers can change without a change in their intrinsic non-dispositional properties (because, they hold, for instance, that some powers are not intrinsic).

Lewis's analysis assumes, of course, that there is no difference between the ascription of a power and the ascription of a disposition. Both are claims about what objects are disposed to do. But if I am right that powers as properties fulfill a different theoretical role than dispositions do, then we can avoid the difficulties and complications induced by the introduction of a non-dispositional causal base. As far as powers are concerned, we can propose a relatively simple non-reductive analysis. And as far as dispositions are concerned, we can rely on some statistical interpretation of the relevant counterfactual that can make the analysis immune to the problem of finks.

5. A NON-REDUCTIVE CONDITIONAL ANALYSIS OF POWERS

When Martin introduced the example of a fink, he meant it as a criticism of reductive analyses of dispositions. But Martin has not shown why, if we reject reduction, some version of the conditional analysis cannot be correct. In fact, before he draws his conclusion (cited in the previous section) about the useless-

---

[17] About the first possibility, see McKitrick 2003; about the second, Smith 1977. About intrinsic finks, see Clark 2008, Clark 2010 and Choi 2012.

ness of counterfactuals in the analysis of powers, he suggests that 'there can be no conditional which is both logically equivalent to a categorical ascription *and* such as to support the elimination of power or dispositional predicates' (Martin 1994, 6, emphasis in the original). But this is not the same as to claim that counterfactual conditionals are only 'clumsy gestures' towards powers.

Since realists about powers do not aim at a reductive analysis of powers, they can include in the circumstances of manifestation that *the object does not change with respect to the power* to be analyzed until the manifestation event occurs. Of course, in case of powers that can be manifested only once by a particular object, the power must be lost by the end of its manifestation, often together with its bearer. No vase can remain fragile after it has been broken; and no sugar cube is soluble after it has been dissolved. But objects can retain a power until its manifestation occurs.

Thus the problem of finks can be avoided if we use a slightly reformed conditional analysis suggested by Hugh Mellor as an improvement on Carnap's account of reduction sentences (Mellor 2000, 7–8). This requires the following minimally revised version of the original formulation of the counterfactual conditional:

For any object o and times t, $t+\delta t$, o has the power P to M at t iff the manifestation event of type M would occur no later than $t+\delta t$, if o were in circumstances of type C as characterized by conditions $\{c_1 \ldots c_n\}$ at t and it retained P at t– $t+\delta t$.

For instance, a sugar cube is (water-)soluble if and only if it would dissolve within a certain period of time if it were immersed in some not already saturated, not very cold etc. water *and* it retained its solubility until it dissolves.

This version of the conditional analysis has several advantages relative to Lewis's. First, it is much simpler. Truth, of course, should not be compromised because of our desire for simplicity, but simplicity without compromise might count as a virtue of an analysis nonetheless. Second, the analysis is compatible with powers being extrinsic. For some, this is perhaps more of a vice than a virtue. However, as I have argued earlier, if we ascribe powers to objects in order to ground natural possibilities, we had better make room for extrinsic powers. And third, this version of the conditional analysis does not assume that every power must have an intrinsic base that would make it nomologically impossible for the object to change its power while it does not change intrinsically; meanwhile, there is nothing in the analysis that would make it incompatible with the assumption that some powers do have intrinsic ground.

Despite such virtues, one might object that the analysis is circular and hence less informative than Lewis's analysis. However, we need to be careful about what we mean by circularity here. If it means that we refer to the yet unanalyzed property in the *analysans*, then the proposed analysis is certainly 'ontologically'

circular. But that circularity is unavoidable since the analysis can be correct only if it involves reference to the same property in the *analysandum* and in the *analysans*.[18] If, however, circularity means that the analysis does not provide any semantic information that can elucidate the content of the *analysandum* because of the reference to the power to be analyzed, then the analysis does not seem circular to me. At the very least, it is unclear why reference to the power in the *analyzans* would make the analysis uninformative.

Certainly, the analysis is not reductive, and this contradicts Lewis's assumption that we do not need properties that are irreducibly powers in our ontology. However, such convictions about whether every genuine property must be purely 'qualitative', or whether the ascription of properties must entail what objects possessing them can do, are independent of the question which counterfactual conditionals are entailed by their correct ascription. But more importantly, if the reference to the power to be analyzed in the *anlysans* is an objection to this version of the conditional analysis, then Lewis's analysis is not in better shape since that analysis also contains tacit reference to the power to be analyzed.

According to Lewis's analysis, one of the conditions of manifestation is that the object retain 'some intrinsic property B' until the manifestation occurs. What is that property? Obviously, it must be the property that grounds the power. Now ordinary objects have many powers at the same time. A knife made of stainless steel, for example, can have the power to conduct electricity, to resist rusting, as well as the power to cut bread (or the more generic power to cut or scratch objects made of material less hard than steel). These are obviously not the same powers, since certain kinds of objects can have one of them without having the other. Soft objects can be good conductors, and hard objects can be very bad ones.

So how can we single out the 'intrinsic property B' the retention of which is necessary in order for the power to become manifest? The only way I see is to say that *whatever property is the ground of the power* must be retained. Of course, if we assume that that property is distinct from the power itself, then there might be ways to detect it independently of the power. But the relevant question is whether or not we can identify it *as* the ground of the power without some implicit reference to the power itself. If our metaphysical conviction is that no property is irreducibly a power, and hence objects' powers must have some non-dispositional ground, then reference to the intrinsic non-dispositional property seems natural, quite independently of the issue of finks or any other objection to the simple conditional analysis. But this is not an issue about how powers are

---

[18] As Michael Jubien says 'if the concept under analysis has a certain characteristic feature, […] then one would think that feature must also somehow be present in the analysans, or else the analysis could not be correct. From this perspective [seeking a 'reductive' analysis] looks like the pursuit of magic.' (Jubien 2009, 95).

connected to conditionals. And certainly, it is hard to see why an analysis that includes the condition 'the power to *M* is retained' is less informative than the one with the condition 'whatever property is the ground of the power to *M*, it is retained'. If there is a difference between the two in terms of their informativeness, it is certainly the second, with its reference to a further unknown property, which is the less informative.

6. DISPOSITIONS, GENERIC POWERS, AND STATISTICAL REGULARITIES

When Lewis introduced explicitly dispositional locutions he meant to specify the meaning of conventional disposition terms so that they fit his reformed conditional analysis. But the introduction of the explicitly dispositional locution can also be interpreted as the first step in a statistical interpretation of the truth-conditions of the corresponding counterfactual conditionals. According to the statistical interpretation, the ascription of dispositions entails counterfactual conditionals that can tolerate exceptions. This means, roughly, that [1] a (kind of) object has the disposition *D* iff it is disposed to *M* in circumstances *C*; and [2] it is disposed to *M* in circumstances *C* iff in the statistically relevant reference class of cases, it would *M* more often than not, were it placed in a circumstance of the same type as *C*.[19] For instance, an object is fragile iff it would break more often than not in the counterfactual circumstances in which it is dropped from an appropriate distance on a hard surface.

It has been argued that the statistical interpretation of the counterfactuals can answer the challenge from finks as well as several other possible counterexamples to the conditional analysis of dispositions.[20] Indeed, this analysis can provide an intuitively adequate account of the connection between the ascription of dispositions and the sort of conditionals that are entailed by their ascription. But the statistical analysis explains this connection so well precisely because dispositions are understood here as behavioral tendencies. It is for this reason that the ascription of dispositions can be shown to entail certain counterfactuals the truth conditions of which can be interpreted in statistical terms. If it is true of an object that it is disposed to behave in certain ways in response to some stimulus, then it must be true as well that in the counterfactual circumstances in which the stimulus event occurs, the difference between the object's displaying and not displaying the behavior would be statistically relevant.

Dispositions the ascription of which involves such statistical regularities play an important explanatory role both in science and in ordinary discourse. But it plays a fundamentally different theoretical role than the ascription of powers.

---

[19] This is a variation of the theory proposed by Manley and Wasserman 2008, 75–76.
[20] See Manley and Wasserman 2008, 76–81.

To recall my earlier example, even if Bob *has the power to*, i.e. *can* swim, or can steal, can hurt other people's feelings etc., this needn't explain what he actually does, and it does not help predict what he is ever going to do. His having those powers is presupposed when we explain his behavior by his being a smoker, a thief, by his being callous, etc.; but ascribing them does not entail any such behavior. In contrast, when we say that Amy is a swimmer it does tell us something about her actual behavior because it explains what she occasionally does, even if the explanation is not particularly interesting or informative.

Of course, we can ascribe a disposition to a particular object even if the object does not actually display the relevant sort of behavioral tendency; if not for other reasons, just because many dispositions are such that particular objects can manifest them only once. For instance, no fragile object can break regularly. However, in most such cases, what grounds the truth of the statistical counterfactual is some actual regularity which is characteristic of the natural kind the particular object instantiates. Things made of a kind of glass are fragile because things made of that kind of glass break more often than not when dropped from a particular distance and when they fall on a hard surface etc. It is this actual regularity which can ground our belief that any particular object made of that kind of glass would break more often than not in the counterfactual circumstances in which it is dropped.

But even if the statistical interpretation of the counterfactuals explains well what it means for an object to be disposed to do certain things, it is ill-suited for the analysis of statements that ascribe powers to objects. For, as we have seen from Bird's sneezing–window-breaking example, if the circumstances in which an object would M more often than not are, statistically speaking, uncharacteristic, then the object is not disposed to M. Similarly, a poisonous material can cure someone in certain special circumstances and a medication can kill. But a poisonous material is not disposed to cure, and a medication is not disposed to kill. Thus, if powers fulfill the theoretical role in our ontology which they do according to my proposal, then the statistical interpretation of the counterfactuals should be understood as an account of objects' dispositions rather than that of their powers.

Despite this, there is an important metaphysical and an important epistemic connection between objects' having powers and their being disposed to behave in certain ways. As mentioned earlier, it is unlikely that every power has a distinct ground. However, it seems extremely plausible that dispositions as behavioral tendencies must have some ground. And their ground is exactly the object's generic power which specifies in which kind of interactions the object *can* participate in which kind of circumstances. Thus, metaphysically, the possession of generic powers grounds objects' dispositions. Epistemically, however, it is on the basis of observed regularities that we ascribe a disposition to certain objects or, more frequently, to certain kinds of objects. And it is through ascribing such

dispositions that we can infer the possession of generic powers. So understood, the ascription of dispositions provides a link between statistical regularities and the ascription of generic powers.

7. CONCLUSION

As I mentioned at the beginning, the concept of a disposition as it is used in contemporary philosophy was introduced as a technical term. As such, any theory of dispositions should be evaluated with reference to the theoretical role that the properties represented by such concepts play in science and philosophy. I tried to show that many analyses of disposition terms hide an important difference between two fundamentally different roles which such terms can play in metaphysics. One of these roles can be well captured by understanding statements of dispositions as expressing what objects are disposed to do in response to certain stimuli. But dispositions understood in this way should be distinguished from powers as properties of objects. For even if dispositions presuppose powers, the ascription of a power does not entail that objects are disposed to behave or interact in certain ways. The ascription of powers has a distinct role in metaphysics. And that role will not be adequately captured by a semantic analysis of the meaning of conventional disposition terms.

The ascription of powers is not constrained in any way by the use of dispositional predicates. It has been observed already in early discussions of dispositions that English has certain grammatical devices with which we can easily generate new disposition terms.[21] Some philosophers have worried that this makes it too easy to introduce new powers into our ontology.[22] This may be a justified worry as far as generic powers are concerned. But the role of specific powers cannot be understood through an analysis of conventional disposition terms for a different reason. Not because we can generate terms for powers too easily; rather, because some powers are so specific that it would be impossible to introduce separate terms for each. This does not mean that we cannot express specific powers, since it is possible to identify them with the help of demonstratives. But the main purpose of the conditional analysis of powers is not to provide an account of the use of conventional disposition terms, but rather to identify powers of varying specificity.

Ascribing such powers to objects serves a special theoretical purpose. Such specific powers are properties that ground natural possibilities. According to many contemporary accounts, properties should be understood in terms of possibilities. My account agrees with this to the extent that properties are not in-

---

[21] See Goodman 1954, 40.
[22] Molnar expresses this worry, which he shares with Quine. See Molnar 2003, 27–28.

dependent of possibilities. However, if I'm right, the order of ontological dependence should be the reverse. Since many properties are powers, it is these properties that ground natural possibilities.[23]

REFERENCES

Armstrong, D.M., 1973. *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.
——, 1997. *A World of States of Affairs*, Cambridge: Cambridge University Press.
Bird, A., 1998. 'Dispositions and Antidotes', *The Philosophical Quarterly* 48: 227–234.
——, 2007. *Nature's Metaphysics: Laws and Properties*, Oxford: Oxford University Press.
Carnap, R., 1936–7. 'Testability and Meaning', *Philosophy of Science* 3: 419–471 and 4: 1–40.
Choi, S. 2008. 'Dispositional Properties and Counterfactual Conditionals', *Mind* 117: 795–841.
——, 2012. 'Intrinsic Finks and Dispostional/Categorical Distinction', *Noûs* 46: 289–325.
Clarke, R., 2008. 'Intrinsic Finks', *The Philosophical Quarterly* 58: 512–518.
——, 2010. 'Opposing Powers', *Philosophical Studies* 149: 153–160.
Crane, T. 2001. *The Elements of Mind*, Oxford: Oxford University Press.
Ellis, B. and Lierse, C. 1994. 'Dispositional Essentialism', *Australasian Journal of Philosophy* 72: 27–45.
Ellis, B. 2001. *Scientific Essentialism*, Cambridge: Cambridge University Press.
Fara, M., 2005. 'Dispositions and Habituals', *Noûs* 39: 43–82.
Goodman, N., 1954. *Fact, Fiction and Forecast*, Cambridge, Mass.: Harvard University Press.
Goldman, A. 1970. *A Theory of Human Action*, Englewood Cliffs, NY: Prentice Hall.
Harré, R., 1970. 'Powers', *British Journal for the Philosophy of Science* 21: 81–101.
Harré, R. and Madden, E.H. 1975. *Causal Powers: A Theory of Natural Necessity*, Oxford: Basil Blackwell.
Heil, J., 2003. *From an Ontological Point of View*, Oxford: Clarendon Press.
Huoranszki, F. 2011. *Freedom of the Will: A Conditional Analysis*, New York: Routledge.
Lewis, D. K., 1983. 'New Work for a Theory of Universals', *Australasian Journal of Philosophy* 61: 343–77.
——, 1997. 'Finkish Dispositions', *The Philosophical Quarterly* 47: 143–158.
Lowe, J. 2010. 'The Individuation of Powers', in A. Marmodoro, A. (ed.) *The Metaphysics of Powers: Their Grounding and Their Manifestations*, 8–27.
Mackie, J.L., 1973. *Truth, Probability and Paradox*, Oxford: Oxford University Press.
——, 1977. 'Dispositions, Grounds and Causes', *Synthese* 34: 361–370.
McKitrick, J., 2003a. 'A Case for Extrinsic Dispositions', *Australasian Journal of Philosophy* 81: 155–174.
——, 2003b. 'The Bare Metaphysical Possibility of Bare Dispositions', *Philosophy and Phenomenological Research* 66: 349–369.
——, 2009. 'Dispositions, Causes, and Reduction', in Handfield, T. (ed.), *Dispositions and Causes*: 31–64, Oxford: Oxford University Press.

Manley, D. & Wasserman, R., 2008. 'On Linking Dispositions and Conditionals', *Mind* 117: 59–84.

Marmodoro, A. (ed.), 2010. *The Metaphysics of Powers: Their Grounding and Their Manifestations*, London: Routledge.

Martin, C.B., 1994. 'Dispositions and Conditionals', *The Philosophical Quarterly* 44: 1–8.

Mellor, D.H., 1974. 'In Defence of Dispositions', *The Philosophical Review* 83: 157–181.

——, 2000. 'The Semantics and Ontology of Dispositions', *Mind* 109: 757–780.

Molnar, G., 2003. *Powers: A Study in Metaphysics*, Oxford: Oxford University Press.

Mumford, S., 1998. *Dispositions*, Oxford: Oxford University Press.

Prior, E., 1985. *Dispositions*, Aberdeen: Aberdeen University Press.

Prior, E., Pargetter, R. & Jackson, F., 1982. 'Three Theses about Dispositions', *American Philosophical Quarterly* 19: 251–257.

Quine, W.V., 1960. *Word and Object*, Cambridge, Mass.: MIT Press.

Ryle, G., 1961. *The Concept of Mind*, 3rd edition, New York: Barnes and Nobles.

Smith, A.D., 1977. 'Dispositional Properties', *Mind* 86: 439–445.

HOWARD ROBINSON

"Are There Any *F*s?": How We Should Understand This Question

1. THE OPTIONS

In this paper I will present an approach to ontology which combines realism and conceptualism in what, I hope, is an intuitively acceptable way. In order to illustrate the advantages of this approach, I shall have to deal somewhat swiftly with a number of hot topics, but my objective is not so much to force conclusions about those topics as to illustrate the virtues of my fundamental approach to the 'are there any *F*s?' question.

I think the question 'are there *F*s?' can properly be interpreted in either of two ways, depending on the *F* in question. One I call the conceptualist interpretation (CI), and it can be read roughly as follows:

> (CI) We have the concept *F*. Is the world so organized that it satisfies this concept in the way that is necessary for the utility of that concept?

If the answer to this is affirmative, then there are *F*s, in the conceptualist sense.

The other interpretation is realist (RI), and goes roughly as follows.

> (RI) Forget about us and our concepts. If there were no conceptualizers around (putting God or Divine minds aside) would there be *F*s?

If the answer to this is affirmative, then there are *F*s in the realist sense. Common sense does not make this distinction, but I do not think that it finds it rebarbative. Why I think this will, I hope, become clear in what follows. Nominalists will find RI objectionable because they will not find the realist view of properties and universals entailed by it acceptable. I will not be concerned here to engage with nominalism of this sort, for I think that, if there is a world, it must be thus and so, and, therefore, it must be characterized by some properties independently of our conceptual practices. The nearest I can come to nominalism is (iv) below.

There are the following positions one might take on adopting the conceptualist and realist positions. .

(i) Realism for all standard concepts, including those for natural objects and those for artifacts.

(ii) Realism applying to natural objects, but not to artifacts, which are treated conceptually.

(iii) Realism for objects at the fundamental level but conceptualism for the rest.

(iv) Conceptualism for everything we know and possibly for everything we are ever likely to know, or even are capable of knowing. How the world is in itself will always evade our grasp: we can only approximate its actual properties.

Most discussion of these issues in modern analytic metaphysics is in what one might describe as a 'mad dog realist' spirit, as in (i), and is only interested in RI. To say with Peter van Inwagen (1990, 109), for example, that there are no tables, but only table-shaped arrangements of simples is to ignore CI as a candidate for answering the 'are there any *F*s?' question. Van Inwagen's position assumes that the only interpretation of this question is the strict realist one. If one took CI seriously, then it is clear that, in that sense, there are tables, for the world is clearly so constructed as to make the application of this concept fruitful.

2. THE USEFULNESS OF CI FOR ANALYTIC METAPHYSICS

There is a lot to be said for taking CI seriously, for if it were applicable, one might solve a whole series of problems that worry modern metaphysicians. These problems arise from applying RI across the board. I have already tried to show (Robinson 2008-9) that CI can be used to deal with vagueness and associated sorites problems. Many, if not most, concepts outside the most exact sciences (which means, I think, outside physics and, possibly, chemistry) are subject to such vagueness. In brief, I argued the following. Vagueness (at least of the sorites-generating kind) is a property of concepts, not of reality. The response to the paradoxes and conflicts with classical logic that vagueness generates should not be to try to develop a logic for those concepts that produces a formally valid natural language, incorporating all its idiosyncrasies, as in a three-valued logic, nor to impute to vague concepts a hidden precision, as does epistemicism, but to use or refrain from using those concepts according to their usefulness, consistent with classical logic. So, for example, when a raised piece of earth is neither clearly a hill nor a mountain, one drops these terms and speaks in terms of particular heights. I argued that one should think of such concepts as belonging to *representational ontologies* use of which can be dropped when cases do not fit their paradigms. The relationship of different ontologies to each other

is not random—there is what I called a *harmonization requirement* between the various ontologies we apply that is weaker than a formal logical compatibility. For example, Newtonian theory and quantum theory are strictly incompatible, but one can show how an essentially quantum world can sustain the application of Newtonian principles at most normal magnitudes. Similarly, one can see how strictly continuous heights can make useful the categories of *hill* and *mountain* for many straightforward cases, without needing to define the terms 'hill' and 'mountain' strictly in terms of exact heights. Where these categories do not fit, the practice of using them is suspended. This can only be done if one treats the relevant concepts in a conceptualist way, for the conceptualist holds that the 'are there *F*s?' question depends, in part, on whether the world will cooperate in the practice of using these concepts, and, where vagueness comes into play, it won't. If one believed these concepts to pick out fundamental constituents of the world, then one would not be free simply to drop them out of convenience and the full rigours of logical consistency would be brought to bear. Some kinds of discourse—such as that of bald-not bald, or hill and mountain—were never intended as basic, but the Newtonian ontology was. To apply it now, it must be downgraded to CI.

Vagueness is not the only problem which CI is helpful in handling. Here are two further examples.

The 'many Fs' problem

The 'many *F*s' problem has two versions, one essentially involving vagueness, one not. The vagueness version—in which the *F* in question is typically 'cloud'—concerns objects with indeterminate borders. Peter Unger (1980) argues that, in these cases, one can draw the border in many different places, thus individuating many overlapping objects. The non-vagueness- involving case goes as follows. Suppose that a complex physical object—a table, an oak tree or a cat - is made of a million atomic parts. There are almost indefinitely many sub-groups of those atoms which would be—or are—sufficient to constitute an object of the relevant kind. So, in the case of the table, imagine all its atoms minus two which are presently integral to the table—these atoms still constitute a table. If pursued, this line of thought leads to the idea that there are a vast number of actual tables contained within this one table, though they massively overlap. Or imagine the tree minus a branch. That would still be an oak tree. But that tree-minus-the-branch is present within the actual tree. Similarly for the cat without one of its paws—or simply without a few hairs or the odd atom. There are a tremendous number of proper part combinations sufficient to be an *F* within the actual *F*, and, as each of these is sufficient to be an *F*, all those *F*s are there and real.

The realist has to struggle with the fact that there are many well qualified candidates for being an *F* of the kind in question. Now I am not entirely convinced that a realist cannot cope with the non-vague cases, but the conceptualist has an easy route. He can simply say that this is not the way we deploy this concept: we deem there to be only one *F* whenever there is an *F* present and the only practical way of treating it is as a single object. This 'deeming' is not a conscious choice. Given our perceptual system, the table presents itself as unitary and we interact with it as one thing. For example, even if there are indefinite number of tables present, you cannot do an indefinite number of different things with them. So our conceptualization is practical as well as—or, perhaps, rather than—intellectual. It is a matter of how we interact with the object.

The 'clay and the statue' problem

A lump of clay is moulded into a statue of a man. The lump of clay and the statue seem to be, in an obvious intuitive sense, the same thing. They occupy exactly the same place and two different physical objects cannot occupy the same place; they each weigh, say, ten pounds, but their combined weight is only ten pounds. Nevertheless, they have different identity conditions. The clay can be remoulded into something else, and, if this happens, it continues to exist but the statue is destroyed.

The realist has a problem with all such relations of constitution, whether or not he treats constitution as a form of identity. If only the base is treated in a fully realist way, then one can treat the composite object as *a convenient way of conceptualizing* some of what is out there, rather than a further entity.

A natural response to this strategy is to say that, in the case of artifacts, perhaps we manufacture and control the concept as well as the object, but this is not so for natural objects. Maybe there is only one table there *because that is how our concept of 'table' works*, and statues only exist, perhaps, because we make them and interpret certain lumps of stone as statues, but the situation is different for natural objects. There is not only one oak tree or one cat there (if there is only one) because of our concepts: we control the reality of tables and statues on all levels, but not the reality of trees and cats.

3. ONTOLOGY AND THE SPECIAL SCIENCES

Coping with this challenge obliges us to face the question of the ontological status of entities dealt with in the special sciences. These seem to be real because they seem to cut nature at its joints, even if not at its most microscopic ones. On the other hand, if we treat them in a strictly realist way, many of the problems

which I have claimed CI solves will appear in their case, for vagueness, constitution and 'many *F*s' problems arise for many natural objects.

Considering the ontological status of the "levels" represented by the different special sciences, the following seem to be the options.

(1a) All levels are equally real.

(2a) All levels are real, but only the lowest one is fundamental: this does not impune the reality of the others. (If you construct something out of real elements, it is real itself, even though derivatively.)

(3a) Only the fundamental level is strictly real, the others are to be understood conceptually.

These options concern the existence of entities, but similar distinctions might be made concerning the genuineness of the causal powers of the different levels. Then it would run

(1b) All levels are equally causally efficacious.

(2b) All levels are causally efficacious, but they all draw their efficacy from the laws operating at the fundamental level.

(3b) Only the fundamental level has any real efficacy; the rest are a mere by-product or appearance of effectiveness.

It is natural to pair these two series off together, especially if one thinks that causal efficacy is a criterion for being physically real. But if one does not hold the latter it is at least possible to claim that all levels are real but that the energies are all micro-physical. My reason for distinguishing them is that Kim states his Causal Exclusion Principle in terms of causal efficacy; but, as we shall see, it is salient for the further question of ontological status. For present purposes, I shall be assuming that reality and efficacy stand or fall together.

For present purposes, I shall not distinguish between (1) and (2). Jonathan Schaffer (2003) has argued that there need not be a basic level and, in that case, all are equally real. I shall not pursue that issue here.

Two influential figures whose positions naturally favour (3) are Armstrong and Kim. David Armstrong (1978) believes that the only real universals are those needed for basic science, the rest are downgraded to the status of "predicates". This means that anything except the scientifically fundamental properties are the creatures of human thought and language. (It does not seem to be generally realized that this has serious consequences for Armstrong's theory of mind; for psychological properties will not be basic, and hence will themselves have the status of predicates, thus relying for their reification on activities of the human mind. This looks like a vicious regress.)

Kim (1998, 2005), too, propounds a principle which can be deployed to suggest that the ontologies of the special sciences have no independent efficacy, and this at least prepares the way for denying that they have an independent real existence. This principle is the *causal exclusion principle*, which goes as follows.

> If an event e has a sufficient cause c at t, no event at t distinct from c can be a cause of e (unless this is a case of genuine causal overdetermination). (Kim, 2005, 17)

Kim initially uses this to argue that a weak definition of physicalism in terms of supervenience cannot avoid epiphenomenalism, but he then considers the *generalization argument*, which extends it, given closure under physics, to the special sciences. Kim rejects this argument, in effect because he does not think that different ontological levels are in competition with each other. This is a major theme to which we shall return.

4. NON-REDUCTIVE PHYSICALISM *METAPHYSICAL* AND *LIGHT*

The ontological status of the special sciences can be approached *via* Barry Loewer's discussion of Jerry Fodor's seemingly extreme realism about the special sciences.

In controversy with Jerry Fodor, Barry Loewer distinguishes between what he calls *Non-Reductive Physicalism, Metaphysical (NRPM)* and *Non-Reductive Physicalism, Light (NRPL).* The latter corresponds to position (3) and the former to the stronger realism expressed in (1) and in (2).

> *NRPM* and *NRPL* agree that the special sciences are conceptually, epistemologically, and methodologically autonomous/irreducible to physics but disagree about what autonomy/irreducibility consists in and how it is to be explained. *NRPM* says that the autonomy/irreducibility is metaphysical and seeks to explain the conceptual and epistemological autonomy in terms of the existence of metaphysically basic special science kinds and laws. On the other hand, *NRPL* attempts to account for the conceptual/methodological irreducibility of the special sciences in terms of facts and laws of microphysics and our conceptual endowment and epistemological situation in the world...

Loewer then raises the question of what the difference would be between a world, w_1, in which *NRPL* held and another, w_2, in which *NRPM* obtained. Given closure under physics, everything would behave in exactly the same way, so the addition of the further entities and their concomitant laws appears to be vacuous.

Fodor appears to leave himself open to this argument by seeming to concede that the existence of two such distinct worlds is possible. Taking psychology as a case of a special science, he says:

> Only God gets to decide whether there is anything, and likewise only God gets to decide whether there are laws about pains; or whether, if there are, the pains that the laws are about are MR ['Metaphysically Real'].

Fodor here seems to be denying the standard physicalist maxim that once God had created all the facts of physics, he had nothing more to do, and this does place him in the bizarre situation that Loewer points out. It is more natural for the realist about the special sciences to deny that w_1 is possible, on the grounds that, once it has been created, the higher order entities and concomitant laws are, *eo ipso* present too. It is a case of what has been called an 'ontological free lunch'.

This last phrase might still leave one puzzled over what the difference between the "light" and the "metaphysical" consists in. Asserting that it means that a further set of entities exist, without any consequences, makes the assertion look suspiciously empty. The expression 'ontological free lunch' itself suggests something very dubious, namely that there both is, and is not, something more.

The idea that there is something more might be summed up by the idea that *F*s are *wholly constituted by* atoms, but that they are *not nothing but* atoms. What then, is the "more" of their nature, over and above how they are constituted?

5. ATTEMPTS TO DEFEND THE REALITY OF ORGANISMS.

I can think of two answers to the question with which the last section ended. One is that the *identity conditions* of the higher order entities are missing from facts about constitution. But the ontological status of identity conditions, considered as real entities "out there" seems very dubious. Such conditions are most naturally seen as *conditions for the application of a concept*, and, as such, play into the hands of the conceptualist and position (3).

Another, more *prima facie* realist option, is what has been called a modern version of *hylomorphism*. The core of this idea is to take *structure* or *organization* as ontologically basic. The term *hylomorphism* of course originates with Aristotle, but the modern theory is simply about taking macro organization at face value and so treating it as being real and efficacious as the microscopic features. Nothing exclusively Aristotelian or scholastic need be invoked.

This idea is very clearly articulated in Jaworski (2010):

> Hylomorphism claims …[t]hat structure is a basic ontological and explanatory principle. (269)
>
> Structure is also a basic explanatory principle in the sense that it explains why members of this or that kind are able to engage in the behaviors they do. It is because humans are organized as they are, for instance, that they are able to speak, to learn, and to engage in the range of activities that distinguish them from other living things and from non-living ones. (272)
>
> Hylomorphism implies…that there are two distinct kinds of properties: properties due to something's structure and properties things possess independently of a broader structure.

The properties of these structures are not idle, according to Jaworski:

> Emergent properties are not epiphenomenal…but make a distinctive causal or explanatory contribution to a system's behavior…

He emphasizes this point:

> Emergent [hylomorphic] properties are not logical constructions out of lower-level properties; they do not represent abstract ways of describing lower-level occurrences or processes. (274)

According to Jaworski, from this follows what might seem to be a direct denial of the causal exclusion principle:

> Hylomorphists endorse **causal pluralism**. They claim that there are causal properties and relations that do not fit the mold set by physics…[this] view is compatible with all forces operating at a fundamental physical level [i.e., none at other levels] and is therefore immune to the empirical objections raised against emergentism. (290-1)

This last quotation is particularly important. In ascribing to causal pluralism, the hylomorphist appears to be denying Kim's causal exclusion principle. But notice that the avowal of causal pluralism is immediately followed by the assertion that "all *forces* are operating at a fundamental physical level"(italics added). What one has, in fact, is an *explanatory* pluralism, with causation adopted into the domain of explanation; the wholly external, mind-free element is *force* and this is exclusively at the micro level. Causal exclusion has been replaced by force exclusion, and explanatory pluralism is now characterized as, or as including, plural causal explanation. But no-one thought that explanations, of all levels, excluded an appeal to causation. Jaworski is really only claiming that, once one realizes that most explanations are causal explanations, explanatory pluralism is

pluralism enough to constitute or ground a full realism about all levels. This sits well with Kim's belief that different ontological levels are not in competition with each other.

Nevertheless, this position might be defended. The truth makers for higher-level explanations are just as real as those for physics. If bricks are real, then so is a house made of bricks. And if bricks have causal powers, so does the house, in virtue of the bricks' powers. So if atoms have real force, combinations of atoms can constitute a real object which has real causal efficacy compounded from the real forces of the atoms of which it is made. This seems to be common sense. Kim states the commonsensical nature of this position emphatically:

> The errant baseball didn't after all break the window, and the earthquake did not cause the buildings to collapse! This strikes us as intolerable. (1998, 81)

This much is true, but the impression that it assists the realist is an illusion, for these truths are neutral between conceptualist and realist interpretations. The same argument could be brought for realism about vague concepts—*house* and *heap* are probably both examples—and vagueness is certainly best handled by treating the concepts in a conceptualist manner. The baseball broke the window, but both 'baseball' and 'window' are to be understood in the conceptualist sense. So if one adopts CI as appropriate for such concepts, there are baseballs etc., but in the conceptualist sense. Similarly for the earthquake and the buildings. Only an extreme realist would feel that common sense was threatened by this reading. Kim's earthquake and van Inwagen's table both exist, but on the conceptualist interpretation of what it is to exist. What is at stake is whether the human perspective has a certain role in reifying what is in fact the micro world in the way we do. There is a sense in which any mereological combination of atoms could be treated as an entity and so could the combined sum of their forces. Which are chosen are a matter of human interest and perspective—not arbitrary, of course, but well-groundedness of conceptual practice does not entail a strict realism. Talk of "human interests" might make it seem too intellectual. One of the most important things is the grain of human perception—what is salient to us and how it manifests itself in our senses. If we can see the independent constituents of an entity, we are less likely to think of that entity as basic. We can see the elements in a crowd, in a swarm of bees or in a weather system, and so are less likely to think of these things as fundamental, even if they seem to have a dynamic of their own. We are generally happy to make a conceptualist interpretation of them. But for most organisms, we see them only whole, for such parts as we do see are essentially parts of the thing—branches, leaves, limbs, teeth etc.—not independent parts. If we saw a plant as a swirling mass of particles passing in and out of an organizational vortex, like a rioting crowd, then, once we came to believe that the organization was a product of the interaction

of the particles following only the laws of physics and not an extraneous imposition, we would probably find it natural to make a conceptualist interpretation of plants. As it is, the nature of our perception seems to endow them with a greater degree of natural integrity than they would seem to possess from a more microscopic viewpoint.

Both as entities and as causal agents, macroscopic objects seem to be by-products of their micro constituents. What does it mean to call the higher order processes 'by-products'? It rests on the premise that everything that happens, happens because of the micro-dynamics. Apparent higher laws, though useful generalizations from our standpoint, do not give the real reason why anything happened. It is like the case of the plant "turning towards the light". Common experience leads us to say that it does so in order to gain more light, because it needs light to survive and replicate. But science tells us that this turning happens because of the chemical reactions involved, without any fundamental teleology. But were not these chemical processes "selected" because they allowed the plant to get the light and thus teleology is restored at the level of biology? Yes, in a sense, but only in the sense that certain micro-processes, from their own dynamics, repeat themselves in a certain way. The micro- processes do not get repeated *because* they lead to the replication of the organism, their repeating themselves *is* the replication. Dawkins' expression "the blind watchmaker" as a label for nature is illuminating. The "blindness" in question is not primarily cognitive, it is volitional. Nature does not intend to produce watches—or eyes or organisms in general—the developing of the quantum field, which is "blind to" its by-products, merely produces things which can be usefully so classified from the perspective of a macroscopic rational animal.

Davidson (2003) rejects Kim's exclusion principle, but, rather ironically, we can draw on a legitimate point of Davidson's in its support. Davidson claims, very plausibly, that it is only at the fundamental level that there are what he calls *strict laws*. Laws at other levels involve *ceteris paribus* clauses and a certain degree of approximation. This strongly suggests that, though they are useful explanatory tools, formulated on the basis of more exact processes that underlie them, the laws of the special sciences are not entities in their own right. It would be natural to argue the same way for the entities to which those laws attach. Davidson's reason for rejecting the exclusion principle rests on some very controversial features of his position. He claims that causal relations are entirely extensional and so events are not efficacious *in virtue of* any of the properties involved in them, so you cannot claim that some of them are active and others idle. The motive behind this is some kind of nominalism that wishes to treat properties as simply "descriptions" under which events fall, and, as such, not agents in the world. At the same time, he wants to treat the mental as "purely conceptual", and the basic physical as, in some sense, more real. For further discussion of Davidson's confusions, see Robinson (2003).

6. CONCEPTUALISM AND RELATIVE IDENTITY

David Wiggins has demonstrated, in a masterfully developed series of monographs (Wiggins 1967, 1980, 2001) that the logic of identity, with Leibniz's Law, rules out the possibility of relative identity. The doctrine of conceptualism for all ordinary entities that I have defended seems likely to leave open the possibility of relative identity, because the same piece of the material world might be conceptualized in different ways for different purposes. But what Wiggins actually shows is that the relativity of identity is impossible within any given representational ontology, but if the concepts in question are ones the use of which one can suspend, they need not be made formally consistent with other representational ontologies that one might choose to employ. One always has the option to withdraw the conceptualization that is leading to trouble and adopt one that is more appropriate or basic. On the other hand, *if the entities in question are taken in a strictly realist sense* then one is bound by Wiggins's argument. So what Wiggins says must be applied if you think that the entities under discussion are fundamental and interpreted according to RI. Wiggins, as a good Aristotelian, thinks that ordinary macroscopic objects—especially biological organisms—are paradigms of fundamental substances and so his logical constraints must apply to them. I have been arguing that they are not, in the appropriate sense, fundamental and that that is why Wiggins's logical discipline does not apply.[1] (I shall suggest at the end that there may not be in the physical world anything of the kind on which Wiggins's arguments can get a substantial grip. As (iv) above suggests, perhaps it is the case that everything on which we can get a grip must be understood according to CI.)

7. CONCEPTUALISM AND THE MIND.

Does this solve the problem of organisms—oak trees, cats, and, worse, other human beings?

Maybe what I have said so far is satisfactory for vegetable organisms, but what about animals, like cats, that are normally thought of as conscious, and what about human beings?

The answer to this will depend to a great extent on whether one is a physicalist. Considered as a nonconscious organism, an animal would be in the same category as a plant, namely a vortex of changing atoms formed entirely in ac-

---

[1] I say 'in the appropriate sense', because, as I argue in Robinson 2004/9, I think that there are two essential components to the traditional conception of substance, one that I call 'descriptive' and the other 'teleological'. Wiggins wishes Aristotle had ignored the latter, whereas what he tried to do was to give it dominance. It is to the descriptive conception, when taken realistically, that Wiggins's arguments most certainly apply.

cord with the laws of microphysics. Seeing this as an entity which is more than a highly organized cloud of particles would depend on our finding ourselves conceptualizing it as such, as with the oak tree. But isn't the cat's consciousness real in a way that is independent of our conceptualization? The correct answer is, I think, 'yes', but it is difficult to see how this could be the case if the cat's subjectivity were not something over and above the organization of elementary particles in its brain. It is not my purpose here to engage deeply with the philosophy of mind, but I shall briefly give reasons for this claim. If it is right that, in general, organic life is to be interpreted conceptually, as our way of making sense of certain patterns which are a by-product of development at the micro level—say in the quantum field—as presented through our senses, then the same will apply to subsystems within organic life. Thus it applies to neural processes, especially as functionally understood. This leads us to the self-undermining position to which (as I said earlier) Armstrong is committed. The very engine that is responsible for conceptualization—the human mind—is itself a unit only within the light of conceptual activity. This is the same problem as the one that faces Dennett's interpretationalist stance, and I have argued against it elsewhere (Robinson 1993: 6 and, in more detail, 2010), as have others (Hornsby 1997, 181-2). So here I shall simply assume that conscious states are fundamental in at least a property dualist sense. These mental states will be "out there" in a fully realist sense, in the same way as whatever constitutes the fundamental level of matter. But will the cat's mind, considered as a complex entity, also be real or will it depend on our reifying it by one of our concepts? Is there, for example, a 'many minds' problem corresponding to the 'many bodies' problem, if one tries to be a realist about minds? Remember that the 'many Fs' problem has two forms. One of them depends on the vagueness of the boundaries of most bodies. It is plausible to deny that minds are vague in this sense. If M is a mental state, then there must be some mind to which it belongs. This will not be true for a pure Humean, for on that theory, an impression can exist independently, detached from any mind and, therefore, presumably, in an indeterminate relation to a given mind—half attached, like a hair that is falling out or a water droplet at the margins of a cloud. I shall simply assume that this cannot be true of mental states. One may be only vaguely aware of some states, but, insofar as it counts as mental, it belongs to some particular mind. This still leaves the non-vague version of the 'many Fs' problem. After all, if you take all the cat's mental states and think away one sensation, you still have a feline mind, so are there not many cat-minds present? I had said when introducing this problem in Section 2 that I was not certain that there might not be a realist solution to it. In the case of minds, I think there certainly is. Insofar as it is determinate whether a certain mental state belongs to a given mind, then one can insist on a maximal criterion for the identity of a mind: it consists of all the mental states that are co-conscious. That very mind *could have* contained one mental state less, but it *does*

contain all the actually co-conscious ones. (See Kovacs 2010 for an argument that this is inconsistent with the supervenience of the mental on the physical.)

So the answer is that the mind as a whole will be real and unitary provided that the co-consciousness relation and its scope are real independently of our conceptualization. The cat will then have only one body *not*—or *not primarily*—because of the way we conceptualize it, but because the one consciousness of the cat acts upon it as a single object. The cat does not have to manoeuvre a set of bodies, as if it were herding sheep. So the cat's agency does for it something parallel to what our intelligent interaction with the world does for us. The same line of argument as the one applied to cats, applies to humans. And the individuality or uniqueness of one's body is the result of the fact that one thinks of it and acts upon it as a unity.

8. CONCLUSION - AND BEYOND.

We have seen that taking a conceptualist stand to most if not all non-basic ontologies makes intuitive sense and contributes towards solving several of the problems that worry contemporary analytic metaphysicians. But it comes at a cost, namely that the mind or mental states must be counted amongst the things that are basic and so this approach is not open to a standard physicalist. The thinking, conceptualizing subject must be amongst those things that are real in the strongest sense.

As a final thought I want to return to the final option that I offered above concerning degrees of realism, which was:

(iv) Conceptualism for everything we know and possibly for everything we are ever likely to know, or even are capable of knowing. How the world is in itself will always evade our grasp: we can only approximate its actual properties.

The discussion so far has been conducted within the scope of the assumption that we can be realists about the basic physical level. This idea is encouraged by our anachronistic but almost automatic assumption that classical atomism of some sort provides the model for the basic level. But we know that this is not realistic and our failure to be able to model intuitively quantum reality—the best we can manage is as a blur of wave and particle—means that our strictly realistic *representational ontology* of the basic level is purely formal and mathematical. All the conceptual categories of our normal ontology, such as that of objects and even of discrete events, may be conceptual impositions on a reality that does not quite fit any of them in itself. The logical discipline of a strict realism may be attenuated for all the categories that we can apply to the external world.

BIBLIOGRAPHY

Armstrong, D. M. 1978. *Universals and Scientific Realism*, Cambridge University Press, Cambridge.

Davidson, D., 2003. 'Thinking Causes', in *Mental Causation*, eds J. Heil and A. Mele, Clarendon Press, Oxford, pp. 3-17.

Jaworski, W., 2010. *Philosophy of Mind: a comprehensive introduction*, Wiley-Blackwell, Malden,

Kim, J., 1998. *Mind in a Physical World: an essay on the mind-body problem and mental causation*, MIT Press, Cambridge Mass.

Kim, J.,   2005. *Physicalism or Something Near Enough*, Princeton University Press, Princeton N.J..

Kovacs, D., 2010, 'Is there a conservative solution to the many thinkers problem?', *Ratio*, vol. 23, 275-90.

Loewer, B., 2008. Why is There Anything Except Physics?", *Synthese* …

Robinson, H., 2003. Introduction to *Objections to Physicalism*, ed. H. Robinson, Clarendon Press, Oxford.

Robinson, H., 2002.'Davidson and non-reductive physicalism: a tale of two cultures', *Physicalism and Its Discontents*, ed. Barry Loewer and Carl Gillet, Cambridge University Press, Cambridge, 129-51.

Robinson, H., 2003/9. 'Substance', *The Stanford Encyclopaedia of Philosophy*, ed. E. Zalta.
    "   2008-9   'Vagueness, Realism, Language and Thought', in *Proceedings of the Aristotelian Society*, vol.CIX,, 83-101.

Robinson, H., 2010. 'Quality, thought and consciousness', in *The Metaphysics of Consciousness*, Royal Insitute of Philosophy Supplement 67, eds P. Basile, J. Kiverstein, P. Phemester, Cambridge University Press, Cambridge, 203-16.

Schaffer, J., 2003. 'Is there a fundamental level?'*Noûs* 37.3, 498-517.

Wiggins, D., 1967. *Identity and Spatio-temporal Continuity*, Blackwell, Oxford
    "   1980. *Sameness and Substance*, Harvard University Press, Cambridge Mass..
    "   2001. *Sameness and Substance Renewed*, Cambridge University Press, Cambridge.

Unger, P., 1980. The Problem of the Many, *Midwest Studies in Philosophy*, 5: 411–67.

van Inwagen, P., 1990. *Material Beings*, Cornell University Press, Ithaca, N.Y..

Peter van Inwagen

"Who Sees Not that All the Dispute is About a Word?": Some Thoughts on Bennett's "Proxy 'Actualism'"

This note is a defense of Alvin Plantinga's "actualist" modal ontology against the criticisms brought against it by Karen Bennett in her (2006) paper "Proxy 'Actualism'"[1] (Familiarlity with both Bennett's paper and chapters IV-VIII of Plantinga's (1974) *The Nature of Necessity* is assumed.)

What I will say in response to Bennett's criticisms of Plantinga's actualism is almost entirely a gloss on the following thesis: It is not essential to Plantinga's philosophy of modality that the word 'actual' (or any word formed from 'actual'—'actually', 'actuality' and so on) occur in its statement. I will defend this thesis in Part I. Having established that thesis (at least to my own satisfaction), I will, in Part II, consider its consequences for Bennett's criticism of Plantinga's actualism. My conclusion will be that these criticisms fail, owing to the fact that they depend on the historical accident that the customary *designation* for Plantinga's position is "actualism"—that if this position had been given a name that did not contain 'actual' or any word formed from 'actual', the criticisms of the position that are presented in "Proxy 'Actualism'" could not even be stated. Finally, in an Appendix, I will discuss a puzzling statement that Bennett has made about the properties Plantinga calls individual essences.

I

I begin with a *précis* of Plantinga's modal ontology. This *précis* will be, I concede, *terminologically tendentious* in that it will contain no occurrences of the word 'actual' (or of any word "based on" 'actual'—'actually', 'actuality', etc.[2]). My position

---

[1] Bennett also criticizes a version of actualism defended by Bernard Linsky and Edward Zalta. The present paper is a reply only to her criticisms of Plantinga.

[2] In order to avoid using the adverb 'actually' in the *précis* of Plantinga's views that follows, I will use the adverbial phrase 'in fact' to perform a certain important function often performed by that adverb. The function of 'actually' to which I allude is illustrated by the result of inserting the word into a famous sentence of Russell's: 'I thought your yacht was longer

is that this *précis*, tendentious though it may be, is an exact expression (in broad outline: much is left out) of the modal ontology presented in chapters IV-VIII of *The Nature of Necessity*. It is an exact statement of the core of Plantinga's modal ontology *in an alternative vocabulary*. It differs *only verbally* from Plantinga's own statement of his ontology in *The Nature of Necessity*. It presents the *same* ontology in (slightly) different words.

The Précis

There are states of affairs.
States of affairs are abstract objects, like propositions. States of affairs and propositions are in fact closely connected. For each state of affairs there is a unique proposition that we may call its propositional analogue—and every proposition is the propositional analogue of one and only one state of affairs. For example, the propositional analogue of the state of affairs *Napoleon's having lost the Battle of Waterloo* is the proposition that Napoleon lost the Battle of Waterloo. Indeed, the connection between a state of affairs and its propositional analogue is so intimate that some—Chisholm (1970) among them—have identified them. (Everything said in the present paper is consistent with the thesis that a state of affairs is identical with its propositional analogue.)

A state of affairs *obtains* just in the case that its propositional analogue is true. (And, therefore, some states of affairs obtain and some don't—for, as I once heard Plantinga remark, "Some propositions are false. Roughly half of them.")

Say that a state of affairs that obtains is *obtentional* and one that does not obtain is *non-obtentional*.[3] Note that 'obtential' applies *only* to states of affairs. Whatever may be the case with 'actual', there can no more be an obtentional human being or pig or neutron star than there can be a human being or pig or neutron star that has a cube root—or is true or false, in the sense in which propositions are true or false.

---

than it actually is'. The best explanation of this function is probably that in such cases 'actually' is being used as a scope-indicator. For example, we might paraphrase 'I thought your yacht was longer than it actually is' in this way: 'Consider the length of your yacht; I thought your yacht had a length longer than *that* length'. (This use of the adverb 'actually' has a parallel adjectival use: 'I thought the length of your yacht was longer than its actual length'. The adjective can, however, be eliminated in favor of the adverb: 'its actual length' $=_{df}$ 'the length it actually has'.) So instead of saying, e.g., 'There could be things that do not actually exist," I will say "There could be things that do not in fact exist" (that is: "Consider all the things that exist; there could be things other than *those* things").

    [3] There really is such a word as 'obtention'; the OED defines it as 'the act of obtaining'. If a state of affairs is identical with its propositional analogue, obtentionality is just truth and non-obtentionality is just falsity.

> A state of affairs is *possible* just in the case that it is possible for it to obtain.
>
> A state of affairs is *impossible* just in the case that it is impossible for it to obtain.
>
> A state of affairs *x includes* a state of affairs *y* just in the case that it is impossible for *x* to obtain and *y* not to obtain.
>
> A state of affairs *x precludes* a state of affairs *y* just in the case that it is impossible for both *x* and *y* to obtain.
>
> A state of affairs is *maximal* just in the case that, for every state of affairs, it either includes or precludes that state of affairs.
>
> A *possible world* (or simply a *world*) is a state of affairs that is both possible and maximal.

We assume without argument that there *are* worlds—that there are states of affairs that are both possible and maximal. We further assume that at least one world obtains. Since it is impossible for two maximal states of affairs to obtain, *exactly* one world obtains;[4] we will call it the *obtential* world.

> A thing *exists in* a world just in the case that it would exist[5] if that world obtained.
>
> A thing *has* a property *in* a world just in the case that it would have that property if that world obtained.

A thing that does not exist is a contradiction in terms. (And so it is and must be, given the "account" of existence presented in footnote 5: since '*x* exists' is equivalent to 'something is identical with *x*', a thing that does not exist would be a thing such that nothing was identical with that thing, and it is an easily proved theorem of standard quantifier logic with identity that there is nothing with which nothing is identical. And it is easy to see why: everything is identical with itself.) It follows immediately that a thing that *could exist but does not exist*—a "modal alien", so to call it—is a contradiction in terms.

---

[4] Well, that's not strictly true—not if there can be necessarily equivalent but distinct propositions, and hence necessarily equivalent but distinct states of affairs (as it may be, *There being eight solar planets* and *The number of solar planets being the cube of the least prime*). I owe this nice point to Kenneth Boyce—and propose churlishly to ignore it and to speak as if there were only one obtential world. The nice point could be accommodated by replacing phrases like 'in the obtential world' with universal quantifications; 'in all obtential worlds', for example.

[5] We understand 'exist' in the following way: '*x* exists' is equivalent to 'Something is (identical with) *x*' (and, therefore, 'Fs exist' is equivalent to 'Something is an F'). We assume that it follows from this definition of 'exists' that the word be applied univocally to things in any logical category or of any sort: to states of affairs, propositions, properties, human beings, pigs, neutron stars…

Let us give the name *existentialism* to the thesis that "modal alien" is a self-contradictory concept[6].

Existentialists do not, of course, deny that there could be things that do not in fact exist. As things stand, for example, there are no million-carat diamonds; there nevertheless *could be* or *could have been* million-carat diamonds. But that modal truth does not imply the falsity of existentialism, since it is (loosely speaking) equivalent to the statement that the properties "being a diamond" and "weighing 200 kilos" are mutually consistent, and that statement obviously does not imply that there are nonexistent things: it refers only to two existent things (two properties) and asserts that these two existent properties stand in a certain relation, to wit, mutual consistency. And statements "about possible existents" can *always* be paraphrased as statements about the relations that hold among existent properties. In some cases, however, such a paraphrase would have to appeal to some rather special properties. Imagine that someone issues this challenge to the existentialists:

> It seems that if Socrates had not existed, it would nevertheless have been true that it was possible for him to have existed. If that's indeed the case, then there is a world w in which (a) Socrates does not exist, and (b) Socrates could have existed. Granted: that there is such a world—such a possible state of affairs—does not follow from existentialism. But it does follow from the contingent existence of Socrates and the symmetry of the accessibility relation. (If a world w in which Socrates does not exist is accessible from the obtentional world, then, by symmetry, a world in which Socrates does exist—the obtentional world—is accessible from w. It follows that it's true in w that Socrates could have existed.) And the symmetry of the accessibility relation is a pretty intuitive thesis—certainly not a thesis that the existentialist is going to want to rule out by definition. But how will speakers in w say that Socrates could have existed if in their world there's no Socrates for them to predicate possible existence of?

One way for existentialists to meet this challenge would be by appeal to a special property, a property that exists in all possible worlds and can be instantiated by and only by Socrates. (We Platonists say that every property exists in all possible worlds—including, of course, those in which it's uninstantiated—and a follower of Plantinga must be a Platonist.) And there is at least one such property. (Depending on how one individuates properties, one may affirm the existence

---

[6] This is not what *Plantinga* means by 'existentialism'—he in fact uses 'existentialism' as a name for a certain thesis he rejects—, much less what Sartre meant by it. Call the thesis Plantinga in fact called 'existentialism' (and which we shall not discuss) something else; 'dependentism', perhaps. Let 'serious existentialism' be the thesis that nothing is true of a thing x in worlds in which x does not exist—not even that it does not exist. (It is a matter of debate whether serious existentialism follows from existentialism.)

of one such property or of infinitely many such properties.) There is at least this property: the property of being Socrates—Socrateity for short[7]. (If Plato and Crito see a figure coming through the mist on a foggy morning in Athens, and Crito says, "Who's that?" and Plato replies, "That's Socrates," Plato ascribes Socrateity to the figure coming through the mist. Or so it seems reasonable to suppose.)

This property, Socrateity, is an *individual essence* (or simply an essence). "Individual essence" may be defined as follows. We begin with two preliminary definitions:

The primary individual essence of a thing x (or simply the primary essence of x) is the property of being identical with x[8].

A property is *an individual essence of* a thing x if and only if it is necessarily coextensive with the primary essence of x.

And we define 'individual essence' in terms of 'individual essence of':

A property is an individual essence (simpliciter, full stop, period) if and only if it is possibly an individual essence of something.[9]

If the inhabitants of a world in which Socrates does not exist want to say that Socrates could have existed, they cannot, as we have said, do it by saying of him that he could have existed, since he's not "there" to have that (or anything else)

---

[7] Socrateity is indeed a property that could be instantiated by and only by Socrates, but it has the property "could be instantiated by and only by Socrates" only in worlds in which Socrates exists. In a world in which Socrates does not exist, Socrateity and "could be instantiated by and only by Socrates" are related as follows: the former does not have the latter, but it *could* have it—and *only* those properties with which it is necessarily coextensive could have it.

[8] The phrase 'the property of being identical with x' is an open term, like 'the mother of x'. It is, however, an *intensional* open term (unlike 'the mother of x'). Phrases of the form 'the property of being identical with…' are thus intensional contexts. If, therefore, we wish to apply universal or existential instantiation to a sentence containing, e.g. the open term 'the property of being identical with z', any singular term that replaces the variable z must be a proper name. Thus, for example, '$\exists y\ y$ = the property of being identical with Socrates' comes from '$\forall x\ \exists y\ y$ = the property of being identical with x' by UI, but '$\exists y\ y$ = the property of being identical with the inventor of bifocals' does not come from '$\forall x\ \exists y\ y$ = the property of being identical with x' by UI. And, of course, '$\exists y\ y$ = the property of being identical with the husband of Xanthippe' does not follow from '$\exists y\ y$ = the property of being identical with Socrates' and 'Socrates = the husband of Xanthippe' by either Leibniz's Law (the principle of the indiscernibility of identicals) or Euclid's Law (the principle of the substitutivity of identicals).

[9] Plantinga (1974, 72–77) gives several equivalent definitions of 'individual essence'. The definition in the text is not one of Plantinga's definitions, but it is equivalent to each of them—and is much simpler than any of them.

said of him. But Socrateity *is* there to have things said of it, and they can say of it that it could have been instantiated.[10]

That is one way for existentialists to meet the challenge imagined above: by affirming the existence of individual essences. But there are other ways. The challenge can be met by affirming the existence of appropriate abstract objects of other sorts than properties. It can, for example, be met by an appeal to propositions like the proposition that Socrates exists (a proposition that exists and is true in exactly those worlds in which Socrates exists, and exists and is false in all other worlds[11]) or the proposition that Socrates could exist (a proposition that exists and is true in all worlds). If there are such propositions as the proposition that Socrates exists and the proposition that Socrates could exist, then the inhabitants of a world in which Socrates does not exist can say of the former that it is possibly true or say of the latter that it is true—or can simply *assert* the latter[12]. (Again, there will be problems about how people could possibly have "referential access" or "cognitive access" to such propositions in worlds in which their subject-terms do not exist—but those problems are no worse than the problem the possibilists face in the matter of referential access to particular nonexistent individuals.)

Here endeth the *précis*.

[10] They can, that is, if they can manage to refer to it. That might indeed be difficult or impossible, for it is hard to see how one could refer to Socrateity if Socrateity existed but was uninstantiated. But if that is a problem for existentialists, it is not a problem that they alone face: it is also a problem for possibilists—that is, for those who affirm the thesis that there are things that might exist (or might have existed) but do not in fact exist. More exactly, possibilists face a very similar problem, a problem that is "parallel to" or "analogous to" this problem, for possibilists will find it difficult to suppose that anyone is able refer to any *given* possible but nonexistent person. They will therefore find it difficult to suppose that any inhabitant of a world in which Socrates does not exist (but might have existed) is able to refer to Socrates.

[11] A serious existentialist must hold that propositions are necessarily existent.

[12] We should, by the way, note that the proposition that Socrates exists is the proposition that Socrates exists only in those worlds in which it's true (i.e., in which Socrates exists). It of course *exists* in worlds in which it's false ("of course" because it couldn't be false in a world in which it didn't exist—or so serious existentialists say) but in those worlds it doesn't have the property "being the proposition that Socrates exists." (Cf. our earlier point about Socrateity and "could be instantiated by and only by Socrates.") In worlds in which the proposition that Socrates exists does not have the property "being the proposition that Socrates exists," however, both the proposition and the property exist, and are related as follows: the former *could* have the latter, and it's the *only* proposition that could have that property. (And *this* should make anyone's head spin: in those worlds, that property—although it exists—*isn't* the property of being the proposition that Socrates exists.)

II

This was prologue. Now the play. The "'actual'-free" *précis* of Plantinga's theory was presented simply to provide a context for the following question:

> How would one argue for the following thesis: A person who accepts all the above definitions and statements—the definitions and statements set out in the *précis*—is a mere proxy existentialist? (What, in point of fact, does the charge, "You're a proxy existentialist" even *mean*?)

It seems to me that the following two propositions are self-evidently true:

> There is no way to argue for that thesis.

> The charge "You're a proxy existentialist" is meaningless.

(These two propositions are, of course, closely connected: if a thesis is meaningless, there is no way to argue for it.) At any rate, I hope that they are self-evident, because I have no argument for either of them. If they are true, I contend, they constitute a strong argument for the following evaluation of Bennett's (2006) criticisms of Plantinga in "Proxy 'Actualism'":

> These criticisms depend essentially on a chapter of historical accidents—(*verbal* accidents):
> • Early informal discussions of the semantics of quantified modal logic (and the objects that made up its universe of discourse) used the terms 'non-actual object' and 'merely possible object' to mean 'object that exists only in non-actual worlds'.
> • It was assumed in these discussions that an object that exists in some non-actual world *w* (and not in the actual world) and the world *w* itself were "non-actual" (or "merely possible") in the same sense.
> • At some point (*c*. 1980), the term 'actualism' began to be used as a name for the thesis that there are no non-actual/merely possible objects (Robert M. Adams was probably the first philosopher to use the word in that sense[13]). At about the

---

[13] Here is an illuminating historical note, kindly supplied by Professor Adams:

"I did say, »Actualism is the doctrine that there are no things that do not exist in the actual world« (Adams 1981, 7). As far as I know, I may have been the first person to define 'actualism' in that way. And whether I was or not, that was the first place in which I did so. As I think you know, I now regard it as a mistake (indeed, an aberration) that I did so. I had given a more careful definition previously: »Actualism, with respect to possible worlds, is the view that if there are any true statements in which there are said to be non-actual possible worlds, they must be reducible to statements in which the only things there are said to be are things which there are in the actual world and which are not identical with non-actual possibles.« On

same time, the term 'possibilism' came into use as a name for the thesis that there are non-actual/merely possible objects.

If, instead of the irrational terminological jumble[14] these accidents left us with, philosophers discussing quantified modal logic and the ontological problems it posed had resolutely applied 'actual' and 'non-actual' and 'merely possible' only to *abstract* objects (such as possible worlds) and, instead of, e.g., 'non-actual pig'/'merely possible pig' had said 'pig that does not exist but could have existed' and had used 'existentialism' instead of 'actualism', and, instead of 'possibilism', some term like 'anti-existentialism' or 'neo-Meinongian possibilism'[15], then

(a) The *real* metaphysical and ontological problems posed by quantified modal logic (and the informal modal discourse of which quantified modal logic is a regimentation) would be exactly as they in fact are,

and

(b) The verbal confusions that have resulted from the above-listed historical accidents—and the *unreal* metaphysical and ontological problems that have resulted from the confusions—would never have existed.

Appendix: The relation between concrete objects and their essences

Bennett (2006, 287) says that she doesn't see much difference between an object and its individual essence:

---

this understanding, it is only »if the notion of possible worlds is to be regarded as primitive,« that »the actualist will not agree that there are nonactual possible worlds« (Adams 1974, 224). D. C. Williams, in his (1959) article »Mind as a Matter of Fact« had used the word 'actualism' before I did, in a related sense, but I think not exactly the same as my sense in either of those essays."

[14] The question of the role played by David Lewis's "genuine modal realism" (Lewis's statements of genuine modal realism involve a wholly idiosyncratic use of 'actual') in creating this jumble raises some very complicated issues. I will not discuss Lewis's view here.

[15] "Neo-Meinongians" (Terence Parsons, for example), unlike Meinong himself and other paleo-Meinongians, do not hold that, e.g. a nonexistent pig has no sort of being whatever: they think that there *are* nonexistent pigs. But neo-Meinongians affirm the being of not only nonexistent pigs but *impossible* pigs, pigs that are nonexistent precisely *because* they have inconsistent (or incomplete) sets of properties. By a neo-Meinongian *possibilist*, I mean someone who affirms the being of nonexistent things, but only of such nonexistents that have complete and consistent sets of properties.

> Perhaps there is a lot of ontological difference between things like the Eiffel Tower and properties like *being made of metal*. Bu… there is not that much ontological difference between things like the Eiffel Tower and properties like *being the Eiffel Tower*.

I think that this statement has to be the result of a confusion of some sort. The Eiffel Tower is a concrete thing, a physical object, an artifact, an edifice, a structure. You can poke it with a stick. The property *being the Eiffel Tower* is an abstract object, and it is as immune to stick-poking as a proposition or a complex number or the middle of next week. And not only can physical objects be poked, but they can also fail to exist—whereas abstract objects exist in all possible worlds (or so we platonists say). Perhaps the confusion has something to do with the fact that 'the property *being the Eiffel Tower*', although it is the name of a property, an abstract object, is formed from a proper name of the concrete object that instantiates it. Let's look at a case of an individual essence that isn't of that sort—an essence that is not what I have called a *primary* essence.

I could, in principle, consider a non-primary essence of the Eiffel Tower, but to do that would place unwise demands on the patience of my readers, owing to the structural complexity of the Eiffel Tower. It will save us all a lot of time if I consider a non-primary essence of a much simpler artifact. Let us say: a table formed by placing a board on a stump[16]. Which properties of an artifact one supposes are essential to it—and, therefore, which of its properties one supposes are among its individual essences—will depend on one's metaphysics of artifacts. (And one's metaphysics of artifacts will presumably depend in its turn on one's general metaphysics of composite material objects). My own metaphysics of artifacts is not going to be of much use in this case, since it can be summed up in the statement that there are none. I'll therefore invent a metaphysics of artifacts that will imply the existence of both my simple table and the Eiffel Tower. I'll "construct" a property that anyone who accepted my imaginary metaphysics of artifacts should concede was an essence (and a non-primary essence) of the table. I think that this property I shall construct will suffice for an illustrative example of a non-primary essence of an artifact, despite the fact that this property is an essence of the table only given a certain metaphysics of artifacts that few if any philosophers would accept.

---

[16] "Let us picture to ourselves a very simple table, improvised from a stump and a board. Now one might have constructed a very similar table by using the same stump and a different board, or by using the same board and a different stump. But the only way of constructing precisely *that* table is to use that particular stump and that particular board. It would seem, therefore, that that particular table is *necessarily* made up of that particular stump and that particular board" (Chisholm 1973, 583).

Our imaginary metaphysics of artifacts consists in the logical consequences of two assumptions.

Assume, first, that *Contact* is the correct answer to the Special Composition Question[17]:

Contact
Necessarily: for all *y* the *x*s compose *y* if and only if no two of the *x*s overlap spatially and any two of the *x*s are in ancestral contact.

(Where '*x* and *y* are in ancestral contact' expresses the ancestral of the relation expressed by '*x* and *y* are in contact'.)

Assume, secondly, that the following thesis about the modal status of the relation between an artifact and certain of its parts is true:

Strong Artifactual Composition
Necessarily: If an artifact *x* is a fusion of certain non-overlapping artifact-parts the *y*s, and if the *y*s are arranged in manner M, then, (i) in every world in which *x* exists, *x* is a fusion of the *y*s arranged in manner M and (ii) anything in any world that is a fusion of the *y*s arranged in manner M is *x*.

I explain the terms of art that occur in this statement of Strong Artifactual Composition as follows:

(a) An *artifact-part* of an artifact is any part of that artifact that was intentionally manipulated in the course of the construction of the artifact by its maker(s). The statement of Strong Artifactual Composition presupposes that for every "artifact" there are non-overlapping artifact-parts, the *y*s, such that that artifact is a fusion of the *y*s—that is: the *y*s are all parts of that artifact and every part of the artifact overlaps some of the *y*s. (It follows from this assumption that neither a statue cast from molten bronze nor a statue made by chipping away at a block of marble is an "artifact". Let's assume that in the present discussion 'artifact' is a term of art, and that it applies only to things that were put together by a procedure that involved their "assembly" out of smaller, pre-existent objects. Our table is an artifact in this sense, and I see no objection to supposing that the Eiffel Tower is as well.)

---

[17] See van Inwagen (1990) §§2 and 3. The *y*s compose *x* (at *t*) just in the case *x* is a fusion or mereological sum of the *y*s (at *t*) and no two of the *y*s overlap (mereologically). '*y* is composed of the *x*s' is a stylistic variant on 'the *x*s compose *y*'.

(b) To specify the *manner* in which the members of a certain collection of non-overlapping artifact-parts are arranged is to specify which of them are in contact with one another, and at what points and regions on their surfaces.)

Now let a *Chisholm Table* be a table that has been formed by placing a board on a stump. (*Contact* implies that placing a board on a stump will bring a fusion of the board and the stump into existence.) Suppose we have a Chisholm Table before us. Let 'Mensa' be a proper name of the table, 'Tabula' a proper name of the board, and 'Truncus' a proper name of the stump. Let 'A' be a proper name of the part of the surface of Tabula that is in contact with Truncus, and let 'B' be a proper name of the part of the surface of Truncus that is in contact with Tabula.

Mensa has the following property (truncotabularity):

> Being a thing x such that Tabula is a part of x and Truncus is a part of x and every part of x overlaps either Tabula or Truncus and A is the part of the surface of Tabula that is in contact with Truncus and B is the part of the surface of Truncus that is in contact with Tabula.

Strong Artifactual Composition implies that truncotabularity is an individual essence of Mensa. (We could in principle write out the name of an analogous essence of the Eiffel Tower. It would be rather long, of course.) Now consider a world ("Nomensa") in which Tabula and Truncus both exist (and have the intrinsic properties they have in fact) and Mensa does not exist.[18] Inhabitants of Nomensa have the semantical resources to assert that it is possible for Mensa to exist, and they have these resources despite the fact that in Nomensa Mensa is not "there" to have possible existence ascribed to it—for they are able to refer to truncotabularity (they can refer to it by using the offset expression above) and to predicate possible instantiation of that property[19]. That is to say, they can assert the possible existence of Mensa by saying that truncotabularity (which is necessarily coextensive with the property that we in the obtentional world call Mensahood or Mensaity or the property of being identical with Mensa) was possibly instantiated—that is, was possibly a property that something had.

---

[18] Some of these worlds are very "close" to the obtentional world; consider, for example, a world in which—although they are never in contact (or are never in contact in the exactly the way in which they are in contact in the obtentional world)—Tabula and Truncus are near each other and it would be an easy matter for someone to place Tabula on Truncus in exactly the manner in which Tabula is placed on Truncus in the obtentional world.

[19] In that respect, truncotabularity is unlike Socrateity or any of Socrates' other essences, which, it would seem, human beings are unable to refer to in worlds in which Socrates does not exist.

One way to assert the possible existence of Mensa in Nomensa, therefore, is to assert of a certain individual essence (an essence that in worlds in which Mensa exists is an individual essence of Mensa) that it is possibly instantiated.[20] Thomas Jager's (1982) "actualist" semantics for quantified modal logic exploits this fact.[21] In an obvious sense, "Jager semantics" does utilize uninstantiated essences as replacements or proxies for "non-actual" objects. And "non-actual" objects certainly require replacements, owing to the fact that 'non-actual object' is either meaningless or means '(possibly existent but) nonexistent object'—and there are no (and could not possibly be any) nonexistent objects. There is therefore an obvious sense in which the property *being identical with the Eiffel Tower* "does duty for" the Eiffel Tower in Jager semantics: the modal ontology that the semantics presupposes does not affirm that, although the Eiffel Tower *is* (that is, has being) in all possible worlds, it is existent in some of them and nonexistent in all the others; it affirms, rather, that, although the property *being identical with the Eiffel Tower* is (and exists) in all possible worlds, it is instantiated in some of them and uninstantiated in all the others.

But why would someone, upon reflecting on a semantics in which uninstantiated essences do duty for nonexistent objects, react to it by saying, "There is not that much ontological difference between things like the Eiffel Tower and properties like *being the Eiffel Tower*?" If that statement were true, there would not be that much ontological difference between Mensa and the property *being Mensa* ("Mensaity" for short). That is, there would not be that much ontological difference between Mensa and the primary essence that is necessarily co-instantiated with truncotabularity. And can't one refute *that* thesis simply by pointing out that while hardly any possible worlds contain Mensa, Mensaity is present in all of them?—and that, therefore, in a world in which Mensa is not present to have possible existence or possible location in Alabama predicated of it, Mensaity *is* present to have possible instantiation or possible co-instantiation with "being located in Alabama" predicated of it? (Of course, that's not the only important ontological difference between Mensa and Mensaity. After all, Mensa is a concrete physical thing, and Mensaity is an abstract object. And one can set

[20] It is perhaps worth noting that there are other ways to assert the possible existence of Mensa in Nomensa than by ascribing possible instantiation to an essence (or possible truth to a proposition like 'A truncotabular thing exists'). For example, an inhabitant of Nomensa might assert the possible existence of Mensa simply by saying "It is possible to place Tabula on Truncus in such a way that A is the part of the surface of Tabula that is in contact with Truncus and B is the part of the surface of Truncus that is in contact with Tabula." If Contact and Strong Artifactual Composition are both true, that sentence expresses a proposition that is necessarily equivalent to the proposition that the sentence 'Mensa possibly exists' expresses in the obtentional world.

[21] See also Plantinga (1974, 123–132).

things down on concrete physical things and one can't set things down on abstract objects: If you want to have a picnic, and the closest thing to a picnic table available to you is Mensaity, you're out of luck.)[22]

BIBLIOGRAPHY

Adams, R. M. 1974. Theories of Actuality. *Noûs* 8: 211–231.
Adams, R. M. 1981. Actualism and Thisness, *Synthese* 49: 3–41.
Bennett, Karen 2006. Proxy 'Actualism'. *Philosophical Studies* 263–294.
Chisholm, Roderick 1970. Events and Propositions. *Noûs* 4: 15–24.
Chisholm, Roderick 1973. Parts as Essential to Their Wholes. *The Review of Metaphysics* 26: 581–603.
Jager, Thomas 1982. An Actualist Semantics for Quantified Modal Logic. *Notre Dame Journal of Formal Logic* 23: 335–349.
Plantinga, Alvin 1974. *The Nature of Necessity*. Oxford: Clarendon Press.
van Inwagen, Peter 1990. *Material Beings*. Ithaca and London: Cornell University Press.
Williams, D. C. 1959. Mind as a Matter of Fact. *The Review of Metaphysics* 13: 203–225.

ZSÓFIA ZVOLENSZKY

Against Sainsbury's Irrealism About Fictional Characters: Harry Potter as an Abstract Artifact

Our commonsense ontology includes ordinary objects like cups, saucers, tea kettles and jars of honey; buildings like the Big Ben; people alive today, like the novelist J.K. Rowling. We consider them part of our reality. We are commonsense realists about objects like cups, saucers, the Big Ben and J.K. Rowling. On the face of it, this is a fairly strong reason to include such objects in the ontology we posit. What about fictional characters like Harry Potter? Our intuitions pull us in opposite directions. On the one hand, clearly, Harry Potter—unlike the Big Ben—doesn't exist, for if we inventoried the objects and people existing today (or that ever existed), Harry Potter wouldn't be among them. On the other hand, we want to say the character didn't exist prior to the 1990s, before J. K. Rowling thought up the novels, but through her imagination and authorial intent to create a fictional world filled with fictional heroes and villains, she brought Harry, Dumbledore and a host of other fictional characters into *existence*. In other words, at first blush, intuitions about authorial creation support including fictional characters in our reality, while our commonsense ontology speaks against realism about fictional characters.

In this paper, I will explore arguments for and against one form of realism about fictional characters: *abstract artifact theory* about fictional characters ('*artifactualism*' for short), the view according to which fictional characters are part of our reality, but (unlike concrete entities like the Big Ben and J. K. Rowling), they are abstract objects created by humans, akin to the institution of marriage and the game of soccer. I shall defend this view against an objection that Mark Sainsbury (2010) considers decisive against artifactualism: "When we think about fictional entities, we do not think of them as abstract. Authors, who ought to know, would fiercely resist the suggestion that they are abstract. Abstract artifact theory entails that producers and consumers of fiction are sunk in error" (111). In other words, artifactualism attributes to people who produce and process sentences and thoughts about Harry Potter massive error, indeed, a category mistake about what kind of thing Harry Potter is. For an abstract object (such as the institution of marriage) isn't the sort of thing that can wear

glasses, ride a double-decker bus, attend school. I shall call this the *category-mistake objection*.

In Section 1, I will distinguish artifactualism from various other forms of realism about fictional characters, and from the position of *irrealism* about the likes of Harry Potter, a view according to which fictional characters don't exist; only the works of fiction portraying them do. In Section 2, I will explore one general and powerful argument for favoring artifactualism over other realist alternatives: it can successfully account for authors creating fictional characters. In Section 3, I will consider and deflect the category-mistake objection, which, according to Sainsbury, gives an edge to irrealism over artifactualism. Artifactualism, I shall conclude (in Section 4), remains a tenable contender.

1. REALIST AND IRREALIST POSITIONS ABOUT HARRY POTTER

We may, along with Mark Sainsbury (2010, 44–114), distinguish three realist alternatives about fictional characters: there really are such things just as there are ordinary concrete objects occupying space and time; but unlike those ordinary objects like cups, saucers and the Big Ben, …

- fictional characters *don't exist*, according to *Meinongianism* about fictional characters;[1]
- fictional characters *are not actual* but merely possible, according to *nonactualism*;[2] and
- fictional characters *are not concrete* but abstract, created by the activities of authors according to *artifactualism*.[3]

---

[1] For brevity's sake, I'll suppress the qualification 'about fictional characters' and will simply talk of realism, irrealism, Meinongianism, nonactualism, artifactualism, Platonism. Whenever these labels appear unqualified, they are shorthand for theories about fictional characters.
   Parsons (1980) is a contemporary proponent of Alexius Meinong's (1904) eponymous theory.
[2] Lewis (1978) put forth such a view. This position is sometimes called possibilism about fictional characters. See also Kripke's earlier (1963) view about Sherlock Holmes.
[3] Kripke (1973), Searle (1974/1979), van Inwagen (1977), Fine (1982), Schiffer (1996), Salmon (1998), Thomasson (1999) are prominent proponents who hold that authors' creative process of writing novels, stories, etc. creates fictional characters. This position is sometimes called creationism about fictional characters.
   There is a position in logical space for holding that fictional characters are abstract but exist timelessly, and authors don't create but discover them—we might call such a view Platonism about fictional characters. Zalta's (1983) unorthodox neo-*Meinongian* proposal can be considered an instance of such an account. The only kind of abstract-object theory I will consider in this paper is artifactualism, given the overwhelming popularity and attention that this position has been enjoying (compared to Platonism), as well as the advantages that I think it has over rival theories (Platonism included) precisely because it treats fictional characters as human-created objects.

One of the chief motivations for Meinongianism is this: plausibly, there are many things that don't exist, things that, while nonexistent, are the objects of our thought and imagination. Harry Potter is one of them, as is the batch of vanilla pudding I considered cooking up this afternoon (from a specific packet of pudding and batch of milk and sugar), but never got around to it. A similar motivation drives the nonactualist position: the range of things that are possible extends beyond things that are actual: the pudding I might have cooked this afternoon is a nonactual, merely possible object; as is Harry Potter.

The artifactualist position raises the intricate issue of deciding what exactly the abstract/concrete distinction consists in. The assumption so far has been that abstract objects (unlike concrete ones) don't occupy space and time. Another option is that abstract objects (unlike concrete ones) lack causal powers. A third option is to identify paradigmatic examples of concrete and of abstract objects in order to illuminate the distinction.[4] I won't dwell on these options here, because the ways in which the abstract/concrete distinction is traditionally drawn are called into question precisely in the light of abstract object created by human activity, *abstract artifacts*, that is—for example, the institution of marriage and the game of chess.[5]

Instead of defining the categories of abstract versus concrete, I will therefore take as my point of departure a broad and fairly uncontroversial range of examples for both concrete and abstract objects. Concrete objects clearly include things like cups, saucers, actual batches of pudding, the Big Ben, J. K. Rowling. Many of those who posit abstract objects count among them numbers, sets, propositions and properties like being tall and being human. Those who consider these abstract objects agree that typically,[6] they are *timelessly existing abstract objects that are mind-independent* in the following sense: their existence at a time t is independent of any mental activity at t.

---

[4] For an overview of these and other ways of drawing the abstract/concrete distinction, see Rosen (2012).

[5] See Rosen (2012) and Fine (1982, 130–131) motivating the claim that there should be room for abstract objects that come into existence contingently:

> …what underlies the platonist's position is a certain ontological prejudice. … These philosophers suppose … that certain features should go together, so that the same entities will be material, will exist in space and time, will exist contingently, etc., and the same entities will be immaterial, not exist in space and time, be necessarily existent, etc. Now although paradigmatic cases of concrete and abstract objects may have exactly the features from one or other of these groups, it must be recognized that there are objects of intermediate status that share features from both.

[6] Barring exceptions like the singleton set of my red mental image upon spotting a strawberry. There are various ways to go on the status of such sets: we could conclude that not all sets are abstract after all or that the notion of mind-independence at work should be revised to allow such sets to be abstract (see Rosen 2012).

There is, however, another type of abstract object one might posit: *abstract artifacts*. Notice that an abstract artifact like the game of chess does have temporal features, after all: the game of chess didn't exist before 1000 A.D. and has been in existence for several centuries (but beyond that, there is disagreement). Nonetheless, an abstract artifact would still be mind-independent in the above sense: the game of chess can exist at a time without anyone having any chess-related mental activity at that time. It's worth giving a variety of examples of abstract artifacts:

- the games of soccer and chess; the chess move of castling;
- the institution of marriage and the office of prime minister;
- religions like Anglicanism or Buddhism;
- traditions like Mardi Gras and Easter celebrations;
- more specific traditions like the New Orleans Mardi Gras festival and the Village Halloween Parade in New York City;
- the tradition of the Easter bunny and its Australian marsupial counterpart, the Easter bilby;
- recipes for preparing treacle pudding and beef Wellington;
- the letters of the alphabet ('A', 'Z', etc.);
- brands like Twinings, Twix and Mini Cooper;
- words and names of a language, including fairly recent additions like 'netiquette' (rules governing polite behavior in interactions on the internet) and 'cot potato' (for a small child spending a lot of time in front of the TV set, that is, a very young couch potato); also the first name 'Dweezil' for boys, coined by Frank Zappa;
- musical works like Mozart's serenade *A Little Night Music* and opera *The Magic Flute*;
- literary works like the seven Harry Potter novels written by J. K. Rowling.

We thus have a long and varied list of candidates for abstract social and cultural (legal, artistic, religious, linguistic etc.) artifacts among which it is natural to make room for fictional characters like Harry Potter also—the defender of artifactualism suggests.

Alternatively, a fourth option is to forgo realism about fictional characters, opting for *irrealism*, which denies all forms of ontological commitment to fictional entities.[7] Irrealism proposes to analyze sentences like (1)-(8) below with the help of an ontology that is committed only to the existence of works of fiction: novels, films, and so on:

---

[7] Sainsbury (2010) favors this view, as does Walton (1990).

(1) Harry Potter doesn't exist.

(2) "Harry had a thin face, knobbly knees, black hair, bright green eyes. He wore round glasses held together with a lot of Scotch tape…"[8]

(3) "From his first days at Hogwarts, the young, green-eyed boy bore the burden of his destiny as a leader, coping with the expectations and duties of his role…"[9]

(4) Harry Potter is a fictional character.

(5) Harry Potter was created by J.K. Rowling.

(6) Stephen King thinks about Harry Potter.[10]

(7) Harry Potter is more famous than Sparrowhawk (the wizard of Earthsea).[11]

(8) Harry Potter is not as reckless as Sparrowhawk.

(I have included (1) here in order to have a complete list of the types of sentences that the various theories about fictional characters have to account for. Of course, capturing the truth of (1) is easy for an irrealist and tricky for realists. I won't discuss realists' proposals for analyzing (1) in this paper.)

(2)-(8), when taken at face value, seem to accrue ontological commitment to characters from fiction, and therefore provide *prima facie* evidence for one or another of the realist positions. It is well to note that (2), (3) and (8) form a separate group on this list: it isn't *literally* true that Harry Potter had knobbly knees, wore glasses etc.; it's true according to the Harry Potter fiction, but if we enumerate all those who wear glasses, Stephen King and the rest, Harry wouldn't be on the list. (2) is quoted from one of the Harry Potter novels. (3) discusses the content of another Harry Potter novel; (8) compares the content of two works of fiction; we can call (2), (3) and (8) *fiction-internal* sentences put forth by authors, readers (including critics). As before, we can say that it isn't *literally* true that Harry Potter is a young green-eyed boy who bore the burden of his destiny as a leader. And it isn't literally true that Sparrowhawk is more reckless than Potter. It is therefore unclear just how great a challenge (2), (3) and (8) pose for irrealism. The irrealist might argue that though each is literally false, they can be replaced, respectively, by the true (2'), (3') and (8'), hence their true ring:

---

[8] From Volume 1: *Harry Potter and the Sorcerer's Stone*, page 20.

[9] The literary critic Michiko Kakutami wrote this in her review of volume 7 in *The New York Times*, "For Harry Potter, Good Old-fashioned Closure", 07/04/2007.

[10] Stephen King has been a major fan of Harry Potter and a defender of the Harry Potter series against criticism from literary theorists.

[11] Ursula K. le Guin began writing her now-classic series of fantasy novels and short stories about Earthsea in the 1960s; they feature Sparrowhawk, a young orphaned boy who discovers he has magic powers.

(2') According to the first Harry Potter novel, Potter had a thin face, knobbly knees, black hair, bright green eyes, wore round glasses held together with Scotch tape.

(3') According to the seven Harry Potter novels, Harry Potter was a young, green-eyed boy who, from his first days at Hogwarts, bore the burden of his destiny as a leader, coping with the expectations and duties of his role.

(8') According to the Harry Potter and Earthsea fiction series, Potter is not as reckless as Sparrowhawk.

(4)–(7) are more complicated, however: they aren't just true according to a body of fiction; they appear to be literally true sentences that authors, readers and critics might *assert* about fictional characters features in fictional works, akin to statements like 'Stephen King wears glasses', and 'Stephen King prefers baseball to soccer'. We can call (4)–(7) *fiction-external* sentences put forth by authors, readers and critics.

The success of irrealism therefore hinges on how compelling, systematic, and general its non-face-value treatment of the various problem sentences is, especially the literally true (4)–(7).[12] The irrealist might analyze these in turn as follows:[13]

(4') There exists a body of fiction according to which Harry Potter is a specific character.

(5') J.K. Rowling wrote a body of fiction according to which Harry Potter is a specific character.

(6') For some property P, Stephen King entertains a propositional attitude with the content that Harry Potter has P.[14]

(7') More people think about Harry Potter than about Sparrowhawk.

We can analyze 'think about' further based on (6'):

(7") There are more people who entertain propositional attitudes with the content that Harry Potter has P for some property P, than there are people who entertain propositional attitudes with the content that Sparrowhawk has Q for some property Q.[15]

---

[12] van Inwagen (for example, 1977, 2000) argues that there is no systematic way to provide irrealist paraphrases for the problem sentences (4)–(7). For ways that an irrealist might respond, see Brock (2002), Caplan (2004).

[13] See Sainsbury (2010, 115–151); for some of these examples, his preferred analysis is different than the one listed here.

[14] I follow Sainsbury in assuming that an irrealist can readily account for a name like 'Harry Potter' occurring in a propositional attitude context.

[15] The properties P and Q may vary from one person to the next.

So far, we have considered four proposals about the nature of fictional characters: irrealism, and three realist positions—Meinongianism, nonactualism and artifactualism. Of these, the nonactualist position relies on a notion of possibility and actuality, which are usually cashed out by reference to possible worlds and the actual world; to understand this position better, it is crucial to enumerate briefly the various stances one might adopt with respect to the nature of possible worlds. According to the nonactualist, Harry Potter is a merely possible object who has a thin face, round glasses, knobbly knees, etc.; so (1) is analyzed as:

(2″) There is a nonactual possible world in which Harry Potter has a thin face, knobbly knees, black hair, bright green eyes, wears round glasses held together with Scotch tape.

Sainsbury points out that nonactualism incurs a commitment about the metaphysics of possible worlds. "Nonactualists wish to locate [fictional] objects in possible worlds; so they need to be realists about possible worlds" (Sainsbury 2010, 74). There are two major classes of realist views on offer about the metaphysics of possible worlds:

- extreme realism about possible worlds, *pw-realism* for short, has it that the actual world is one among a plurality of possible worlds that are causally and spatiotemporally isolated from one another.[16] This view takes (2″) at face value.[17]
- ersatz realism about possible worlds, *pw-ersatzism* for short, has it that a possible world is abstract, for example, a maximally consistent set of propositions representing a way the world could be;[18] this yields the following analysis of (2):

(2‴) There is something abstract, a set of (maximally consistent) propositions representing Harry Potter as having a thin face, knobbly knees, black hair, bright green eyes, wearing round glasses held together with Scotch tape.

---

[16] For a long time, Lewis (1973, 1986) remained the lone proponent of pw-realism.

[17] There is a problem, however (Sainsbury 2010, 85–87): in the novels, Harry Potter's isn't given a complete description, down to his last detail about sock color; Potter is thus incomplete. But all possible objects are complete (they have the same ontological status we do, it's just that some of them are nonactual, inhabiting merely possible worlds). One of the more tenable options for the nonactualist is to relate the incomplete Harry Potter to various possible objects—Potter-surrogates—that have all the properties that the novels ascribe to Potter, but are complete (down to the last detail about sock color). Given that on this option, the nonactualist has to quantify over Potter-surrogates to account for (2), she cannot take (2″) at face value in the end.

[18] Pw-ersatzism has had many proponents. Adams (1974) held this particular, proposition-based version of pw-ersatzism.

Sainsbury (2010, 222, fn. 8) points out that pw-ersatzism collapses into a view according to which fictional characters are abstract objects (in our case: a representation comprising propositions about Harry Potter, for example). "For then the true metaphysical nature of a supposedly nonactual fictional character is actual": an actual set of propositions. It is only in conjunction with pw-realism that nonactualism offers a distinct alternative to a theory like artifactualism, according to which Harry Potter is abstract.

Nonactualism therefore comes in a package bundled with pw-realism, a controversial proposal. In addition, we are considering two other realist contenders—Meinongianism and artifactualism. Analyzing negative existential claims like (1) poses a challenge to all of these views. Accounting for (2)–(8) poses various degrees of difficulty to the realist contenders. Let us now turn to how much weight (5) carries, and more generally, how important it is to maintain, as artifactualism does, that authors create fictional characters.

2. THE IMPORTANCE OF AUTHORIAL CREATION

Why insist that authors *create* fictional characters? It does seem natural to say (5):

(5) Harry Potter was created by J. K. Rowling.

Artifactualism, positing Harry Potter as an artifact created by J. K. Rowling, takes (5) at face value. But we have already seen that this is not our only option; we could also accept an irrealist analysis of (5) that doesn't take it at face value: "J. K. Rowling wrote a body of fiction in which Harry Potter is a specific character". Quite independently of irrealism, several philosophers have had serious qualms about taking (5) at face value: Brock (2010, 338) sets out to "explain why creationism about fictional characters [the view that fictional characters exist by being created by their author(s)] is an abject failure. It suffers from the same problem as theological creationism: the purported explanation is more mysterious than the data it seeks to explain" because it cannot offer a satisfactory account of the spatial and temporal dimensions of fictional characters, for example, their moment of creation. Yagisawa (2001, 154) argues that the most influential creationist views (by Searle and van Inwagen) "are ultimately unsuccessful in establishing creationism"; more generally, he thinks no view on which fictional characters exist can do justice to our intuition that a claim like "Harry Potter doesn't exist" is true and is entailed by the true "Harry Potter is a fictional character". In the light of such doubts about creationism, it is worth homing in on an argument for artifactualism (a form of creationism), showing that it is the best

form of realism one could adopt *precisely because it incorporates authorial creation*.[19] The goal of this section is to expound such an argument.

How might the various forms of realism handle (5)? On this point, artifactualism shows a clear edge relative to its two rivals. A negligible point of advantage is that according to neither rival theories is Potter *created*—going from nonexistent to existent. According to the Meinongian, Potter isn't created—brought into existence—because he doesn't exist (he just is). And according to the nonactualist, Potter had existed all along as a merely possible object and continues to exist as a merely possible object after the novels are written. According to Sainsbury (2010, 61–63, 82–85), the real advantage of artifactualism concerns its response to the so-called *selection problem*: upon introducing the name 'Harry Potter' in her novel, how does J.K. Rowling manage to select one rather than another among the countless candidate objects? According to Meinongianism, there are countless nonexistent candidates; according to nonactualism, there are countless merely possible, nonactual candidates. Sainsbury (2010, 63) doesn't see "how a Meinongian can offer any sensible account of how an author's or reader's thoughts are supposed to engage with one rather than another nonexistent entity". We are about to see that a more decisive objection emerges against the Meinongian once we consider the difficulties that the *nonactualist* encounters when it comes to the selection problem and other problems.

In the "Addenda" to his "Naming and Necessity" lectures, Kripke (1972, 156–7) motivates two theses for expressions like 'unicorn' and 'Harry Potter':

- The *metaphysical thesis*: There is no basis for counting any *merely possible* object as Harry Potter, Sherlock Holmes, a unicorn, etc.
- The *epistemological thesis*: There is no basis for counting any *actual* object as Harry Potter, Sherlock Holmes, a unicorn, etc.

In the metaphysical thesis, Kripke's target seems to be the nonactualist. At the end of this section, we will see, however, that both theses bear on Meinongianism also. Along the way, we will also see that the two arguments are at root intimately connected.

Behind the metaphysical thesis is what we might call the *insufficient-specificity problem*.[20] The Harry Potter novels specify many details about Harry; but they

---

[19] An argument for fictional characters as objects created by people is noteworthy in the light of Brock's (2010, 340–342) criticism. He calls this the Fundamental Thesis: "Fictional characters, to the extent that there are any, are genuinely created by the authors of the works in which their names (or designating descriptions) first appear." Brock then remarks that "arguments in support of the fundamental thesis are almost completely lacking". In this Section, I set out to produce precisely this sort of argument.

[20] Kaplan also emphasizes insufficient specificity as an obstacle to naming nonexistents (1973, 506; 1989, 609).

also leave a lot of other details unspecified, for example, which of various parental cells Harry came from. Due to such lack of specificity in the novels, we have no basis for deciding between two distinct merely possible candidates (they originate from distinct sperms, say) that are just like Harry is described in the novels, which of them is Harry Potter. Notice that it is in part due to insufficient specificity in the novels that Sainsbury's selection problem arises—for the Meinongian as well as the nonactualist.

The epistemological thesis turns out to generate an even deeper problem for the nonactualist, one that we shall see (at the end of this section) affects the Meinongian also. Behind the epistemological thesis is what we might call the *coincidental-resemblance problem*, which Kripke discusses in connection with the mythical species of unicorn:

> …the mere discovery of animals with the properties attributed to unicorns in the myth would be no means to show that these were the animals the myth was about: perhaps the myth was spun out of whole cloth and the fact that animals with the same appearance actually existed was mere coincidence. In that case, we cannot say that the unicorns of the myth really existed; we must also establish a historical connection that shows that the myth is *about* these animals. (Kripke 1972, 157, emphasis in the original)

Kripke is making two points here: even if we find animals qualitatively like the unicorns of the myth, that wouldn't justify counting them as unicorns given (i) the lack of historical connection between the newly found species and the use of the expression 'unicorn'; and given that (ii) the unicorn myth was "spun out of whole cloth", not created in the right way, to make the term apply to the newly found species. The upshot of (i) and (ii): we would have no more than mere qualitative coincidence between unicorns as described in the myth and the actual species discovered. And for a proper name, reference takes more than coincidental resemblance, so we don't have any candidate actual objects to count as unicorns.[21]

In the case of the expression 'unicorn', the coincidental-resemblance problem thus arises as a result of two distinct problems: (i) *historical unconnectedness* and (ii) *unsuited mode of introduction*. Pure myth-making mode and pure fiction-writing mode both give rise to expressions that aren't introduced in the right way to refer to actual objects. It is natural to expect the *intentions and beliefs* of language users to be highly relevant in determining the mode in which they introduce expressions of their language. On this point, it is customary to note a key difference between myth and fiction (which I will take for granted for the purposes of this paper):

---

[21] Kaplan quotes Harry Deutsch: "reference is no coincidence" (Kaplan 1989: 608).

> The difference between authors and myth-makers is one of propositional attitude: authors *make-believe* their works of fiction, whereas myth-makers do not make-believe their myths; rather, they genuinely *believe* their myths. (Caplan 2004, 334, emphasis in the original)[22]

According to this, those who created the myth of the unicorn had unicorn-related *beliefs* (not just pretended beliefs). Still Kripke does raise the issue of an unsuited mode of introduction: myth-spinning. Right after the passage above, Kripke (1972, 157–158) repeats the same point with respect to 'Sherlock Holmes' also: "it is theoretically possible though in practice fantastically unlikely, that Doyle was writing pure fiction with only coincidental resemblance to [an] actual man". A crucial consideration emerges from these fleeting remarks about unicorns and Sherlock Holmes: given (ii) the way the myth/fiction was created, and (i) the fact that we encounter historical unconnectedness, the result is that we find no more than coincidental resemblance to actual objects.

The unsuited-mode problem would arise even if we had at hand a myth or a novel specifying mythical beings/characters completely, down to the last bit of information about sock color and origin (it would be mind-numbing to read such a novel).[23] So even in special cases of names from complete fictions in which the metaphysical thesis is circumvented, the epistemological thesis would still present problems. (Given the focus of this paper, in what follows, I will focus on characters from fiction, setting myths and mythical beings to the side; the points I make about the various problems can be generalized to names from myths also.)

Both theses and all the problems considered so far have taken it for granted that the candidate objects to count as Harry Potter are *concrete, spatiotemporal objects*. It is therefore well to keep this qualification in mind. For example, for (ii) we get: the fiction-writing mode in which the expression 'Harry Potter' had been introduced into the language is unsuited for the name to refer to an actual concrete, spatiotemporal object. For (i) we get: actual, concrete, spatiotemporal objects as potential referents for the name are historically unconnected to the introduction and subsequent use of 'Harry Potter'.

It's crucial to note that of the two problems (i) and (ii), unsuited mode of introduction is the more fundamental one, explaining historical unconnectedness *of the relevant sort*: for all we know, there could be an actual person who provided inspiration for Rowling's Harry Potter; this would make for *a* historical connection between Rowling's use of the name and the actual boy. But it wouldn't be the *reference-determining* kind of historical connection we're interested in, the sort of historical connection that would circumvent the coincidental-resemblance

---

[22] See also Salmon 1998; Braun 2005.
[23] Kaplan (1989, 609) makes this point.

problem by determining the reference of 'Harry Potter'. Rowling's authorial intentions, her fiction-writing mode precludes historical connections that are reference-fixing.[24] Kripke (1972, 92) also gives an example of an irrelevant kind of causal/historical connection: the "causal chain from our use of the term 'Santa Claus' to a certain historical saint"; despite such a link, children, when they use the name 'Santa Claus' "by this time probably do not refer to that saint".

Being an actual person who is the spitting image of Harry Potter as he is described in the Rowling novels would make for no more than coincidental resemblance to Harry Potter. Here is why: J. K. Rowling's intention was to introduce the name 'Harry Potter' for a fictional character rather than an actual person who fits a certain set of descriptions. And for a name to refer to an actual person takes more than coincidental resemblance; reference is shaped (i) in part by causal-historical connections between uses of the name and an object (whether that be a concrete or an abstract object), and (ii) in part by the mode of introduction. Given that (ii) Rowling's intention was to create a *fictional character* rather than refer to a flesh-and-blood person with introducing 'Harry Potter', (i) the name 'Harry Potter' was never historically linked (in the relevant way) to an actual orphaned boy wearing glasses, with a Z-shaped scar on his forehead, growing up in suburban England learning wizardry in a boarding school, and so on, and the name cannot refer to any actual concrete boy with spatiotemporal dimensions.

Not only is the unsuited-mode problem more fundamental than (i); it is also more general. Notice that it readily generalizes to concrete, spatiotemporal objects of all sorts, merely possible ones included; this way, we get:

> *the unsuited-mode problem generalized*: the fiction-writing mode of introducing proper names into the language is unsuited for them to have as their reference concrete, spatiotemporal objects, whether they be actual or merely possible.

It is well to generalize in the same way the coincidental-resemblance problem also:

> *The coincidental-resemblance problem generalized*: there is no more than mere qualitative coincidence between concrete, spatiotemporal objects (whether they be actual or merely possible) and fictional characters as described in works of fiction.

---

[24] Notice that I am assuming here that authors have *the final word* on whether they are introducing a name for a historical figure or a fictional character. Suppose that Tolstoy, upon asked about Napoleon in his *War and Peace* were to have sincerely said: "I intended the figure of Napoleon in *War and Peace* to be a fictional character; I drew a great deal of inspiration from the French military leader, but still, resemblance between the character and the historical figure is pure coincidence". In this imagined scenario, the position I am assuming is that Tolstoy would have introduced 'Napoleon' in *War and Peace* in fiction-writing mode, so the reference of the name would not have been the historical figure, and historical connections to the first Emperor of the French would not have been of the relevant, reference-fixing sort.

Therefore, as we dig deeper, the pair of problems behind the epistemological thesis turn out to target nonactualism.

As before, in the case of 'Harry Potter', the unsuited-mode problem generalized underlies the generalized coincidental-resemblance problem. And both apply to the metaphysical thesis also: the generalized unsuited-mode problem provides the following *additional* reason for holding the metaphysical thesis. *If* the character of Harry Potter is not fully specified in the novels, *then* what grounds do we have at all for choosing between two distinct merely possible concrete, spatiotemporal objects which to count as Harry Potter when, given J. K. Rowling's fiction-writing mode of introducing 'Harry Potter', it would be a matter of sheer coincidental resemblance for the name to refer to either of those candidate objects? *With respect to names from fiction, the unsuited-mode problem (and in its wake, the coincidental resemblance problem) therefore raises a key issue underlying both the metaphysical and the epistemological theses discussed by Kripke*; this is a striking detail to bring to the surface given that Kripke mentions the unsuited-mode problem in passing only (saying no more than the two half-sentences quoted above), devoting far more attention to the metaphysical thesis.

Just how bizarre the idea of reference based on coincidental resemblance is—the conception of reference for 'Harry Potter' to which the nonactualist is committed—can be brought out based on considerations about *nonfictional* names that fail to refer. The French astronomer Le Verrier put forth a hypothesis about the existence of an intra-Mercurial planet which he named 'Vulcan', to explain perturbations in the orbit of Mercury. There were various independent sightings mistakenly believed to be of Vulcan before enthusiasm dwindled; By 1916, Einstein's general theory of relativity confirmed that the perturbations were produced by the gravitational field of the Sun; there was no intra-Mercurial planet at all; the Vulcan-hypothesis was refuted; 'Vulcan' turned out not to refer to anything.

What about a counterfactual situation in which the Vulcan-hypothesis is a success story? Imagine a counterfactual scenario with the laws of physics slightly different, and there being an intra-Mercurial planet affecting the orbit of Mercury; Le Verrier puts forth his hypothesis; there are sightings converging on the planet, which comes to be called 'Vulcan', the name featured in Le Verrier's prior hypothesis. But that is not *our* term 'Vulcan' that comes to name the counterfactual planet, but a different one. It is preposterous to think that in coining the name in the actual world, Le Verrier managed to name *that counterfactual object even though his naming attempt failed in the actual world*. 'Vulcan' might have been a success story just as 'London' might have been introduced as a name for a river instead of a city; but all that is irrelevant to how and whether these strings, as parts of *our* language, were introduced and subsequently used.[25] Le Verrier

---

[25] See Kripke (1971, 145; 1972, 77, 102–3, especially fn. 51).

strove to name an actual concrete, spatiotemporal object; due to his failure to do so, he didn't *by coincidence* name a nonactual concrete, spatiotemporal object (as the nonactualist would have it); doing so was no part of his intention. So 'Vulcan' doesn't refer to any concrete objects in any counterfactual situations. Kaplan (1973, 506–508) makes this point eloquently with respect to a fictional name like 'Pegasus'. But what is far more interesting is that the point holds for 'Vulcan'! We can say the following about this name, as well as other proper names intended for concrete objects or for fictional characters:[26] *if it cannot make it here, it cannot make it anywhere.* If the name doesn't refer to a concrete, spatiotemporal object here, in the actual world, it doesn't refer to such an object in other possible worlds either. Elsewhere (Zvolenszky 2007), I call this the *inverse-Sinatra principle for proper names.*[27]

The inverse Sinatra principle is quite general, covering names like 'Vulcan', 'Pegasus', and 'Harry Potter'. And the reason why these names cannot make it anywhere given that they cannot make it here (in the actual world), is because nonactual concrete objects are, at best, coincidentally similar to the descriptions given for Vulcan, Pegasus and Harry Potter. We thus have a nonfictional variant of the coincidental resemblance problem: there is no more than mere qualitative coincidence between merely possible concrete, spatiotemporal objects and nonreferring names featured in failed hypotheses.

Notice that 'Vulcan' and 'Harry Potter' differ in one crucial detail: for the case of 'Vulcan', the unsuited-mode objection doesn't arise. Le Verrier's intention had been to introduce 'Vulcan' for a concrete, spatiotemporal object; so a historical connection, if there had been one, linking uses of the name to an actual concrete object, could have served to fix the reference of 'Vulcan', circumventing coincidental-resemblance-related qualms. A historical connection can be secured in the actual world only—there is absolutely no historical connection between *our* use of 'Vulcan' and a merely possible concrete, spatiotemporal object. And in the absence of an actual historical connection, qualms about coincidental resemblance do arise, leading to the metaphysical thesis about 'Vulcan': if the specification of Vulcan isn't complete, allowing that several distinct merely possible concrete objects fit the specification equally, then we have no basis for counting any one of them as Vulcan. (Notice that here, as before, my argument leading to the metaphysical thesis for Vulcan was crucially linked to considerations about coincidental resemblance and historical unconnectedness,

---

[26] Even an irrealist about fictional characters can, based on the considerations about Vulcan and unicorns above, accept the inverse-Sinatra principle.

[27] Frank Sinatra sang about New York City: "If I can make it there, I'll make it anywhere".

which were originally identified behind the other thesis—the epistemological one. With respect to 'Vulcan', too, we see that the two theses are intimately connected.)

The foregoing observation allows us to highlight a more general point of advantage for the artifactualist position over both Meinongianism and nonactualism.

According to artifactualism, Harry Potter is an actual object; it's just that unlike concrete objects like the Big Ben, Harry Potter is abstract. Yet the fact that he is an actual artifact makes room for a certain kind of causal-historical dependence on the physical world: in the 1990s, J. K. Rowling's creative activities bring it about that Potter is an actual (rather than a merely possible) abstract object. The sort of dependence in place allows Harry Potter qua abstract artifact to be the kind of referent for Rowling's name 'Harry Potter' with respect to which issues having to do with historical unconnectedness and, in turn, coincidental resemblance, and, in turn, the epistemological thesis, do not arise. (Notice that before, we noted that for names of fictional characters, no historical connection to concrete, spatiotemporal objects is of the relevant, reference-fixing sort. Meanwhile, the point made here is that for the artifactualist, a historical connection to an *abstract artifact* is precisely what fixes the reference of 'Harry Potter'.)

By contrast, alternative realist accounts that make Harry Potter a concrete object whose existence does not causally depend on us either because the object is nonexistent (according to Meinongianism) or because it is nonactual (according to nonactualism), face a challenge. First, these theorists have to explain why those objects are candidates of the right ontological status to count as the referents of 'Harry Potter'. As we have already seen, on this point, the nonactualist founders already. The Meinongian can get past this hurdle: he may suggest that his nonexistents are objects of thought and hence have just the right sort of ontological status to be suitable targets of authors' intended reference. But on the next hurdle the Meinongian stumbles: if his nonexistent objects are of a suitable sort as objects of fiction-writing, what historical connection is there to account for Rowling's 'Harry Potter' referring to one of countless nonexistent candidate objects (each equally faithful to the way Potter is depicted in the novels but varying in details left unspecified—about sock color, etc.)? The Meinongian cannot provide such a historical connection: causal-historical connection between his timelessly nonexistent objects and actual concreta (like authors) is extremely problematic, downright unintelligible even. And because of historical unconnectedness, the Meinongian is confronted with qualms about having to work with no more than coincidental resemblance between Harry Potter as specified in the novels, and various qualitatively identical Meinongian nonexistents. And, on the one hand, coincidental resemblance does not suffice for reference, according to the epistemological thesis; and, on the other hand, with insufficiently

specified characters like Harry Potter, coincidental resemblance leaves room for
the metaphysical thesis (and also the selection problem) to arise.[28]

Once fleshed out, Kripke's (1972) fleeting remarks about fictional characters
can be summarized as follows: qualitative resemblance is insufficient to deter-
mine the reference of a proper name; a causal-historical connection between
names and their referents is necessary to determine to whom or to what proper
names refer. For names of actual objects like 'J.K. Rowling' and 'London', this
overarching lesson transparently emerges from the second lecture of *Naming and
Necessity*. It is considerably less transparent that Kripke reiterates the very same
lesson for names of fictional characters. Of the forms of realism considered, arti-
factualism is the only one that can heed this lesson.

3. DEFLECTING THE CATEGORY-MISTAKE OBJECTION

Sainsbury, an advocate of irrealism, agrees with the conclusion of the previous
chapter: among realist contenders, artifactualism has the edge. Unlike Yagisawa
(2001), he doesn't think that the major challenge artifactualists face concerns
accounting for the truth of negative existential claims like "Harry Potter doesn't
exist". Instead, Sainsbury (2010, 111) writes:

> I see the problems for abstract artifact theory lying elsewhere… They have the
> form: on abstract artifact theories, fictional characters just are not the kinds of
> things we want them to be. We want them to be as they are said to be in the sto-
> ries, to be detectives and to play the violin, but they are said to be something of an
> entirely different kind.
> …
> Fictional characters do not have the properties they are ascribed during their crea-
> tion. This is mysterious: Conan Doyle stipulates that Holmes wears a deerstalker,
> there is such an entity as Holmes, yet that entity does not end up having (i.e. ex-
> emplifying) the property of wearing a deerstalker. He does end up having (exem-
> plifying) a genuine property, that of encoding wearing a deerstalker, but this is not

---

[28] This line of argument brings to the fore why the only abstract-theory contender we con-
sidered for fictional characters was artifactualism: it is the only view according to which Harry
Potter is created and hence historically linked to goings on in the actual world. Platonism,
a theory according to which Harry Potter is a *timelessly existing* abstract object (akin to num-
bers, sets), would, like Meinongianism and nonactualism, run into problems with historical
unconnectedness and hence coincidental resemblance, and, in their wake, the metaphysical
and epistemological theses. For an attempt to combine the advantages of artifactualism and
Meinongianism, see Zalta's (2000, 2006).

a property that's intellectually accessible to most authors. People can, of course, fail to understand what they are doing, but it's surprising to be told that so many authors, perhaps all, fail so often and so seriously.

Sainsbury is here relying on a distinction between exemplifying and encoding originally suggested by Meinong's student Mally (1912): a concrete object like J.K. Rowling doesn't encode any properties; but she does *exemplify* being British and fails to exemplify wearing glasses. Meanwhile, Harry Potter *encodes* wearing glasses and being British, but exemplifies neither of these properties. He does, however, exemplify being abstract and being a fictional character.

Part of Sainsbury's objection then is that according to artifactualism, fictional characters are of the wrong ontological category—abstract rather than concrete—to exemplify the sorts of properties ascribed to them by the authors who create them. I call this *the category-mistake objection*. A consequence of the category-mistake objection is that artifactualism attributes massive error to those who create, read and discuss fictional characters.

My aim is to show that the strategy behind the category-mistake objection, if it were to work, would show far too much with respect to a broad range of metaphysical debates. The strategy is therefore objectionable. I will formulate three arguments to show that the category-mistake strategy does not withstand scrutiny.

My first argument is about the metaphysics of possible worlds. In Section 1, we distinguished two positions in the debate about the nature of possible worlds: realism versus ersatzism about possible worlds (pw-realism and pw-ersatzism). According to pw-ersatzism, possible worlds are abstract, for example, maximally consistent sets of propositions representing ways the world could be. And merely possible individuals are likewise abstract (as Sainsbury acknowledges): representations comprising propositions about the individual. Now, when I consider a counterfactual scenario in which I dye my hair green today, I am ascribing to myself the property of having green hair, or so it seems to me when I reflect on my mental episode. Yet the category mistake objection could be raised here: according to pw-ersatzism, possible objects are of the wrong ontological category—abstract rather than concrete—to exemplify the sorts of properties ascribed to them by those who entertain counterfactual scenarios. This objection would apply to all forms of pw-ersatzism, regardless of whether they construe worlds in terms of states of affairs, universals or sentences. On all these versions, possible objects are the wrong kinds of things to be ascribed the properties we ordinarily ascribe to them. Anyone who thinks pw-ersatzism cannot be dismissed quite so

quickly has reason to consider the strategy behind the category-mistake objection (as targeting artifactualism as well as pw-ersatzism) specious.[29]

Another point casts further doubt on the category-mistake strategy. The category-mistake objection against pw-ersatzism, if it were to work, would seem to leave the rather controversial position of pw-realism, famous for eliciting incredulous stares (Lewis 1973, 86), as the only realist account of possible worlds. According to pw-realism, the actual world is one among a plurality of possible worlds that are causally and spatiotemporally isolated from one another. A moment's further thought reveals that an objection closely related to the category-mistake objection affects pw-realism also. If an ordinary speaker were asked if she thought there are countless merely possible worlds and countless merely possible objects, and if she thought such things have the same ontological status as the actual world and actual objects, respectively, she would answer in the negative to both questions. Hence the incredulous stare that confronts pw-realism. Yet, contrary to people's intuitions, pw-realism posits that possible worlds have the same ontological status as that enjoyed by the actual world, and possible concrete objects have the same ontological status enjoyed by actual concrete objects. Call this the *mistaken-ontological-status objection to pw-realism*. It is unclear why this objection should have any less force than the category-mistake objection against pw-ersatzism. But if that objection were effective against pw-ersatzism while the mistaken-ontological-status objection were effective against pw-realism, then irrealist accounts of possible worlds would be left as the only alternatives standing. This conclusion seems much too quickly and easily obtained for irrealists about possible worlds (including Sainsbury). The pair of objections seem, from the outset, to rig the stakes against *all* forms of realism about possible worlds. Anyone who thinks that realism about possible worlds cannot be dismissed quite so easily has reason to consider at least one of the two objections specious. Until the irrealist about possible worlds provides special reasons that discredit the mistaken-ontological-category objection, both objections remain suspect.

My second argument is intended to show that for someone who finds a form of pw-ersatzism (a not unpopular view about the metaphysics of possible worlds) independently plausible, there is little reason to *resist* admitting fictional characters as abstract objects, the category-mistake objection notwithstanding. Here is

---

[29] An argument from authority (whatever its merits might be): interestingly, while Lewis (1986) carefully considered a long list of arguments against pw-ersatzism, he did not address the category-mistake objection against it.

Of course, one person's *modus tollens* is another *modus ponens*; I have motivated the following conditional: if the category-mistake objection is effective against artifactualism, then it is effective against pw-ersatzism. I have taken the *modus tollens* direction and concluded that the objection is ineffective against artifactualism. Someone else might take the *modus ponens* direction and conclude that the objection is a new and effective one against pw-ersatzism.

why. Imagine a certain spool of yarn I knit into a sock: Sock1, based on a specific set of knitting instructions. Imagine another specific, actual spool of the same yarn that I could have used to knit a qualitatively identical (or very similar) pair to Sock1: Sock2. As things stand, I never got around to knitting Sock2. So Sock2 doesn't exist, but it might have.[30] According to the pw-ersatzist, Sock2 is abstract: we might take it to correspond to a set of propositions representing Sock2 as having a certain color, shape, size, pattern, person knitting it; the proposition set can then be said to encode color, shape, knitter, etc., and exemplify being abstract and consisting of propositions. This set of propositions actually exists; but the scenario it represents is unactualized (because I never get around to knitting Sock2). Now consider the fictional sock that is featured at the end of the second Harry Potter novel: Harry pulls off one of his "slimy, filthy" socks, tricking Lucius Malfoy into unwittingly giving it to his long-suffering house elf, Dobby, thereby releasing Dobby from serving the Malfoy family (house elves are freed when their masters give them an article of clothing). J.K. Rowling doesn't give this sock a name, but for easy reference, let's call it DobbySock. According to artifactualism, J.K. Rowling created DobbySock; she specified it as having been worn by Harry Potter, as being slimy and filthy, but she didn't say what color it was. Whatever form of pw-ersatzism we might opt for, we can go the same way with DobbySock; for example, we can take DobbySock to correspond to a set of propositions representing it as being filthy, slimy, etc.; this set of propositions encodes DobbySock as being filthy, slimy, etc. (in the second Harry Potter novel); and the same set exemplifies being Rowling's creation, a fictional character, a famous fictional character even. We must realize that by taking these parallel approaches to Sock2 and DobbySock, there isn't that great a difference in the nature of the merely possible Sock2 and the artifact DobbySock: both correspond to sets of propositions representing socks, encoding certain properties (like being a sock) and exemplifying others (like being abstract, containing propositions). One noteworthy difference is that the first set *encodes* being created by me (the knitter), while the second set *exemplifies* being created by J.K. Rowling. But this is a difference we expect, and the similarities are otherwise striking. For a pw-ersatzist, with merely possible objects on board, it would be *ad hoc* to resist what is mostly parallel treatment for fictional characters: a form of artifactualism. Given how costly such an *ad hoc* move would be, the pw-ersatzist should embrace artifactualism and not worry about the category-mistake objection.

It is well to address three worries at this point. First, there is a crucial difference between Sock2 and DobbySock according to someone who combines

---

[30] Notice that 'Sock2' is a special name to which the inverse Sinatra principle discussed in Section 2 does not apply. And that's all well: 'Sock2' doesn't make it here but makes it in the possible worlds in which it gets knit.

pw-ersatzism and artifactualism: (on at least one plausible view of propositions), the proposition set for Sock2 exists timelessly, as do the pw-ersatzist's possible worlds; by contrast, DobbySock is an artifact and hence not a timeless existent. Why should the pw-ersatzist be moved to admit the latter kind of beast then, an object that is unlike her possible worlds and objects? Three reasons: (i) beyond this difference, there are crucial similarities between Sock2 and DobbySock, ones that make it plausible to treat both as abstract; (ii) operas, novels, the institution of marriage, etc. are overwhelmingly plausible candidates for abstract artifacts already; so the burden of providing a workable alternative is on those who want to deny that these are abstract artifacts; (iii) with operas and novels on board as abstract artifacts, between the Platonist and artifactualist alternatives, the latter is a far more tenable choice. For the Platonist view (mentioned in footnote 3 above)—according to which DobbySock is a timelessly existing abstract object—is affected by the selection problem, insufficient specificity, unsuited mode of introduction and coincidental resemblance; problems (discussed in Section 2) that artifactualism (singularly among realist contenders) avoids.

Second, notice that in contrasting encoding and exemplifying above, I have talked about *proposition sets* encoding and exemplifying properties like being knit by me and created by J.K. Rowling. Proposition sets represent ways the world might be; they are representational devices. And "[t]he distinction between encoding and exemplifying is one that is properly available for representational vehicles, but that's not what fictional characters are. They are what's represented", Sainsbury (2010, 112) objects. The worry is that fictional characters qua abstract objects aren't the right sorts of things to be representational devices and to be encoding properties. This worry is easily responded to: although possible worlds seem at first like really big particular objects, like all-encompassing, gigantic galaxies, the ersatz-realist does not balk at construing them as sets of propositions or as structural universals[31]. In the same way, the artifactualist should not worry about taking fictional characters to be sets of propositions (or as structural universals). That it is the proposition set about Harry Potter that encodes and exemplifies various properties is not a problem given that fictional characters correspond to such proposition sets. On the version of pw-ersatzism we are considering, "Sock2 encodes shape, size, pattern and exemplifies being abstract" is loose talk for "The 'Sock2' proposition set represents Sock2, encoding shape, size, pattern, and exemplifying being abstract". Likewise, on the version of artifactualism we are considering, "DobbySock encodes being filthy and slimy and exemplifies being abstract and created by J.K. Rowling" is loose talk for "The 'DobbySock' proposition set represents DobbySock, encoding filthiness,

---

[31] A paradigmatic example of a structural universal is *being a water molecule*: for an object to instantiate this universal, it has to have the right kinds of parts in the right kind of arrangement.

sliminess, and exemplifying being abstract and created by J.K. Rowling." Perhaps some of the pw-ersatzist's alternatives are conceptually more satisfying in some way than proposition sets; but the point stands: whatever kind of abstract objects the pw-ersatzist might posit as her possible objects, she has good reason to extend her theory to some very similar beasts: fictional characters as abstract artifacts. And her choice of construal for these objects can then accommodate the encoding/exemplifying distinction in much the same way as the proposition set construal did.

A third worry arises: aren't we multiplying abstract objects that are qualitatively identical to one another? Imagine a merely possible sock that is qualitatively identical to DobbySock, as specified in the second Harry Potter novel, call it JustLikeDS. The 'JustLikeDS' proposition set encodes the same properties as the 'DobbySock' proposition set encodes. The two sets exemplify some of the same properties: being abstract, being sets, consisting of propositions. Now, isn't it an extravagant proliferation of objects to hold that with JustLikeDS already in existence, J.K. Rowling creates a qualitative duplicate, DobbySock, upon conjuring up the second Harry Potter novel? We can see that this outcome is not worrisome at all if we reflect on some perfectly ordinary scenarios that are analogous.

Consider another abstract object, say, Mozart's opera *The Magic Flute*. Consider the collection of musical chord sequences, timing, order, etc. for the various singers and instruments, which the total score of *The Magic Flute* comprises. This is a type, which is a paradigmatic instance of an abstract object that can have specific performances of the opera as its tokens. Now, *The Magic Flute* qua abstract type was created by Mozart. But now consider a type—call it JustLikeMF—of chord sequences, instructions, various specifications qualitatively just like the score of *The Magic Flute*. If we take sets, properties and propositions to be timelessly existing abstract objects (a widely held position), then clearly, a type such as JustLikeMF should be regarded as an abstract object, plausibly, a timelessly existing one. But then when Mozart came along, he ended up proliferating qualitatively identical types by writing the score of *The Magic Flute*, in addition to the timelessly existing JustLikeMF. Proliferation of this sort is inevitable, yet it isn't taken as cause for concern for those who posit types as abstract, and musical pieces as types created by composers.

Consider a similar example: the swiftly created and enacted new Hungarian constitution ("Fundamental Law" it's called) didn't always exist; it came into existence in 2011 only, when it was drafted; indeed, beforehand, many considered it unfathomable that an object like the Fundamental Law should ever be created; but it was. The Fundamental Law is a type that can have instances: printed and electronic copies, a reading event of the text. The Fundamental Law didn't exist before the current government came into power, but it exists now. Yet a qualitatively identical type, an ordered sequence of propositions, is

plausibly an abstract object that existed well before 2011, if not timelessly[32]. Proliferation of this sort is inevitable if we want to maintain that the Fundamental Law is an artifact created in the recent past while types are abstract also. And if proliferation is no cause for concern here, it isn't worrisome in the case of DobbySock and JustLikeDS either.

Consider a third example: words being added to the English vocabulary. For example, a fairly recent addition to the English language is the expression 'cot potato', meaning a very young child who spends a lot of time watching television. The expression type 'cot potato' can have hand-written, typed, electronic, spoken, mouthed or signed tokens. Linguists tend to take for granted that expression types are abstract objects, specifically, abstract artifacts that didn't always exist. But (relative to 'cot potato') a qualitatively identical phonological type, orthographic type, and semantic type qua abstract types have been around for much longer,[33] so with the addition of new words like 'cot potato' to the English language, we get a the very same kind of proliferation that DobbySock and JustLikeDS had presented; and this sort of word proliferation is rampant: for any expression type of any language, there is a qualitative duplicate that is an antecedently existing abstract object. If that isn't worrisome, nor is the case of DobbySock and JustLikeDS.

The upshot of these examples is that proliferation of qualitative duplicates is inevitable for abstract artifacts across the board. If (like many theorists) we still want our ontology to make room for works of art, social and legal institutions, games, words within a language, traditions, festivals, religions as abstract artifacts, then we should have no qualms about including fictional characters on the list. And overall, the upshot of my second argument has been that for pwersatzists, resisting artifactualism would be an *ad hoc* move.

My third argument is that the category-mistake objection can be extended to a broad range of objects that are commonly regarded as abstract artifacts. If we don't balk at the objection there, we should pay no head to it with respect to artifactualism either. Were those who were coming up with and modifying the rules of chess thinking of themselves as creating something abstract? Were the writers of the Fundamental Law of Hungary thinking of themselves as creating something abstract? Was Mozart, when composing *The Magic Flute*? Were those who coined the term 'cot potato'? If these people were interviewed, they would

---

[32] Whether or not we take the ordered sequence of propositions to exist timelessly depends in part on our view of propositions, an issue on which I'd like to maintain neutrality. Either way, the qualitatively identical type enjoys prior existence relative to the Fundamental Law.

[33] I avoid talking about timelessly existing types here for the sake of neutrality on various matters. I want to leave open the possibility that the orthographic type 'cot potato' doesn't exist timelessly because it didn't exist prior to the English orthographic system coming into existence. Likewise, I want to leave open the possibility that the semantic type 'cot potato' doesn't exist timelessly because it didn't exist prior to the existence of television sets.

likely be baffled by the idea that they were aiming at creating abstract objects. Yet regarding the game of chess, the Fundamental Law, *The Magic Flute*, and words of English as abstract artifacts is a plausible option, more plausible than its alternatives, and one that many consider platitudinous. Then why worry about the category-mistake objection against artifactualism? [34]

A defender of the category-mistake objection can make a comeback: all the abstract artifacts considered here are types capable of having concrete, spatio-temporal instances such as specific games of chess, copies of the Fundamental Law, performances of *The Magic Flute*, utterances of 'cot potato'. The creators of these abstract artifacts *were specifying conditions for the instances of the types they were creating*. But there is a crucial disanalogy between the type-like abstract artifacts just considered and fictional characters qua abstract artifacts: the latter are not the kinds of things capable of having instances. Hence the application of the category-mistake objection to the latter.

The artifactualist can respond to this in three ways. Granted: types can have instances. But it is not so outlandish to think of fictional characters as having something a bit like instances: for example, an opera singer singing Papagena in *The Magic Flute plays* Papagena; though admittedly, she *isn't* Papagena. In the Harry Potter movies, there were several actors *playing* Professor Dumbledore; though admittedly, none of them *was* Dumbledore. We might insist, however that the opera singer and the actors *are*—within the fiction—instances of the characters they play, so the disanalogy isn't as great as the defender of the category-mistake objection makes out.

A second and more substantial line of response from the artifactualist: consider, again, merely possible objects like Sock2, which are plausibly construed within a pw-ersatzist framework as, for example, a set of propositions. This proposition set can be instantiated in a sense: it can be *actualized*—it is actualized when I end up knitting Sock2. We can think in terms of this model of actualization for fictional characters like DobbySock qua abstract artifact also. It is just that for reasons explained in Section 2, fictional characters are forever unactualized, indeed, unactualizable. The inverse Sinatra principle sums up a crucial feature of names, fictional names included: if they don't make it here, they don't make it anywhere; that is, if they don't refer to concrete, spatiotemporal objects in the actual world, then they don't do so with respect to other possible worlds either. Fictional characters by their very nature are barred from being actualized. But

---

[34] I am symphathetic to Thomasson's (1999; 2009, 16) point "that those who accept the existence of such ordinary social and cultural objects as laws, marriages, symphonies, and works of literature themselves are apparently already committed to the existence of created abstracta, so that no special problems arise in accepting created abstracta to account for fictional characters", and "it is not obviously more parsimonious to do without fictional characters if we must posit abstract artifacts in some other arena, e.g. to make sense of our talk about novels, symphonies, laws of state, and the like."

this is a feature of theirs due to considerations about coincidental resemblance being insufficient for reference. This feature does not make for a decisive point of disanalogy between them and merely possible individuals. So granted, there is a special reason why in one sense, fictional characters qua abstract objects aren't ever instantiated (that is, actualized); but that need not make them radically different from types.

Third, consider the tradition of the Easter rabbit and its Australian marsupial variant, the Easter bilby. What exactly might be concrete, spatiotemporal instances of these traditions—events involving nest-making, egg-painting, candy-hiding, candy-hunting?—is somewhat puzzling. So fictional characters aren't the only abstract artifacts for which it isn't straightforward what their instances might be. But saying that the tradition of the Easter bunny is therefore not an abstract artifact while the other instantiable types are, invites the challenge for the defender of the category-mistake objection: on what basis will she count among her abstract artifacts the tradition of the Village Halloween Parade in New York City and Mardi Gras festivals, but exclude the tradition of the Easter bilby?

The upshot of the third argument is that the difference between types and fictional characters qua abstract artifacts isn't as great as it initially appears. Yet the category-mistake objection applies to a broad and varied range of types: the game of chess, *The Magic Flute*, words. If the objection is ineffective there, we have little reason to worry about it when it comes to fictional characters conceived as abstract artifacts.[35]

4. CONCLUSION

My aim in this paper has been to motivate taking at face value claims about fictional characters being created by their authors. This requires a form of realism about fictional characters: they are supposed to *exist* once created. Among the contending realist theories about fictional characters that Sainsbury (2010) considered, artifactualism, which takes Harry Potter to be an abstract artifact, is the only one according to which fictional characters are created. Based on Kripke's

---

[35] Notice that the third argument is not defeated by those who contest the abstract artifact status of one or another among the various candidates I have enumerated. As long as these philosophers are moved to grant abstract artifact status to *some* of the examples mentioned—operas, novels, words or the game of chess—the third argument affects them already. Meanwhile, for philosophers holding that *none* of the items listed in Section 1 as abstract artifacts are in fact such (because they suggest, say, that these aren't abstract, after all), the burden of proof is on them to explain how operas, novels, words, etc. are nonabstract.

(1972) brief considerations about expressions like 'unicorn' and 'Sherlock Holmes', I constructed a general argument showing that artifactualism is superior to its realist rivals. Sainsbury accepts this conclusion, yet argues that one should ultimately reject realism about fictional characters because artifactualism faces insurmountable difficulties due to the category-mistake objection. I gave three arguments showing that the category-mistake objection is problematic because if it were to work, it would show too much: first, it would show ersatzism about possible worlds to be a nonstarter; second, it would prevent the ersatzist from taking on board fictional characters as abstract artifacts, an *ad hoc* move for her; and third, it would cast doubt on the abstract artifact status of a broad range of social and cultural artifacts like the game of chess, words of English, bodies of law, novels and operas. *Pace* Sainsbury, artifactualism about fictional characters remains unscathed by the category-mistake objection.

In the process of disarming the category-mistake objection, I aimed also to demystify what Harry Potter as an abstract artifact might be. Let me close with two further demystifying considerations. Brock (2010) deems a position like artifactualism "an abject failure" because it cannot offer a satisfactory account of the spatial and temporal dimensions of fictional characters, for example, their moment of creation. This line of attack ignores that quintessential examples of abstract artifacts like the institution of marriage and the game of chess are just as difficult to locate in space and time as Harry Potter is. In particular, when each came into existence is a thorny question: there is extensive debate as to which stage of rules for a board game to count as the birth of chess—the 15th century, the 17th, or the 19th? Most likely, no agreement will be reached on this issue; but that hardly casts doubt on the abstract artifact status of the game of chess. Nor has Brock given us reason to doubt the abstract artifact status of Harry Potter.

Consider a related point about criteria for individuating abstract artifacts: does a board game very much like contemporary chess but with different rules about stalemate count as *chess* at an earlier stage? How much meaning change and/or pronunciation change can a *word of English* undergo and still remain the same word? Can the institution of *marriage* be modified such that the two parties may be of the same sex? The considerations that could decide such questions seem unclear and arbitrary. Yet the lack of clear answers doesn't call into question the abstract artifact status of the game of chess, words of English and the institution of marriage. Abstract artifacts can and do change over time as humans shape and modify them. In the light of this, it is only expected that Harry Potter qua abstract artifact may change as the film adaptations fill in details (for example, the sock Harry sneaks to Dobby is black) or contradict the original novels (for example, in the novel, as Harry gets ready to trick Dobby's master into giving Dobby a sock, Harry hides the diary of Tom Riddle inside the sock, while in the movie, he hides the sock inside the diary). Or maybe the Harry of the novel is a distinct artifact than the Harry of the movies? Again a clear and nonarbitrary

answer seems doubtful. But those who, upon encountering comparable issues with quintessential abstract artifacts like words of English and the institution of marriage did not speak up, should make peace with Harry Potter as an abstract artifact.[36]

BIBLIOGRAPHY

Adams, Robert M. 1974. Theories of Actuality. *Noûs* 8: 211–231.
Braun, David 2005. Empty Names, Fictional Names, Mythical Names. *Noûs* 39: 596–631.
Brock, Stuart 2002. Fictionalism about Fictional Characters. *Noûs* 36: 1–21.
Brock, Stuart 2010. The Creationist Fiction: the Case against Creationism about Fictional Characters. *Philosophical Review* 119: 337–364.
Caplan, Ben 2004. Creatures of Fiction, Myth, and Imagination. *American Philosophical Quarterly* 41: 331–337.
Fine, Kit 1982. The Problem of Non-Existents. I. Internalism., *Topoi* 1: 97–140.
Kaplan, David 1973. Bob and Carol and Ted and Alice. In K.J.I. Hintikka, J.M.E. Moravcsik, and P. Suppes (eds.) *Approaches to Natural Language*. Dordrecht: Reidel, 490–518.
Kaplan, David 1989. Afterthoughts. in Joseph Almog *et al.* (eds.) *Themes from Kaplan*. Oxford: Oxford University Press, 565–614.
Kripke, Saul 1963. Semantical Considerations on Modal Logic. *Acta Philosophica Fennica* 16, 83–94.
Kripke, Saul 1971. Identity and Necessity. In Milton K. Munitz (ed.) *Identity and Individuation*. New York: New York University Press, 135–164.
Kripke, Saul 1972. Naming and Necessity. In Donald Davidson and Gilbert Harman (eds.) *Semantics of Natural Language*, 2nd edition. Dordrecht: D. Reidel, 253–355; Reprinted with Preface, *Naming and Necessity*. Cambridge, MA: Harvard, 1980, page references are to the reprint.
Kripke, Saul 1973. Reference and Existence (The John Locke Lectures), The Saul Kripke Center Archives, The CUNY Graduate Center. (Copies available from The Saul Kripke Center upon request.)
Lewis, David 1973. *Counterfactuals*. Oxford: Blackwell.
Lewis, David 1978. Truth in Fiction. *American Philosophical Quarterly* 15: 37–46. Reprinted in David Lewis, *Philosophical Papers* (Volume 1), Oxford: Oxford University Press, 1983, 261–275.
Lewis, David. 1986. *The Plurality of Worlds*. Oxford: Blackwell.
Mally, Ernst 1912. *Gegenstandtheoretische Grundlagen der Logik und Logistik*. Leipzig: Barth.
Meinong, Alexius 1904. Über Gegenstandtheorie. In Alexius Meinong (ed.) *Untersuchungen zur Gegenstandtheorie und Psychologie*, Leipzig: Barth. Translation by I. Levi, D.B. Terrell, and R.M. Chisholm, in Roderick Chisholm (ed.) *Realism and the Background of Phenomenology*. Glencoe: Free Press, 1960, 76–117.
Parsons, Terence 1980. *Nonexistent Objects*. New Haven: Yale University Press.

Rosen, Gideon 2012. Abstract Objects. In Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2012 Edition), URL = <http://plato.stanford.edu/ archives/spr2012/ entries/abstract-objects/>.

Sainsbury, Mark 2010. *Fiction and Fictionalism*. London: Routledge.

Salmon, Nathan 1998. Nonexistence. *Noûs* 32, 277–319.

Searle, John R. 1979. The Logical Status of Fictional Discourse. In P.A. French *et al.* (eds.) *Contemporary Perspectives in the Philosophy of Language*. Minneapolis: University of Minneapolis Press, 233–243.

Schiffer, Stephen 1996. Language-Created Language-Independent Entities. *Philosophical Topics*, 24, 149–166.

Thomasson, Amie L. 1999. *Fiction and Metaphysics*. Cambridge: Cambridge University Press.

Thomasson Amie L. 2009. Fictional Entities. In Jaegwon Kim, Ernest Sosa, Gary Rosenkrantz (eds.) *A Companion to Metaphysics*, Second edition. Oxford: Blackwell, 10–18.

Van Inwagen, Peter 1977. Creatures of Fiction. *American Philosophical Quarterly* 14: 299–308.

Van Inwagen, Peter 2000. Quantification and Fictional Discourse. In Anthony Everett – Thomas Hofweber (eds.) *Empty Names, Fiction, and the Puzzles of Nonexistence*. Stanford: CSLI Publications, 235–247.

Walton, Kendall L. 1990. *Mimesis as Make-Believe*. Cambridge, MA: Harvard University Press.

Yagisawa, Takashi 2001. Against Creationism in Fiction. *Philosophical Perspectives* 15: 153–172.

Zalta, E.N. (1983). Abstract Objects: An Introduction to Axiomatic Metaphysics, Dordrecht: D. Reidel.

Zalta, Edward N. 2000. Pretense Theory and Abstract Object Theory. In Anthony Everett – Thomas Hofweber (eds.) *Empty Names, Fiction, and the Puzzles of Nonexistence*. Stanford: CSLI Publications, 117–147.

Zalta, Edward N. 2006. Deriving and Validating Kripkean Claims Using the Theory of Abstract Objects. *Noûs* 40/4: 591–622.

Zvolenszky, Zsófia 2007. Naming with Necessity. In *Modality, Names and Descriptions*. Ph.D. thesis, New York University.

Contributors

GÁBOR BÁCS • assistant professor, Department of Social Science, University of Kaposvar, Hungary. Areas of specialization: metaphysics.
E-mail: bacs.gabor2@gmail.com

THOMAS M. Crisp • associate professor, Department of Philosophy, Biola University, USA. Areas of specialization: metaphysics, epistemology.
E-mail: thomas.m.crisp@gmail.com

FERENC HUORANSZKI • full professor, Department of Philosophy, Central European University, Hungary. Areas of specialization: metaphysics.
E-mail: huoransz@ceu.hu

HOWARD ROBINSON • full professor, Department of Philosophy, Central European University, Hungary. Areas of specialization: philosophy of mind, metaphysics.
E-mail: robinson@ceu.hu

PETER VAN INWAGEN • full professor, Department of Philosophy, The University of Notre Dame, USA. Areas of specialization: metaphysics, epistemology.
E-mail: vaninwagen.1@nd.edu

JÁNOS TŐZSÉR • associate professor, Department of Social Science, University of Kaposvar, Hungary. Areas of specialization: philosophy of mind, metaphilosophy.
E-mail: jantozser@gmail.com

ZSÓFIA ZVOLENSZKY • associate professor, Department of Philosophy, Eötvös Loránd University, Hungary. Areas of specialization: metaphysics, philosophy of language.
E-mail: zvolenszky@nyu.edu