Issue 12 (1) Annual Conference 2018

British Journal of Undergraduate Philosophy

Editor-in-chief: Alexander Raubo
London School of Economics

Journal of the British Undergraduate Philosophy Society

British Journal of Undergraduate Philosophy

British Journal of Undergraduate Philosophy

The Journal of the British Undergraduate Philosophy Society

Issue 12 (1) Annual Conference 2018

www.bups.org.uk

British Journal of Undergraduate Philosophy

www.bups.org.uk

Cover image used with the kind permission of Sophia Ben-Achour.

Editorial

The work is the death mask of its conception.

– Walter Benjamin

Editorials, for those who do not write them frequently, appear as rhetorically ambiguous texts. The aims seem clear: to introduce or discern the theme binding the proceeding texts together, to reflect on the events that preceded publication, to acknowledge the work of colleagues and contributors. The (possibly) verifiable effect would be a reader that reads the journal attuned to the parts that are novel and the parts that are instructive, conscious of their value. But what is the best way of producing this effect? How do you weigh the competing aims?

Some writers approach an editorial as a eulogy, but there are not always things to praise. Nor, for that matter, is the hyperbolic nature of praise (which many counteract with the qualification that in this case, hyperbole is more than appropriate, to no rhetorical avail) always understood as sincere or discerning. Others –more mundanely –see them as bulletin boards for recent activities, but in doing so, miss out on a rare opportunity to signal guiding values not easily discernible in reportage. A third kind take the editorial as an opportunity to opine and provoke, often with an air of glory and the feebleness of a literary debutante. It should be acknowledged that such partiality toward ersatz monumentality is rarely afforded the philosopher, lest they eschew disinterested contemplation for a quick hit of journalistic buzz. Let us disregard this last style and proceed, modestly (a virtue for the analytics), by recounting some novelties.

Edmund Smith has ascended from his commissioning days to assume the Presidency of BUPS. His indefatigable enthusiasm for philosophy and splendid erudition inspires unqualified admiration. Few people are able to competently broach such diverse subjects as Adorno, anthropology, Hegel, pragmatism, and set theory and make them all look worthy of serious study. Perhaps most laudable is his generosity as an educator, which can be gauged by reading the many pieces he has authored for our Facebook page.

Under Edmund's presidency, BUPS has adopted a change in its constitution: our goal is no longer to publish a journal that would be indistinguishable from a professional philosophy journal. Of course, no one has ever expected a representative sample of undergraduate work to be on the appropriate level for such indiscernibility to occur. Then again, our admission standards have always been too high for any such representative sample to ever appear in our pages. And here lies the crux. For if we were at any point under the illusion that our reach was so broad and deep as to consistently unearth the rare undergraduate work so complete and sophisticated as rival something published in professional journals, the ambition to build a reputation on quality and quality alone would seem perfectly sensible. But would we then still be an undergraduate philosophy journal?

Reading some of our first editorials written by Robert Charleston reveals how concerned our founders were about discussing and reflecting upon the runtiness of the undergraduate experience. The topics of Charleston's pieces in our first edition gave advice about how to deal with rejection (by funding bodies, universities, even BUPS), how to write philosophy, and charted the reasons for studying philosophy. None of this would creep into a professional journal, or at least not with the felt urgency and sincerity, yet its educational value is clear. So, concurrently with the spirit of our founders, our revised constitution now features the aim of complementing undergraduate philosophical education. The main means by which this aim is to be realised has always been at the core of what BUPS does: publishing a journal that provides students with experience in writing and submitting philosophy, organising conferences that encourages its participants to become better at presenting and discussing work, facilitating networking of undergraduates across Britain that share a love of knowledge. But with this change we hope to emphasise this educational aspect.

This emphasis has been most visible on our Facebook pages, where we have begun to post regular philosophy related content (such as Edmund's short pieces mentioned earlier). We are very happy that these updates have sparked some responses and hope that we will extend the reach of BUPS by providing this new means of philosophical engagement.

Speaking about our online presence, Sophie Osiecki, our most senior committee member, has been absolutely indispensable in her efforts at securing and managing the funding required for our operations (including our website) as well as running our marketing efforts. It has become a bit of a BUPS trope to say that Sophie's role, which according to the masthead of the journal is Manuscript Editor and Peer Reviewer, touches *de facto* on every aspect of the committee work, but–since it is true–it is worth repeating.

On the journal production side of things, a near-trope (or at least, a slight variation on a standard BUPS trope) is to describe Nathan Oseroff as the consummate Editorial Officer. Again, the trope would not be a trope if the description was not apt. His consistency and skill in typesetting and proof-editing seems to save our face every time we publish by making the journal look as presentable as it does.

Yusuf Tayara, trope-less–but only due to having joined the committee only recently–has proven himself in running the commissioning system. All the papers that you see in these pages have at multiple stages been under his watchful, coordinating eye. Many thanks as well to our Peer Reviewers – Matei Gheorghiu, Anne Deng, and the eminent BUPS emeritus Farbod Akhlaghi-Ghaffarokh – as well as our Manuscript Editors – Benedetta Delfino, (again) Matei Gheorghiu, and Oli Woolley – for their work.

Concludingly, it is worth remarking that editorials are also–and perhaps primarily–valedictions: to the time spent preparing the journal, to our younger enthusiasms, sometimes even to the topic that follows. We are finished and need to move on, perhaps in the same vein, but never in the same way.

AR

Contents

59. On the Structure of Mental Processes: The Systematicity Debate

Ellen Jin

Knowledge and Power in Relation to Thomas Kuhn's Incommensurability of Paradigms[*]

Matthew Colin Sayce
St. Mary's University

The term 'paradigm', which was popularised–but not invented–by Thomas Kuhn, is argued to have been used in 'several different ways' [4, 77] in his book, *The Structure of Scientific Revolutions* (hereafter, *Structure*). The definition most commonly used, however, seems roughly equivalent to a paradigm which comprises 'methods for gathering and analyzing data, and habits of scientific thoughts and action', and these methods and habits 'when combined, make up both a view of the world and a way of doing science' [4, 77]. That is to say, a paradigm is the currently existing, accepted way of practicing science within a particular framework. The way in which scientists approach scientific problems, how they are identified, and how they are solved, is dictated by the paradigm[1].

The problem of incommensurability was 'discovered' by Kuhn early in his career: he noticed 'conceptual differences' between current and historical work, when the terminology used was the same [9]. He concluded from this that the differences 'indicated breaks between different modes of thought', or paradigms, and that these different modes were entirely distinct from each other, and 'significant both for the nature…and…the development of knowledge' [9]. Incommensurability is more than mere incompatibility; incommensurability 'refer[s] to various factors that make the evaluation of competing theories problematic' [8, 216]. For paradigms to be incommensurable, then, is for two competing methods of conducting scientific practice, to hold different conceptions and induce different ways for their adherents to conceive of the world.

It has been argued that Kuhn's work has attacked the 'traditional viewpoint that scientific knowledge is certain, stable, and progressive' [10, 37], and, as noted above, that the incommensurability of modes of thought has consequences when ascribing knowledge to scientists. I argue that through his work on the incommensurability of paradigms, Kuhn has revealed the temporary nature of scientific knowledge, and that even with an alternative representation of a paradigm, this impermanence of knowledge remains

[*] Delivered at the BJUPS Annual Conference, 14–15 April 2018 at London School of Economics.

[1] There are other uses of the term 'paradigm' in *Structure*, although with what is, arguably, the most important definition established, there is no need to delve further.

true. Furthermore, I argue that the existing paradigm is maintained by those academic, hierarchical power structures–from the top-down and the bottom-up–and that the benefits of this are bidirectional. I consider objections to Kuhn's work on the incommensurability of paradigms and conclude that, despite some scholarly claim, the incommensurability of paradigms is still a notion that affects knowledge and power in science.

The incommensurability of paradigms in relation to knowledge, it could be argued, demonstrates that scientific knowledge is not permanent or immovable. In virtue of the fact that paradigms are how scientific problems are approached, it has been argued that '[w]hen a problem is solved there is knowledge production' [1, 588], which is to say, scientific knowledge is produced when a problem is successfully solved according to the current paradigm. This seems to corroborate common-sense thinking; a successful scientific experiment, or breakthrough, contributes some new knowledge to that which is already known.

However, in *Structure*, Kuhn writes that 'the proponents of competing paradigms practice their trades in different worlds' [6, 150]. This claim is both the 'most fundamental aspect of incommensurability' and, arguably, 'the least intelligible' [5, 483]. What Kuhn's explanation of this statement claims, is that 'two groups of scientists see different things when they look from the same point in the same direction' [6, 150]. As such, scientists on either side of a scientific revolution, which is when 'an older paradigm is replaced...by an incompatible new one' [7, 86], would view the same phenomena in two distinct ways. If the two paradigms are incommensurable in this way, that is to say, they are 'mutually exclusive' [5, 483] with little or no translatable content, then the kind of knowledge that is produced will be different.

Indeed, within the new paradigm, there is a possibility to produce knowledge which could not have been produced in the previous paradigm. Kuhn provides the example of the disbelief towards curved space in 'Einstein's general theory of relativity'; there was such disbelief because according to the previous Newtonian paradigm 'space was necessarily flat' [6, 149]. In this case, the competing paradigms demonstrate two different kinds of knowledge, insofar as the knowledge produced by either paradigm is of a distinctly different subject matter. Although the concepts share the name of 'space', what is being referred to, or what the knowledge is of, is fundamentally different in each paradigm. The knowledge that it is possible to produce, then, differs according to each paradigm, meaning that not only does ones understanding, or perception, of the world change, but the knowledge that is created by scientific discovery also does.

The way in which [1, 588] consider the paradigm, however, is striking. The authors discuss the paradigm as a 'stock of ideas' that can be 'harvested, to produce science', but can also be 'exhausted and depleted', and when this stock runs out, they argue that the 'paradigm becomes unable to solve problems' [1, 588]. This notion is unconvincing; a paradigm is a method, 'a way of doing science' that allows problems to be solved [4, 77], it is not often considered as a stockpile of ideas that can be taken, or 'exploited' [1, 588]. The authors seem to represent the paradigm as containing a finite, tangible amount of ideas, which does not seem correct. One could ask, at what point is the paradigm's last idea harvested? It is not obvious that the authors could provide a convincing answer. This alternative representation of the paradigm notwithstanding, the argument that the incommensurability of paradigms brings to a halt the production of a particular kind of knowledge, remains. Whether the paradigm is represented as a method, or as a stock of ideas, the incommensurability of two, competing paradigms, raises the possibility of production of a different knowledge.

There is an argument to be made that the relationship between power and the dominant paradigm is bidirectional. That is, the paradigm is maintained by those in positions of power, and those in positions of power benefit from the paradigm. [1] argue that '[p]aradigms transform groups of researchers into a profession, and lead to the formation of professional bodies' (p.588). In this instance, then, the paradigm is beneficial to researchers and professionals, who work within the framework and concepts of the paradigm, because it can elevate them within the academic hierarchy. Furthermore, this hierarchy is 'developed ... to organize and select what is in accordance with the paradigm, i.e. the hierarchy among scientists serves to preserve the paradigm' [1, 588]. What the argument appears to be, here, is that the scientists who adhere to the dominant paradigm are more likely to subsequently find themselves put into positions of power, and once in power, those individuals seek to preserve the dominant hierarchy. This is an interesting argument, but one might still ask how this distribution of power affects the distribution of knowledge.

The positions of power, in which the proponents of the dominant paradigm find themselves, are those such as 'the editors of academic journals... who decide the direction of research' [1, 588]. This means that the proponents of the paradigm have control over what is, arguably, the primary source of knowledge for their particular academic community. Indeed, the editors of academic journals can select research they want to publish, and thereby influence the research of many others. The editors have the final say on 'what constitutes new and valuable knowledge' [1, 589], and because there is a chance that they will be in that position of power due to their adherence to the paradigm, it seems likely that they will favour research that adheres to the same. With near exclusive control over the direction an academic journal can take, it is argued that the editors 'ultimately, decide upon the life or death of a paradigm' [1, 589]–although, one should be sceptical as to whether one can necessarily wield this much power with one journal. By adhering to the paradigm, then, an individual can gain power within their field, and with that power, they can strengthen the dominance of the paradigm.

The relationship between power, knowledge, and paradigms has been demonstrated–but what of the incommensurability of paradigms? It seems to follow from the fact that proponents of the paradigm find themselves in positions of power, that if a competing paradigm were to manifest, it would be relatively easy for the proponents of the dominant paradigm to dispel the idea of the competitor as progress. That is, of course, not to say that this is the case, or that this should be the case, but there is the possibility for it. This may not exclusively be because the editors of the academic journals make the decision to ignore research from the competing paradigm, or prioritise research from the dominant; the resistance could also come from the bottom of the academic hierarchy. [1, 589] claim that 'researchers aim at gaining professional promotion, peer recognition, respect and reputation by publishing their research in academic journals'. If a researcher wants to be published by the journal, they would likely have to subscribe to the same paradigm–perhaps against their better judgement. In this case, those with power are not the only parties affecting the dominant paradigm, it also comes from the lower tiers of the hierarchy. If a researcher wishes to contribute to her field of study, it is very likely that she will have to adhere to the dominant paradigm.

It is noteworthy, that a physicist–with a sufficient knowledge of two paradigms that are frequently posited as competing, namely 'Newtonian mechanics' [2, 57] and 'Einstein's special theory of relativity' [2, 58]–might object to the claim of their incommensurability. Initially discussed by [2], and further developed some years later [3], the author provides reasons to claim that through a 'procedurally defined experiment' using 'theory neutral' terms and procedures, there can be a 'neutral observation language' [3, 157], which distinguishes between the two competing paradigms. It is claimed that because adherents to either paradigm would 'agree on the respective predictions and on the measurement', both 'paradigms can be compared' and are 'commensurable' [3, 158]. This demonstrates how an individual from another discipline might begin to deal with the relationship between paradigms and knowledge. They may simply claim that knowledge is not necessarily affected by Kuhn's incommensurability of paradigms, because it is not always the case that competing paradigms cannot be comparable or commensurable. Although it is likely that any such response from Franklin would be less modest–his argument concludes with the claim that the 'philosophical problems associated with both theory-ladenness and incommensurability have been solved' [3, 165]. One might object that the argument proffered can not necessarily be extrapolated to all instances of incommensurability, and his conclusion is a little bold. Only with substantial scientific knowledge of at least two competing paradigms, however, could one begin to posit the kind of response Franklin may offer, or indeed the subsequent rebuttal.

It should be acknowledged, also, that in the *Postscript to Structure*, Kuhn writes that 'philosophers have seriously misconstrued the intent of...parts of my argument', and that the philosophers have accused him of believing that 'proponents of incommensurable theories cannot communicate with each other at all' [6, 198]. It seems like this note from Kuhn is similar to Franklin's response, and that the discipline of philosophy has presupposed a total incommensurability across all elements of the paradigm. It seems that this may not have been what Kuhn was aiming to convey, and a scientific approach has shown it to be not the case.

It has been demonstrated that the paradigmatic scientific landscape that Kuhn discussed has a direct effect on knowledge and power. Specifically, the incommensurability of paradigms means that a particular knowledge is produced, while another knowledge is never actualised, because another paradigm is never dominant. Further to this, paradigms are perpetuated by individuals who attain positions of power by adhering the methods and beliefs propounded by that same paradigm–these individuals can, for example, become editors of academic journals and exert a large influence on the way research is then conducted. Paradigms perpetuate power, while those in power continue to perpetuate the paradigm. From a scientific perspective, however, the entire notion of incommensurability of paradigms can be brought into question by demonstrating that competing paradigms can be commensurable; they can share results and some measurements. This demonstrates the importance for philosophy to engage with other disciplines; while interpretation of Kuhn's exact use of terminology is, undoubtedly, important, it is just as important to establish whether the theory is watertight. Scientific research has demonstrated that it is not. Despite this, the effects of the incommensurability of paradigms on knowledge and power in the scientific landscape are evident.

References

[1] Faria, J. R., Besancenot, D., & Novak, A. J. (2011). Paradigm Depletion, Knowledge Production and Research Effort: Considering Thomas Kuhn's Ideas. *Metroeconomica*. 62(4): 587–604.

[2] Franklin, A. (1984). Are Paradigms Incommensurable? *The British Journal for the Philosophy of Science*. 35(1): 57–60.

[3] Franklin, A. (2015). The Theory-Ladenness of Experiment. *Journal for General Philosophy of Science*, 45(1): 155–166.

[4] Godfrey-Smith, P. (2003). *Theory and Reality: An Introduction to the Philosophy of Science*. Chicago: University of Chicago Press.

[5] Hoyningen-Huene, P. (1990). Kuhn's Conception of Incommensurability. *Studies in History and Philosophy of Science*, 21(3): 481–492.

[6] Kuhn, T.S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

[7] Kuhn, T.S. (1998). The Nature and Necessity of Scientific Revolutions. In *Philosophy of Science: The Central Issues*, M. Curd & J.A. Cover (Eds.) . New York: W. W. Norton & Company, 86–101.

[8] Lipton, P. (2003). Kant on Wheels. *Social Epistemology*, 17(2-3): 215–219.

[9] Oberheim, E., & Hoyningen-Huene, P. (2016). The Incommensurability of Scientific Theories. *The Stanford Encyclopedia of Philosophy*, Edward A. Zalta (Ed.). Available from: https://plato.stanford.edu/entries/incommensurability/#Int. [Accessed 15th May 2017].

[10] Phillips, D.L. (1975). Paradigms and Incommensurability. *Theory and Society*, 2(1): 37–61.

The ω-Stone: A Set Theoretical Approach to Omnipotence

Theodor Nenu
Hertford College, University of Oxford

Introduction

The Paradox of the Stone (also known as the Riddle of the Stone) is a paramount example of a dilemma belonging to the family of those which aim to undermine the omnipotence of God, enjoying a great deal of popularity in the Philosophy of Religion literature. The problem is to answer the question: Can God create a stone too heavy for Him to lift?

In the standard conception of God, this entity is portrayed to be omniperfect: that is: omnipotent, omnipresent, omniscient, morally perfect and so forth. My aim in this paper is to argue that the believer should not be alarmed by this apparent threat to omnipotence, since the paradox fails to hinder this attribute.

Prima facie issues

Proceeding by an exhaustion of possibilities, assume that God exists and that indeed he is omnipotent:

> *Case 1*: If he cannot create such a stone, the very fact that he is not up to the task exposes one weakness on his side, one thing that he *cannot do*. Since, presumably, an omnipotent being can do anything, the negative answer to the riddle *seems problematic*.
>
> *Case 2*: God is in a position to create such stone and assume he performs the feat. By hypothesis, he now cannot lift the freshly-created stone and, therefore, he just rendered himself non-omnipotent. Since these cases cover all possibilities, the dilemma has been presented and we are now in a good position to offer a hopefully satisfactory solution.

Before I properly embark on this task, I shall mention that the argument that will follow *does not purport to establish the existence of God* or to show that He is indeed omnipotent. Rather, *my aim is to give a new account of omnipotence*, one that captures the infinite aspects of the concept, together with a direct solution to the riddle above.

Omnipotence is not limitless

The guilty party in this riddle is a faulty understanding of three key philosophical notions: omnipotence, infinity and modality. So, before proceeding further, we shall briefly clarify some common misunderstandings of the these concepts. God, irrespective of his greatness, cannot interfere with necessary truths, class which includes the truths of mathematics and logic.

Although it perfectly suffices to use these, we may extend that list if we wish to analytic and a priori truths as well, or simply label them as relations of ideas, in a Humean[1] spirit. Plenty[2] of philosophers of religion, including St. Thomas Aquinas or Richard Swinburne, argue that people sometimes assume logically impossible 'events' to be on par with those that are logically possible. In Swinburne's words [1, 151]: 'But *they are* not. A *logically impossible event* is not an event, just as a dead person is not a person. It is something described by a form of words that purport to describe an event, but do not describe anything that it is conceivable to suppose could occur.' (This view is shared by many others, including [2, 486], who regards Mathematics and Logic as being truths as involving God, calling them 'secular' necessary truths.)

After all, the believer shouldn't even want to classify God as above logic, as [4, 54] argues: 'If an omnipotent being could make contradictory claims be true, then an omnipotent being could make it the case that it both exists and does not exist, that it could be both infinitely good and infinitely evil simultaneously, and so on. Surely this is an incoherent view if ever there was one!'

Thus, omnipotence of God does not mean that he can bring about round squares and the like. It might seem that we cheaply solved the paradox by postulating restrictions that resist it, but *we did not*. There is *no self-evident contradiction* in our enquiry and, after all, it wouldn't be any contradiction in thinking that *omnipotent beings do not exist*, which is what the paradox aims for. The question that we are looking at is definitely not like asking: *Can God create an object that is not self-identical?*

We have said what omnipotence is *not* about. But omnipotence of a deity should include outstanding feats such as, say, creating universes at will, making solid things pass

[1] [3] made a classification of the objects of knowledge into what's now known as Hume's Fork: the distinction between *Relations of Ideas* and *Matters of Fact*. For lack of space, we may view the former as an umbrella term that captures necessary truths, analytic truths and a priori truths. For more, see [9].

[2] Descartes is the most well-known philosopher who argues otherwise, his views being very controversial. He holds that God, if he wished so, could interfere with Mathematical truths: 'God could have brought it about... that it was not true that twice four make eight.' [7, 2:294]. This makes the Stone Paradox go away at a very high cost.

through each other, or even giving life to other *smaller* divine entities, to name a few. Hence, His powers range from the ordinary feats of creating a rock to grand feats such as creating an angel, say.

God can perform feats from modest, *finite* impressiveness to grand, *infinite* (a detailed discussion of the infinite will follow) impressiveness, without going into the realm of logical impossibilities. In short, God's omnipotence is defined as the maximal collection of possible actions that are not logical impossibilities, in particular *disguised logical impossibilities*.

A straightforward initial answer

Before proceeding further, I want to disregard some obvious solutions to the paradox, not because they are not valid objections, but because they are no intellectual fun. One of them deals with modality, particularly necessity. Just as $(\phi \rightarrow \Box * \phi)$ is not a valid formula of modal logic, omnipotence does not entail necessary omnipotence. The literature often discusses God's existence being necessary,[3] but rarely omnipotence is discussed as a necessary attribute.

Therefore, whoever takes a second look at what goes on when the positive route of answering the riddle is taken, will notice that the contradiction does not occur at the *present* time, but at a later future stage when God attempts to lift his new creation. Thus, it is perfectly consistent to assume that *an omnipotent being at time t can render himself non-omnipotent* at t', *in virtue of his omnipotence* from time t. So it can be the case that God is omnipotent at time t and not omnipotent at a later time t', for instance if He kills himself.

This is a simple solution to the paradox that settles it without much trouble, with a 'Yes' answer to the question, without affecting (non-necessary[4]) omnipotence [5]. Nonetheless, my ambition is to reconcile the paradox with the necessary omnipotence, for the supporters of the claim that omnipotence is a necessary attribute of God. I will show later that in this case, the matter will be settled with a 'No' answer to our question, without affecting necessary omnipotence.

[3] One such example is Alvin Plantinga's take on Anselm's Ontological Argument, from his definition of God it follows that if God exists, then his existence is necessary. (However, it might be the case that one can argue in a similar fashion, following Plantinga's template, that omnipotence is also a necessary feature of God.)

[4] Joshua Hoffman and Gary Rosenkrantz discuss something along these lines [5, 245], distinguishing between *accidental* omnipotence and *essential* omnipotence.

A second complaint concerns the terms used in the dilemma, namely ‘heavy’, ‘lift’, et al, which are terms that make sense exclusively in the physical realm. For beings that reside outside space-time, I don’t even know if it is meaningful to use these transitive verbs. Nevertheless, later on, I’ll touch on this key aspect in more detail, and spell out the problems with it, together with a fix for this issue.

Infinity and Divinity

Infinity is a concept that is of uttermost importance in the Philosophy of Religion. St. Thomas Aquinas was amongst the first who attempted to reconcile infinity with the Christian religion; nonetheless, the combination is a quite an unstable one. Aquinas was a firm believer in idea that God is infinite, which was contradictory with Artistotle’s teachings (by which nothing can have an infinitely complex structure). What Aristotle allowed for is a potential infinite, found in processes that go on forever, such as the revolution of the heavens, but not an actual infinite, the one that is posited by the teachings of Christianity.

When we talk of God’s goodness, His power, His knowledge, we do not think of a process that is taking place, we think of a completed totality. Being a follower of Aristotle’s teachings, he was eager to find a way to marry the two incompatible conceptions of infinity. There is no better subject that does more justice to the concept of Infinity than Mathematics, and because of this we must turn to one of the foundational branches of Mathematics, namely Set Theory.

A Mathematical Treatment of Infinity

A restatement of the riddle in set theoretical terms casts light on the deep issues of the problem, for we argued that God’s omnipotence should not interfere with Mathematics, and presumably all of Mathematics can be reduced to Zermelo-Fraenkel Set Theory[5]. Also, it is crystal clear that the concept of omnipotence makes us consider of what it means for a deity to have *infinite* power, which is a notoriously vague concept and perhaps impossible to fully grasp by the human mind. Nevertheless, even in that case, it doesn’t mean at all that we cannot comprehend some *aspects of infinity*, and branches such as Mathematical Analysis and Set Theory, through works of Weierstrass, Cantor and many others, have done a remarkable job in clarifying matters.

[5] Category Theory may be an exception, but this is a debate which does not affect the argument.

Furthermore, I consider Set Theory to be *a tool which allows us to give a more elegant and elaborate account of omnipotence* than natural language allows for. Unaided human intuition has, from time immemorial, proved faulty when it attempts to give an account of infinity, giving rise to countless classic paradoxes, such as *The paradox of Achilles and the tortoise*, originally stated by Zeno of Elea in 5 BC, which was an influential argument at that time that motion cannot exist. However, with the advances of Mathematics, almost any student of the discipline can make use of mathematical tools, e.g. infinite series, to shed light on Zeno's paradox, which may even look silly in the modern age[6].

So, Mathematics made admirable progress in this quest for infinity and its foundational branch, Set Theory, truly is *a symphony of the infinite*, in David Hilbert's words [6]. Works of Georg Cantor (1845-1918) revealed unintuitive facts about infinity and about sets of numbers, such as:

The set of natural numbers and the set of rational numbers are equinumerous (i.e., having the *same number of elements*), both having cardinality $\aleph_0$ (first size of infinity, pronounced 'aleph zero'). A further interesting fact is that both of them have fewer elements than the set of real numbers[7] or, surprisingly, the set of uncomputable[8] numbers. All this was possible by means of what is now known as Hume's Principle: the number of fs is equal to the number of *G*s if and only if there is a bijection between the fs and the *G*s.

But initial attempts for a theory of sets were not so clean, showing that sometimes Mathematical impossibilities can hide in hindsight without being spotted for a good while. According to Cantor, a set is a 'many, which can be thought of as one, i.e. a totality of definite elements that can be combined into a whole by a law.' Under this Naive Conception of Set, any predicate ϕ has the property (by use of the law of the excluded middle), that however we pick an object, ϕ either applies to it or not. So all predicates have a set of things that they apply to, namely their *extension*. Mathematicians like formality, so that translates into us asserting that the formula: $\exists y(Sy \wedge \forall x(x \in y \leftrightarrow \phi))$

[6] For a whole treatise on Infinity, see [10].

[7] A wonderful proof of this fact is called Cantor's Diagonal Argument. Instead of showing that $\mathbb{R}$ itself has more elements than $\mathbb{N}$, it shows that even the interval (0, 1) cannot be put in the bijection with $\mathbb{N}$, showing that no matter how we'd try to list $\mathbb{R}$'s elements, we'd always leave some number out. A version of this will be used later on in the paper.

[8] A computable number is one that belongs to the language of some Turing Machine. For further reference, check Alan Turing's 1936 paper: 'On Computable Numbers with an application to the *Entscheduingsproblem*'. In simpler terms, a computable number is one that can be generated digit by digit by means of a computer program. It does not matter if our number Is irrational, for the square root of 2 or pi are computable, as long as there's a procedure of generating their digits correctly.

is a logical truth, with y being ϕ's extension.

We obtain thus the validity of the following formula:

$\exists y(Sy \wedge \forall x(x \in y \leftrightarrow (Sx \wedge \neg x \in x)))$.

What follows is what is famously known in the Mathematical world as *Russell's Paradox*:[9] Let S be the presumed extension of ϕ in the foregoing formula.

What is the answer to the question: Does $S \in S$?

If $S \in S$, then S contains itself. Since we chose S to be the set of those sets which do not contain themselves, it fails to meet the membership criterion for S and, therefore, $S \notin S$. On the other hand, if $S \notin S$, then S does not contain itself, so it satisfies the membership condition for S, thus S must capture it as well, therefore $S \in S$. Since all possibilities lead to a contradiction, there cannot be such an S. This is an argument that does not rely on any specific conception of set theory, which proves the validity of

$\neg\exists y(Sy \wedge \forall x(x \in y \leftrightarrow (Sx \wedge \neg x \in x)))$,

which contradicts our previous formula which was rendered valid by the naive conception of set: it follows that we need a better conception of set.

Main point from Section 6: Not everything that syntactically looks like a set has putative reference to an actual, well-defined set. Sometimes the natural way of thinking about things does not instinctively reveal logical absurdities. Pure, logical impossibilities may be extremely hard to spot and sometimes they are disguised in a deceptive way.

God's Actions, Mappings and Sets

We turn back again, to our question of omnipotence. We may want to assert that some actions are (at least conceptually) more difficult than others. To make a first attempt of capturing this conceptual difficulty, we assign numbers to actions (with greater numbers implying greater difficulty). Since our stock of numbers is infinite, we have no trouble assigning any *finite* number to particular, non-sophisticated* action.

[9] Russell came up with this paradox as a reply to Gottlob Frege's assumption that any concept has an extension, which was the foundation on which the Logicism programme in the Philosophy of Mathematics was built.

It is not essential at all to pinpoint the inner workings of this assignment function and to have a perfect understanding of what gets assigned to what. Perhaps the act of creating an ordinary pencil (or an *ordinary stone*, to stay in the spirit of the problem) would get assigned value 0. Perhaps to create a not so ordinary pencil, say one of the size of the Universe, we'd need the number 23.

What about more sophisticated tasks, such as creating a *conscious celestial eternal being* such as an angel? Intuition suggests that limiting this action to a finite natural number is problematic. No matter if this number is $10^{10^{10^{10}}}$, it still seems like we limited the action when we should've gone further. Does this mean that God cannot perform this (more sophisticated) action? Absolutely not. This action is assigned *infinity*; but what are we supposed to do then with actions that are more difficult than this one? Are we forced to assign all of them the same thing, namely infinity? It would be highly unsatisfactory if we couldn't distinguish between sophisticated tasks of various conceptual difficulties. So, we have an infinite stock of natural numbers for unsophisticated, *trivial* actions (for Him), such as *creating a planet that's bigger than the whole Milky Way*. But we cannot deal so straightforwardly with some other actions and we do not want to trivialise them by placing them all under the same umbrella labelled *infinity*.

As I said, in virtue of Cantor's works, we now know that $\mathbb{R}$ has a greater cardinality than $\mathbb{N}$, two infinite sets of different sizes. What if we have actions that belong more or less to the same class (say we assign $\aleph_0$ to both), but we still want to assign them in a way so that we know which one is (slightly) more difficult? But, ultimately, what if we want to do a truly ambitious task (e.g. creating a *smaller God*)? Arguably, this is an action that we want an omnipotent God to be able to perform. In case we need a size of infinity that's greater than $\mathbb{R}$'s to assign this action to, does such size exist? Nothing of what we said so far indicated whether sizes of infinity go beyond that of $\mathbb{R}$'s cardinality. Fortunately, the answer is a clear yes, and we shall prove this.

> Theorem C: Starting with $\mathbb{N}$, we can define a sequence $(Sn)_n$ of sets where each set has a strictly greater cardinality than its predecessor, for instance the sequence defined by: $S_0 = \mathbb{N}$ and $S_{n+1} = P(S_n)$, where P stands for power set (A is the power set of B iff it contains all its subsets, including $\emptyset$ and B itself).

A Cantor-style Proof: Why is it the case that if we have an infinite set A, then the set B given by $P(A)$ has a greater cardinality? Well, suppose (for a contradiction) that it doesn't: that is, by Hume's Principle, there is a bijection f between the sets and that we could pair off every element $a \in A$ with an element f$(a) \in B$, in such a way that no elements of B were left over.

Consider then the following set $S = \{x \in A : x \notin f(x)\}$. Clearly, S is a subset of A, so a member of B. Since f is a bijection between A and B, there is an element $m \in A$

such that f(m) $=$ S. The problem comes when we ask ourselves: Does $m \in S$

1. . If it does, then $m \in f(m)$. Since $m \in f(m)$, it follows that the membership condition for S does not hold for m. It follows that $m \notin S$.
2. If it doesn't, then $m \notin S$. Thus, $m \notin f(m)$. Therefore, m satisfies the membership requirement for S, and it follows that $m \in S$.

Since these cases are exhaustive and all of them lead to a contradiction, it follows that there couldn't have been such a bijection in the first instance. Thus, for every infinite set, the power set has an even bigger cardinality. In conclusion, there is no greatest size of infinity.

The cardinalities of these sets are called the Beth numbers. Thus, the zeroth Beth number is $\aleph_0$, the first Beth number is the number of elements in $P(\mathbb{N})$, and so on, those numbers giving us an infinite sequence of infinities[10]. Therefore, when people ordinarily say that God's power is infinite, we are running into a language problem, for they use the word infinity as a generic[11]. This is because they cannot mean (or at least they shouldn't) *the infinite*, for there is no such thing (in light of our proof); we can always find another Beth number that overtakes it.

Thus, a description of what an all-powerful being can perform is no straightforward matter and all those infinite aspects that the task comes equipped with, I believe, can be captured by the following picture:

We showed that the Naive Conception of Set needs revision, so we are going to present the Iterative Conception of set by means of the von Neumann hierarchy [8]: The objective of this is to define a way to generate all and only *pure sets*. We showed that some *things*, at first sight, appear to be sets, but they are not (Russell's Paradox). Therefore, we must be careful what we allow to sneak in.

The last theoretical background we need before we can draw a conclusion is: An ordinal number, or ordinal, is one generalisation of the concept of a natural number that is used to describe a way to arrange a collection of objects in order, one after another. It is easier to illustrate them, starting off with the well-known naturals: 0, 1, 2, ...

We then *define* something that's greater than all these, the first ordinal number, ω . We keep counting: $\omega+1$, $\omega+2$,... After all these, we reach 2ω , and then we keep counting:

[10] The question whether the first Beth number denotes the size of the set of real numbers is called the Continuum Hypothesis. Kurt Godel has shown that ZF Set Theory + CH is consistent, whilst, unexpectedly, Paul Cohen has proved through a technique called forcing that ZF Set Theory.+ the negation of CH is consistent.

[11] In linguistics, a generic is a word with a less specific meaning.

$2\omega + 1, 2\omega + 2, \ldots$ and so on. After *a while*, we'll reach ω^2, followed by $\omega^2 + 1, \ldots$ After *a bigger while*, we reach ω^ω and so on, in a never ending process, going one by one in a beautiful river of infinities.

This beautiful sequence that walks through every stage of infinity will be used to label von Neumann's iterative process of constructing well-founded sets, which takes place in stages. We start off with the Empty Set, at Stage 0:

* $S_0 = \emptyset$
* For any ordinal α, $S_{\alpha+1} = P(S_a)$
* For any limit ordinal[12] β, S_β is the union of all the S-stages so far.

Those are all called pure sets.

By Theorem C, after we reach S_ω the sizes of the sets start to increase in a spectacular beauty, and whilst iterating in the sequence above, we have ever-increasing sizes of infinity and greatness.

Omnipotence restated in Set-Theoretical terms

Now comes the crux of the paper. Let's map God's feats or possible feats not to numbers, for numbers sizes are limited in a sense, but to sets who are to be judged by their cardinality. Again, it's extremely difficult and *completely unnecessary* to give a precise description of this mapping. So, to settle previous issues, sophisticated actions of similar difficulty will be mapped to distinct sets of the same cardinality. Some examples of this mapping:

Perhaps the previous basic act of creating an ordinary pencil would map to $\emptyset$, whose cardinality is 0 and maybe the act of creating a conscious being such as a human would map $\mathbb{N}$, whose cardinality is $\aleph_0$, which can first be found at Stage ω. Perhaps creating a celestial creature maps to a set first encountered at Stage $\omega^\omega + 7\omega$, thus mapping to beyond a very large Beth number. We can endlessly expand the sizes without worry.

We are benefiting from a fountain of infinities (of actions), in the broadest sense of the word *infinity*, for every action performable by an omnipotent being corresponds to a set and every set translates as an action that can be performed by such a deity.

[12] Limit ordinals are those first ordinals that come after a countable sequence of ordinals. Omega is such ordinal, or five times omega, or omega squared, etc.

Thus, I'll sometimes speak of 'performing sets' meaning 'performing the action that corresponds to that set', but I'll also speak of an action being on the Hierarchy, meaning 'the set corresponding to that action is on the Hierarchy'. In that myriad of Stages that generate sets, we'll find actions ranging from the creation of new universes to creating mini gods, actions that were first generated at some stages indexed by stupendously large ordinals.

Thus, the omnipotence of God is defined by being able to perform any action on The Hierarchy, for The Hierarchy contains the maximal collection of actions that are not logical impossibilities.

Whatever logic-preserving statements the human mind can come up with[13], will be situated somewhere on the Hierarchy. We talked in Section 4 about the fact that the transitive verb 'to lift' is not appropriate for deities. The things that God can do should be more theoretical, and not based on a physical action. So, we want to reformulate the riddle into something that makes sense and which preserves the aims of the original question, whilst still having the same logical form. The logical form of the riddle of the stone's statement is the following: $\exists x(Sx \wedge Gx \wedge \phi(x))$.

Instead of Stones, we'll make Sets as the object of our enquiry, for they are closer to the nature[14] of God than Stones. One possible set-theoretical sentence that captures the essence of the riddle, keeping its logical form is: Can God create a pure set such that He cannot create a bigger pure set?

The question, at first sight, seems perfectly legitimate and paradox-free. It doesn't *prima facie*[15] have the same self-evident logical impossibility of, say, that the action of creating the largest natural number, where you can always add +1 to get a bigger one.

We know what it means to *create* a stone, but what does it mean to create a mathematical object? For, in a way, mathematical objects were always *there* (at least if we adopt a realist view of Mathematics). It is not as though God can create the first natural number greater than 3. 4 has always been *there*, an eternal, mind-independent abstract object. But this number can be *exemplified* in nature by various collections of

[13] It is remarkable how few actions the human mind is capable, even in principle, to come up with if we are to adopt this exposition. We wouldn't even touch aleph-zero conceivable actions, which is nothing in this scheme of things.

[14] Plato would say that Stones reside in the Realm of Becoming, whilst Sets, Numbers, the Form of Good, and suchlike reside in The Realm of Being.

[15] I argued in Section 6 about the non-obviousness of realising the logical impossibility that there cannot be a set of all sets.

four objects. For the sake of argument, assume that all abstract objects are somehow 'dead', but that God can bring them any of them to life (i.e. make them concrete) if he wishes, as long as their mathematical nature is not contradictory. So, can God bring to life $\emptyset$ in this conception? The answer is a clear yes. To achieve the task above, i.e. for God to create a pure set such that he cannot create a bigger pure set, he'd essentially have to bring to life the set of all ordinals, a set that has as members everything in the von Neumann hierarchy, because it is straightforward to deduce that if the biggest set would reside on some stage at the hierarchy, the next stage would give us a bigger set (by Theorem C).

Perhaps we are wrong and we are in for a counterproof that God can perfectly well do actions outside the hierarchy. Perhaps it doesn't contradict classical logic to create the set of all ordinals, which is not situated on the hierarchy. Again, he can do this if and only if the set of all ordinals is a nonparadoxical mathematical concept.

But now we are faced with what's called the Burali-Forti paradox: There is no set of all ordinals. The collection (notice that we didn't call it set) of von Neumann Ordinals, just like Russell's Paradox, cannot be a set in any set theory that operates on classical logic. Suppose the class of all ordinals is a set, call it S.

If $x \in S$ and $y \in x$, then $y \in S$, because any ordinal contains just ordinals. S is well ordered by $\in$, just as its members, so we have an order '$<$' on S. Since all ordinal classes that are sets are also ordinal numbers, it follows that S is both an ordinal class and an ordinal number. So $S \in S$. Thus, according to the way the hierarchy is defined, S¡S. But no ordinal class is less than itself. So, $\neg S < S$. Contradiction.

Therefore, we are forced to answer the modified riddle in the negative: God cannot create a set such that He cannot create a bigger set. Does this threaten his omnipotence? No, because creating this set would lead to unforgiving mathematical inconsistency, it would be an action of the same nature as making $2+2$ not equal to 4 and omnipotence is a concept that is consistent with necessary truths, as argued in Section 3.

Conclusion

In this paper, we gave an account of omnipotence intended to capture the infinite aspects integrated in the concept together with a motivation for this account. Pictured this way, an omnipotent being can do an imagination-expanding number of things, as long as logical impossibility is not encountered. Our paradox asks God to perform disguised logical impossibility, which we argued that it cannot be done even by an omnipotent being. Therefore, God cannot perform the task in question, but omnipotence is not under threat.

References

[1] Swinburne, R. (1977). Omnipotence. In *The Coherence of Theism*. Oxford: Oxford University Press.

[2] Leftow, B. (2012). Non Secular Modalities. In *God and Necessity*, Oxford: Oxford University Press.

[3] Hume, D. (1748). *Enquiry Concerning Human Understanding*, T.L. Beuchamp (Ed.), Oxford: Oxford University Press: Section IV, Part 1.

[4] Meister, C.V. (2008). Chapter 1. In *The Routledge Companion to Philosophy of Religion*. Meister, C.V. and P. Copan (Eds.). London: Routledge, Taylor & Francis Group.

[5] Hoffman, J., Rosenkrantz G. (1980). Omnipotence. In *The Routledge Companion to Philosophy of Religion*. Meister, C.V. and P. Copan (Eds.) London: Routledge, Taylor & Francis Group.

[6] Hilbert, D. (1984). On the Infinite. In *Philosophy of Mathematics: Selected Readings*. P. Benacerraf & H. Putnam (Eds.), Cambridge: Cambridge University Press: 183–201.

[7] Descartes, R. (1984). *The Philosophical Writings of Descartes*. Vol. II. Translated by Cottingham, J. and R. Stoothoff and D. Murdoch. Cambridge: Cambridge University Press.

[8] Boolos, G. (1971). The Iterative Conception of Set. *Journal of Philosophy* 68(8): 215–231.

[9] Millican, P. (2007). Ontological Arguments and the Superiority of Existence: Reply to Nagasawa. *Mind* 116: 1041–53.

[10] Moore, A.W. (2016). *A History of The Infinite*, London, United Kingdom: BBC Radio 4.

The Walls of Our Cage: A Critical Investigation into the Role ofLimits and Restrictions in*Tractatus Logico-Philosophicus* and*Philosophical Investigations*

Ruby Main
University of Dundee

Introduction: A Continuation in Theme

Wittgenstein's *Tractatus Logico-Philosophicus* and *Philosophical Investigations*, emblematic of two distinct phases in his philosophy, can both be shown to have a fixation on limits; on setting them and realising their inevitability. By analysing his approach to this theme, the apparent disconnect between the two works can instead be viewed as an evolution. The preface to the *Tractatus* sets out that the project's purpose is to draw the limits of meaningful language. The text tackles the limits of language from several angles: as a consequence of the 'picture theory' of propositions and as a result of language's logical form; furthermore, by arguing that limits are created by the notion that no fact can have a greater value than another. In Wittgenstein's *Lecture on Ethics* he uses the phrase 'running against the walls of our cage' to describes a limit similar to the latter, suggesting that using language to try to describe ethics is like trying to overcome the inevitability of 'walls', which is futile. While the later work in *Philosophical Investigations* does appear to working against limits by showing that even definitions and concepts are not bound by rigid restrictions, it also directly and indirectly keeps the limits of language in focus. In refusing to attempt to apply a rigid theory, *Investigations* asserts its own inability to view language from 'outside' its walls. Limits are particularly evident in how Wittgenstein approaches philosophy. In both texts the correct use of philosophy is in the analysis and examination of language. Even highly theoretical understandings are always based inside of language. Philosophy, in using language–either as a picture or as a tool–is bounded by its limits. The 'walls of our cage' can be used conceptually to understand Wittgenstein's early and later work, and their relationship to one another.

The Limiting Nature of Picture Theory

The *Tractatus Logico-Philosophicus* is more explicit in its dealings with limits. Its preface promises that it will set and discuss the limits of language. Wittgenstein asserts that as he will be using language to do this: 'It will therefore only be in language that the limit can be drawn, and what lies on the other side of the limit will simply be nonsense' [9][1]. Many of the Tractarian limits arise from his 'picture theory' understanding of proposition. 'Picture theory' is not the metaphorical notion of language creating pictures of the world, but rather a more solid assertion that propositions are expressions of thoughts, which are literal 'logical pictures' of facts, mapped onto them in some sense. In everyday language, the 'logical from' of a proposition is hidden, but total analysis would reveal the structure. Anthony Kenny gives the example of the proposition 'my fork is to the left of my knee' to demonstrate how such analysis would take place. When fully analysed to reveal its logical form, this simple proposition involving two apparent objects would account for all the relations which constitute the fork, the owner of the knee and the world at large. The analysis would be completed once it had reached indivisible objects that could no longer be described by means of their relations to other objects. Indivisible objects are directly attached to 'names' which form the lowest unit of language. The relations of these indivisible objects expressed via their names is what Wittgenstein coins the 'elementary propositions'. To fully analyse in this manner would be practically impossible 'but the thought expressed by the proposition already has the complexity of the fully analysed proposition' [2, 4-5]. Picture theory explains that factual language is the expression of facts, which are arrangements of elementary propositions. Wittgenstein never alludes to what these propositions or indivisible objects may be but instead attempt to prove their existence a priori in the logical structure of language. Factual language cannot make propositions outside of the bounds of these objects which constitute the word. Therefore, factual language is restricted: it cannot make reference to possibilities which could not exist in the word. Wittgenstein suggests that the entirety of language can therefore be hypothetically measured; 'suppose that I am given all elementary propositions: then I can simply ask what propositions I can construct out of them. And then I have all propositions and that fixes their limits' (TLP 4.51). The indivisible objects constitute objective reality, 'the world', and propositions describe their given arrangement. 'The limits of my language means the limits of my world' (TLP 5.6) is the full iteration of this concept. Moving on from 'picture theory', Wittgenstein also introduces the limits set by logic. Logic for early Wittgenstein is used to navigate philosophical confusions about language; it exists outside of the confines of language and is the medium in which the propositions can occur. Being the medium in which propositions occur, it fixes the limit of their use [7, 210]. Logical form, the structure which attaches propositions to the world, is present, or shown in language, but it cannot be the subject of a proposition with sense. Propositions set out the arrangement of objects, but cannot discuss

[1] Abbreviation TLP and proposition number will be used henceforth.

the structure that makes the relation between reality and proposition possible [4, 236]. It is shown that 'what finds its reflection in language, language cannot represent' (TLP 4.121).

Paradox and Nonsense

A.W. Moore points out that Wittgenstein's own discussing of logical form 'cannot be interpreted in a way that is constant with his own views and must therefore, by his own lights, be regarded as nonsense' [4, 236]. Wittgenstein is acutely aware of this fact, going so far as to describe the propositions of the *Tractatus* as nonsensical; 'anyone who understands me eventually recognises them as nonsensical, when he has used them–as steps–to climb up beyond them' (TLP 6.54). The *Tractatus* in trying to set limits to the possibilities of meaningful language, finds itself overstretching its own prescribed restrictions. Many of its propositions, being the kind that he proposes cannot be made logically in language, become paradoxical. Wittgenstein, despite acknowledging the futility of using language to describe the abstract, the structures beyond the objective world, maintains that there is much importance in the abstract. 'There are, indeed, things that cannot be put into words. They make themselves manifest. They are what is mystical' (TLP 6.522). Philosophy cannot transcend the limits of language to handle the concepts which are often discussed by it. The practice of philosophy for the early Wittgenstein can, while residing in meaningful language, do nothing other than introduce clarity into language through analysis. In establishing this correct use of philosophy. Wittgenstein was able to implicitly show that the things beyond factual language with sense; aesthetics, ethics and religion, are inherently more important than narrowly defined philosophy. Paraphrasing Pears [5, 88-89], these parts of language cannot be discussed using facts. They may have value, but no fact can have can higher value than another (TLP 6.41), and so the ethical or aesthetic, which are dependent on value judgements, are not adequately expressed in propositions. This leads to the final proposition of the *Tractatus*: 'What we cannot speak about we must pass over in silence' (TLP 7). Limits, similar to those instigated by the limits in the 'picture theory', logical form, and the problem of equal value all arise in Wittgenstein's later work.

Lecture on Ethics and the Walls of our Cage

Wittgenstein's *Lecture on Ethics* can be used as to analyse the transition between the *Tractatus* and *Investigations*. It is not a lecture in ethics, despite its title and introduction; rather, it is a lecture on our incapacity to truly discuss ethics. It has several threads of thought in common with both the early highly theoretical work, and the later more descriptive *Investigations*. In the lecture, the limits of language get the fitting title of 'the walls of our cage' and our desire to use language to talk about that which is beyond it, such as ethics, is seen as running up against these wall. When he asserts it 'is nonsense to say that I wonder at the existence of the world, because I cannot imagine it not existing,' [10, 8], Wittgenstein shows a fundamental issue of using language to describe that out with the world, such as considering its existence. 'Wondering' in this way is not like wondering at an ordinary state of affairs before us; we cannot wonder when we cannot envision the states being otherwise. Wonder is nonsensical in this context, because the alternative situation cannot be envisioned.

The word is used as it has some similarity to 'wondering' in the usual sense, but is more of a metaphor than an accurate term. He notes that when we describe any ethics using language, we are being metaphorical. In ordinary circumstances a metaphor can be removed and a fact still remain, however, in ethical metaphors 'as soon as we try to drop the simile and simply to state the facts which stand behind it, we find that there are no such facts. And so, what at first appeared to be, simile now seems to be mere nonsense' [10, 10]. He argues that even our use of words such as 'good' to describe a person are analogies. The way we might use 'good' to describe an item of monetary value, or an item that performs its function well, are very different to how we use 'good' in an ethical sense. Without a context, such as the ability to perform a function or trading value, it appears that it is impossible to have an absolute definition for any concept.

Underneath the analogy there is no clear fact, which is the problem of ethical language. Lois Wolcher describes this sentiment: 'A philosopher who thinks that there exists something essential about the good or ethical is a captive of his own dogmatic insistence that the words 'good' or 'ethical' must have one core meaning despite what the evidence of their actual use shows him. Such a philosopher is confused about the way language works' [11, 7]. The limitation of the equality of facts from the *Tractatus* continues in the lecture. He suggests that if one were to put every true fact into a book, and to read in this book about a murder, it would 'be on exactly the same level as any other event, for instance the falling of a stone' [10, 6]. This book would only picture the practical truths of the world. Ethical considerations involve values which cannot be expressed by facts, and thus language at all, without the help of metaphor. Wittgenstein suggests that his own and 'the tendency of all men who ever tried to write or talk Ethics or Religion was to run against the boundaries of language' [10, 11-12]. Language cannot handle the discussion of absolutes, due to the dependence of all words on contexts. And in factual language there is no way to assign additional value to one event over another. Wittgenstein does however not try to discredit or disallow the human need to run against the walls, he acknowledges the importance that we all put on trying to express those things out with the limits of language. The *Lecture on Ethics* sets up how *Investigations* will handle limits. Context-dependency in the latter becomes an even greater source of restriction.

The Hidden Limits of *Philosophical Investigations*

The *Philosophical Investigations*, being so radically different to the *Tractatus*, and being primarily concerned with the many possibilities of language, at first seems to be firmly against limits as a concept. However, by expanding upon the main problems raised in *Lecture on Ethics*, it ends up reaffirming that there are limits to language, and that our desire to push these limits is confused and useless. He states, 'For I may give the concept 'number rigid limits in this way, that is uses the word 'number' for a rigidly limited concept, but I may also use it so that the extension of the concept is not closed by a frontier. And this is how we do use the word 'game'. For how is the concept of a game bounded? What still counts as a game and what no longer does?' (PI 68). In the following remarks, Wittgenstein addresses the different ways in which one might use the word 'game' to describe a host of different activities without any apparent unifying element between them. The meaning of the word 'game' is subject to its context and the intentions of the speaker or writer. This initially seems to be in direct contrast to the restrictions of a 'picture theory' understanding of language. In 'picture theory', names in all propositions were directly attached to objects, in other words, names had distinct and unchangeable definitions. In this case, the meaning of a word is not a linked to an object, but rather its application in language gives it meaning. This is alluded to in *Lecture on Ethics* in which Wittgenstein analyses the language games surrounding the use of the word 'good', though his later work would imply that even 'absolute good', as we mean it, would still be a kind of language game, subject to context. In contrast to the restricted factual language of the *Tractatus*, more can be done with language than forming propositions. Indeed, even philosophy had usually only considered three types of sentences: assertions, questions and commands, yet in *Investigations* a countless number are presented [3, 57]. The limits of language therefore are not bound to a total number of objects in the universe. Language games are still subject to limits of a different kind. To paraphrase Paul Standish, the limits of *Investigation* apply to different games with particularity [6, 223]. There are general rules which determine the uses of words in specific contexts. Language then is not like in the *Tractatus* where a full set of propositions would fix its limit.

Language to its speakers is the point from which understanding is shaped–we cannot imagine a language which did not follow similar conventions to our own, or had utterances which did not relate to the rules of our commonplace language games [3, 50]. This can be seen in a thought experiment: if one were to encounter a group of people that appeared to speak a language, but the sounds they made had no correlation to other 'forms of life' as we understand, including gesturing, daily activities, or greeting: 'language' would be untranslatable. Furthermore, it would not even be considered a 'language' in our sense of the word. We would need to be at a viewpoint in with our own language to attempt to consider this phenomenon 'language' [1, 134-35]. Language as we understand or interpret it is thus limited by 'The common behaviour of mankind' (PI 206). By reflecting human life, in all its forms, language cannot exist outside of it. The *Tractatus*, being unable to provide a model of language which did not make itself nonsense, can be described as a 'limit of reflexivity', in that its theory 'bends back' on itself. The later works do not attempt to obtain an impossible perspective outside of language because they do not attempt to provide a full external theory [6, 234]. In sum, *Philosophical Investigations* acknowledges another reflexive limit, for reasons of the inescapability of language as a form of life. Language can never express what it is to be outside itself, so a total theoretical perspective is unachievable.

Wittgenstein's Role for Philosophy

The later Wittgenstein being open to the idea of multiple uses of words, that is, abandoning the rigidity of his earlier 'picture theory' and an overarching theoretical framework for language. The *Investigations* will stop passing over things in silence. However, in drawing perhaps the most obvious parallel between this book and the *Tractatus*, some things are still beyond meaningful discussion. The treatment of philosophy in both texts is probably the most striking similarity between the two and the greatest continuity. In the *Tractatus*, philosophy is a tool for analysing language: to 'set limits to what cannot be thought by working outward through what can be thought' (TLP 4.114). Compared to the *Investigations*, where 'Philosophy may in no way interfere with the actual use of language; it can in the end only describe it' (PI 214) and 'simply puts everything before us, and neither explains nor deduces anything' (PI 126), there is an obvious continuation. A.W. Moore suggest that the key difference in approach is that later Wittgenstein does not merely describe the ideal role of philosophy but 'practices what he preaches' [4, 257]. The *Investigations* is a collection of the practical applications of this kind of linguistic, clarity-finding philosophy, which can be called, as it analyses the application of language 'grammatical investigation' [3, 13]. McGinn calls the desire to answer questions about the form of the world by prescribing a specific framework the 'theoretical attitude' [3, 16]. Wittgenstein's opposition to this mode of thinking is evident in his desire to examine language as it occurs, and to understand why we fall into our current assumptions about the way language works. 'The real discovery is the one that makes me capable of stopping doing philosophy when I want to.–The one that gives philosophy peace, so that it is no longer tormented by questions which bring itself into question' (PI 133). In other words, philosophy is a kind of therapy that reveals why these circular questions have arisen from the misapplied or misunderstood use of language. Unlike in the *Tractatus*, where philosophy was the tool of total analysis, breaking propositions into smaller and more elementary parts, 'grammatical investigation', a replacement for philosophy, examines the use, and the particularity of language. Anthony Kenny's 'fork and knee' example would be investigated in quite a different manner. It would examine how we name directions based on the names given to our hands, and indeed how there is a need for a much linguistic training to account for how we 'name' things at. Setting out language clearly also reveals where our former misconceptions and confusions have originated. It reveals not only language as it stands, but also where our 'theoretical attitude' has come from, illuminating the darkness that this unsatisfied attitude leaves [3, 21]. When Wittgenstein acknowledged much of the *Tractatus* as nonsense, he acknowledged that his model of language and philosophy was composed of propositions which had not been assigned meaning. He asserted that the correct use of philosophy would be setting out propositions with clarity through analysis, making meaning clear. Unclear, however, was how such a, method could be applied in practice to philosophical questions, instead it advocated silence, and appreciation for those concepts which exist outside language. In Investigations, he shows that philosophical problems arise from a natural tendency for us to confuse language games, for framing questions in ways that are inconsistent with our working understanding of words [2, 130]. In the *Lecture on Ethics* this was demonstrated by the phrase 'I wonder at the world's existence.' The limit of language in this case then, is not simply a wall, hit because of a lack of adequate definition, but a wall struck because the fundamental features of language lead to confusion and asking questions which try to transcend it.

Both texts are aware of the imminent frustration which comes from the restrictions of language. They both attempt to quell our desire to express what language cannot. It is the incapacity to get outside the structures of language, be they tied to indivisible objects, or a product of the humans' form of life. In the *Tractatus*, he showed, 'A picture cannot, however, place itself outside its representational form' (TLP 2.174). Even if language is not based on 'picture theory', it is still a range of different practices which are used to describe, act on and live in the world. Understanding language as a resource, which is used in a wide variety of situations, helps to show why it remains limited [4, 269]. Despite not being tied to the world through exact definition, language is a product of being in the world, it is a 'form of life'. When Wittgenstein's answers, 'What is the aim of philosophy? - To shew the fly the way out of the fly bottle' (PI 309) he does not mean that philosophy will allow us to leave the limits of language, but rather that it will put the causes of philosophical confusion into perspective, and perhaps abandon thinking about the wall as something to be broken through in the first place. B.A. Worthington suggest that the *Tractatus* in its closing propositions recommends the adoption of a non-reflective language which was not able to reference its own structure as a solution to the metaphysical 'problem of life' (the philosophical questions about ethics, value, and higher purpose) [12, 495]. 'Non-reflective' in the case of Investigations means descriptive. Investigations advocates grammatical investigation until the need to get outside of language is no longer meaningful where the *Tractatus* advocates its appreciative silence to remain non-reflective. Wittgenstein characterises former approaches to philosophy as 'the uncovering of one or another piece of plain nonsense and bumps that the understanding has got by running its head up against the limits of language. These bumps make us see the value of the discovery' (PI 119). This is incredibly reminiscent of the *Lecture on Ethics*. Language is not able to adequately express these problems, but the problems can be minimised by considering language from a different perspective, namely a more descriptive one. The limits of language are present, but the solution to the metaphysical questions is understanding why we have arrived at them.

Concluding Remarks on Inescapability

Despite striking differences in method and structure, Wittgenstein's early and later work draw similar conclusions about the possibilities of language, its relationship to the world, and to concepts beyond the world. As the *Lecture on Ethics* shows, the restrictions of language are a difficulty that shaped the transition process between the two texts, and several modes of thinking about language stretch through all three works. The limits in the *Tractatus* are the result of the way it approaches the structure of language. The text itself, being one which sets out to describe this structure, undermines its own limits. In many respects these self-referential problems are solved in *Philosophical Investigations* via a radically different approach. By moving away from the theoretical and into the descriptive, Wittgenstein could not only show the limits of language, but explain why we become so fixated on trying to escape them. The most striking similarity and the most vital aspect of his philosophy is that we cannot, despite trying, get outside of language, and philosophical questions are our attempts. In *Investigations*, we cannot escape the form of life that language represents and in the *Tractatus* we cannot make propositions without distinct meaning. Either way, the walls of our cage are utterly inescapable.

References

[1] Hanfling, O. (1989). *Wittgenstein's Later Philosophy*. Basingstoke: Macmillan.

[2] Kenny, A. (2006). *Wittgenstein*. Revised Edition. Oxford: Blackwell.

[3] McGinn, M. (1997). *Wittgenstein and the Philosophical Investigations*. London: Routledge.

[4] Moore, A.W. (2012). *The Evolution of Modern Metaphysics: Making Sense of Things*. Cambridge: Cambridge University Press.

[5] Pears, D. (1971). *Wittgenstein*. Frank Kemode (Ed.). Fontana: Collins.

[6] Standish, P. (1992). Beyond the Self: Wittgenstein, Heidegger and the limits of language. Aldershot: Avebury.

[7] Stiers. P. (2000). Meaning and the Limit of the World in Wittgenstein's Early and Later Philosophy. *Philosophical Investigations*, 23(3): 193–217.

[8] Wittgenstein, L. (2001). *Philosophical Investigations*: The German text with a revised English translation, 3rd edn. Translated by G.E.M. Anscombe. Oxford: Blackwell.

[9] Wittgenstein L. (1961). *Tractatus Logico-Philosophicus*, Translated by D.F. Pears and B.F. McGuiness. London: Routledge & Keagan Paul.

[10] Wittgenstein, L. (1965). A Lecture on Ethics. *The Philosophical Review*, 74(1): 3–12.

[11] Wolcher, L.E. (1998). A Meditation on Wittgenstein's Lecture on Ethics. *Law and Critique*, 9(1): 3–35.

[12] Worthington, B.A. (1981). Ethics and the Limits of Language in Wittgenstein's 'Tractatus'. *Journal of the History of Philosophy*, 19(4): 481–496.

Taking Man-Made Morality Seriously: Moral Artefactualism

Steven Diggin
Hertford College, University of Oxford

This essay argues for the controversial thesis that moral kinds are fruitfully analysed as artefact kinds. I outline the implications of this for moral realism, and show how the normativity of moral terms can be captured.

G.E.M. Anscombe famously argued that the history of common moral terms provides good reason to abandon them [1]. J.L. Mackie similarly argued that the social impetus towards the objectification of morality is ground for establishing a moral error theory [3]. These are powerful arguments against a natural or non-natural metaethical account of morality. The connection between moral terms and the history of society makes it implausible to suppose that the terms refer to an objective (non-)natural kind. It seems as if moral kinds were invented, and not discovered.

A possible, and commonly-drawn, implication of the artificial history of moral kinds is that they are a social kind, like laws or money. If this were the case, then morality would only depend on collective belief for its existence. If we stopped believing in it, it would cease to exist. Furthermore, as long as we believed in it, there would be no natural boundaries as to what form it may take. As such, it would fall very far short of a realist paradigm. This may be grounds for establishing a moral error theory. However, one appeal of such an account lies in its ability to explain the apparent normativity of morality in a simple way: the force of moral terms has its roots in our collective belief, and moral sanctions are just social sanctions. I do not intend to show such an account to be incoherent in this essay, but it is clear on reflection that it is at least implausible. There is a significant difference between our obligations due to morality and our obligations due to, for example, the law. Morality does not seem to depend on collective belief for its existence. If it exists, then it is more than just a convention. However, social kinds are not the only way of accounting for the invention of morality.

This essay begins an analysis of morality as an artefact kind, just like furniture or cutlery. Artefacts are human kinds without being social kinds. Although artefact kinds depend on human intention in a way congruent with morality's social history, they are closer to the realist paradigm than social kinds. We are inclined to agree that tables and chairs do exist in the relevant way, and therefore, so does morality. In this essay, I first outline the metaphysics of artefacts. I then show how a social history of moral terms may be especially appropriate to an artefactual interpretation. I discuss several problems which a full explication of such a view would need to overcome. Finally, I outline the implications of this analysis for metaethics, and indeed moral realism. I conclude that moral artefactualism is an attractive metaethical account.

Artefacts

Several questions must be answered in this section. What exactly are artefacts? How are they dependent on minds, and how does this affect their ontology and objectivity? How might artefacts have normative properties and from where do these properties come? Firstly, and perhaps unsurprisingly, there is no strong philosophical consensus on the metaphysics of artefacts. Indeed, this is a section of philosophy much ignored until around two decades ago. However, the field has bloomed in recent years. We certainly know which kinds of objects are artefacts (furniture and cutlery and tents and knives and paperweights and so on). However, it is difficult to establish uncontroversial necessary and sufficient conditions for artefacthood. One of the only uncontroversial statements which can be made is that artefacts have something to do with human intention. My understanding of artefacts is much indebted to the work of [4, 5]. She describes a central necessary condition of artefacthood as follows:

> Necessarily, for all x and all artefactual kinds K, x is a K only if x is the product of a largely successful intention that (Kx), where one intends (Kx) only if one has a substantive concept of the nature of Ks that largely matches that of some group of prior makers of Ks (if there are any) and intends to realize that concept by imposing K-relevant features on the object [4, 600]

My analysis does not rely exclusively on this account. Rather, I draw the plausibility and implications of an artefact account of morality from analogy rather than through a fully developed metaphysics of artefacts. Nevertheless, I think that the above account is largely correct, and is useful to bear in mind. A full explication of the below analysis would require a substantially further developed metaphysics of artefacts.

There are interesting questions raised regarding the ontology of artefacts. Ordinarily, almost everyone would be prepared to defend the existence of tables as objective in the relevant way, whatever that may be. However, the way in which artefacts are dependent on human intention has led several metaphysicians to deny the existence of artefacts (for example, [6, 127]). I do not intend to weigh in on this metaphysical debate here. However, while the ontological status of artefactual kinds may not be as secure as that of natural kinds (whatever this might mean), only a radically sceptical account of metaphysics would deny their existence in any significant way. In Thomasson's words [4, 605], artefacts are undoubtedly more than mere 'mental constructs'. I explore how the ontology of artefacts contrasts with the ontology of the purported entities of the established schools of metaethics below.

One especially interesting analysis in the metaphysics of artefacts is of public artefacts (as distinct from private tools). [5, 63-65] argues that two central conditions for something to be a public artefact are that 1) the creator intends for the artefact to be subject to certain norms of treatment and 2) that people (prima facie) treat the artefact in the way that the creator intends. Therefore, for any public artefact kind, there is a way in which one should treat it or regard it. In simple cases, chairs are for sitting on, plates are for eating on, and so on. More complex artefacts have more complex norms, for example, tables are for, inter alia, sitting at, working on, etc. Furthermore, these norms can be specified as applying in particular contexts to particular types of persons. This normativity may be objective. It is just part of the meaning of the term 'chair' that we have a reason to sit on it. Indeed, anyone at all would have reason to sit on it, rather than on something else. However, from where this normativity comes is unclear. It would seem odd if, in creating something, one could somehow give one's creation spooky normative properties. Thomasson argues that 'there must be widespread intentional states within the relevant society', prior to the creation of an artefact [5, 67]. A plausible view might hold that something like patterns of normativity existed before the artefact was created, and the creator merely captured and strengthened these patterns of normativity. However, this is substantial question best left to later investigation. I show how this analysis is important for an artefactual account of morality below.

A Social History of Morality as an Artefact Kind

Let us consider the first person to ever use a particular moral term. They are inventing a totally new term, not translating or altering one which already exists. Let us call this term 'X'. Let us suppose that there is some chieftain or high priest, who announces that some action, for example, bravery in battle, is X. Is it the case that bravery in battle X? Under which conditions is this the case and why? The moral realist would answer that bravery in battle is X if and only if there is some natural or non-natural kind, X-ness, and bravery in battle really is an instance of this kind. However, there is a reason that the arguments of Anscombe and Mackie above seem so powerful. The apparent history of moral kinds just does not seem like a discovery of some mind-independent kind. The realist may overcome these arguments, but only with a substantial complication of an otherwise elegant metaethical account. The robust realist view therefore becomes unattractive.

A moral error theorist would say that it is never true that bravery in battle is X. This may be either because the truth conditions of X-ness are never reached in practice, or could never be reached in principle. A non-cognitivist would claim that X isn't a predicate which can have truth-conditions. This seems particularly odd in the case of invention of moral terms. The inventor certainly intends the term to have truth-conditions. Both of these accounts face substantial problems.

An artefactual account of morality argues that the statement, 'Bravery in battle is *X*', is true under exactly those conditions in which an inventor successfully creates the first instance of an artefact kind. This is because moral terms do not differ significantly from artefactual terms.

What are the conditions in which an inventor successfully creates the first instance of an artefact kind? Let us suppose that chairs do not yet exist (people have always sat on stools). I am currently seated on a hunk of matter, which I shall call Felix. I announce to the world that Felix is a chair, where 'chair' is a word which no one has ever spoken before. Under which conditions is this true? There may be many ways of my going wrong. For one, I may be deceived as to what Felix is. Perhaps I think that Felix is a piece of plastic and cloth, but he is actually a living organism! In that case, and in similar ones, it seems that I am wrong to call Felix a chair. This is because Felix did not match the substantive concept of chairhood which I was trying to realise. This is the key condition for a successful invention. As Thomasson argues, I need to be largely successful in realising a large part of my substantive concept of the nature of the artefact. Now, one important part of the nature of a public artefact is the way in which people will treat it. My conception of a chair is as something for sitting on. So, if I announce that Felix is a chair, and no one sits on him, or desires to sit on him, then my conception has not been largely realized. That is, if the norms to which I intend Felix to be subject do not exist, then he is plausibly not a chair as I conceive of the artefact kind.

Let us apply this analysis to term *X*. What are the analogous truth conditions for the sentence, 'Bravery in battle is *X*'? Well, in the same way as Felix had to largely match the conception that I had of him, so too does bravery in battle have to match the conception that this inventor has of it. For example, if part of the inventor's conception of *X* is that things are *X* if they maximise happiness, and she believes that bravery in battle maximises happiness, but it is not the case that bravery in battle actually does so, then bravery in battle is not X. More importantly, if moral kinds are artefacts, then they are public artefacts. Therefore, if part of the inventor's conception of *X* is that it is for doing, then it must be normatively correct to be brave in battle in order for bravery in battle to really be *X*. In this way, the inherent normativity of moral terms is captured.

After invention, the moral kind, *X*, is propagated in the same way that ordinary artefact terms are propagated. The next person to use the term does so correctly (and the resultant sentences are true) only if they are largely successful in realising a substantive conception of X which largely matches that of contemporary or previous makers of *X*. The next creator of an instance of a moral kind must intend for that instance to be of the same kind as the chieftain's instance, and must be largely successful in getting this intention across. That is not to say that their conception must be exactly the same as the chieftain's. It may differ in some respects, insignificant at first, but becoming more substantial as the artefact is propagated further.

In this way, the meaning of the term (and the extension of the kind) can change over time. Moral sentences which were once true can become false and vice versa. For example, the hopeful creators of many historical instances of the kind 'morally right' may have had, as part of their conception of rightness, that acts are right which obey God's will. Atheists may happily establish a moral error theory for such uses of the term, since the conception of the kind is not realized. However, this is not to say that current applications of any moral term are all false. Therefore, moral kinds may actually exist.

The above interpretation of the history of morality and the truth conditions of moral sentences is attractive. Morality seems to exist in an ordinary way, but a robust realist metaethical account does not seem to capture the complex artificial history of moral terms. An artefactual analysis of morality accounts for this history without sacrificing too much objectivity.

Problems to be Overcome

In this section, I outline how a full explication of an artefactual account of metaethics might proceed. I identify problems which may either be overcome in metaethics, or in the wider field of the metaphysics of artefacts. I conclude that this metaethical account cannot be immediately dismissed.

Firstly, the analogy between artefacts and moral kinds may break down through problems with creatorship. In my fanciful history above, there was a clear inventor of a moral kind. However, in many contemporary supposed instances of moral kinds, there is no clear creator. Furthermore, when I make moral judgements, I make them for myself and by myself, rather than referring back to the intentions of some apparent creator of a moral instance. It seems plausible that no such identifiable creator exists. This problem arises in the metaphysics of artifacts as well. When I see a chair, I make a judgement that it is a chair, independent of whatever some obscure creator's intentions may be. Furthermore, even in the case of ordinary artefacts, there may be no clear creator. For example, a factory can operate without any of the workers, supervisors or managers having any direct intentions as to what kind of artefact the product will be. Thus, the exact conceptual connection between creators and artefacts may be more complex than we previously thought, but that is no reason to reject an artefactual account even when there is no clear, identifiable creator. Nevertheless, a full explication of this metaethical account requires a more developed metaphysics of artefacts, and this is where this problem must be solved.

Next, what kind of things are the instances of moral kinds? For example, if bravery in battle is right, then we seem metaphysically committed to some abstract action ('bravery in battle'), of which there are individual performances. When a chair is created, the creator's intentions are directed towards that thing, right there. No such concrete object exists in the case of moral kinds. We will have to be metaphysically committed to abstract actions, states of affairs, characters of persons, dispositions and so forth. This likely breaches an attractive principle of metaphysical parsimony, but we may tentatively accept these entities. The metaphysical commitments of this account must be further outlined.

What exactly does creation of an instance of a moral kind entail? In the case of ordinary objects, matter is molded into the correct form with recognizable features, and so forth. This is paradigm creation. It does not seem possible to shape the abstract objects in the case of moral kinds. However, an answer to this may be found in the metaphysics of artefacts. An artefact may be 'created' just by changing the context in which it is. For example, moving a pebble from the beach to a desk creates a paperweight. Similarly, the creation of an instance of a moral kind may just involve changing the context of a certain abstract object (for example, by bringing it to public attention). I am confident that this problem can be solved.

What are the implications for moral relativism? It depends on what conceptions are held of moral terms. For example, a potential creator of an instance of a moral kind may only intend for persons in his community to view it in a particular normative way. However, other moral kinds may be envisaged as universally normatively applicable. If the latter, then a kind is only successfully created in the absence of significant relativism. Perhaps this shows the difference between morality and mere manners. Morality is not intended to be relativistic, whereas manners are. Such an account may be developed further.

How exactly are we to account for the normativity of moral kinds? If an acceptable analogy between moral kinds and artefact kinds can be maintained, then the solution will be found in the metaphysics of artefacts. Above, I sketched an answer to this question which relied on the preexistence of some normative attitudes which are captured by the creator of an instance of a moral kind. These are then strengthened after the invention of the moral kind. Perhaps the reason that many historical instances of moral terms seemed to be controlled by those in power is that these persons could also control the normative patterns of a society. In any case, this question must be investigated further. Nevertheless, this is a promising way to account for the normativity of moral kinds.

Metaethical Implications

Artefacts, while they do not quite reach the realist paradigm, do seem to exist in an important respect. Tables and chairs and knives and buildings are all real things. Therefore, if morality can be fruitfully analysed as a group of artefact kinds, it seems to have a strong ontological status. It seems to be a species of a realist metaethics, even though it is somehow mind-dependent.

We might have classified this account of morality as some form of constructivism. However, morality is a creation, not a mere construct. The ontological status of artefacts seems more secure than that of contracts and other constructivist kinds. Furthermore, an artifact conception of morality can capture normativity without recourse to some fundamental normative principle (like 'rationality' or 'equality'). This weakness is one of the major problems facing constructivist accounts of metaethics. Therefore, an artefactual account keeps the appealing aspects of a constructivism in metaethics without sharing its major problems.

On the other hand, moral artefactualism is not a quite a form of moral naturalism either. While it sacrifices the robust realism which a naturalist account might attain, artefactualism also has several advantages over moral naturalism. Firstly, normativity can be captured in a non-spooky way. Second, an error theory is less plausible in an artefactual account (a global normative error theory must deny the existence of public artefacts). Finally, Harman's [2, ch. 1] criticism of moral naturalism does not apply. Harman argued that moral terms do not aid explanation. One can know everything about a particular situation without recourse to moral terms. However, in the same way as artefacts are not explanatorily redundant, Harman's criticism may not apply to an artefactual account of morality. Therefore, despite the sacrifice of robust realism, moral artefactualism is an eminently more desirable metaethical account than moral naturalism.

In conclusion, this essay has sought to open debate on a controversial thesis. It has attempted to reconcile the artificial history of morality with its seeming objectivity by arguing for an analogy between moral kinds and artefactual kinds. It has sacrificed the mind-independence of morality in order to account for its social history. This allows for the normativity of moral kinds to be captured in a natural way. However, such mind-dependence has implications for moral realism. Such a preliminary account raises as many questions as it answers. Many problems are left unsolved. However, I am optimistic that this account can be accepted as realist, and can reconcile the prominent schools of metaethics.

References

[1] Anscombe, G.E.M. (1958). Modern Moral Philosophy, *Philosophy*, 33(1).

[2] Harman, G. (1977). *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press

[3] Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*, Hammondsworth: Penguin.

[4] Thomasson, A.L. (2003). Realism and Human Kinds, *Philosophy and Phenomenological Research*, 68(3).

[5] Thomasson, A.L. (2014). Public Artifacts, Intentions and Norms. In Pieter Vermaas et. al., (Eds.)*Artefact Kinds: Ontology and the Human-Made World*.Springer: Synthese Library: 45–62.

[6] Van Inwagen, P. (1990). *Material Beings*, London: Cornell University Press.

The Utility of Idle Explanations

Jack Johnson
St Edmund Hall, University of Oxford

Introduction

One of the things humans do is try to find explanations for why events or phenomena happen. While many of these explanations have a practical motivation, many do not; examples are plentiful of explanations that have been pursued at a great cost in time and resources but have no known practical application. I call these types of explanation idle explanations, and the purpose of this essay is to argue that while they serve no direct purpose, they provide indirect benefits to those who pursue them. I shall also contrast our attitudes to idle explanations with those to other sorts of information. In the first section, I define some of the terms I use, and will lay out more carefully the questions I will be addressing. In the second section, I introduce my chosen theory of explanation, namely Lewis' causal theory. Finally, in the last section, I make clear why explanations are practically useful, and argue that idle explanations provide an especially important utility.

Active and Idle Explanations

Explanations are often pursued with a specific practical goal in mind. An example of this is the early thermodynamicists, who were driven to provide accurate descriptions and explanations of the behaviour of gases in large part by their desire to improve the efficiency of steam engines. The purposes of explanations are not always so technical. If my friend is upset, I will likely ask why he is upset. That is, I will ask for an explanation, the purpose of which will be to allow me to try and rectify the thing causing his distress.

We often, however, attempt to explain things even if we don't think such explanations will be useful to us (I mean useful in a general sense, since my friend's explanation was useful to me in that it allowed me to help him). An excellent example of this is in people's interest in history. Many of the questions of history that we find the most captivating are why-questions: Why did Hitler invade Russia? Why did the US stock market crash in 1929? Why did Rome fall? Many of us would like to possess explanations of these events, even though such information we know will never be of practical use to us (some might want these explanations in order to guide policy, but these individuals are a tiny minority of the interested crowd). A similar situation exists for scientific explanations. While these seem to promise more practical applications than those of history, the majority of people's interest is not based upon these practical considerations.

Explanations which are consciously pursued for practical purposes I shall call *active explanations*. Explanations which are pursued not for practical reasons I shall call *idle explanations*. Note that these are not supposed to be intrinsic features of the explanations; these labels are relative to the individuals pursuing an explanation. One explanation may be both active and idle relative to two investigators, and an explanation nobody is pursuing is neither active nor idle. If I wish to refer to explanations, regardless of whether they are idle, I shall use the term *general explanation*.

I think it is plausible that the majority of the explanations we pursue are idle explanations. Whether or not this is so, it is a mystery why we should feel driven to do something which offers no practical benefit to us. It will soon be shown that while the idle explanations we pursue may not provide us with any practical benefit, the drive to find explanations for their own sake has been highly useful to humans in the past, since idle explanations played an important role making complex predictions possible.

Factual Information

Whatever use explanations have, there is also great utility in the possessing of factual information. This is the sort of information one gives as answers to *who*, *where*, *when* or *what* questions. Examples of factual information are 'the capital of the United Kingdom is London' and 'The Magna Carta was signed in 1215'. Factual information does not differ intrinsically from explanatory information (which will be discussed later), both are fundamentally the same stuff: information. Factual information is simply distinguished by its not being presented as a response to explanation.

Many instances of factual knowledge are clearly useful to have. There is a parallel to idle explanations, however, in that there is factual knowledge people compile despite it having no practical purpose (other than use in pub quizzes). I shall call this kind of factual information *trivia*. A secondary question this essay will attempt to answer is why we do not feel the same intrinsic interest in compiling trivia as we do in finding idle explanations. Indeed, most of us are positively bored by gathering trivia, and it has received the derogatory name of *stamp collecting*.

What Explanations are

Let's return to explanations. To assess the utility of idle explanations, we first need a theory of what they are. This will be provided by theory of general explanations. There has been much said about what explanations consist of over the last seventy years, and whatever choice I make will likely be controversial to some extent. The theory of explanation I find by far most convincing, though, is David Lewis' causal theory [1, 214-41]. In this section I will give a much-condensed account of this theory. The concepts and terminology introduced here will allow me to show in section 3 the precise practical role of idle explanations.

Causal Histories

The first concept that needs to be understood is that of a causal history. This can be thought of as the network of events which leads to some particular event. The causal history H of an event E is a relational structure, and the relata are events. The events present in H can be defined recursively as follows:

1. E is in H
2. X is an event in H if X caused an event in H

A description of H is not exhausted by a list of events in it, however. Also required are the causal relations each event bears to other events. From knowledge of this local causal structure can be deduced the form of large substructures of H, or the structure of H as a whole. We can imagine H as having the structure of the branches of a tree, E being situated at the base of the trunk.

Event is used here in an everyday sense, including everything from highly local events like flashes or impacts, to events extended far in time and space, like an airborne invasion.

Causation and 'cause' are used also in the relatively everyday sense, in which one might say that A's causing B means that if A had not occurred, nor would B. Regardless of what relation one thinks causation might consist of, though, so long as it relates two events, we can substitute it into (2) for 'caused' and we still generate causal histories.

Explanations as Information

We should begin by distinguishing two different senses of explanation. Firstly, there is the act of explanation. This is the sort of explanation about which it makes sense to ask, 'how long was it?', 'who gave it?', 'when?'. The second is the kind this essay is concerned with, and is the kind of which it instead makes sense to ask, 'who thought of it?', 'is it complicated?', 'when was it discovered?'. For the remainder of this essay, 'explanation' will refer to this latter kind.

Lewis' thesis can be stated as follows:

> An explanation of an event is a chunk of information about the causal history of that event

For information to count as about a causal history H–to be *explanatory information* (relative to H)–it must rule out some possibilities regarding H. Information that rules nothing out about H is not explanatory and cannot on its own constitute an explanation. Apart from this, though, there are no restrictions on the subject or quantity of explanatory information.

Subject

Explanatory information can fundamentally consist of two types of information: information about the events in H, and information about the causal relations between events in H. This being the case, though, there are a huge variety of ways that this information can be more or less explicitly presented. Below are some examples.

1. A particular event X is present in H (e.g. the Great Fire of London)
2. There is an event of kind K in H (e.g. a fire)
3. There is a substructure of kind K in H (e.g. a plot)
4. There is a process of kind K in H (e.g. chemical, biological)
5. H has global structure S (e.g. negative feedback loop)
6. There is a process in H of a kind to produce a kind of effect F in H (e.g. 'we have lungs for the same reason we have a heart')
7. There is no event of kind K in H (e.g. a CIA plot)

This names only a few. The point is that information about H can be delivered in any number of more or less direct ways, and this goes a long way to accounting for the wide variety of kinds of explanation it is possible to give.

Quantity, Relevance and Explanation Requests

Any information which rules out a possibility for a causal history H of event E counts as an explanation of E, regardless of how little it rules out. It is clear, though, that explanations can differ drastically in their informativeness. Part of the way which we judge an explanation's value is in its informativeness. That is, how many possibilities about H it rules out. We can see a variety of informativeness in the previously mentioned cases, with direct specification like that of 'A particular event X is present in H' tending to be more informative than the indirect specification of 'There is a process in H of a kind to produce a kind of effect F in H' or the negative specification of 'There is no event of kind K in H'.

Informativeness rarely corresponds directly with practical usefulness, however. For explanatory information to be useable it must also be relevant to what a person wishes to know. We have all heard someone respond to the question 'why didn't you tell me' with the unhelpful 'because you didn't ask'. This response *is* an explanation, according to Lewis' theory. However, it is just one consisting of irrelevant and unwanted information.

When requesting explanations, we typically specify what sort of explanative information we are looking for. I will call the information given in response to a request *response information*. It should be understood that requests need not take place between individuals; a person can form a request internally, and then attempt to respond to it themselves.

A request begins usually with a question of the form *why* E? (though *how* E? is often used, or the request may be of neither form and implicit). This limits response information to explanatory (of E) information. Beyond this, we can impose any sort of restriction on response information we like. This might involve just specifying the kind of event or process it should be about, or it might be less direct.

There is a common kind of indirect restriction which makes more explicit use of the concept of a causal history, and tends to be more efficient in isolating the desired information. This restriction takes the form of a contrast class, which is a set C of events. If a contrast class is employed, the response information should include no information which is itself an explanation of any event in C. That is, the desired explanation should include no information about the causal histories of any event in C. The events of C may be real events, but much more often the events and their causal histories are imaginary.

An imaginary causal history H^* for an imaginary event E^* is not fully and objectively defined as with 'E is in H' and 'X is an event in H if X caused an event in H', but rather is specified (only partially) by the individual(s) involved, and this may be done explicitly or implicitly. I will suppose that H^* consists of some kind of representation similar enough to real causal histories in its structure to play the role demanded by the class restriction. For the contrast class to work properly, H^* generally has to be as realistic as possible. This means that, despite the fact that H* includes imaginary events, these events should be related as much as possible by the causal relations which actually apply to those kinds of events.

As well as this, imaginary causal histories (called ICHs from here) should be as close to real causal histories as possible while still including E^*. Thus, an imaginary causal history for a Soviet moon landing should include as much of actual history as possible, and depart from it only where necessary. We should include in such a causal history a better funded Soviet space program long before we include a pro-Soviet extraterrestrial intervention.

Thus, if I ask why the US, *rather than the USSR*, landed on the Moon, we can understand this as being the question 'Why did the US land on the Moon' (A) accompanied by the singlet contrast class {'The USSR landed on the Moon'} (B). This signals that responses should consist of no information that would equally explain B, were it actual. Thus, I rule out much irrelevant explanatory information like 'The Moon is not defended by alien air defence missile batteries' since this would also have to be the case for a Soviet landing in all the most realistic ICHs. Equally, if I ask why the US landed on The Moon, rather than on Mars, the explanatory information disallowed by the new contrast class shifts to be different from that of B.

This concludes my account of Lewis' theory. This last subsection has concentrated less on the theory itself, than the methods by which we can use explanations, as defined by the theory, to make highly specific requests for information. We will see the importance of this below.

Practical Utility

In this section I will first show how general explanation serves to aid prediction. I will then argue that active explanations are incapable of doing all the work required, and thus idle explanations play an important role.

Prediction

The importance of prediction for survival is beyond clear. The ability to form accurate beliefs about the future allows one to prepare for it in advance, and thus be better placed to overcome the problems it brings. Sometimes prediction is easy. Certain features of nature are regular enough that their accurate prediction is a simple exercise: lighting precedes thunder, hitting someone will anger them, and unsupported objects fall.

Sometimes, however, predictions are not so easily made. These instances of prediction might concern the chance of death from an illness, the chance of frost ruining a crop, or the chance of raiding by bandits. There is no simple way to judge what the future will bring in these situations. This may be due to there being little regularity involved, the situation being completely new, or being too complex to judge intuitively, or a combination of these.

Prediction of complex future events involves two main parts. First, one has to recognise what the present situation is. One needs to know what the important elements are, and how they are arranged in the world. Second, one needs to be able to project the causal behaviour of these elements into the future. One needs to know the way in which events usually causally relate to one another. I will call this information *causal information*.

In the case of judging whether there will be a crop-killing frost, one must first consider the present situation. What is the weather, the season, the state of the crop? Has the wicked witch from the wood behaved herself recently? Then one must consider what causal principles will determine how the situation will evolve. Does a clear sky cause frost? Do the witch's curses cause the crops' resistance to frost to weaken? In this manner one can judge how the future might look, assuming the causal information they have is correct.

Imaginary Causal Histories

We have seen an activity like this before in 2.4. What we do when we predict is we attempt to construct an imaginary causal history of an event in the future which fits as closely as possible the actual causal history that event will have. Instead of beginning with knowledge of this event, however, we begin with several events we know to be in the causal history (because they are present to us) and we try to identify what a causal history containing these events might look like, and what event might lie at the end of this history. The purpose of this imaginary history differs to that which we saw before; it is now for prediction, rather than specification. The skills involved in its construction, though, are precisely the same.

Here, then, we can see the first use of idle explanations: they give us practise in constructing realistic ICHs. As was demonstrated previously, explanation requests very often involve contrast classes including imaginary events. This is as much the case for idle explanations as for active ones, and so it is the case that the frequent pursuit of the former will require one to become well accustomed to holding in one's mind the complex structures of ICHs.

There is more to constructing realistic ICHs than just a familiarity with their general features, though. Producing realistic ICHs about specific sorts of events requires a great deal of knowledge about *real* causal histories involving those kinds of events.

The causal relations between kinds of events in an ICH should match as closely as possible the causal relations which *actually* exist between events of that kind. These relations may differ very much depending on the kind of event in question.

Take the crop frost example. Constructing the necessary ICH here will require detailed knowledge of the causal relations between events of frost, crop failure, clear skies and witches' curses. As any researcher knows, actually discovering the causal relations existent in the world is a difficult thing to do. There are so many different factors at play at any time, some relevant, some not, that it is a hard thing indeed to spot when and how one event actually causes another. Modern research groups have a hard time doing this, and they have access to vast amounts of data and or experimental apparatus. One can scarcely imagine, then, the difficulty which humans throughout history had in obtaining accurate causal information.

The Utility of Explanations

Yet, to some extent they were successful. They were able to compile enough causal information to make the predictions necessary for their survival. I think it is clear to see that the causal information required for realistic ICHs is the information one gets through explanations. Previously, I demonstrated the variety of explanative information which exists for any one causal history and then demonstrated the ways which we can isolate the bits of explanatory information we want. With these means at our disposal we can examine with great precision the causal histories of the past, and so predict more accurately causal histories of the future.

The Utility of Idle Explanations

At first sight, it would appear this gathering of causal information would be done by active explanation. After all, a person is conscious of the predictions they have to make, and so will try to obtain the specific information they think they need via explanations. There is problem with relying upon active explanation for this, though. I said above that causal information is especially difficult to extract from the world. Thus, it takes a long time to compile any useful amount of it. People generally simply do not know far enough in advance the predictions they will have to make for them to be able to obtain the information required for them by active explanation.

The process of gathering causal information has to begin a long time in advance of when that information might be needed. Since the explanations by which this is done cannot be directed towards any particular prediction (the predictions will not be known at the time at which the explanations are pursued), these explanations must be idle explanations. But wait, are these really idle explanations? They *are* directed towards a practical purpose: prediction. Even though they are directed at no *particular* prediction, they can be directed towards the project of prediction *in general.*

This objection would be correct if it were the case that humans think in terms of overarching goals like general ability to predict. They do not, however, or at least not in the absence of a developed external scientific and educational pressure to do so. Humans tend to consciously think mainly in the short term; long-term projects are usually taken care of unconsciously and indirectly. We can see this with the project of reproduction. Some humans have a conscious interest in children, but many do not, or have only a partial interest. If humans were relied upon to consciously pursue reproduction as its own end, we would probably have gone extinct long ago during some period or another of hardship. Instead, we indirectly and unconsciously pursue reproduction by *directly* and *consciously* pursing sexual intercourse. Thus, our genes have developed a way of achieving an overarching 'goal' (using the standard metaphor in natural selection) by getting our short-term conscious minds to pursue an auxiliary goal, which to those pursuing it seems like idle leisure.

So it is with explanation. Humans simply lack the foresight to consciously prepare for prediction by active explanation. Instead, we are endowed with the drive to explain, whether or not we ourselves see a practical purpose to these explanations. Thus, we pursue idle explanations. However, these explanations are not really idle at all; it is through them that we, over long periods, compile a body of causal information that is complete enough to be used for prediction. This is the utility of idle explanations.

1 Trivia

The features of causal information which led to a need for idle explanations are not found in factual information. Whereas causal information is famously difficult to obtain, questions of who, when, where, or what rarely require extended and careful observation. Of course, not all factual information is easy to obtain. For example, discovering what the first word uttered in the twenty-first century was would be incredibly difficult. But discovering what the cause of this first word was is a whole other level of in terms of difficulty. For any given topic, the factual information will typically be available much faster than the causal information.

This means that factual information can generally be found quickly enough that its collection need not pre-empt the predictions which use it. Thus, there is little need for us to gather factual information we have no known use for. Consequently, we lack a *trivia drive*; we tend to find less interest in the compilation of factual information–or stamp collecting–than we do in explanatory information.

Conclusion

In this essay I have attempted to show that explanations, understood as chunks of information about a causal history, provide precisely the material needed to carry out successful prediction. I then argued that the process of gathering causal information is too slow to be left until when such information is needed, and so it done over long periods by idle explanations. Finally, I argued that, in contrast, we lack a trivia drive because factual information is easier to obtain quickly.

References

[1] Lewis, D. (1986). Causal Explanation. In *Philosophical Papers*, Vol. II, Oxford: Oxford University Press: 214–240.

[2] Hempel, C. (1965). Aspects of Scientific Explanation. In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press.

[3] Friedman, M. (1974). Explanation and scientific understanding, *The Journal of Philosophy*, 71(1).

[4] van Fraassen, B.C. (1980). *The Scientific Image*, Oxford: Clarendon Press.

[5] Kitcher, P. (1981). Explanatory Unification, *Philosophy of Science*, 48(4).

Deriving Basic Equality from Self-Respect

Alexander Arridge
St Anne's College, Oxford University

That all human beings are fundamentally equal is an assumption that predicates almost all Western moral and political thought. Yet precisely because of its universally uncontested status, it is a subject that receives very little direct philosophical scrutiny. This essay is concerned with showing that this statement of equality is more than mere assertion. I show that an original position of individual self-respect necessarily entails the equal moral status of all other human agents: respecting oneself as a being of moral worth is fundamentally inconsistent with denying equal respect to beings of the same type. This conclusion is reached through the development of a heuristic procedure for proper introspection. Both the process and the conclusions derived from it have far-reaching implications for moral and political philosophy. I demonstrate the robustness of this procedure by considering some of its implications, notably for interrogating the moral basis behind our treatment of non-human animals and 'marginal cases'.

I begin by outlining the general structure of what any argument for fundamental equality must look like; with this structure in mind, I proceed to develop a heuristic for discovering the properties or capacities shared by the set of all human beings. This heuristic takes the form of an iterative process of introspection. After laying this out in some detail, I then shift focus to the nature of self-respect. Laying out an intuitive and uncontroversial characterization of self-respect, I then synthesize the strands of argument by exploring the implications of an individual position of self-respect in light of the results of my heuristic. I find that the set of capacities constitutive of a human being possesses normative value independently of the fact that a human being possesses them; from here, I argue that a position of individual self-respect is coherent if and only if we afford equal moral respect to beings with the same set of capacities (i.e. other human beings). As such, I show that human equality is necessarily entailed by any one being of that type coherently holding themselves with self-respect. To finish, I consider briefly some implications of the heuristic process and the conclusions it entails.

The basic equality of human beings, if it is to hold on anything more than the level of assertion, must obtain as the result of all humans sharing some set of *morally relevant* properties. Thus the central question this essay seeks to answer is this: What property, or set of properties, do human beings share such that the fact of our sharing them entails a fundamental equality between us on a normative level? In order to answer this question, it seems obvious that we must first come to know precisely what constitutes a human being? More specifically, what are the essential characteristics the set of human beings share? Only from this starting point can we discover which of these shared characteristics is *relevant*, in such a way so as to demand our being considered fundamentally equal. For example, suppose that all human beings, without exception, possessed a lock of golden hair. By definition, this strand of golden hair would be a property that, simply by virtue of being human, all human beings possessed. Yet this fact seems not to bear at all on their equal moral status: nothing described in the above situation suggests that possession of a golden hair is a *morally relevant* property of the human character, such that it entails moral equality. Confronted with a problem of this type, I start this essay by constructing a heuristic capable of ascertaining such properties.

In order to answer the prior question of what constitutes a human being, we need only look inwards. An essential feature of our human agency is a capacity for self-reflection; with the correct procedure and under the right conditions, this capacity is sufficient in itself to discover the set of necessary characteristics that constitute the reflective agent as a human being. It is important to note at this point that this process is concerned with discovering the essential capacities or characteristics of an agent as a *human being*, rather than as an *individual* of the type 'human being'. To a great extent, the moral status afforded to us in virtue of being both of these things is different. Morality applies to us first in our role as a human being; moral concern following from our status as a unique individual is derivative of the more general moral concern afforded to us as a human being. In this sense, equality obtains between human beings: equality between individual beings of that type is derivative. Thus it is imperative to stress that the introspecting agent at the centre of this procedure is concerned only with their identity as a *human being*, rather than the more contingent and specific set of properties that characterise them as an individual.

Thus, by asking oneself the question 'What changes can I endure before I become something else?' one can distinguish between one?s inessential and essential capacities as a human being: essential properties are those for which possession (to some sufficient degree) is necessary for one to retain one?s identity as a *human being*. For example, perhaps I would find that my capacity to experience pain and pleasure is necessary for me to hold myself as a human being: without it I would be nothing more than a rational robot, and so I would not be able to suffer this loss and still identify myself as a human. Indeed, it seems obvious that the set of capacities identified by this process number greater than one: our identity as a human being is surely not reducible to a single characteristic. Indeed, our capacity for the experience of pain and pleasure is a likely candidate; our capacity for autonomy and self-definition, without which we would be nothing more than directed robots, and some degree of cognitive capacity are all equally likely candidates. I am not here concerned with comprehensively delineating the results of this process: I am only concerned to construct and defend the structure of the process itself, and draw implications from this form. Filling out the basic structure is unnecessary at this stage.

At this stage of the process, then, in answer to the question 'What could I lose and still identify as a human being?' we have identified a set of necessary capacities: if the loss of some capacity would disincline or preclude me from identifying as a human being, then this must constitute a *necessary* part of my identity as a human being. Indeed, we must conclude as such both if the loss of a capacity would render us *incapable* of self-identifying as a human, or simply would disincline us from *freely identifying* ourselves has human. For example, permanently losing the capacity for conscious thought would render me incapable of identifying as a human being, whereas losing the capacity to feel pain or pleasure would not rob me of the conscious capacity necessary for self-identification, but it would likely disincline me from identifying myself fully as a human being.

With the above question answered, and a set of necessary human capacities obtained, the heuristic requires us to answer the question: '*How much* of this capacity could I lose and still consider myself a human being?' This question is essential, as we must be able fully to capture the full range of difference *within* the set of human beings: some humans have a greater capacity for abstract thought than others, or are more acutely sensitive to pain or pleasure, yet are still equally as human, and thus equally as entitled to moral respect as those others who are lesser in capacity. Thus, it seems that the introspective process is necessarily iterative: only when we think of removing some characteristic from ourselves little by little can we discover the *sufficient* level at which possession of that characteristic still allows me to identify myself as human. In the terminology popularised by Rawls, the set of capacities necessary for my identity as a human being is a set of *range properties*: as Waldron defines it, '*R* is a range property with respect to *S* if *R* is binary and there is a scalar property, *S*, such that *R* applies to individual items in virtue of their being within a certain range on the scale connoted by *S*' [1]. For example, the property of 'being in Oxford' is a range property, as it is equally satisfied whether one is exactly in the center of town or on the extreme outskirts (or indeed anywhere within that range). Similarly, the human capacity to feel pain or pleasure is one characterized by a range: anywhere within the range, conceived as being above the 'sufficient level' specified by the iterative second-stage of the heuristic, facilitates one?s free self-identification as a human being.

So where do we find ourselves now? A human being, in isolation, has embarked on the procedure described above, and established a set of necessary range-properties, which are together necessary and sufficient for them to identify themselves as a human being. This set is a comprehensive exposition of the essential capacities of a human being. Yet, as mentioned in the opening section, we are still not much closer to establishing which of these properties is *morally relevant* in such a way so as to entail both a human being?s moral worth, and the *equal worth* of all beings of this type. The answer to this problem follows from considering the full implications of an original position of self-respect. The meaning of the term 'self-respect' is highly contested, so I now intend briefly to clarify how I intend to use the term. My conception of the term is intuitive and minimally contentious. It is the essence of 'recognition self-respect' [2], as opposed to 'evaluative self-respect'. To help elucidate this distinction, let us return to my earlier characterization of the duality lying at the core of every person?s identity, that of them being simultaneously a human being *and* a unique individual of that type. Evaluative self-respect applies primarily to our status as a unique individual: we evaluate our actions or our character against standards prescribed both individually and socially, and value ourselves as an individual in accordance with this evaluation. We may respect ourselves based on our talent for mathematics or football, or our proclivity to do good for others. Yet the form of self-respect underscoring my argument, and the form that will be shown necessarily to entail equality between human beings, is recognition self-respect. This is the fundamental value we assign to our status *as a human being* (part of which individuality or autonomy may constitute necessary capacities). As a being fulfilling the set of necessary range-properties constitutive of human personhood, we assign ourselves some ultimate value. Independent of our manifestation as an individual of that type, and independent of circumstance, we endow ourselves with a special type of value that is non-evaluative in nature. This value operates in the way we treat ourselves, and the way we expect others to treat us. The relevant fundamental feature of self-respect is that it represents an assignment of moral value to oneself by oneself: the subject and object of valuation are the human individual. This notion of self-value lies at the center of the forthcoming analysis: I show that a position of self-respect, conceived essentially as a position of self-valuing, necessarily entails a moral equality between human beings.

With this established, we must note that as each range-property is a necessary condition for the agent's identity as a human being, and as the agent is aware of this, then the agent cannot achieve this state of self-respect without any one of the properties. Each property, by virtue of it being a necessary condition, is equally vital for the agent?s own perception of their value *as a human being*. Thus, we can say (or, more precisely, the introspecting agent can say) that each property is an equal store of value for that individual; without any one of these properties, the agent ceases to value themselves as a human being. They may still of course value themselves as some other non-human being, but it would be impossible for them coherently to assign themselves value as a creature of the type human being, as they would, necessarily, not be a being of this type. For a human being to value themselves is for them to assign overall value to the set of capacities that constitute them as a being; as possession of each capacity is necessary for them to identify as a human being, then it is also equally necessary for their self-respect as a being of that type. As such, each capacity contributes equally to the individual?s self-assigned value; each is an equally necessary constitutive part of human self-respect. The important move to make here, however, is to realise that these properties are worthy of respect, or contain *relevant value, independent* from the fact that a human agent possesses them. These capacities are not valuable because we possess them, but we are valuable because we possess these qualities. We shall presently explore this claim.

Consider the following situation: one day you are breaking rocks, simply to pass the time. Suddenly, just before striking one of the rocks, it begins talking to you, and asks you not to smash it to smithereens (we can be sure that you aren?t hallucinating). The rock asks you not to end its existence as a rock by breaking it into tiny pieces. The rock presently displays more than adequate evidence for possessing a degree of conscious awareness that we would consider sufficient for a human being. Would you ignore the rock and proceed to hit it anyway? Would you think to yourself: 'the possession of this characteristic only carries value when I possess it, so the rock is endowed with no more relevant value as a result of possessing this characteristic than it would otherwise have as just a rock ? thus I should feel absolutely fine about continuing to smash it to pieces'. Such reasoning is deeply intuitively wrong. We would not smash the rock, and thereafter treat it with far greater respect than had it not possessed this characteristic. A similar situation would obtain if we knew that a particular bush was capable of feeling pain or pleasure, or if an ant had a deeply felt self-conception of the good life. This seems adequately to demonstrate that the characteristics or properties that we believe give value to us as a being *have value independently* of us possessing them. As they not only facilitate our own self-respect, but also demand our respect when exhibited in entities external or otherwise distinct from ourselves, then they must exist as vessels of value independently of the fact that we possess them. The implications of this are explored below.

Most obviously, if we deny proper respect to other entities that share these characteristics (for example if we had proceeded to smash the talking rock) then we thereby deny the value of these characteristics themselves. In turn, we thereby undermine to a greater or lesser extent our own self-worth as a human being. Let us again consider the example of the rock. Why, intuitively, do we flinch at the idea of breaking the talking rock, while happily proceeding to break an otherwise identical non-talking rock? It seems that the answer is this: we identify a property in the talking rock (a degree of human-like consciousness) that we know is a necessary constitutive capacity of us as a human being. This capacity, when present in us, forms part of the basis of our self-respect: without possession of that capacity to some sufficient degree, we would be unable to respect ourselves as a human being. We attempt to preserve that capacity in ourselves as it has value as a constitutive part of our self-respect: destruction or degradation of that capacity is an assault on our fundamental value as a being. So, if we proceed to destroy the talking rock, we deny its conscious capacity (relevantly similar to our own) any normative weight. In doing so we undermine the basis for our own self-respect: denying the value of some capacity when it is external to us strips it of any value it could possibly possess for ourselves. If we value these capacities in ourselves, then to be consistent we must value them in others: denying these capacities proper moral consideration wherever they are found undermines the basis of our own fundamental worth as a being: thus, in order to preserve the basis of our *own* self-respect, we must spare the rock from destruction. What implications does this have for relations between humans? Well, if we treat other entities in possession of some of the characteristics that we find necessary for our self-respect with some degree of the respect with which we treat ourselves, then seeing another being in possession of precisely a sufficient degree of *all* of these characteristics (i.e. another human being) then the possession of these characteristics endows that other with a precisely equal claim to respect. Thus, as the set of human beings (by definition) all share a certain set of characteristics, then we must afford them equal moral status if we are to retain our own status in an original position of self-respect. To treat other human beings with any less respect than we treat ourselves would be to make impossible holding *ourselves* as beings worthy of respect. In order to preserve our own worth, we must assert the equal worth of others sharing the same essential characteristics. Thus, it follows from a position of original self-respect that human beings necessarily enjoy a fundamental equality.

We have shown that human beings enjoy a fundamental normative equality by considering the necessary implications of any one individual of that type holding themselves with self-respect. There are two related and immediately obvious objections to some important implications of this approach, which I shall presently consider, and another that requires only minor clarification. The first is that this approach is *speciesist*. One could argue that the foundations of this approach, and the moral considerations it encourages are human-centric: our moral compass is firmly orientated around human capacities, and so conceiving of moral value in the way this approach encourages may lead us to overlook the moral claims of non-human animals. Such criticism is deeply misguided: my heuristic and its results give a rich and compelling account of why we must assign moral worth to non-human animals. Think back to the example of the talking rock. My account gives a rigorous philosophical explanation as to why the rock should be the subject of a moral obligation not to harm it. The same reasoning applies to non-human animals. Animals such as a cow, for instance, display clearly numerous 'human' capacities, albeit perhaps to a lesser degree. That they possess the capacity to feel pain and pleasure is beyond doubt, as is the fact that they possess a considerable degree of conscious awareness. Thus, like the talking rock, to treat animals with any less moral respect than they are due is not only to contravene *their* dignity as beings, but also to undermine our own worth as beings capable of self-respect. We are obliged to afford them a degree of moral respect for the same reason we are obliged to treat other humans with equal respect: to do otherwise would undermine the basis of our own self-respect. Indeed, although a full substantive formulation of the practical implications of my view is not here necessary, it is worth saying that such an approach would certainly entail a very high moral standard for the treatment of non-human animals. The raising of farm-animals for the meat industry would clearly be morally prohibited: to murder and eat any animal that displays high levels of numerous human capacities subverts the value of the capacities, and thus undermines the basis of our own self-respect. Treating animals better than we do presently would be necessarily entailed simply by us holding ourselves as a being worthy of respect; in this sense, my account is far from speciesist.

The next objection is that the account of moral obligation following from the heuristic outlined in this essay struggles to deal with 'marginal cases', such as severely disabled people or persons in a vegetative state. In essence, the objection runs like this: persons in vegetative states, or otherwise severely mentally handicapped persons, clearly do not possess to a sufficient degree the same set of capacities constitutive of a human being. Thus under this view, they are seen as less morally valuable than typical humans, and thus less worthy of respect. One could maintain that this is morally objectionable. Yet such a criticism is again profoundly misguided, as is its central assumption that treating 'marginal cases' with less respect than typical human beings is objectionable. First, consider the situation wherein one is forced to kill one of two people, who are identical in all but the fact that one is in a permanent vegetative state, and the other is not. Which would we decide to kill? Our intuition says that it would have to be the vegetative individual. Of course, this is not to say that that individual demands *no* moral respect: quite the opposite, in fact, given that they demonstrably still possess or easily potentially possess some human capacities. Far from being incapable of adequately dealing with marginal cases, my approach provides grounding both for our intuition that such cases demand a high degree of moral respect, and that this respect should be less than that commanded by a typical human being. The grounding of moral reasoning in the necessity of consistency in our position of self-respect clearly, at first glance, adequately deals with some preliminary objections against it.

Lastly, I am aware that many philosophers distinguish strongly between notions of moral 'concern' and 'respect'. Conventionally, moral concern for an object is said to arise from its having interests, whereas an object is worthy of respect in virtue of its having the capacity for autonomous agency (in the Kantian view). I am aware that I have, to an extent, conflated the two notions throughout the essay. This was suggested as an objection to, or at least a problem with, the view expounded in this essay. Yet I see no such problem. The longstanding distinction between concern and respect is not dwelled upon precisely because I attempt to construct a novel account of human equality, and in so doing reveal some implications of such an account for moral theory more generally. As such, in considering some of the implications of deriving moral equality from the bedrock of self-respect, we can see that the distinction between respect and concern seems to evaporate: to show something proper moral concern, such as other humans or non-human animals, is simply to show it the level of respect it is due, as a proportionate corollary of the respect we show ourselves. The reasoning of this essay seems to suggest that moral concern is derivative from an original position of human self-respect, as a necessary implication of its coherence, and as such there is no unjustified conflation or confusion of terms.

To conclude, the central objective of this essay has been to show that there is a sound basis for the claim that human beings are fundamentally morally equal. This has been demonstrated by carefully drawing out the implications of a human individual coherently holding themselves with self-respect. As we have seen, I construct a heuristic to determine the set of necessary range-properties that constitute a human being; a position of self-respect entails that each of these properties possesses normative value, regardless of who or what possesses them. As such, for our own self-respect to be coherent, we must show equal moral respect to beings in possession of the same properties as ourselves (ie. other human beings) and proportionate respect to those beings in possession of some of those capacities. The implications of this view of course require more detailed expounding in future: yet at the present time it is clear that the heuristic constructed in this essay represents a powerful tool, and one which can be deployed to show coherently that human beings enjoy a fundamental moral equality.

References

[1] Waldron, J. (2008). Basic Equality. *NYU School of Law, Public Law Research Paper* No. 08-61. Available at SSRN: https://ssrn.com/abstract=1311816 or http://dx.doi.org/10.2139/ssrn.1311816.

[2] Dillon, R. (2014). Respect, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (Ed.) http://plato.stanford.edu/entries/respect/#ResForNatOthNon.

Aphantasia: An Unimaginative Defence of the Transparency Argument

Benjamin Evans
BPP Waterloo

Much debate in philosophy of mind concerns the nature of our perceptual experience and, more specifically, whether the phenomenal content of perceptual experience supervenes on its representational content. One argument advanced in favour of this thesis, known as *representationalism*, is the transparency argument. However, upon deeper analysis, this argument seems as much a consideration in favour of reducing representational content to phenomenal content as the reverse. The central contention of my argument is that an analysis of the content of other mental states, especially in the light of aphantasia (a condition whose sufferers lack mental imagery), decisively rules out this line of rebuttal.

But before any analysis of arguments can be embarked upon, the terms of discussion must be set out. Of the two types of mental content, phenomenal content is the most intuitive to understand; it concerns '"what it is like" for the subject to be in a particular mental state' [1]. This can be contrasted with representational content, which 'represents the world to be a certain way' [2, 203]. To help draw a distinction, consider your perceptual experience of reading this essay, either on a computer screen or in paper format. The particular perceptual experience you are currently subjectively experiencing'feels' a particular way; it also 'represents' particular facts about the world as being the case: notably, that there is (by supposition) a computer in front of you, with a philosophy essay on its screen. The former element of your perception is its phenomenal content, the latter element is its representational content. Representationalism, then, argues 'the propositional [representational] content of perceptual experience in a particular modality (for example, vision) determines their phenomenal character' [2, 206]; in other words, the phenomenal content of perceptual experience is reducible to its representational content.

Of all the arguments marshalled for representationalism, the most potent is the *transparency argument*, as first raised by Gilbert Harman. The argument runs something like this:

1. In experience it is relatively easy to 'distinguish between the properties of a represented object and the properties of a representation of that object' [3, 5]. (To take an illustrative example, imagine you are looking at a painting of a unicorn. By turning your attention to the represented object, namely the unicorn, you discern that it has a horn. However you can also turn your attention to the rep-

resentation itself, namely the paint, and discern that it does not have a horn.)

2. Although one can separate the properties of representation and represented object in the case of a painting, one cannot do this in perceptual experience, because all plausible candidates for the intrinsic properties of the experience (that is, for the phenomenal content of the experience) are features of the represented objects of experience rather than the experience itself. (To appreciate this fully, consider colour, a clear element of the phenomenal content of perceptual experience. Colour is always present as a feature of objects represented in experience, for example, a white car, a black pen, a yellow dog, etc., rather than being discernible as having a separate, purely phenomenal character. To employ Harman's delightful phrase, we are never aware of *mental paint* of phenomenal content going beyond the representational content of our experience.)

3. If phenomenal but non-representational elements of perceptual content exist, they must be introspectible. This may seem a contestable premise at first, but in truth it is a highly reasonable one, once consideration is given to the fact that what marks out perceptual experience from other mental states is that *perceptual experience is, or is capable of being, perceived*[1]. Consequently, to posit an element of perceptual experience which is incapable of being perceived at all is to posit a contradiction.

4. Were perceptual experience to exist that possessed phenomenal content but lacked representational content, we would have to be able to introspect it (premise 3), however we do not (premise 2), therefore such experience does not exist.

5. * Therefore, the phenomenal content of perception must be reducible to the representational content of perception. (After all, were it not, the fact that phenomenal content never existed outside of instances of representational content would be a vast and inexplicable coincidence.)

This argument seems convincing at first, but there is a notable logical leap between premise four and the conclusion. Specifically, there seems no to be immediately apparent reason why transparency should lead us to reduce phenomenal content to representational content, rather than leading us to reduce representational content to phenomenal content. Indeed, although transparency may demonstrate that we never perceive phenomenal content going beyond representational content, it also establishes that we never perceive representational content going beyond phenomenal content. Following premise three–elements of perceptual experience need be actually perceived

[1] This is a fundamental intuition that also poses a threat to the direct object account of perception, although a discussion of this issue will have to be left for a later paper.

to exist–we arrive at the belief that (at least so far as perceptual content is concerned) representational content independent of phenomenal content does not exist. Upon arriving at the conclusion, we then face two distinct possibilities: reducing phenomenal content to representational content or representational content to phenomenal content. Neither is obviously better than the other, and only one is conducive to the representationalist thesis.

There does, however, exist a means by which the representationalist can demonstrate that their preferred form of reduction is the more plausible, and thereby turn the transparency argument in their favour. That is, by demonstrating areas outside of perceptual experience where representational content exists absent of phenomenal content, they can argue that representational content need not, and does not usually, reduce to phenomenal content. Once that is established, we have good reason not to believe that representational content reduces to phenomenal content in the area of perceptual experience and for believing in the only alternative: that phenomenal content reduces to representational content.

One potential example of such a representational but non-phenomenal mental state is posited by Tye, who 'maintains that beliefs do not have phenomenal character. People suppose otherwise just to the extent to which various other states with phenomenal character (such as some memories and imaginings) accompany beliefs' [4, 619]. Medical evidence for such a view about belief is, however, present in the curious condition known as *aphantasia*. This is a condition involving 'reduced or absent voluntary mental imagery' [5, 2]. Sufferers are outwardly normal, to the extent that many go decades before they realise the nature of their impairment. However, whereas ordinary individuals are capable of constructing mental images (for example, visualising a beach or the faces of their parents), those with aphantasia are unable to do so. The condition is a rare one; however in a 2009 study involving 2,500 participants, 2.1-2.7% were found to 'claim no visual imagination' whatsoever [6]. Equally, individual cases such as that of 'MX', an Edinburgh architect who reported 'he was missing his mind's eye' [7] after an operation, are compelling in themselves. As interesting as aphantasia is in its own right, its implications for the nature of belief are even more interesting.

Crucially:

1. The only plausible phenomenal content for belief's representational content to reduce to is mental imagery of some kind (be it pure imagination, or mental imagery connected to memory).

2. Sufferers of aphantasia possess beliefs whilst lacking mental imagery of any kind.

3. The representational content attached to the beliefs of aphantasia sufferers therefore does not supervene upon, or reduce to, the phenomenal content of aphantasia sufferers' mental states; given there is no relevant phenomenal content for

such representational content to reduce to.

4. If beliefs amongst aphantasia sufferers can possess representational content without this reducing to/supervening upon phenomenal content, the same should be true of belief more generally, unless we are to suppose that belief amongst aphantasia sufferers is a fundamentally different kind of mental state to belief amongst ordinary individuals.
5. Given that aphantasia sufferers appear and act almost identically to ordinary individuals, and that there are reported instances of individuals gaining the condition without also reporting a fundamental change to the nature of their beliefs (cf. the "MX" case), the contention that belief in aphantasia sufferers is fundamentally different from ordinary belief is highly dubious.
6. We therefore have a clear example of a widespread and fundamental type of mental state (namely belief) possessing representational content that does not reduce to phenomenal content (although it may be accompanied by such content in ordinary individuals).
7. Through argument by analogy, we now have a good reason to doubt that representational content reduces to phenomenal content in the case of other mental states.
8. Accepting the initial premises of the transparency argument, as applied against both phenomenal content separate from representational content and representational content separate from phenomenal content, we have compelling reason to reduce one of phenomenal or representational content to the other in the case of perceptual experience.
9. * Conclusion: we should uphold the initial conclusion of the transparency argument, and reduce phenomenal perceptual content to representational perceptual content.

Thus our overall conclusion is a simple one: the transparency argument for representationalism rests on the contention that it is more plausible to reduce phenomenal content to representational content than to do the reverse. A crucial way to support such an argument would be to provide examples of mental states where representational content does not reduce to phenomenal content. aphantasia demonstrates that belief is one such state[2]. Therefore a refined transparency argument can be upheld against its objectors, and provides a strong basis to ground a wider representationalist position.

[2] Or, at the very least, that a subset of beliefs are such states.

References

[1] Lycan, W. Representational Theories of Consciousness, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (Ed.), https://plato.stanford.edu/entries/consciousness-representational/.

[2] Byrne, A. (2001). Intentionalism Defended. *The Philosophical Review*, 110(2): 199–240.

[3] Harman, G. (1990). The Intrinsic Quality of Experience. *Philosophical Perspectives*, 4:31–52.

[4] MacPherson, F. (2003). Review of *Consciousness, Color and Content*, by Michael Tye, *The Philosophical Quarterly*: 619–621.

[5] Zeman, A., Michaela Dewar, M. and S.D. Sala, (2015). Lives Without Imagery: Congenital Aphantasia, *Cortex*: 3–5.

[6] Faw, B. (2009). Conflicting Intuitions May Be Based On Differing Abilities: Evidence from Mental Imaging Research. *Journal of Consciousness Studies*, 16(4): 45–68.

[7] Zimmer, C. (2010). The Brain: Look Deep Into the Mind's Eye, *Discovery*.

On the Structure of Mental Processes: The Systematicity Debate

Ellen Jin
Pembroke College, University of Oxford

The structure of our mental processes is an insistently penetrating and, to date, enduring mystery. The latent philosophical and scientific value of this question has prompted many projects across and between neuroscience, linguistics, computer science, and the other cognitive sciences. The lack of consensus derives, in large part, from the competing theoretical assumptions that inform the ways in which cognitive scientists conceive of cognition schematically and/or understand the indirect indications of cognition empirically (e.g. through fMRI scans of or verbal reports from subjects). Two such schools of thought are the classical model of cognition and connectionism, which models intellectual capacities or specific mental phenomena as the emergent processes of interconnected networks of discrete units, i.e. a neural net. The structure of mental processes underpins and incites a narrower problem concerning whether or not and to what extent a model can explain the phenomenon of systematicity, the idea that the ability to produce or understand a proposition is intrinsically connected to the ability to produce or understand others of a related structure. For example, if one understands 'John loves Mary', then one, by nomonological (syntactical) necessity, understands 'Mary loves John' as well. Fodor and McLaughlin, in their revisitation of Smolensky's reply to [1], argue that the connectionist model cannot display systematicity and productivity as it fails to exhibit the *classical constituents* that form the domain for structure sensitive mental processes [2, 183]. This debate constitutes the central concern of this following discussion.

In this paper, I first assess whether or not and to what degree the connectionist model displays systematicity and productivity. I argue that connectionism is not systematic as it merely displays signs of approximate systematicity that is, in actuality, the result of frequent association and inferential relations. Secondly, I discuss the extent to which Fodor and McLaughlin's rejection of the connectionist model as a systematic and causally-sensitive model is a strong objection against the claim that connectionist models nevertheless realize various psychological states. My broader position is that while connectionism is not systematic in the way that the classical model is, Smolensky's sub symbolic approach is able to provide a satisfactory explanation for the claim that various psychological capacities are or can be realized by connectionist systems.

The Classical vs. Connectionist Model

The classical model of cognition is in part an extension of Jerry Fodor's Language of Thought Hypothesis (LOTH) [2, 183] This theory holds that cognition should be understood by analogy with classical computational architectures as a system of rules and representations. Assuming that the causal syntactical, systematic, and productive characteristics of natural language are true descriptions of how natural language operates, Fodor reasons by inference that the same structure can be mapped onto most cognitive processes such as concept learning and perceptual recognition. Fodor had in his mind that the LOTH must be a presupposition for all psychological models. This assumption rests on the more fundamental assumption that there is a representational medium in which concept learning and perceiving occurs [3] This is the sticking point: in light of the idea that cognitive processes are by nature computational, there must be a representational medium (much like the syntactical network of natural language) for the process to occur in. The 'evidence' for a representational medium of thought – i.e. a mentalese – is, as mentioned, inferred from the operational characteristics of natural language. It does so, he argues, in virtue of satisfying the following two conditions:

1. its representations must have a combinatorial syntax and semantics. This is to say that complex representations of objets are composed of simpler, discrete units, and that the process of composition is systematic;
2. the cognitive processes defined within the LOT must be 'causally sensitive only to their syntax', meaning the mental process carried out is a causal output or effect of the process's own compositional structure[4, 143].

For example, 'for a pair of expression types E1, E2, the first is a classical constituent of the second only if the first is tokened whenever the second is tokened'. This means that when a representation is tokened, the constituents of that representation are automatically tokened as well. The ability to express a proposition is just the ability to token a representation whose content is that proposition. As Fodor posits, it is only in virtue of having classical constituents can mental representations be structure-sensitive, that is that the cognitive process that composes and expresses propositions 'have access to their constituents' [2, 187]

The classical model of cognition, as implied by its two central tenets, explain systematicity as an effect of the constituent structures and the semantic relations between those structures. Under the classical interpretation, thoughts are systematic in virtue of their discrete atomic units and the rules of composition. Systematicity is demonstrated when I can have the thought that 'girl loves John' in virtue of having the initial thought 'John loves the girl', so long as 'John', 'loves', and 'the girl' are all atomic units of thought within my mental vocabulary.

This points to the reversible and infinitely re-organizable characteristic of compositionality given that the relations between the atomic units are logical inferences [2, 186]. Systematicity also implies inferential coherence. Our capacity to follow a pattern of inference, according to the classical view, is intrinsically connected to our capacity to draw certain other inferences. Thoughts are also productive because in principle there is no upper bound on the number of sentences a mind can construct as we can generate infinite, non-repeating composite representations by reorganising its atomic constituents. Together, the systematicity, productivity, and coherence of thought suggest to the classical model that mental representations possess a constituent structure.

Against the classical model, Paul Smolensky posits the 'distributed view of connectionist compositionality' that allows connectionism to instantiate the two classical requirements for systematicity without resorting to a LOT [4, 151]. Connectionism understands the mind in terms of an interconnected network of mechanisms. Its proponents contend that cognitive properties can be explained in terms of their emergent properties from the collective behavior of simple interacting mechanisms and *adaptive* characteristics. Smolensky posits that the fundamental difference between the two approaches are as follows: in the classical model of cognition, the principle of (1) composite structure and (2) that mental processes are sensitive to this composite structure are 'formalized using syntactic structures for thoughts and symbol manipulation for mental processes'. Under connectionism, these two principles are 'formalized using distributed vectorial representations for mental states and the corresponding notion of compositionality, together with association-based mental processes that derive their structure sensitivity from the structure sensitivity of the vectorial representations engaging in those processes' [4, 150-51]. The argument is that the connectionist approach satisfies the two principles but differs in how it instantiates them formally. Proponents of the classical view reject this. Firstly, they argue that connectionist models do not provide mental representations with *classical constituents*. Accordingly, there is no indication as to how mental processes can be structure-sensitive if it lacks classical constituents. As a result, there is no way that the model can be systematic if mental processes are not structure-sensitive' [2, 188].

Does Connectionism have a constituent structure/display systematicity?

It is Smolensky's central defense that Fodor and Pylyshyn are mistaken in their claim that connectionist models lack constituent structure because they failed to understand its distributed representation system [4, 137]. Smolensky offers an alternative to the classical account of systematicity, one that corresponds to ways in which complex mental representations can be distributed. According to this model, representations of higher or macro level conceptual entities such as propositions and sentences are 'distributed' or spread out over nodes (as opposed to being confined to a single atom), while the nodes might simultaneously participate in the representation of multiple mental processes. The connectionist model's goal is then to show that by employing the distributed representation system, it can ascribe to mental states the compositional structure demanded by the classical model [4, 144]. In the following discussion, I provide a recount of one of Smolensky's defenses. I come to the conclusion that though the output (i.e. composite proposition) of the cognitive process may appear systematic, there is a distinction to be made between appearing systematic and behaving systematically.

Smolensky considers a distributed representation of a 'cup with coffee'. By subtracting from it a distributed representation of a 'cup without coffee', what remains is the connectionist representation of 'coffee'. In order to produce these representations, Smolensky uses a set of micro-features, such as 'upright container', 'hot liquid', 'glass contacting wood', and so on. When we have a distributed representation of 'cup with coffee', the *active units* are the ones which correspond to the micro features and part of the description of a cup with coffee. When we combine the representations of 'cup without coffee' and 'coffee', we have, in effect, constructed the representation 'cup with coffee' from a representation of 'cup' and a representation of 'coffee' [4, 145-46]. This is the way in which distributed representation attempts to satisfy the compositionality requirement. The argument is that it exhibits an inference mechanism that takes as input the vector representing 'cup without coffee' and 'coffee' and produces an output vector representing 'cup with coffee' as a mechanism that extracts a part from a whole. In this sense, Smolensky argues that it is no different from a 'symbolic inference mechanism that takes the syntactic structure A and B and extracts from it the syntactic constituent A' [4, 150].

This attempt to satisfy the compositionality and constituent structure requirements is rather self-defeating for the connectionists. Fodor's objection is not just that connectionist models are unable to account for higher cognition. Rather, it is that they can do so *only if they implement the classicist's symbolic processing tools*. To revisit, we may be reminded that the LOTH requires that the logical relations and the satisfaction requirements that hold between discrete mental 'propositions' are causally accessible to the subject. The classicist view of systematicity thus claims that the phenomenon of systematicity occurs in virtue of the above two tenets. In this case, it seems that Smolensky, in his attempt to establish a parallelism between his inference mechanism and Fodor's symbolic mechanism, has not demonstrated connectionism's own systematicity but merely, as Fodor argues, 'implemented classical architecture' [4, 145].

There is an extent to which connectionist models can display thought that appear systematic, but it does not do so systematically. Given that under the connectionist conception of cognition, mental processes are stored *non-symbolically* between discrete units within a neural network, cognitive processes explained through connectionism are still largely associationist exercises. Even if some systematicity is displayed, it will be a result of consistent exposure and confirmation (i.e. association). This is to say that the thought displayed is not a direct manifestation of the organization principles, which have the inherent ability to formulate and reformulate thoughts systematically. As such, the connectionist model may only claim to have superficial systematicity in its output, but cannot attribute this superficial systematicity to any deeper structure of the model itself. The displayed systematicity is superficial because its apparent systematicity is merely a sign of increased out-put producing reliability which is, as Smolensky himself posits, the result of a series of *inferential* and *statistical* inferences. So long as the system operates on principles of statistical inferences, the output can only be probabilistic and is therefore not systematic. Here Fodor's complaint against connectionists comes through clearly is preferable, while connectionist models may implement systems that exhibit systematicity, they will not have explained the cognitive process's systematicity unless the model itself takes systematicity as a nomological necessity [2, 188] The sticking point is that such demonstrations as exhibited by the coffee example only show that networks can be accustomed to exhibit systematic processing, and not that they cause it in virtue of the internal structure of the system itself. This distinction can easily be seen when we understand that on the classical account, the same rules that govern one mental process *automatically* and likewise govern any of its compositional variants (i.e. any reformulation of a composite's atomic units). This is evidently not the case for the coffee example. In short, connectionism inevitably fails to provide a truly systematic explanation of cognition insofar as it conflates the intrinsically systematic nature of thought with a system of associations, regardless of its predictive powers and general reliability.

Can Psychological Capacities Be Realized By Connectionist Systems?

Thus far, I have argued that connectionism is not systematic as it merely displays signs of approximate systematicity that is, in actuality, the result of frequent association and inferential relations. That said, the strength of Fodor and Plyshlyn's criticism depends on how much of the classical model's beliefs surrounding the classical model of cognition we are willing to accept as putative facts. Only if we take systematicity and relevant stipulations from the LOTH to be the de facto descriptions of how mental representation operates can we confidently reject the fact that various psychological capacities are realized by connectionist models as evidence for their systematicity. The fact that various psychological capacities are or can be realized by connectionist systems is not an indication that they did or do so systematically. It merely demonstrates their adaptive capacity.

That said, Smolensky's cognitive correspondence principle presents a challenge for the classical model of cognition. To discuss this problem, we may briefly visit what Smolensky terms the 'Structure/Statistics Dilemma', which describes the explanatory tendency whereby focusing on rules governing higher level cognition pulls us toward 'structured, symbolic representations and processes', whereas variance and lower-level descriptions of cognition requires that we employ 'statistical, numerical descriptions' [4, 138]. Smolensky subscribes to a sub symbolic approach that conceives of the cognitive system as a 'fundamentally soft machine that is so complex that it is sometimes appears hard when viewed at higher levels' [4, 138]. This is a counterintuitive but nevertheless interesting suggestion that appears quite formidable upon further examination. It is conceived of in part to address what Smolensky calls the cognitive correspondence principle, which states that 'when connectionist computational systems are analyzed at higher levels, elements of symbolic computations appear as emergent properties' [4, 152].

Even though it does not operate systematically, the sub-symbolic approach is most able to accommodate the fact that various psychological capacities are realized by connectionist systems as it does not assume that the lower level activities of all organisms or psychological states can be subsumed under one psychological law or are bound by logical relations. Rather, in virtue of its belief that cognitive capacities can be modeled in terms of their emergent properties from the collective behavior of interacting nodes and adaptive mechanisms, it is able to explain, in its own terms, how certain capacities do not obey a certain psychological law, without casting it aside as an anomaly.

Conclusion

In brief, I have argued that the connectionist model may only claim to have superficial systematicity in its output, but cannot attribute this superficial systematicity to any deeper structure of the model itself. The displayed systematicity is superficial because its apparent systematicity is merely a sign of increased out-put producing reliability which is, as Smolensky himself posits, the result of a series of inferential and statistical inferences. So long as the system operates on principles of statistical inferences, the output can only be probabilistic and is therefore not systematic. Strictly speaking, connectionism fails to provide a truly systematic explanation of cognition insofar as it conflates the intrinsically systematic nature of thought with a system of associations, regardless of its predictive powers and general reliability. If we accept composite structure and structure-sensitivity as putative facts about the nature of cognition, then the same objection advanced by Fodor can be extended toward the claim that psychological capacities are/can be realized by connectionist systems. With that said, I have briefly discussed in the previous section how Smolensky's cognitive correspondence principle may limit the extent to which the Fodor's objection is a wholesale rejection of connectionism's explanatory power.

References

[1] Fodor, J. and Pylyshyn, Z. (1988). Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, 28: 3–71.

[2] Fodor, J. and McLaughlin, B. (1990). Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work. *Cognition*, 35: 183–204.

[3] Fodor, J. (2005). The Language of Thought. In *Philosophy of Psychology*, José L. Bermúdez (Ed.). New York: Routledge: 101–26.

[4] Smolensky, P. (1987). The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pylyshyn. *The Southern Journal of Philosophy*, 26 (Supplement): 137–161.

Durham University

MA Philosophy
Durham University

MA Philosophy

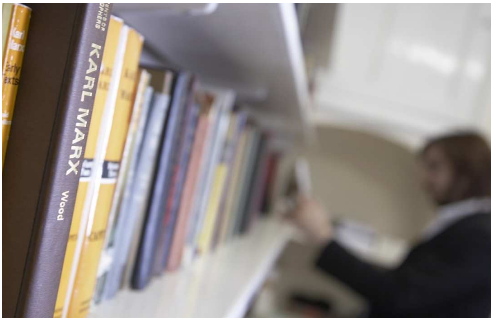
Course structure

This one year programme (two years part-time) comprises taught modules and a substantial dissertation. Modules are taught via group seminars and one-to-one tutorials. Alongside core compulsory modules you can choose between a range of topic modules allowing you to develop your own research focus throughout the course. Examples of topic modules include:

- Current Issues in Ethics
- Philosophy of the Social Sciences
- Mind and Action
- Philosophical Issues in Science and Medicine
- Current Issues in Aesthetics and Theory of Art
- Current Issues in Metaphysics
- Phenomenology and the Sciences of Mind
- Gender Theory and Feminist Philosophy
- Ancient Philosophers on Necessity, Fate and Free Will
- Environmental Philosophy
- Business Ethics

Modules offered are subject to change.

For up-to-date information, see www.durham.ac.uk/philosophy/postgrad/taught/taughtphil



Career Opportunities

This MA course provides students with a wide number of important skills, including sophisticated analytic skills and intellectual clarity. This kind of study involves a high level of commitment, motivation and discipline, inherently employable qualities. It also provides an ideal academic environment for those who would like to study the subject at a higher level in preparation for a PhD.

MA students can benefit from a range of other activities in the department, including the department's postgraduate philosophy society (EIDOS), weekly research seminars and reading groups, and occasional conferences, workshops and Royal Institute of Philosophy lectures.

Entry requirements

A typical 2:1 classification or higher at undergraduate level or equivalent qualification with a substantial philosophy component.

If candidates wish to take the MA course with a research focus on Science, Medicine and Society, they may be considered if they have an undergraduate degree or equivalent qualification with another appropriate component, for example science-related subjects. For details, see www.durham.ac.uk/philosophy/postgrad/taught/taughtphil/resfocusscimedsci

For English language requirements, please see www.durham.ac.uk/philosophy/postgrad/taught/taughtphil

How to apply

You can apply online at www.durham.ac.uk/postgraduate/apply

Enquiries

For postgraduate enquiries, please contact: +44 (0)191 334 6553 or philosophy.pgsec@durham.ac.uk

All information provided was correct at the time of print, but is subject to change - for the latest information, please see www.durham.ac.uk/philosophy/postgrad/taught/taughtphil

Issue 12 (1)
Annual Conference 2018

Journal of the British Undergraduate Philosophy Society