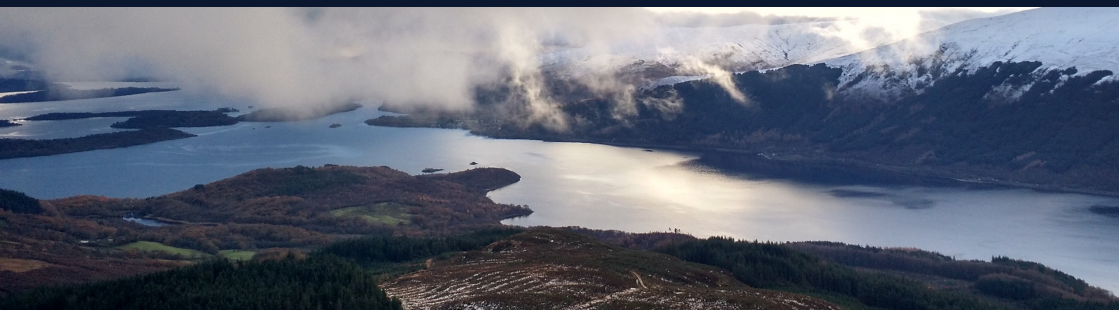# *British Journal of Undergraduate Philosophy*

Editor-in-chief: Geoff Keeling
London School of Economics

Journal of the British Undergraduate Philosophy Society

British Journal of Undergraduate Philosophy

The Journal of the British Undergraduate Philosophy Society
Issue 10(1) Spring 2016

www.bups.org/bjup-online

British Journal of Undergraduate Philosophy

www.bups.org/bjup-online

Editorial

> A serious and good philosophical work could be written consisting entirely of jokes.
>
> —— Ludwig Wittgenstein

December 29th 2013: there I am, cocooned in blankets, nursing the ritual hangover that's part-and-parcel of celebrating Christmas at my mother's. I stretch out a hand to grab the old BlackBerry, and through sleepy eyes wrestle through a well-articulated email from someone called Dan Treger, who seems to be offering me a job. Despite completing the Oxford Reading Tree as a child, and thus being a 'free reader' now, then and for evermore, I can't make head or tail of it. So I adorn some jim-jams, graced, of course, with the family tartan, and stagger downstairs to find the more literate members of the family.

It turns out that I'd applied *months ago* to the journal and forgotten all about the application in the hustle and bustle of advent calendars, presents and lavish over-eating. Whether by divine intervention or human error, the application had been successful. I now had a job – sort of. It was my responsibility to take submissions, anonymise them, and send them to smart people who provide feedback. In the months that followed, and with only one incident that resembled a *coup d'état*, I climbed the ranks of BUPS until I ended up sitting like a dishevelled budgie on the top perch.

I then had a hand in producing four journals, including the remarkable 10th anniversary edition. This couldn't have happened without the BUPS team, who really are the most talented philosophers, editors, organisers and administrators that I could have hoped for. So it's with a heavy heart, but a smile on my face, that I'm throwing in the towel. It's time to give up the mantle, and it gives me no greater pleasure than to announce the inimitable Ella Langham as my successor. Ella makes history by being the first female Editor-in-Chief of *BJUP* since its founding over a decade ago (it's about time!), and I have every confidence that Ella will be instrumental in bringing our work to a new generation of undergraduate philosophers.

I must thank all those who submitted papers to the Journal this year. BUPS is immensely grateful to all the undergraduates who took the time to produce the outstanding papers that greet us in the submissions inbox each year. There are also many people deserving of a mention for their contribution to the production of this edition. Of course, no amount of thanks can begin to capture the tremendous dedication of the *BJUP* team, which manifests in the considerable time and energy put in to the development of each edition. But, in partial fulfilment of the debt I owe to the wonderful members of this team, here are

some special mentions.

I first met Sophie Osiecki at UCL in January 2014. It was the first BUPS meeting that I was invited to, and I had absolutely no idea where to go. I'm not quite sure how Sophie worked out that I was the new Assistant Commissioning Editor, because I wasn't even trying to look like a philosopher, but she approached a very lost-looking me, and asked whether I was headed for the BUPS meeting.

Over the last two years, Sophie has always been a 'Manuscript Editor'. But this title doesn't begin to capture the quality and quantity of the work that Sophie has performed. For example, when the BUPS website went down shortly before the conference associated with this edition, Sophie went without instruction above and beyond the call of duty, and took it upon herself to build a temporary BUPS website to ensure that conference information was available to those who wanted to come. It's this kind of dedication that makes Sophie such a wonderful asset to the BUPS team.

Sophie deserves a second mention, in conjunction with Nate Oseroff and Bede Hager-Suart. Sophie, Nate and Bede are collectively responsible for the spectacular presentation of the *BJUP* that is consistently delivered year after year, after year. I am still unsure about who, or what, this LaTeX thing might be. But Nate, Bede and Sophie most certainly know how to put together an academic journal. Twice a year, these three give literally days of their time to transform a collection of outstanding papers into a readable publication. If I had spent less time studying analytic philosophy, and more time cultivating the art of verbosity, I might have more creative phrases up my sleeve. But I hope it suffices to say that Sophie, Nate and Bede are incredible assets to the team, and that it's an absolute pleasure to have worked alongside them.

I recruited Achilleas Sarantaris as Commissioning Editor in a drunken conversation at the UCL Student Union. Though this might sound like a recipe for disaster, things turned out relatively well as I soon learned that Achilleas is familiar with at least three disputes in contemporary analytic philosophy. Achilleas did a stellar job of co-ordinating the review process, despite the severe time constraint of being the drummer in an up-and-coming J-Pop band. Achilleas is leaving BUPS to focus on his studies and his music, and I am sure he'll be greatly missed by all.

Edmund Smith joined us in February as an Assistant Commissioning Editor. I'm always a little sceptical about Hegelians, but recruiting Edmund proved to be a great decision. Edmund has, without a shadow of a doubt, compiled the most beautiful commissioning spreadsheet that I've ever had the good fortune of working with. His capacity for organisation is second-to-none, and he

certainly knows how to impress a man by sending the sum of his labours in a neatly wrapped zip folder. I'm delighted to announce that Edmund will be taking over as Commissioning Editor for the next edition, in what I am sure will be the start of a very fruitful career at BUPS.

I now turn to some personal thanks. Farbod Akhlaghi-Ghaffarokh, the former President of BUPS, has provided immense support and friendship throughout my time here. Matej Križnar, my pocket anthropologist, has provided a measured and sober response to every deep-breaths-in-a-brown-bag crisis that has erupted throughout the years at BUPS. Žiga Dernač, whilst perhaps not doing much to calm things down, has certainly been an ever-present force throughout the calamities, and I think we're all the better for it. It's not an understatement to say that these guys, collectively and individually, have prevented many breakdowns and meltdowns, and made my time as Editor all the more worthwhile.

Finally, I am indebted to the Faculty of Philosophy at the University of Cambridge, who have kindly agreed to host our conference. I am also grateful to Professor Richard Holton (Cambridge), who is our invited keynote speaker. And now, for the last time, I am delighted to welcome you to this new edition of *BJUP*. Stay safe – and don't run with scissors.

GK

Contents

Nietzsche on 'Character' and 'Translation': Proposal for an Interpretative Strategy

David Kretz
Bard College Berlin

Nietzsche's writings pose a range of challenges for interpreters, both because of what he said and how he said it. His literary style refuses easy categorization into any genre of text. Often one finds very different, even contradictory thoughts on a particular topic in his vast *oeuvre*, only some of which seem to be explained by a development of views. At times he remains deliberately allusive and enigmatic.[1] And the controversial post-mortem redaction of some of his most influential writing by his Nazi-affiliated sister does do not make things easier either.

Different interpretative strategies have been adopted in response to these challenges. Some have attempted *rational reconstructions* that aim to distill of his texts a coherent body of arguments and theses on a range of topics. While such a strategy seems promising when dealing with some philosophers (say, Kant or Aristotle), for the above cited reasons, their potential in the case of Nietzsche seems humble, at least *prima facie*. Nevertheless, some people tried and some with considerable success. Alexander Nehamas' *Nietzsche: Life as Literature* could perhaps be cited as an admirable example of this tradition in its attempt to defend a unified and coherent thesis about Nietzsche (namely, that the literary text is Nietzsche's primary model for understanding the world).[2]

Most scholars have qualified and limited their reconstructions. Rather than going for the global picture of defending one big theory of Nietzsche, they attempt to reconstruct only his views on one particular sub-topic, and perhaps at only one point in time (say, 'the early Nietzsche's theory of art'). This introduces degrees of *eclecticism* and *pragmatism* – if the whole is not to be had, we must (eclectically) take parts of Nietzsche and (pragmatically) put them to use somehow. While scholars are often reluctant to apply such terms to their approach, some people, like Bernard Williams and Michel Foucault, are nevertheless quite straightforward about it: "I agree with a remark made by Michel Foucault in a late interview, that there is no single Nietzscheism, and

---

[1]The end of the second essay of the *Genealogy* comes to mind, for example.

[2]Nehamas [1] p. 3.

that the right question to ask is 'What serious use can Nietzsche be put to?'"[3]

The most sophisticated accounts of Nietzsche, on both sides of the analytic–continental divide, propose very carefully elaborated views, combining sensitivity to historical and philological context with erudite acquaintance with the whole of Nietzsche's *oeuvre*, and subtle argument with literary sensibilities, to qualify and justify their eclectic reconstructions. The accounts of people like Robert Pippin or Gilles Deleuze cannot be easily dismissed as overly ambitious rationalism, nor as cheap 'anything goes' eclecticism. While each of their readings might on some high level have some shortcomings,[4]

and while they do not venture to construct a single coherent theoretical body taking into account *all of* Nietzsche's writing, they can give reasons for the trade-offs and interpretive choices they make, and unless we can do better than them the proper thing to do might be to accept their readings, or what we deem the best among them, as authoritative.

It is beyond the scope of this essay to criticise thoroughly and decisively even a single one of those thoughtful interpretations. However, I want to take my dissatisfaction with both 'rational reconstruction' and 'eclectic pragmatism', and the fact that even those who blend the two in sophisticated ways remain in disagreement with each other, as motivation to explore the possibility of another interpretative strategy. The merits and potential of such a strategy will not be sufficiently tested unless a much longer application of it has been undertaken. And I should admit at the outset that until then any draft cannot wholly avoid the charge of being itself either overly rationalistic or overly eclectic. I am aware of the problem, yet think that this will not be an obstacle to the essay's main goal: to sketch out in rough strokes a different interpretative strategy, that will elicit enough curiosity to motivate further exploration, and prove in-

---

[3]Williams [2] p. 5. More on Foucault below.

[4]Pippin's readings of Nietzsche, and of Nietzscheans like Bernard Williams, often come in a distinctly Hegelian flavour, for example. Nor does he try to hide the fact very much (cf. Pippin [14] pp. 46, 49, 51, 61; [12] p. 125; [13] p. 185. For his take on Williams' Nietzscheanism see his 'Final Reply to Geuss', available in draft on his homepage). The general thrust of such a reading is to find in Nietzsche a dialectical movement that opposes a positive, affirmative side to the negative-critical side of his project. Yet at the culmination of Ch. 3 of *Nietzsche, Psychology, First Philosophy* Pippin admits that "the positive, erotic side of the project he [Nietzsche] proposes is only just barely on view and remains merely suggestive, tantalizing in the way he probably intends" (Pippin [14], p. 65). We can agree with the last part of the claim or not, but note that gently pushing Nietzsche into a Hegelian direction remains difficult, if not impossible, for even the most qualified and motivated.

structive in itself. I also want to add that my goal is not, of course, to replace previous approaches, but to offer considerations that helpfully complement them.

The proposal I want to outline – provisionally and imaginatively – unfolds in three steps. First, I offer a close reading of Nietzsche's thoughts on 'character', arguing that the literary character can serve as a model for synthesizing a whole out of Nietzsche's thought, albeit a dynamic and open-ended one (1). Secondly, following Duncan Large's speculation, and taking my cues from (i) Nietzsche's own writing, especially on 'character', (ii) etymology, and (iii) the phenomenology of translation, I argue that translation is indeed the key method by with such a synthesis could proceed (2). Lastly, I offer an axiological consideration, why such an interpretative strategy expresses precisely those values that are appropriate when dealing philosophically with Nietzsche, and conclude with a discussion of one value in question, the value of fidelity (3). The first part deals with an *ontological* condition, the second part with *methodology*, the last part with *axiology*.

(1) My *ontological* thesis is that trying to reconstruct from Nietzsche's writings a character, rather than a theory, allows for the reconstruction of a whole, avoiding some pitfalls of eclecticism – but a whole with all the dynamics, developments, and contradictions even, such as are typical of us human, hence avoiding some pitfalls of rationalist reconstructions. The notion of character is understood here more like we would use it when referring to a character in a theatre play or a novel, rather than to an entrenched moral or psychological disposition.[5]

Would Nietzsche be open to such a use of the term? He uses the word in different ways and with changing connotations throughout his writing. For once, he sometimes uses it to designate persons, sometimes he uses it for objects. For example, he speaks of the character of the genre of opera, or the character of a people, *Volkscharakter*.[6] In *The Birth of Tragedy* he notes that the term character, understood in its theatrical context, is at the root and origin of (Dionysian) drama: "This process of the tragical chorus is the original phenomenon [*Urphänomen*] of drama – this experience of seeing oneself transformed before

---

[5]The claims of this paragraph might seem to align themselves with Nehamas' theses that for Nietzsche literature is the model for life, and that the self's relation to itself is like an author's relation to her text (Nehamas [1]). While I am generally sympathetic towards much of this, I remain neutral on those theses here. What's at stake for me is not how Nietzsche thinks that we relate to ourselves or our lives, but rather how serious readers of Nietzsche shall relate to his work.

[6]Nietzsche [8] §23.

one's eyes and acting as if one had really entered another body, another character" (§8, p. 43). [7] Following common usage, a character in a play is seen as something one can take up and transform oneself into at will.

Yet more often in his very early writings, he conceives of it as a deeply entrenched ethical disposition and as radically opposed to the theatrical realm of mere appearance. In the *Untimely Meditations* for example he criticises David Strauss as lacking character while pretending otherwise: "This union of audacity and weakness, of rash words and cowardly acquiescence...this lack of character and strength masquerading as strength and character...".[8] With a similar usage of the term he celebrates the purity of Wagner's character, which protected him from corrupting parties and false friends.[9] In short, far from a *role* one takes on, one's character is exactly that which is *not* staged and played, one's true self in a way, that is opposed to mere appearance, role, image.[10] For example, Shakespeare's Romans, he contends, "are nothing but flesh-and-blood Englishmen", but *even so* they are more authentic characters than "our contemporary men of letters, popular figures, officials or politicians" who are but "creations of historical culture [*Bildungsgebirge*], wholly structure [*Bildung*], image [*Bild*], form, without demonstrable content and, unhappily, ill-designed form, and, what is more, uniform [*Uniform*]", as he puts it memorably.[11]

His next work onwards however, *Human, All-too Human*, we begin to see a shift in Nietzsche's understanding of character: from the true centre of a strong and admirable self, character becomes seen as something more unstable, malleable, and constructed. Section 41 holds that "the character is unalterable is not in the strict sense true; this favourite proposition means rather no more than that, during the brief life-time of a man, the effective motives are unable to scratch deep enough to erase the imprinted script of many millennia". And §222 radicalises the point: "Man as a whole has become and is changeable and

---

[7] Ibid. p. 43.

[8] Nietzsche [9] pp. 33–34.

[9] Ibid. p. 244. Cf. Nietzsche [10] §182.

[10] In these early text, the notion at times seems to take on an even biological rigidity. For example in the 4th meditation he refers to "the character and bone-structure of human nature" (cf. Nietzsche [10], §56). Later on such remarks become rare, but references to physiology never completely stop. As Robert Pippin puts is in an apposite remark, "studying properly Nietzsche's position on nature and naturalism is quite difficult.... There is another essay one could wish for from Nietzsche in *Ecce Homo* – 'How I avoided becoming a 'Gedankenlose Naturalist'" (Pippin [14] p. 55).

[11] Nietzsche [9] pp. 85–86.

even the individual man is not something firm and steadfast." The idea of character as an entrenched moral disposition is even viewed as a perverse ideal: "The man of strong character lacks knowledge of how many possibilities of action there are and how many directions it can take; his intellect is unfree, fettered...."[12]

But with this change in the ontology of character comes an interesting development of his value judgments. On the one hand, there is an intensification of his disregard for all kinds of dramatically enacted characters. For example, when he holds in §160 that, because characters are changeable, their expression through visual and literary art is "just as superficial as is our knowledge of men" and it is therefore valued only by "those who comprehend even an actual human being only in crude and unnatural simplification". The anti-theatrical sentiments peak in the later published fifth volume of *The Gay Science*, where he notes that acting and masquerading, the exploration of different subjectivities, is for the weak and mischievous, for "lower-class families", "diplomats", "Jews", and "women", and also notes – his early fascination for Wagner long gone – that "I am essentially anti-theatrical – but Wagner was, conversely, essentially a man of the theatre and an actor".[13] The entire institution of theatre is most decisively condemned:

> No one brings the finest sense of his art to the theatre; nor does the artist who works for the theatre: there one is people, public, herd, woman, pharisee, voting cattle, democrat, neighbour, fellow man; there, even the most personal conscience is vanquished by the levelling magic of the 'greatest number'; there, stupidity breeds lasciviousness and is contagious;

However, there exists in parallel to the persistent strand of anti-theatrical sentiments a range of different views in his writings. As he holds in a famous passage from the earlier parts of *The Gay Science*:

> *One thing is needful* – to 'give style' to one's character – a great and rare art! It is practiced by those who survey all the strengths and weaknesses that their nature has to offer and then fit them into an artistic plan until each appears as art and reason and even weaknesses delight the eye.[14]

---

[12] Nietzsche [11] §228.

[13] Nietzsche [7] §361, §368.

[14] Ibid. §290.

Here character appears as something in need of artful fashioning. How are we to understand the two side by side? Nietzsche is not simply contradicting himself, because the main difference, it seems, is the time-span in question. Professional actors are suspicious for taking on too many roles in too little time, and in democratic societies, Nietzsche thinks, everyone becomes such an actor:

> But there are... ages, truly democratic ones... where the individual is convinced he can do just about anything *and is up to playing any role*, and everyone experiments with himself, improvises, experiments again, enjoys experimenting.... Precisely because of this, another human type becomes ever more disadvantaged and is finally made impossible; above all the great architects: the strength to build is now paralysed... who still dares to undertake works that *require* millennia to complete?[15]

Nietzsche seems to value people who show persistence and seriousness in their long-term attempts to fashion great selves like great artworks, who have the passion in themselves to life great lives, or even become "greater than life", as we would say now. A key passage from the first part of *Human, All Too Human* puts it nicely, where Nietzsche praises those who have the "seriousness of the effective craftsman which first learns to construct the parts properly, before it ventures to fashion a great whole; they allowed themselves time for it".[16] While the fleeting play of moderns seems like an expression of nihilistic incapacity to love, commit, and truly value and desire, the great artists will act with a seriousness and dedication that dissolves the boundaries between reality and imitation, the hallmark of all Metaphysical Dualism.[17]

There are other further passages that indicate that Nietzsche did in fact appre-

---

[15] Ibid. §361.

[16] Nietzsche [11] §163.

[17] Nietzsche [7] §290. The above cited passages on the democratic habits of experimenting with roles, invite some reflections on Foucault's reading of Nietzsche. It can be doubted that Foucault would have taken aboard any ofthe rather anti-democratic, anti-Semitic, or misogynist expressions that one finds in Nietzsche. To the contrary, he seems very supportive of the democratic habit of experimenting with roles. If anything he seems to have taken on Nietzsche's *method*, rather than his *values* (e.g. pp. 266, 315–316). Such a choice, eclectic though it is, makes sense given his larger philosophical aims of criticizing present institutions. It would make less sense if one's goal were primarily to interpret Nietzsche, given the prominence of value questions in his largely anti-methodological writings (*pace* Deleuze, pp. 78–79). In the words of Albert Camus, cited in an epitaph by Bernard Williams, "quand on n'a pas de caractère, il faut bien se donner une méthode" (Williams [3]).

ciate good acting, which does not merely experiment with roles but builds real characters. To see this we need to introduce another notion that will be central to a character-based interpretative approach.

(2) My *methodological* thesis is that the 'royal road' to reconstructing a character from texts that were not written for the stage is the practice of *translation*. Several considerations can be adduced to support this thesis, of which I want to mention three in particular: underlying the idea is first of all the author's intuition (ii), confirmed by practical experience, that the process of acting and the process or translating are phenomenologically profoundly alike. In both cases the actor/translator sets out to become deeply acquainted with another subjectivity; they learn to speak with someone else's voice, take up foreign thoughts and a foreign mindset. A good actor knows how Polonius would have said something, similarly to the way a good translator could tell us how Nietzsche would have said something in another language. Secondly, (ii) we can see how this insight has found its way into ordinary language by noting that the term 'interpretation' designates not only the explanation of some text or object, but also both theatrical acting and translational practice. For example, we speak of a director's interpretation of a play, an actor's interpretation or a role, and it is a commonplace to say that every translation is always also an interpretation.

Lastly and most importantly, (iii) whether one agrees to the parallel drawn or not, it seems that Nietzsche himself supported it. In a key passage from *Beyond Good and Evil*, which discusses translation, after noting how Germans are favourably inclined towards everything "ponderous, lumbering, solemnly awkward, every long-winded and boring type of style", he adds that "Lessing is an exception, thanks to his *actor's nature* that understood and excelled at so much".[18] Here we see how, in one of his last works, Nietzsche speaks sympathetically of a great actor – who is not actually an actor but manifests the skill in his excellent translations!

A translator who has extensively translated an author would thus seem best able to fully engage this author, not unlike the way the best actors engage their characters.[19] She would have learned to speak and think like that author, not consciously by knowing all the arguments, though that might be part of it, but also unconsciously. Nietzsche would have become like a second first lan-

---

[18] Nietzsche [6] §28.

[19] Duncan Large has suggested this much, "speculating that the translator might represent the best approximation to Nietzsche's ideal reader" (Large [4] p. 1). To the best of my knowledge the point has not yet been argued in more detail.

guage to that person. Moreover, like a good actor, who studied a character thoroughly and can convincingly enact that character even in situations outside the play, creatively applying the role to new situations, such a translator could perhaps even, to some degree, *extrapolate* and tell convincingly and authentically though not, of course, decisively, what an author would have said on questions that he did not in fact encounter or write about.

Immediate objections abound and qualifications are in order. For the philosophical purpose of this essay let us side-step for a moment the practical obstacles, such as time constraints on the translator's side or limited access to the author's text, and focus on more theoretical problems: for example, what if two different interpreters come to different conclusions, different enactments, different translations, as seems bound to happen? On this point I contend that the translational interpretative strategy will be on par with the best rational reconstructions. Different readings come to different conclusions under both methods. Partly this might be due to genuine ambiguities in the text, partly it might be because the interpreters brings their own prejudices or agendas to the reading.

(3) However, rightly considered, this challenge brings out the greatest promise of the translational approach, and with this I am moving to the *axiological* analysis of the proposed method. The problem of prejudices distorting interpretations is unavoidable, for rational reconstructivists and translational interpreters alike. But it is the latter that is better able to own up to this fact, affirm it as it were, and integrate it positively into the approach. Far from it being a weakness, framing the interpretative process in terms of acting, translating, and performing opens up the possibility of engaging with and connecting to a vast range of work that has been done on the ethico-political aspects of performing, acting, spectatorship, and related concepts. Rationalist reconstructions seem, by design if not in fact, bound to the pretence of a scholarly gaze, dissecting an author's body of work with disinterested objectivity – an illusion if ever there was one, as Nietzsche would be the first to tell us.[20] A translator's interpretation, on the other hand, understood also as creative and recreating performance, is open in principle to a wide range of stances *vis-à-vis* an author's works: reconstructive and eclectic ones, but also critical, erotic, playful, comic or tragic, to name just a few options, as seems appropriate to the author and the audience. Beyond and in addition to traditional epistemological values that we expect from rationally reconstructed theories (e.g. coherence, clarity,

---

[20]For example, Nietzsche [5] §§3.23–5.

simplicity, and charity), the translational approach invites considerations of a range of ethical, political, or aesthetic values.

What values would, could, or should such an approach express? It is a decisive question, and by no means an easy one. Often it seems easier to say which values the interpretation should *not* express. A satisfactory affirmative answer, in absence of a decisive one, which seems likely to remain unattained (perhaps for the better, one might add), still has to wait until a more extensive attempt at such an interpretation has been carried out.[21] But I would like to briefly discuss one value in particular, and by way of conclusion point to a methodological consequence of this discussion.

One value that has been traditionally associated with the practice of translation is the value of fidelity.[22] Against the well-worn Italian adage "traduttore, traditore", a good translator is often considered to be one that is faithful to an original – text, author, meaning or intention. It seems doubtful to me that Nietzsche would much respect such an approach to his work. In §83 of *The Gay Science*, Nietzsche's most sustained discussion of translation, he distinguishes two different approaches to translation that can be found in different cultures, depending on their "historical sense".[23] Against the Germans of his time and their sense of history, their concern for and fidelity to the past, he opposes the practice of the revolutionary French, as well as the ancient Romans:

> And Roman antiquity itself: how violently and yet naively it laid its hands on everything good and lofty in the older Greek antiquity! How they translated things into the Roman present! . . . They seem to ask us: "Should we not make new for ourselves what is old and put *ourselves* into it? Should we not be allowed to breath out soul into their dead body?"

The paragraph itself does not make Nietzsche's preference entirely clear. To-

---

[21] For example, as Pippin points out in his discussion of Nehamas' theses, classical notions of beauty and harmony seem no fit candidates for Nietzsche as guiding selves-constituting self-narrations. Similar considerations might apply for Nietzsche-constituting translational interpretations (cf. Pippin [12] pp. 122–125).

[22] A value that seems to me is not necessarily much cherished neither by rational reconstructivists, nor eclectic pragmatists. Because both of them care first and foremostly about establishing theories that work the former perhaps within certain, mostly textual, boundaries. A rational reconstructivist, for example, would have in principle no reason to object to a reading that is compatible with the text, rationally coherent, but arguably not what the author meant. In fact, such readings often seem necessary to make past authors speak to the present.

[23] Ibid.

wards the end of Book IV he refers to the historical sense as a "disease" and "something poor and cold", thus indicating that it is indeed the Roman approach he prefers, and thus not any kind simple fidelity to an orthodox original.[24] There might of course be good reasons to reject this approach: reasons external to one's views on Nietzsche – for example, if one does not like the clear link he establishes between this kind of translational practice and imperialist aspirations ("one conquered by translating . . . not with any sense of theft but with the very best conscience of the *imperium Romanum*")[25] – and reasons internal to one's views on Nietzsche – for example, if one prefers a more Hegelian reading as described above, according to which neither the old Romans nor the new Germans might have it 'right'.

Be that as it may, the rejection of fidelity and skepticism towards imperialism, could give rise to the following methodological consideration: one way of thinking about the relation between the translational interpreter and her character/author would be to say that the former has eventually to be both performer and audience in this interpretation – that she has to oscillate between actively and creatively being in-character and a reflective, listening stance out-of-character. As with a translator speaking two languages perfectly, the interesting part of the interpretation would not be the *result*, the finished interpretation in either of the two languages, but rather the *process* of going back and forth, of a plastic *being-in-between* – short, the Dialogue, in which the interpreter takes on different voices and stages a conversation between herself and the characters she learned, through translation, to speak with and to speak as.

But "enough, enough! At this stage there's only one thing appropriate for me to do: keep quiet".[26] Such a dialogue will have to wait until a thorough attempt at a translational interpretation has been undertaken. A prolegomenon to such an undertaking, to thematize the ontological, methodological, and axiological premises of such an approach, has been the hope and goal of this essay.

References

[1] Nehamas, A. (1985) *Nietzsche: Life as Literature*. Cambridge: Harvard University Press.

---

[24]Ibid. §337.

[25]Ibid. §83.

[26]Nietzsche [5] p. 77.

[2] Williams, B. (1993) 'Nietzsche's Moral Minimalist Philosophy', in *European Journal of Philosophy* 1/1.

[3] —— (1985) *Ethics and the Limits of Philosophy*. Abingdon: Routledge Classics.

[4] Large, D. (2012) 'Nietzsche and/in/on Translation', in *Journal of Nietzsche Studies* 43/1.

[5] Nietsche, F. (2009) *On the Genealogy of Morals*, I. Johnston. Arlington: Richer Resources Publications.

[6] —— (2002) *Beyond Good and Evil*, J. Norman (trans.). Cambridge: Cambridge University Press.

[7] —— (2001) *The Gay Science*, J. Nauckhoff (trans.). Cambridge: Cambridge University Press.

[8] —— (1999) *The Birth of Tragedy* R. Spiers (trans.). Cambridge: Cambridge University Press.

[9] —— (1997) *Untimely Mediations*, R.J. Hollingdale (trans.). Cambridge: Cambridge University Press.

[10] —— (1997) *Daybreak*, R.J. Hollingdale (trans.). Cambridge: Cambridge University Press.

[11] —— (1996) *Human, All Too Human*, R.J. Hollingdale (trans.). Cambridge: Cambridge University Press.

[12] Pippin, R. (2014) 'Self-Interpreting Selves: Comments on Alexander Nehamas' *Nietzsche: Life as Literature*', in *The Journal of Nietzsche Studies* 45/2: pp. 118–133.

[13] —— (2013) 'Heidegger on Nietzsche on Nihilism', in T. Pangle & J. Lomax (eds.) *Political Philosophy Cross-Examined*. New York: Palgrave MacMillan.

[14] —— (2010) *Nietzsche, Psychology, First Philosophy*. Chicago: University of Chicago Press.

Filial Obligations: Revisiting the Friendship Theory

Daniel Choi
University of Toronto

Most of us agree that we have some special duties to our parents, and these duties seem weightier than what we owe to our friends. We might have a duty to pay our parents' medical bills, but we do not have the same duty to pay our friend's medical bills; such a friend would be acting from generosity and supererogation, or beyond the call of duty. But why do we owe these things to our parents? You might say that you owe it to them because they sacrificed just as much in raising you, or you might point to the special relationship you have with you parents. Following this intuition, there have been three main suggested theories of filial duty: the debt theory, the gratitude theory, and the friendship theory. Recently, Simon Keller proposed a special goods theory, stating that our filial duties arise from parents and children being able to provide 'special goods' to one another – that is, goods that can be provided only by child or parent.[1] Brynn Welch identified some counterintuitive cases in Keller's theory, and added some amendments which resulted in the gratitude for special goods theory.[2]

However, both of these 'special goods' theories, along with their predecessors, have the issue of being too stringent. On these views, filial duties are 'weighty'; they demand much, and interfere with our plans and projects; in other words, these theories suggest that they "override almost everything else".[3] This issue requires a closer look into the nature of these filial duties, which consequently requires a closer look at their structure and motivation. Diane Jeske offers a unique perspective of how the reasons behind intimate relationships can be 'objective' and 'agent-relative' by using an analogy to one's relation to one's future self.[4] This paper will use Jeske's concepts to supplement the friendship theory with the 'weightiness' of filial duties. On this new model of the friendship theory, filial duties are analogous to duties of friendship, and therefore

---

[1]Keller [2].

[2]Welch [3].

[3]Welch [3] p. 726.

[4]Jeske [1].

less weighty; however, the intuition that filial duties ought to be more strin-gent are captured by reducing them to projects (I call this the 'familial project'). Although projects are within the domain of permission and supererogation,[5] the intuition that filial duties ought to be more stringent is not undercut if, following Jeske, we see this 'familial project' as 'objective' and 'agent-relative', which are 'weightier' than other projects. This model captures our normative intuition of filial duties within the friendship theory, making this theory ten-able and comprehensive.

1   Debt Theory, Gratitude Theory, and Friendship Theory

Before looking at the special goods theory and the gratitude for special goods theory, I will quickly gloss over the other competing theories prominent in the literature. My exposition of these other theories is only aimed at providing context for the contemporary 'special goods' theories. Starting with the debt theory, it claims that parent–child relation is analogous to the creditor–debtor relation.[6]Parents provide various resources, like financial support, time or ef-fort to rearing a child. The debt theory says that all of these goods provided by a parent are on credit, and that the child is in debt to the parent which they must repay. In essence, filial duties are the child's duties to pay their debts back to their parents.

Filial duties, however, seem to be different from a duty to repay a debt in several ways. First, filial duties do not look like they can be "discharged, once and for all".[7] Imagine if a parent's services were reduced to a monetary amount, say, one million dollars; it seems wrong to think that a child can pay one million dollars to this parent and be released from any subsequent familial duties. Second, filial duties do not have the same proportionality concerns as duties of debt might have. That is, you may have been a "healthy and angelic child, undemanding and a delight to nurture",[8] but that does not mean you have weaker filial duties, whereas on the debt theory you might have a smaller

---

[5]This claim might be contentious to some, but nothing of importance hinges on my narrow use of 'projects'.

[6]By 'analogous', I mean that they give rise to the same duties; I do not mean that they are the same type or kind of relationship.

[7]Keller [2] p. 256.

[8]Ibid. p. 256.

debt.[9] Third, filial duties are sensitive to the relationship itself. The parent–child relationship can be affected by some irreconcilable differences, and this can leave the child with fewer filial duties, whereas duties of debt would be insensitive to these changes in the relationship. By and large, duties of debt do not seem to be adequately analogous to filial duties.

The gratitude theory claims that to "fulfill your filial duties is to perform appropriate acts of gratitude in response to the good things your parents have done for you".[10] The gratitude theory meets the proportionality concerns of the debt theory because gratitude is less strict about proportionality. If somebody saved your life, duties of gratitude do not demand you to be their personal servant or take a bullet for them. The duty of gratitude might be satisfied by something as small as a nice dinner or at least something proportionally smaller than the price of one's life. The gratitude theory seems much more plausible than the debt theory so far.

The criticism with the gratitude theory is that it, too, fails to capture our intuitions of filial duties. First, there is a similar proportionality concern with the gratitude theory. Gratitude might be lax in its demands to reciprocate proportionally, but it still asks to give gratitude in some proportional manner. Again, if filial duties were analogous to duties of gratitude, they would demand less if you were "undemanding and a delight to nurture", and demand more if you were a little hell-raiser. This seems counterintuitive to notions of filial duties because filial duties are thought to be indifferent to these proportionality concerns. Second, filial duties are "ongoing and open-ended, and can be very demanding",[11] whereas duties of gratitude are not. The general thought here is that gratitude fails to capture the weight of the demands of familial duties. If a parent fell deathly ill, filial duties might demand that the child invest significant amounts of time and money for their parents, even more than what is demanded by the duties of gratitude. In contrasts, duties of gratitude might be discharged with a card, a bottle of wine, or whatever, and one might be subsequently released from the duty; however, filial duties seem to go on for a lifetime, and cannot be discharged and released. Again, duties of gratitude are not analogous to filial duties.

---

[9]In response to the proportionality concern, some have tried to propose an 'insurance theory', which says that investments from the parent can yield duties indifferent to the proportionality of good bestowed by the parent. This theory fails for the same reasons: filial duties are not analogous to exchanging goods and services.

[10]Keller [2] p. 257.

[11]Keller [2] p. 260.

The friendship theory claims that filial duties are analogous to duties of friendship. Friendships are aimed at sustaining an ongoing relationship rather than meeting duties of debt or gratitude. This view meets the proportionality concerns because friendships do not require strict reciprocity. It would be odd to think that friends keep a record of every benefit bestowed on one another; in fact, this would be the sign of a bad friendship. Friendships rather have the general positive duty to care for the well-being of one another. By and large, friendship theory is more tenable than the gratitude theory and the debt theory.

The friendship theory seems like the most plausible account of filial duties. But again, there are reasons to think that the analogy fails. Like families, friendships can have fallouts and disagreements which change their respective duties. However, families are more demanding in that reasons which affect friendship duties do not apply with familial duties. For instance, paying the medical bills of a parent appears to be a duty whereas paying the medical bill of a friend appears not to be a duty. Friendship duties are less demanding and do not seem to be completely analogous to filial duties.

2   Special Goods Theory and the Gratitude for Special Goods Theory

Keller's special goods theory does away with analogies altogether. According to Keller, filial duties should be understood on "its own distinctive terms".[12] The other theories fail because the family relationship is a different kind of relationship than the friendship or debtor–creditor relation. The family relationship is unique and gives rise to its own *sui generis* duties. The idea here is there are unique ('special') goods and needs which can only be provided and satisfied by reciprocity between parent and child. This is not to be confused with generic goods, like your mother's cooking you dinner or your helping your mother move furniture. A special good, for instance, can be a parent's "sense of continuity and transcendence, a feeling that they will, in some respect, persist beyond their own deaths".[13][14] Another is a child's "special value

---

[12]Keller [2] p. 265.

[13]Ibid. p. 267.

[14]The idea here is that parents (*qua* existential beings) gain some solace in their legacy through their child. Keller is not completely clear on the 'specialness' of special goods, but there seem to be plausible examples of goods only attained by the relationship between children and parents

in having a parent from whom to seek advice".[15] A less starry-eyed example is the comfort of having someone who is "committed to ensuring that your needs and interests will be met".[16] A child then has filial duties because the parent–child relationship bestows special goods to them and they have duties within that relationship to bestow special goods to their parents.

It looks like this theory does better than its alternatives. It can explain why filial duties are "ongoing and open-ended" because parents will have special needs as long as they are alive. It explains the proportionality concern by making it irrelevant; that is, as long as the conditions of a good (on the 'special goods' picture) family are met, the aim becomes providing special goods without tally, like the friendship theory. It can also explain the demands of filial duties and why they cannot be easily escaped. If the aim of filial duties is to provide special goods, then it does not matter how great the special needs are; moreover, terminating the relationship does not change the duties arising from these special needs.

However, Welch raises problematic cases for the special goods theory. An "elderly woman suffering from dementia requires medical care, and she has a wealthy daughter who can provide such care for her".[17] According to the special goods theory, the daughter has no duty to care for the mother since the mother cannot reciprocate the contributions of special goods with the daughter. Welch's complaint is that Keller's view "cannot offer a reasonable, theoretical limit to filial obligations".[18] Another problem case, the 'abandonment case': a mother who selfishly abandons her daughter is reunited to now provide her daughter with special goods.[19] Here, the daughter seems to have filial duties to this negligent mother. It seems counterintuitive to say that the daughter owes less in the first case than the second.

Welch offers an alternative hybrid theory to settle these issues with the special goods theory: the gratitude for special goods theory. This theory stipulates the value of providing special goods in the past, and further adds the following

---

(or *vice versa*). (Although, in principle, 'special goods' relations can be instantiated in friendships and romantic relationships.)

[15] Ibid. p. 267.

[16] Ibid.

[17] Ibid. p. 728.

[18] Ibid.

[19] Ibid. p. 727.

new condition to the special goods theory: "Expressing gratitude by meeting the parent's needs would not undermine the mutual respect on which moral relationships are based."[20] Applied to the abandonment case, it releases the daughter from her duties to the mother because an expression of gratitude would violate the new clause.

A criticism Welch must deal with is that filial duties are too weighty and demand too much from us. Filial duties may ask us to devote unreasonable amounts of effort to our parents. A move to deal with this problem is to argue that excessive self-sacrifice is a failure of self-respect. Perhaps we want to say that some 'serious'projects take priority over filial duties.[21] Keller writes that "the child's duties to provide special goods to the parent should not be such as seriously to impede the child's ability to live a good life"[22]. But little argument is provided for this, and we can think of problem cases – for instance, where a child makes various serious projects to escape filial duties. Where does this leave the force of filial duties? Surely some serious projects ought to be put aside for filial duties, but which ones? It seems there is some worry to one's autonomy if we purport that the push of filial duties override all other endeavours. The rest of this paper will explore this issue.

3   'Familial Project' Theory

It seems counterintuitive to say that all instances of serious projects are precluded by filial duties to provide special goods, as 'special goods' theories suggest. Perhaps familial duties really should override everything else, and it might be an entirely western view which places undue importance on individualism. This seems farfetched, as autonomy is generally very important to us. Or perhaps one might want to endorse some perfectionist account of the value of being a virtuous person outweighing the value of personal projects – this, however, seems equally unconvincing. On the other hand, to allow serious projects to take priority over filial duties may be a serious precedence to set, and may undercut the force of filial duties. This is an issue from the friendship theory, namely, that it fails to reflect the robust nature of familial duties. Friendships can change, grow apart and the duties of friendship correspondingly fade, but

---

[20]Welch [3] p. 730.

[21]'Serious' interests are not 'basic' enough to impede on a good life, but not trivial ('peripheral') enough that they are obviously outweighed by duty.

[22]Keller [2] p. 269.

filial duties are not like that – filial duties appear to be more robust in the sense that they more resilient to change and have stronger demands.

The friendship theory may not match our intuition of filial duties, but perhaps it is too early to jettison this theory altogether. There is a worry that familial duties are not acquired voluntarily. Nobody chooses to be in a family yet everybody has filial duties thrust upon them, and these filial duties can affect choices, plans, and projects, and (*ipso facto*) impede autonomy. Duties of friendship, on the other hand, seem to be voluntarily taken when one voluntarily enters into the relationship. This might be further motivation for taking the friendship theory, but we need an explanation of its counterintuitive results if it is to be a comprehensive theory. Before sketching this new friendship theory, we must first explicate some concepts offered by Jeske.

Jeske analyses the reasons for entering and maintaining friendships, and recognises two sets of distinctions: 'subjective' and 'objective', and 'agent-relative' and 'agent-neutral'. First, subjective reasons are reasons for an agent to desire (or to bring about) a state of affairs, whereas objective reasons are agent neutral.[23] Subjective reasons are thought of as 'agent-relative', meaning one's reasons (*per se*) for promoting a state of affairs are not reasons for anybody else. My liking for a particular chair, for instance, might be subjective and agent-relative because my reason for liking it are not reasons for you to like it. In contrast, objective reasons are thought of as 'agent-neutral', meaning they are reasons for everyone; to go back to the illustration, perhaps there is a chair with universally appealing features, like perfect lumbar support. Jeske suggests that the reasons for friendships are unique because they are 'objective'and 'agent-relative'.

What does it mean for a reason to be objective and agent-relative? Jeske uses a parallel case to illustrate. She relates the relationship one has to one's future self as a case of objective agent-relative reasons. Private projects have reference to a particular agent, and one's self in the distant future will have private projects that differ from one's present self. One's present projects are subjective agent-relative, but it would not be appropriate to call the projects of one's future self 'subjective', because subjectivity is defined by current values.[24] Reasons one has to one's future self are objective agent-relative reasons, and can be analogous to reasons of intimacy. Taking into account the subjective agent-relative reasons of a friend in one's deliberative process is to treat them as

---

[23]Jeske [1], p. 330.

[24]Ibid. p. 342.

objective because they belong to your friend and not yourself. As it follows, reasons can be both objective and agent relative.

Jeske's picture offers a richer foundation to reply to the criticism of the friendship theory. My view is not committed to Jeske's overall view of friendships and duties. Rather, I borrow her distinctions, particularly objective agent-relativity. To reiterate issue at hand, the main objection to the friendship theory is that it does not match our intuitions of filial duties. Filial duties seem more stringent than duties of friendship; again, duties of friendship can be released on laxer grounds than filial duties, and duties of friendship seem to be less demanding. Let us begin by distilling our notions of filial duties from our duties of friendship, demarcating the duties of friendship from the (*sui generis*) filial duties. Let us further generalise duties of friendship to something along the lines of 'to care for each other's well-being', and call everything else (the weightiness, the robustness, the stringency, etc.) 'mere familial duties'. My suggestion is that 'mere familial duties' are to be understood an objective agent-relative 'familial project', although projects are commonly understood to be subject and agent-relative. This of course makes filial duties not duties at all; rather, they are within the realm of permissions. This leaves only the duties of friendship to be duties proper, meaning they are objective and agent-neutral. Let us take a closer look at this view.

To begin, it may be helpful to clarify the features of a 'project'. Projects have a feature of intentionality, meaning there is some goal or aim which generates actions; moreover, this goal can be open-ended and subject to change. For instance, a private project of mine could be to become a better philosopher, and this leads me to read more philosophy. There is no 'end' to becoming a better philosopher, and it is entirely possible for me to change this goal to becoming a better philosopher of religion. Goals or aims direct action in prudential ways, and can generate *pro tanto* rules or constraints. If my aim was to train to win a marathon, it might generate rules which are prudent to follow, such as running ten miles every day. However, perhaps I only have time to run five miles today. I would then feel disappointment or compunction. Projects can also be shared among people – for example, when a community comes together to raise money for a cause. Again, failure of prudential actions is met with a feeling of guilt, not in moral failure but a failure to bring about better results.

The familial project is a project of shared narrative. The aim is to preserve the family unit and foster its continuation. Let me explain this without appealing

to neo-Darwinian principles or some collectivistic normative theory.[25] I think this is where the special goods theory can help. 'Special' goods, previously distinguished from 'generic' goods, are particular to family members having special needs. There is a mutual understanding that special needs can only be satisfied within the family unit. With this mutual understanding, a common project is formed to meet these needs through the shared project. With this project, there are prudential rules to follow, and these are the 'mere familial duties' which are conflated with duties of friendship. Due to the diversity of special needs, the family unit must be sustained and preserved. One might begin with a special need of a parent with wisdom and knowledge, then a special need to be taken care of, and a further special need to have some continuity after death through children or grandchildren.

The familial project is properly understood as objective and agent-relative, and not subjective and agent-relative. The shared project is agent-relative in that it only applies to the particular members of a family, yet it is objective because the familial project is not a project contrived by me. There is an element of externality to the familial project because I am merely a subscriber to the project and not the provider. Private projects are subjective because they are built around what I value. There is a sense in which I similarly value the familial project, but this seems incidental because other private projects often conflict with the familial project.[26]

When projects conflict, we assess the motivation for subscribing to projects and see how important they are to us. This weighing of the options and subscribing to projects is within the realm of permissions and has no (*ceteris paribus*) moral bearing. Again, thinking of a daughter who chooses to travel the world rather than help with her mother's medical bills, it seems the daughter is callous and doing something wrong. But one must keep in mind that weighing the options is difficult; there is still a realm of regret, doubt, and compunction. On this view, she has no strict duty to choose the option of helping her sick mother, and her choosing to do so would be beyond the call of duty. This might be a novel case since the daughter's private project of travelling the world seems obviously outweighed by her familial project, but more serious cases may shed light on the importance of having the freedom to choose between projects.

---

[25] I do not want to be committed to some innate need to protect the family unit or some culturally relative zealous value of the family unit.

[26] The ontology of the familial project is not a concern here.

The motivation for taking this view is that on the 'special goods' theories filial duties seem stringent enough to violate one's autonomy. Familial duties can preclude other projects that are genuinely important to us and to living a good life. Arguments from self-respect, which say that being concerned only about meeting familial needs violates respect to oneself, leave room for problem cases where self-respect is not violated but familial duties ask us to sacrifice important projects. This is the motivation for casting off familial duties and reframing them as mere projects. Weighing projects leaves room for autonomy. The familial project is also important to add to the picture because the friendship theory seems too lax. Duties of friendship alone are often not enough to motivate certain actions.

It is possible to make a further ambitious claim that is not an essential piece of my argument, but might suggest another motivation for taking this view. Jeske suggests at the end of her paper that "a mark of moral reasons"is the fact that they are objective rather than subjective.[27] Since, on this view, the familial project is objective, it may suggest that they have a quasi-ethical role in affecting our actions. This might sound like a case of having your cake and eating it too, as this might fall into the pitfalls of the other theories of filial duties and violate autonomy; however, we still have the choice of subscribing to the familial project, whereas we would not have the choice to take on filial duties on the other theories. In any case, this is just a complementary motivation for my argument, and its only use would be to give more weight to familial projects when they conflict with private projects.

Theories preceding the 'special goods' theories have run counter to our intuitions of filial duties in one way or another. The 'special goods' theories, however, have the issue of being too weighty and plausibly restricting autonomy. The friendship theory's fault was that it was not stringent enough, but we can build a model which contains the intuitions of mere familial duties within the familial project. My suggested theory amounts to a return to the friendship theory and while explaining away the counterintuitive aspects. This is more than just a descriptive account of how to settle the autonomy worry. It makes the stronger claim that other theories of filial duties are mistaken in their normative grounding. Filial duties proper are analogous to the mere duties of friendship.

---

[27]Jeske [1] p. 345.

References

[1] Jeske, D. (2001) 'Friendship and Reasons of Intimacy', in *Philosophy and Phenomenological Research* 63/2: 329–346.

[2] Keller, S. (2006) 'Four Theories of Filial Duty', in *The Philosophy Quarterly* 56/223: 254–274.

[3] Welch, B.F. (2012) 'A Theory of Filial Obligations', *Social Theory and Practice* 38/4: 717–737.

What would a Marxist evaluation of the Haitian Revolution look like?*

Azzan Akbar
UCL

Introduction

> The history of all hitherto existing society is the history of class struggle. —— Karl Marx

Despite the impact these words came to have in the coming centuries after the publication of *The Communist Manifesto*, one might suggest that Marx has overlooked another kind of struggle – that of race. However, it is not the case that Marx was unaware of race issues; indeed, "Marx took great care to . . . unravel . . . how racialized slavery contrasted with wage slavery".[1] In addition, Marx's various writings on slavery, particularly the American case, show that Marx was aware of the 'peculiar institution'. Given this, and from what we know of Marx's writings on revolution, it is noteworthy that Marx wrote little, if anything at all, on what has been described as "the largest and most successful slave rebellion in the Western Hemisphere" – the Haitian Revolution.[2] It is in light of this that I now seek to ask – what would a Marxist evaluation of the Haitian Revolution look like?

Although this essay will seek to collate some of the existing literature on the topic, it is most peculiar that such a revolution has not already been revised and understood by Marxist academics over time. As Lee notes, "Even on the left, the Haitian Revolution does not get the recognition it merits".[3] Grüner adds that "many other Marxist historians . . . do not ever account for this issue".[4] Why is it that not only Marx, but contemporary Marxist historians, who are

---

*Delivered at the BJUPS Spring Conference, 23–24 March 2016 at the University of Cambridge.

[1]Lee [10].

[2]Blackpast [1].

[3]Lee [10].

[4]Grüner [7].

better placed to access information, have not properly evaluated the Haitian Revolution? Lee suggests

> [The Haitian Revolution] directly threatened the slave empires in the new world. In the nineteenth and twentieth centuries, it offered hope of insurrection for independence to the colonies subject to the European empires. It has always been a challenge to liberals and their counsel of piecemeal reform and gradualism, which rarely if ever delivers change, and instead promises a counter-model of class struggle and revolution.[5]

The history of Haiti has been hidden from the western academic tradition – let us now unveil it, and consider how Marx could have responded to the events of the Haitian Revolution. To this end, I shall be presenting the pro-Marxist view which argues that the Haitian Revolution be characterised as Marxist in tendency. It should be noted that when we say 'revolution', we not only refer to the 'physical' struggle, but also to the underlying ideologies of those leading the revolution, the conditions prior to its outbreak and some of the short-term consequences.

My essay will begin by presenting a few key factual features of the Haitian Revolution and some key aspects of Marxist philosophy that will be important throughout the paper. I will go on to present the pro-Marxist case that supports the view that the Haitian Revolution can be characterised as 'Marxist', before presenting some limitations. I will end by considering responses to these limitations, whilst seeking to conclude, on the whole, that the Haitian Revolution can be characterised as Marxist but, as with many events in history, has certain features that make it stand uniquely as a revolution.

Background

It is proper to present a brief historical introduction to the Haitian Revolution but concurrently highlight any relevant philosophical ideas that will arise later on in this essay. I will also present some key elements of Marxist philosophy that will be key. During the 18th century, Haiti was a French colony known as Saint-Domingue. Laden with plantations, the labourers of which were black slaves from Africa, the colony brought great profits to the French bourgeoisie. By 1789, the same year as the outbreak of the French Revolution,

---

[5]Lee [10].

Saint-Domingue was "the most profitable colony the world had ever known".[6] In addition, the slave population was growing rapidly; by 1791, "the black slaves outnumbered their white masters and overseers".[7]

Over the next twelve years, the slaves of Saint-Domingue, led by former slave Toussaint L'Ouverture, fought off the armies of the most powerful empires in Europe at the time; France, Spain and Britain, all of whom attempted to seize control of the colony. Throughout this time, France was also contending with its own revolutions as power shifted between monarchy, nobility and Jacobins. Finally, by the end of 1803 Saint-Domingue declared independence from France, naming the new country 'Haiti'. By 1805, a constitution was promulgated by the successor of Toussaint, Jean-Jacques Dessaline, who became Emperor Jacques I with the establishment of the constitution.

In saying that the Haitian Revolution was 'Marxist' in its tendency, I am essentially claiming that the Haitian Revolution was an instance of the kind of revolutions Marx discussed and predicted would occur. This will allow us to see what a Marxist evaluation of the Haitian Revolution looks like.

I believe that there are three key elements of Marxist thought to consider if we are to identify a revolution as Marxist in its tendency. These are: [8]

1. Dialectical class struggle – typically this may include a struggle for power between the bourgeois, property-owning class and the proletariat class who work for the bourgeoisie.

2. Primitive accumulation – "the process by which precapitalist modes of production, such as feudalism and chattel slavery, are transformed into the capitalist mode of production." [9]

3. Historical materialism – "...the idea that forms of society rise and fall as they further and then impede the development of human productive power." [10]

These features (F*) are, in my view, jointly necessary and sufficient conditions to characterise a revolution as Marxist. That is to say that having F* guarantees

---

[6]James [8] p. 57.

[7]Lee [10].

[8]For the purposes of word economy, I shall henceforth refer to these features collectively as F*.

[9]Blunden [14]

[10]Wolff [18]

that said revolution is Marxist, and if said revolution lacks F*, then we could not truly call it Marxist. However, I do not wish to restrict myself to these features; indeed, other features of Marxist philosophy may also arise without compromising these features as necessary or sufficient conditions.

The pro-Marxist view

Given what we have said above, I articulate the pro-Marxist argument that attempts to prove that the Haitian Revolution was Marxist in tendency:

(P1) A revolution can be characterised as 'Marxist' if it displays F*.

(P2) The Haitian Revolution displays F*.

(C) The Haitian Revolution can be characterised as Marxist.

If the Haitian Revolution can be characterised as Marxist, (thereby being Marxist in tendency), then I will have shown what a Marxist evaluation of the Haitian Revolution would look like.

(P1) follows largely from what was said above regarding F* as necessary and sufficient conditions of a revolution being characterised as Marxist. Since F* are key features of Marxist philosophy, it follows that for a revolution to be Marxist, it would have to display these features.

The proof of (P2) is now what follows extensively; I shall take each of the individual conditions of F* and show how the Haitian Revolution displayed this.

Condition 1

The first condition was articulated as 'dialectical class struggle'. Indeed, the Haitian Revolution can be framed as an example of a dialectical class struggle between the bourgeoisie and the proletariats. According to Lawrence, we can describe slavery as an instance of capitalist relations:

(1) Normally wage labour arises out of the dissolution of slavery and serf-dom.

(2) In some instances, slavery survives, as an anomaly.

(3) Under these circumstances, the slave owners are capitalists, the slave are proletarians, and the products of slave labor are commodities.

(4) These commodities enter the cycles of industrial capital in the market; at the same time the slave-owning capitalists realise their surplus value.[11]

This is shown by Marx's own comments that "a commodity produced by a capitalist does not differ in any way from that produced by an independent labourer or by communities of working-people or by slaves".[12] If Marx shows that slavery as an economic system can be capitalist in nature, then it is possible that this same relation be translated into the Haitian Revolution. A bourgeoisie–proletariat dialectic persists in a capitalist system, and we can move to argue that the Haitian Revolution is an example of this. Indeed, according to James, "the slaves worked on the land, and, like revolutionary peasants everywhere, they aimed at the extermination of their oppressors . . . they were closer to a modern proletariat than any group of workers in existence at the time."[13]

However, it might not be proper to simply reduce the struggle of the slaves of Saint-Domingue to a mere class relation. Indeed, there is a certain racial feature that is relevant to the slavery of Saint-Domingue – the fact the slaves were brought in specifically from Africa. This reveals a certain belief that 'blacks' were in some way inferior. If the slaves of Saint-Domingue were enslaved partly due to the belief that blacks were in some way 'non-human', does this suggest that the struggle of the enslaved was not one of class, but instead a race struggle?

Marx himself was not unaware of race issues. Lee argues that because Marx makes a distinction between race and class slavery in Capital Vol. 3, which "has not been as widely read", this created a (mis)conception that Marx was a class reductionist. Instead, Marx used slavery to demonstrate how dominance played a factor in class struggle. He compared the outrageousness of factory managers defending their higher pay and economic position based on class relations to the "outrageousness of slave plantation masters defending their socio-political and economic positions in terms of innate talent based on racial grounds". Such comparisons show that there was a distinction between class and race slavery. Despite this, there are some similarities also. According to Marx, "The wage labourer, like the slave, must have a master to make him

---

[11]Lawrence [9].

[12]Marx [13].

[13]James [8] pp. 85–86.

work and rule him."[14]

The distinctions and according similarities between the two types of struggles show that the Haitian Revolution was an example of 'dialectical struggle'. If we do not wish to subsume race relations into class relations (as they are separate struggles) then it might not be the case that we have a 'dialectical class struggle', but rather a 'dialectical race struggle' between white planter and black slave. However, this does not necessarily translate into the Haitian Revolution failing to be a Marxist revolution in tendency. The Haitian slaves struggled against white planters and French bourgeoisie; we can characterise this struggle in the sense Marx discussed – Marx lists "freeman and slave" as the first example of historical dialectical struggle in the Manifesto. [15] Given this, we can still adequately consider the dialectical struggle between the slaves and the planters as the sort of thing Marx considered important. Consequently, the Haitian Revolution fulfils the first condition of being a Marxist revolution.

Condition 2

The second condition was expressed as 'primitive accumulation'. Looking at Marx's account of modern slavery and also through the features of Haitian Revolution, it can be suggested that the Haitian Revolution was an example of this primitive accumulation in work. Grüner writes "African slavery in general (and Haiti's…specifically) is an essential component of the process of primitive accumulation of capital….Thus African slavery will be, inevitably, just as essential in the construction of the 'modern-bourgeois' world-economy."[16]

It is clear to see that the Haitian Revolution is a key example of primitive accumulation, as we have defined above, occurring when pre-capitalist modes of production such as chattel slavery are transformed into capitalism. The colony of Saint-Domingue, as has been mentioned, was one of the most profitable slave colonies of the French Empire. Thus, the colony's economy was funded entirely by slavery. However, following independence from France in 1804, slavery was abolished. Both the constitutions of 1801 and 1805 "protected the right to private property".[17] This seems to suggest a movement towards a capi-

---

[14]Lee [10].

[15]Marx [13].

[16]Grüner [7].

[17]Smith [17].

talist economy, especially given that the 1801 Constitution declared "The commerce of the colony consists only in the exchange of goods and products of its territory."[18] Given this, we might appropriately say that the Haitian Revolution or more specifically, the economic system of Haiti had gone through some process of 'primitive accumulation' – moving from chattel slavery to a capital-based economy.

There are some limitations to this characterisation, of course. We only need to look to the years following the establishment of Haiti as a nation; according to Smith: "Isolated and starved of capital, the Black and free men of colour were unable to develop a vibrant capitalist economy, but slipped back into subsistence farming and the consequent underdevelopment that has plagued Haiti ever since."[19] Despite Haiti even now being defined as a 'free market economy', "Two-fifths of all Haitians [still] depend on the agricultural sector"[20] demonstrating that although it made efforts to move from chattel slavery to a capitalist economy, it remains relatively pre-capitalist. Indeed, if Lawrence is correct in his summation of Marx that "the capitalist mode of production appears first in industry, later in agriculture",[21] then Haiti has not thoroughly demonstrated a capitalist mode of production, remaining reliant on agriculture without a proper industrial sector flourishing.

Such limitations, however, should not prevent us from accepting that the Haitian Revolution *failed* in its achievement of primitive accumulation. First, it did at least move away from chattel slavery. Secondly, despite its reliance on agriculture and subsistence farming, Haiti remains, structurally, a free-market economy. The economy is thus run by capitalist mechanisms, even if the core of such mechanisms rely on agriculture. In this way, the Haitian Revolution somewhat succeeds in being an example of primitive accumulation, forcing the economy to move from a system of slavery to an economy with buyers, sellers and commerce. What remains important for our argument is that in principle, the Haitian Revolution displays this principle of primitive accumulation that is at the core of what we have defined as a Marxist revolution.

---

[18][14].

[19]Smith [17].

[20][4].

[21]Lawrence [9].

Condition 3

The third and final condition of F* was expressed as historical materialism. Essentially, this feature seeks to establish if the Haitian Revolution is an example of social change occurring through the tension between the productive forces in society and the economic systems in place. The conditions are summarised by Fromm as such:

> When a mode of production or social organisation hampers, rather than furthers, the given productive forces, a society, if it is not to collapse, will choose such forms of production as fit the new set of productive forces and develop them.[22]

Before the outbreak of revolution, the productive forces of Saint-Domingue were the black slaves, and the mode of social organisation as chattel or forced slavery. Indeed, as Marx observed that the "the bourgeoisie forged the weapons that bring death to itself",[23] so too did the French bourgeoisie forge the weapons that would bring death to its colonisation of Saint-Domingue. As James describes:

> With every stride in production the colony was marching to its doom. The enormous increase of slaves was filling the colony with native Africans, more resentful, more intractable, more ready for rebellion.... Of the half-a-million slaves in the colony in 1789, more than two-thirds had been born in Africa.[24]

Indeed, the flourishing economic production of the colony of Saint-Domingue would merely be the 'calm before the storm' – as the production relations, namely slavery, became untenable for the productive forces, the slaves, revolution began. Between 1783 and 1789 production nearly doubled[25] – in August 1791 the slave insurrection in Saint-Domingue begins. Clearly then, a Marxist analysis of historical change based on the production relations in society is applicable to the events of the Haitian Revolution.

---

[22]Fromm [5].

[23]Marx [13].

[24]James [8] pp. 55–56.

[25]Ibid. p. 55.

Summative Comments

To summarise the argument presented in favour of the Marxist analysis, I showed above that the three separate conditions of F* could be seen to persist in the Haitian Revolution. Given the argument we expressed in this section, in particular (P2), it can confidently be argued that the Haitian Revolution was a Marxist revolution in tendency, since it conforms to the conditions of F*.

Limitations

There are some limitations to the pro-Marxist view that should be expressed. Firstly, the argument asserts, almost as if to beg-the-question, that a revolution can be characterised as Marxist if it displays F* then concludes that, given this and the fact that the Haitian Revolution demonstrates F*, that the Haitian Revolution was therefore Marxist in tendency. Can we be so sure that a revolution is Marxist if it displays F*?

I note here that the first premise does not seek to limit the characteristics of a Marxist revolution solely to F* – indeed other aspects of Marxist theory have been omitted (not purposefully) in the analysis. For example, F* does not discuss the alienation of labour. To what extent does alienated labour characterise a Marxist proletariat that could equally be applied to the Saint-Domingue slaves? While such a concept is important in Marxist theory, I argue that in leaving this out from discussion I have not greatly hampered my ability to adequately argue that the Haitian Revolution was Marxist in theory. What remains is a debate on the core tenets of Marxism, a debate for which there is no relevance or capacity in this paper.

A further limitation of the pro-Marxist case is that it may not adequately allow for the Haitian Revolution to be a revolution of its own kind, unique and un-analysable in eurocentric terms. Such a view has been adopted by Grüner:

> Against euro-centrist common sense...such revolutions are not an exclusive product of European socio-economical, political and cultural developments either. The Haitian Revolution illustrates this outstandingly, it entails its own specificities.[26]

Indeed, such a response is persuasive. Why should we even attempt to 'white-wash' the Haitian Revolution? The approbation one should give to its events

---

[26]Grüner [7].

should stand not in its analogy to Marxism, but as its own, unique kind of revolution. For example, the Constitution of 1805, uniquely declares in own of its articles:

> All acception [sic] of colour among the children of one and the same family, of whom the chief magistrate is the father, being necessarily to cease, the Haytians [sic] shall hence forward be known only by the general appellation of Blacks.[27]

As Grüner argues, "the first and most radical answer to all the false philosophical-political 'universalisms' entailed by Enlightenment thought... was given by the Haitian Revolution."[28] Assuming such an argument to be in the very least, possibly true, we see that the Revolution did have features that made it stand uniquely.

I believe, however, the objection that the Haitian Revolution was unique does not disprove the pro-Marxist view. For even if there were features unique to the Haitian Revolution alone, this possibility does not contradict the outstanding thesis that the Haitian Revolution was Marxist in tendency. The Haitian Revolution can be unique in many aspects as well as Marxist in its tendency. Much like the historically-debated Russian Revolution, the Haitian Revolution was a Marxist revolution with features that made it stand above as a historical event.

Conclusion

This essay has discussed extensively the pro-Marxist view that claims that the Haitian Revolution can be characterised as Marxist in tendency. I began with a distillation of what I believe to be core tenets of Marxist philosophy, namely F*. In so doing, I brought together the key features of what a Marxist revolution generally entails. I then showed that the Haitian Revolution does display the features that were distilled, concluding that the Haitian Revolution was Marxist in tendency. I then presented some of the limitations with the pro-Marxist view and showed that the view does not attempt to beg the question, and it allows for other tenets of Marxist philosophy to be included in such an analysis of the Haitian Revolution – or indeed any other revolution. I also presented the objection that the Haitian Revolution was unique and then argued that such an objection still allows for one to conclude that the Haitian Revolu-

---

[27] [2].

[28] Grüner [7].

tion can be characterised as Marxist in tendency, and does not undermine this view. I did, however, accept that there are features of the Haitian Revolution that make it stand uniquely to other revolutions and events in history.

In conclusion, a Marxist evaluation of the Haitian Revolution would claim that the revolution was Marxist in tendency, but also had many features that made it stand in isolation to similar revolutions in history.

References

[1] Blackpast.org, 'Haitian Revolution (1791–1804)|The Black Past: Remembered And Reclaimed'. [Online] Available at `http://www.blackpast.org/gah/haitian-revolution-1791-1804`

[2] Constitution of Haiti (1805). Available at `http://www2.webster.edu/~corbetre/haiti/history/earlyhaiti/1805-const.htm`

[3] Dessalines, J. (1805) 'Haiti: 1805 Constitution'. [Online] Available at `http://www2.webster.edu/~corbetre/haiti/history/earlyhaiti/1805-const.htm`

[4] Forbes, 'Haiti – Forbes'. Forbes Web. [Online] Available at `http://www.forbes.com/places/haiti/`

[5] Fromm, E. (1961) *Marx's Concept Of Man*. New York: Frederick Ungar Publishing.

[6] Garfield, J. (2013) 'I Have Avenged America'. *Haiti and the Atlantic World* [Online] Available at `http://haitidoi.com/2013/08/02/i-have-avenged-america/`

[7] Gruner, E. (2010) 'Haiti: A (Forgotten) Philosophical Revolution' A. Marchi (trans.), in *Sociedad (B. Aires)*. [Online] Available at `http://socialsciences.scielo.org/scielo.php?pid=S0327-77122008000100004&script=sci_arttext`

[8] James, C.L.R. (1989) *The Black Jacobins*. New York: Vintage Books.

[9] Lawrence, K. (2015) 'Karl Marx on American Slavery'. [Online] Available at `http://www.sojournertruth.net/marxslavery.pdf`

[10] Lee, S. (2015) 'Until We Are All Abolitionists: Marx On Slavery, Race, And Class'. *The International Marxist-Humanist* [Online] Available at `http://www.internationalmarxisthumanist.org/articles/abolitionists-marx-slavery-race-class-salome-lee`

[11] Marx, K. & Engels, F. (2015) 'Address of the International Working Men's Association To Abraham Lincoln, President Of The United States Of America'. *Marxists.org* [Online] Available at `https://www.marxists.org/archive/marx/iwma/documents/1864/lincoln-letter.htm/`

[12] Marx, K. & Engels, F. (2015) 'Manifesto of the Communist Party'. *Marxists.org* [Online] Available at `https://www.marxists.org/archive/marx/works/1848/communist-manifesto/`

[13] Marx, K. (2015) 'Economic Manuscripts: Capital: Volume Two'. *Marxists.org* [Online] Available at `https://www.marxists.org/archive/marx/works/1885-c2/`

[14] Marxist Encyclopaedia, Available at `https://www.marxists.org/glossary/terms/p/r.htm`

[15] L'Overture, T. (1801) 'Constitution Of 1801 By Haiti'. *Marxists.org* [Online] Available at `https://www.marxists.org/history/haiti/1801/constitution.htm`

[16] Robinson, J. (1983) *Black Marxism*. Chapel Hill, N.C.: University of North Carolina Press.

[17] Smith, A. (2009) 'The Black Jacobins: A Review Of C.L.R. James's Classic Account Of Haiti's Slave Revolt'. *International Socialist Review* [Online] Available at `http://isreview.org/issue/63/black-jacobins`

[18] Wolff, J. (2011) 'Karl Marx', in E.N. Zalta (ed.), *The Stanford Encyclopaedia of Philosophy* [Online] Available at `http://plato.stanford.edu/archives/sum2011/entries/marx/`

Does the Extended Phenotype Exist?*

Elizabeth MacFadyen
University of Leeds

Richard Dawkins' extended phenotype (hereafter EP) was conceived as a new way of looking at genetics and evolutionary biology, intended to raise new questions and further existing understanding. While Dawkins likely intended to enhance the explanatory power of evolutionary theory, subsequent scientific advancements may render EP outdated and unhelpful. Here, it is considered in what sense EP exists and if niche construction theory (hereafter NCT) may provide a better way of thinking about evolution.

Dawkins' conclusion, that animal artefacts and behaviour should be considered phenotypes, is reached by arguing that the only direct consequence of a gene is the synthesis of the protein it codes for; all other biological process or features are only indirectly caused by genes. As such, Dawkins reasons, far more indirect effects of genes should be considered phenotypes, rather than arbitrarily drawing a line between internal and external consequences, as long they in some way impact on the fitness of the gene causing the specific effect. This has clear links to Dawkins' selfish gene theory, a metaphor Dawkins uses to argue for a gene-centric view of evolution; that is, that natural selection acts on individual genes, causing genes to be in competition with each other to be passed to the next generation. However, more recently, EP has been linked to an organism-centric view of evolution instead, leading to NCT. NCT supports the idea that artefacts, such as a beaver dam, could be considered phenotypes; however, it also supposes that these phenotypes may be heritable, and that, over time, may change the environment to such an extent as to impact on the fitness of other organisms, and even change the selection pressures on future organisms.

In 'Three Kinds of Constructionism: The Role of Analogies and Metaphor in the Debate over Niche Constructionism', Elizabeth Archetti conceives of three ways a theory may be true: literally, analogically and figuratively.[1] How literally true EP is depends on the definition of phenotype. For Dawkins, phenotype is "[t]he manifested attributes of an organism, the joint products of its

---

*Delivered at the BJUPS Spring Conference, 23–24 March 2016 at the University of Cambridge.

[1]Archetti [1].

genes and their environment during ontogeny".[2] This in fact contains two very different definitions of phenotype: firstly, that the phenotype encompasses specifically an organism's observable traits; secondly, that the phenotype is the causal consequence of genes and their interactions.[3] Under this second definition, EP undoubtedly literally exists, provided it is used to describe something that is, however indirectly, demonstrably a consequence of an organism's genes. Even the first definition does not prevent a literal interpretation: while it means that, say, the house of a caddis fly could not be considered to be in itself the caddis phenotype, the act of building the house could be. It is surely not controversial to suggest that behaviour, as far as it is genetically determined, can be an attribute of an organism.[4]

Unfortunately, this literal sense gives the gene far more reach than Dawkins would perhaps like: for him, an effect can only be an EP if it impacts on the fitness of the organism.[5] In truth, a phenotype does not need to impact on selection. One can only imagine, for example, that curly and straight hair are both present in human populations precisely because there is no selective advantage of one over the other, yet surely Dawkins would accept that hair type is a phenotype. Eye colour may provide another example: once early humans moved north, blue eyes may have become more common not because they gave their bearer an advantage, but simply because they ceased to be a disadvantage. The true criterion for a phenotypehood seems only to be that there be variance correlating with genotypic variation. As such, Dawkins is inconsistent when he says, for example, that buildings are not an EP of human beings.[6] By his own reasoning, that humans today build buildings must mean that some of our ancestors did not, and the fact that not all species construct shelters indicates a genetic component.[7] While obviously, as Dawkins says, the building does not increase the fitness of the architect, we can imagine that in prehistory constructing one's own shelter was advantageous compared to depending on naturally-occurring shelter. Thus even Dawkins' more stringent EP definition has been met.

EP may also be true analogically, say if an artefact is standing in for an animal's

---

[2]Dawkins [3] p. 299.

[3]Wells [10].

[4]Dawkins [3]

[5]Dawkins [3] & [4].

[6]Dawkins [4].

[7]Dawkins [3].

phenotype, as according to the first definition of phenotype as an attribute of an organism. Examples of this could be the brush turkey mound or bower bird bower.[8] In both cases, the artefact provides the female with information regarding the quality of the male constructor as a mate, in the same way that a peacock's tail feathers might. The mound or bower is representative of the male's phenotypic traits, without the traits of the mound or bower being in any sense traits of the creator. This contrasts with, say, a spider web. The existence of a spider web indicates that its creator was able to construct webs, but does not give an indication of any other aspect of its creator's phenotype.

In many respects, however, the figurative view of EP may be more accurate. In order to enhance evolutionary understanding, the traits of an artefact should be treated as if they were traits of the builder. For example, it may be better to treat a beaver dam as if it were the beaver's phenotype so that its evolutionary origins can be investigated in the same way any other beaver trait would be. It seems likely that Dawkins intended EP to be this sort of ontological metaphor, and he states that it is not intended as a testable hypothesis, but rather as a different perspective of the same facts designed to further understanding.[9] More generally, it may be impossible for a theory built on metaphors to be itself anything else. Darwinian natural selection contains several metaphors, including 'selection' itself.[10] In EP, the concept of a gene is a metaphor, used to explain phenotypic variation arising from variation in both genes and other replicators.[11] Thus, the gene 'for' a characteristic is a catch-all term for all the systems of inheritance and their interactions that culminate in that characteristic. Finally, for Dawkins a phenotype is just a way for the gene causing it to increase its chances of being carried forward to the next generation, and it makes no difference whether that phenotype is internal or external to the organism housing the gene.[12] This means that EP, as Dawkins means it, is an extension of his selfish gene metaphor, and so must surely be a metaphor itself.

However, if the selfish gene metaphor were established to be invalid, then it would follow that the concept of EP cannot be true even metaphorically. A general criticism of the selfish gene theory is that the concept of a single gene

---

[8]Wells [10].

[9]Paton [8], Dawkins [4] p. 1.

[10]Paton [8].

[11]Jablonka [5] & Dawkins [4].

[12]Wells [10].

for a single trait is an oversimplification, one which Dawkins acknowledges.[13] However, new research in fields such as epigenetics – referring to heritable changes in phenotype without changes in genotype, caused by factors such as age or health – suggests that the pathways that lead from genotype to phenotype are so complex that Dawkins' simplification may now be unhelpful, or even detrimental, even if it wasn't at the time of his writing.[14] This has led to the suggestion that such theories should focus on 'heritable variable traits'.[15] As these could theoretically include animal artefacts, e.g. termite heuweltjies in South Africa,[16] genes, behaviours or any other phenotype, it removes the need to consider just one of those to be the main focus of evolution.

Furthermore, a phenotype cannot necessarily be viewed as the gene's way of passing itself to the next generation. Sexual reproduction means that no matter how effective a phenotype is, there is a strong element of chance involved in whether or not it is successful in passing itself on to the next generation. During the development of gametes, genes are randomly separated and recombined, a process called independent assortment. As such, while genes that produce successful phenotypes may have a better chance of being replicated than genes that don't, there is a possibility, however remote, that none of an organism's genes may be present in its grandchild. Additionally, just as the only thing the gene 'cares' about is replicating itself, so also the only thing natural selection 'cares' about is how well a particular organism fits with its environment. To illustrate, suppose all members of a population have gene A, but due the effect of environment or epistatic interactions (the effects of other genes), gene A results in trait 1 in some individuals and trait 2 in others. Supposing trait 1 was a better adaptation to the environment than trait 2, it follows that after many generations, all members of the population would have trait 1, yet the proportion of individuals with gene A would be unchanged. Now suppose in another population, half of the individuals have gene A and half have gene B. But suppose that again due to epistatic or environmental interactions, both genes produce trait 1. According to the selfish gene theory, A and B are in competition with each other; however, neither actually has a selective advantage over the other. Both A and B would, theoretically, continue to exist in the same proportions for many generations. This would suggest that in both cases

---

[13]Dawkins [3].

[14]Ball [2].

[15]Jablonka [5].

[16]Turner [9].

natural selection is selecting for trait 1, not the genes that cause it. Moreover, it is not necessarily 'up to' the gene what trait it produces. To use the example of a beaver dam, while it's true that the dam, depending on its quality, increases the chance of the gene 'for' it to replicate, the dam will also have feedback effects on the fitness of other genes.[17] A gene that increases the fitness of other genes risks decreasing its own relative fitness, which seems incompatible with the selfish gene theory.

An alternative to EP that allows for this more complex feedback model is NCT. NCT is the process by which an organism changes the relationship between its features and its environment's factors, thus potentially altering selection pressures on itself and other organisms.[18] It includes the same principle as EP: that genes can be expressed in the environment or in the behaviour of hetero- or conspecifics, as in the case of some parasitic infections, but also takes into account the impact EP may have on the evolution of both the organism to which it belongs and others. Like EP, NCT can be perceived as literal, analogical and metaphorical. Here, literal may concern artefacts like beaver dams or spider webs that are clearly engineered; analogical concerns a construction process, for example learnt behaviour or migration, that might enhance the survival of an organism and its offspring but does not result in a physical product; and, more generally, metaphorical could concern any organismal modification of the environment that does not otherwise fit with the connotations of 'construction'.[19]

However, NCT is in many respects a flawed theory. The evolutionary consequences of NCT are debatable. The 'fly in soup' objection – that the effect of an individual or even a single species would be too small to have any significant impact on the environment – is disputed by some proponents of NCT,[20] but may still be valid. While it is easy to see the consequences that, say, photosynthetic oxygen production has on the world, it is uncertain whether other forms of NCT could be as significant. Odling-Smee, Laland and Feldman suggest that some bird species' incorporating spider silk into their nests is evidence for spider web construction's having widespread consequences.[21] However, it is difficult to imagine what selective advantage this would provide in an environment

---

[17] Laland [6] & Wells [10].

[18] Odling-Smee, Laland & Feldman [7].

[19] Archetti [1].

[20] Turner [9].

[21] Odlling-Smee, et al. [7].

with plentiful nest building materials. Unless birds only began building nests once spiders started making webs, there does not seem to be any evolutionary significance to this. Analogically, NCT is also unconvincing. One could argue, for example, that migration is in no sense construction. Through natural selection alone, an organism living in a seasonally changing environment must be adapted to two quite different niches: summer and winter. The same holds for a migratory animal – the only difference is that its niches are separated spatially rather than temporally. While some proponents of NCT argue that adaptations organisms have for different seasons – e.g. food hoarding for winter – also count as NCT, such a broad interpretation leads to other problems.[22] If NCT refers to any and all modification of the environment, then it becomes a necessary consequence of existing.[23] If every living being niche-constructs, then arguably the concept lacks any explanatory power.

There is only a very narrow sense in which EP exists. Taken literally, it encompasses far more than Dawkins ever intended; if its literal definition renders it so different to Dawkins' original conception then it may be better not to consider the two versions to be the same. While it may have some explanatory power as a metaphor, new understanding of genetics suggests that both EP and its selfish gene theory underpinning are oversimplifications such that they may hinder rather than help understanding. EP may exist as an analogy, but this does not give it as much explanatory power as Dawkins may want. As such, it seems that there is no real or helpful sense in which EP exists. NCT may initially seem to provide a better explanation of the same facts; however, if EP is too narrow, then NCT is too broad and its explanatory power too extended. In its current state it is no more successful at explaining the complexity of the different inheritance systems, their interactions and their consequences than is EP.

References

[1] Archetti, E. (2015) 'Three Kinds of Constructionism: The Role of Analogies and Metaphor in the Debate over Niche Constructionism', in *Biological Theory* 10/2: 103–115.

---

[22] Odlling-Smee, et. al. [7].

[23] Archetti [1].

[2] Ball, P. (2011). *Nature*. [Online] [Accessed 12 November 2015] Available at: `http://www.nature.com`.

[3] Dawkins, R. (1999) *The Extended Phenotype: The Long Reach of the Gene*. Oxford: Oxford University Press.

[4] Dawkins, R. (2004) 'Extended Phenotype – But Not Too Extended; A Reply to Laland, Turner and Jablonka'*Biology and Philosophy*,19(3): 377–396.

[5] Jablonka, E. (2004) 'From Replicators to Heritably Varying Phenotypic Traits: The Extended Phenotype Revisited', in *Biology and Philosophy* 19/3: 353–375.

[6] Laland, K. (2004) 'Extending the Extended Phenotype', in *Biology and Philosophy* 19/3: 313–325.

[7] Odling-Smee, J., K. Laland & M. Feldman (2003) *Niche Construction: The Neglected Process in Evolution*. Princeton, NJ: Princeton University Press.

[8] Paton, R.C. (1992) 'Towards a Metaphorical Biology', in *Biology and Philosophy*, 7/3: 279–294.

[9] Turner, J. (2004) 'Extended Phenotypes and Extended Organisms', in *Biology and Philosophy* 19/3: 327–352.

[10] Wells, D. (2015) 'The Extended Phenotype(s): a Comparison with Niche Construction Theory', in *Biology and Philosophy* 30/4: 547–567.

Can Externalism Refute Cartesian Scepticism?*

Anna McAlpine
University of St Andrews

1 Introduction

The aim of this paper is to argue in favour of the claim that externalism is successful in refuting Cartesian scepticism. In order to prove this I shall argue against the Cartesian sceptic who believes that phenomenal identity of good cases and bad cases entails indiscriminability as far as which one obtains. Externalism is a theory that has a lot of sub-categories within it but due to space restrictions I shall not be able to address all permutations of the theory in this essay. I will be focussing my investigation around externalist disjunctivism as well as bringing in a slightly stronger externalist view from Williamson when I examine externalism and the mental state. I will structure my argument around three key areas within each case: (1) belief discrimination, (2) content discrimination and (3) the difference in mental states. After proving that we can discriminate between the good and bad case by differentiating these three factors in each case, I will briefly discuss my view that the ability to comprehend second-order knowledge should not affect our right to first-order knowledge.[1]

2 Cartesian Scepticism: The Good, the Bad and the Phenomenally Identical

One of the key aims of Cartesian scepticism is to deny us the possibility of asserting beliefs as knowledge. The sceptic's argument revolves around two cases referred to as the 'good case' and 'bad case'. There can be different understandings of how exactly these cases can be defined. I shall begin this essay by thinking of the good case as being a world in which I am a brain within a person who is directly interacting with the external world and the bad case as

---

[1]In this essay I will not argue that externalism can bring us second-order knowledge. I discuss higher knowledge in this essay only to block a rebuttal from the sceptic that may threaten our ability to ascertain first-order knowledge.

being a brain in a vat (BIV) which is being fed information from a computer program. Both the bad case and the good case can be said to be phenomenally alike. This can be accepted by both internalists and (some) externalists, as in each case it is possible that we are receiving visually similar images. Although this point is widely accepted, it is this point that the Cartesian sceptics will use as a springboard for their argument that we cannot assert beliefs as knowledge (and as such we should not hold these beliefs or assert them).

I will illustrate this by outlining the Cartesian sceptic's main arguments and conclusion below.

P_1 The good case and the bad case are phenomenally alike.

P_2 The good case and the bad case are indiscriminable in virtue of which one obtains.

P_3 I cannot know that I am not a BIV.

P_4 If I cannot know that I am not a BIV, then I cannot know p.

C I should not assert or believe p if I cannot know p. (silent stage)

The Cartesian sceptical argument can be very hard to dismiss once you first accept it, as it does seem on first glance to contain reasonable premises that lead us to the conclusion. However, I will aim in this essay to block the Cartesian argument before it gets a grip on us. I will do this by stopping the argument in the early stages. I will use externalism to refute the link between premise one and premise two – I do not think that the phenomenal alikeness of the good case and bad case will then entail that they are indiscriminable. In my first attempt at refuting this claim I will turn to externalist disjunctivism (a specific form of externalism) for a solution. A disjunctivist can claim that, whilst the good case and bad case may be phenomenally alike, they are not indiscriminable. There are many ways in which disjunctivism can do this using its different forms. In the following sections, I shall focus on disjunctivism of belief, thought content, and mental states in both the good and bad cases.

3 The Belief Factor

The most basic way for a disjunctivist to claim that we can discriminate between a good case and bad case is to claim that we do not have the same belief in each case. It seems that the most straightforward way to try and challenge the sceptical view is to say that we have a true belief in the good case and a false belief in the bad case – this would provide us with a distinction between the cases to show that they are discriminable. If we take, for example, the belief

that Paul is eating an apple in my room. The disjunctivist may claim that in the good case you have the true belief that Paul is eating an apple but in the bad case you have a false belief – say that Paul is eating a pear. The reason a disjunctivist could make such a claim is due to a certain unreliability factor that they can play on. It seems that if you are a BIV in the bad case then your beliefs will be formed on purely internal methods as you are not linked up to the external world. You are instead linked up to a computer program, and so it seems that if there is a scientist who feeds you data to form beliefs then on some days he may feel mischievous and input false beliefs to the BIV.

I think then that the disjunctivists' first assault of the sceptical argument is a natural one – the reliance upon the computer program and scientist for the BIV to form beliefs seems to make the BIV's claim on belief a very uncertain one in terms of truth and falsity as it relies on this ambivalent scientist.

4   The Sceptic's Reply and a Change in 'Case'

I think that, on the surface of the matter, the disjunctivist's claim seems to be a reasonable one; although she is not saying that the BIV will *always* have false beliefs, it does seem that it would be very easy for it to not have a true belief due to its dependence on the scientist operating the computer program. However, it seems that the sceptics can manipulate the example, and strengthen their case in the process, in order to alleviate the concern regarding true and false beliefs. The sceptic can change the bad case so that the scientist in control of the input data to the BIV is a benevolent man who will always feed the BIV true beliefs. In virtue of this, it will also be implied that it will never be the case that the BIV is deceived.

If this were so then the difference between the good and bad case based around the notion of the falsity of beliefs in the bad case would no longer be a counterexample to the sceptic's argument. The BIV now has veridical beliefs – we can call it VBIV. In this case I must find something stronger than this to counter the sceptical premise of indiscrimability.

Before I continue, though, it is important to note that this strengthening of the bad case also seems to change the nature of relations between each case. Whereas before I was examining each case in terms of the good case being a brain in a person's body that is a part of the external world and the bad case being a BIV hooked up to a computer program, it seems now that although the basic idea in each case is the same, as the BIV is now under the control of a benevolent scientist who will always give it justfied true beliefs (JTBs). It therefore seems there must be a restructuring. I think that, due to the fact that

both can now have true beliefs, it must be taken that, in the good case where I am a person in the external world, I can have knowledge (K) and in the bad case I am in a state of ignorance (−K).

5 Content Disjunctivism

In response to the strengthened Cartesian case we can use content disjunctivism to counter the indiscriminability principle. This view proposes that we simply have different thought content in each case. We can now argue that we are in a state of ignorance in the bad case because the type of belief and thoughts that we have are but mere illusion. I can state this more clearly by using my example again. It may seem that in both cases there is a thought containing the content 'Paul is eating an apple'. This is a proposition which the externalist will state has object-dependent content. So, through applying the externalist principle, we can argue that in the good case this object-dependent content is apparent to us and so we are able to have this thought and belief because we are causally connected to the external world in which this content resides (e.g. in the good case I am externally connected to Paul eating the apple which produces a thought). This is a strong externalist principle by which Paul is actually a constituent of my thought. However, the externalist can argue that we are in a state of ignorance in the bad case because we are not able to form a thought. I cannot form a thought in the bad case because I am not connected to the external world and cannot have the experience of Paul – so he is literally lacking from my thought.[2]

You might subscribe to this but then ask what is happening then when it appears that I form a thought or belief in the bad case. The externalist answer is that I am really only having an illusion of content or thought. In this respect, the two cases can still appear phenomenally alike but remain distinct and discriminable because via co-operation and interaction with the external world we achieve a thought – otherwise, we are deluding ourselves that we have thought.In the bad case where Paul never existed, the relevant content from which we could form the belief that Paul was there eating an apple is simply not available to us.

A sceptic may counter that what has been said may hold in worlds where Paul does not exist in the bad case – but they have strengthened their case so that the BIV will always have a JTB and so that, in actuality, Paul has to exist in the

---

[2]See, in particular, Crawford [1] p. 201. This thought is also discussed in McDowell [5].

bad case. An externalist can still expand the scope of their argument, though, to include cases where Paul exists. Even in the state of ignorance, where Paul exists, the singular content of 'Paul' is not available because we could not have demonstrative access to Paul himself.[3] It is clear, then, that the reason why we can argue that we can discriminate between each case is that in the good case, in virtue of being a part of the external world, we will have demonstrative access to the singular content, which the BIV will not have and so will not be able to have the same content of thought that the good case experiences.

6 Sceptic Response

I believe that content disjunctivism is a good way to allow discrimination of the good and bad cases. However, the sceptic can respond that, even if they grant that it can bring in a distinction in cases where there are singular objects, the same analysis cannot hold for cases where there are thoughts that do not have this singular content. For example, a Cartesian sceptic may argue that we cannot apply the same content logic to the thought 'there is a person eating an apple in the room' which we derive from the belief that Paul is eating an apple in the room. In the more generalised sentence, the good and bad cases seem to be once again equalised and indiscriminable due to the fact that there is no longer the issue of their thought's relating to a specific object in the external world.

In light of this issue, I will now argue in favour of a stronger disjunctivist view of mental states – a natural follow-up to the discrimination of content. It has been implied by the distinction between belief and content in both cases that, in the bad case, we can have an illusion or hallucination of belief and content, as opposed to being hooked up in some correct way to the external facts in the good case. To take this distinction further, I will examine what I believe is a distinct difference in mental states in the good and bad cases, which I will look upon based on the differences of perception that have so far been implied.

7 The Disparity of Mental States

Following from the fact that the externalist view must now extend their theory to show the disparity between a good case and a bad case, I will now analyse

---

[3]Hawthorne & Sturgeon [4].

the mental states in each case. It seems that, if I can argue that in each case the two brains are in separate mental states, then I can show that the two cases are discriminable.[4] This argument aims to exploit the difference in perceiving in both cases and the way this difference affects the mental states in them. Pritchard wrote a lot on disjunctivism and has claimed that on the issue of mental states, unfortunately, the disjunctivist view seems somewhat divided. He states that "whether epistemological disjunctivism is committed to mentalism depends on whether you count an agent's seeing that *p* as a mental state".[5]

Pritchard uses the term 'metaphysical disjunctivism', which, as he puts it, seems the most natural view on which *perception* of *p* counts as a mental state.[6] This metaphysical disjunctivism claims that the mental states differ between the two cases because the supervenience base also so-differs.[7]

However, it seems that this distinction on the basis of supervenience may not be so useful after all. A principle of local supervenience allows the different cases to replicate each other so that the bad case can replicate the phenomenology of the good case just by reproducing its activities without containing any of the objects of external experience.[8] With respect to this principle, it could be argued that the metaphysical disjunctivist account fails because the use of perception and a factive mental state could be argued to be a factor that can characterise the good case but is not essential to the good case's being different from the bad case. The crux of the counterargument presented by Ludwig and Thalbard would be that perception, whist being a property of the good case, is not an essential property and that there are lots of other common properties shared between cases that make them indiscriminable.[9]

---

[4]I argue this in virtue of the fact that a recognisable distinction of separate mental states would give rise to criteria for discriminability.

[5]Pritchard [2] p. 3. By 'mentalism' Pritchard means that if in the two cases we have the same internal support for a belief then we are in the same mental state.

[6]Ibid.

[7]By 'supervenience' Pritchard is drawing on the relationship between epistemic properties and their asymmetric necessitation from non-epistemic properties (cf. Pritchard & Kallestrup [3]).

[8]Pascal & Thalabard [7].

[9]I suppose the sceptic who would favour this account would say that the property of perception is not essential to the good case because we might be hallucinating – just as we are under an illusion in the bad case. If we imagine that someone has taken a copious amount of drugs then they might start to hallucinate that Paul is eating an apple in the room when in fact they are on their own – just like the BIV could have the same hallucination. In virtue of this, the disjunctivist must make a stronger claim if they are to defeat Cartesian scepticism.

8   Linking Mental States and Perception

The problem then seems to be that the disjunctivist view may not be able to interpret the difference in perception within a bad case and a hallucinatory experience in the good case. However, "disjunctivists...maintain that a genuinely perceptual experience is of a different fundamental kind from a hallucinatory experience".[10] I think, then, that it is a weak rebuttal to counter that the disjunctivist argument by claiming that there are still common factors amongst the good and bad cases, as perception is not an essential property of the good case.[11] I believe this because it seems that there may be times where I will be in the good case, insofar as I am not a BIV (I have direct experiences of the external world) yet I hallucinate. Yet in such a case the disjunctivist would not hold that I obtain knowledge; this would still be a state of ignorance due to the fact that there would be no genuine 'constituent' of my thought linking me to the external world; so it seems that the disjunctivist can still align their arguments to prove that perception is an essential property of the good case.[12]

I do think that a harder line can be drawn to guard against the possible mix up of hallucinations and genuine perception to secure the thought that the good case and bad case are discriminable and have different mental states. I think this can be done using an extra-strong externalist understanding of perception and mental states. To show this I shall be using ideas from Williamson, an externalist philosopher. He may not fully subscribe to a disjunctivist view but I think that his ideas on mental states and knowledge are relevant in that they can be adopted by the disjunctivist theory in order to provide a robust refutation of Cartesian scepticism.[13]

---

[10] Overgaard [8] p. 53.

[11] The idea of the 'highest common factor' theory is explained by McDowell as it is used by sceptics to advance the idea that the experience available to us is the same for both the good and bad cases due to a common traits in both (McDowell [6] p. 386).

[12] Overgaard [8] p. 53.

[13] Williamson himself states that "[a]lthough the letter of disjunctivism accounts have been rejected, the spirit may have been retained", which suggests that it is possible to couple the two to defeat scepticism (Williamson [9] p. 228).

9   A Stronger View

Williamson is an externalist who argues that knowing is a factive mental state.[14] Williamson's argument is basic – you cannot further compartmentalise knowledge – and knowledge is a factive mental state. What I want to take from Williamson's paper is that perceiving is a basic factive mental state. If this is the case then it can be seen that in the good case we are in a completely different mental state than in the bad case in which we are not perceiving (and as such, not knowing). Even if in the bad case we are a VBIV (veridical brain in vat), this would still not affect the externalist argument because one of the more controversial points of Williamson's argument is that truth is not a component of knowledge.[15] Williamson argues that knowledge cannot be reduced further, i.e. we cannot have VJTB = K as this would imply that these extra things all come together to make K; Williamson's point is that K is the base element.

Williamson's view seems to go much further than the view of the disjunctivist. This is because he claims that there can be no 'common factors'between the good case and the bad case as the subjects in them are in completely different mental states. Through Williamson's externalist view, the subject will not be perceiving in the bad case and as such will not gain knowledge. In the bad case, at best, they can be said to have a belief which is a non-factive mental state (e.g. believing). This would prove that there is no shared mental state and so we are able to discriminate between the good and bad case, refuting Cartesian scepticism.

10   Knowing That We Know

Some may question the last point of the last section, namely,that in the good case we have knowledge as a factive mental attitude and in the bad case only a belief which is a non-factive mental attitude. Some might want to argue against Williamson's argument by using the following summation:

P₁ Knowing is a mental state.

P₂ Mental states are transparent (i.e. if you are in state *S*, you know you are in state *S*).

---

[14]Ibid. p. 213.

[15]Ibid. p. 218.

P_3 $K \Rightarrow KK, \neg K \Rightarrow \neg KK$[16]

What seems to be a major problem for externalists is that, even if a sceptic is to accept my argument that through the essential differences we can show that it is possible to discriminate the bad case and the good case, the sceptic can still take this further. The sceptic could now argue that we are still stuck within the sceptical argument as we have no way of knowing that we know in the good case or not knowing in the bad case. This then incorporates a worry about second-order knowledge. It seems that, when we are in the good case and we are externally hooked up to the world in the right way so that we have knowledge, this will have no indication on us knowing that we know. In relation to the above premises, based on this lack of higher-order knowledge, it would seem that the sceptic will invite us to reject P_1 and bring us back into the fold of thinking that the good and bad case are indiscriminable.

I think that what really needs to be done is for us to reject premise two and the idea that mental states must be transparent. I think we can reject this as we do not need to know what mental state we are in to be in that state. This can be illustrated more concisely with an example.[17] It is a natural assumption to think that when someone is in pain they will know that they are in pain – after all it is happening to them so how could they *not* know? This might lead us to think that if someone does not know that they are in pain then they cannot claim to be in pain. Yet there could be a circumstance where you are preoccupied and even though you are in pain you do not realise it because of your distraction. Or in another case you are feeling very upset with your life and will mistake even a small itch for feelings of pain. Therefore, although I have not made a claim on second-order knowledge, the sceptic cannot use the concept of higher knowledge to defeat the claim that the good case is in a different mental state from the bad case just because we cannot know when we are in each case. This would lead us to the conclusion that this strong externalist principle when used in conjunction with disjunctivism can indeed be used to deny the Cartesian sceptic the right to assert that the bad case and good case are indiscriminable.

---

[16] That is, if *A* knows that *p* then they know that they know that *p*; if they do not know that *p* then they do not know that they know it either.

[17] Williamson [9] p. 215.

11 Conclusion

To conclude, I have argued that externalism can refute the Cartesian sceptic by blocking the sceptic's argument in the initial stages through a denying the phenomenal identity of the good and bad case entails indiscriminability as to which obtains. This has been shown to be the case due to the analysis of the disjunctivist argument, which could be used to refute each sceptical scenario even when it was repeatedly strengthened. This was then coupled with the strong externalist view, put forward by Williamson, which proved that we are in two different mental states in the good and bad case (and such an argument is not put under threat by the worry of second-order knowledge).[18] Due, then, to the combined factors of disjunctivism and a stronger externalist conception of mental states I believe that they can show the discriminability of the good and bad case and restore us to a state of knowing and asserting utterances, *contra* to the wishes of the Cartesian sceptic.

References

[1] Crawford, S. (1998) 'Discussions in Defence of Object-Dependent Thought', in *Proceedings of the Aristotelian Society* 98/2.

[2] Pritchard, D. (2012) 'Epistemological Disjunctivism and the Epistemic Externalism/ Internalism Distinction', in *Epistemological Disjunctivism*. Oxford: Oxford University Press.

[3] —— & J. Kallestrup (2016) *Epistemic Supervenience, Anti-Individualism and Knowledge First Epistemology for Knowledge First*. Oxford: Oxford University Press.

[4] Hawthorne, J. & K. Kovakovich (2006) 'Disjunctivism', in *Aristotelian Society Supplementary* 80/1.

[5] McDowell, J. (1986). 'Singular Thought and the Extent of Inner Space', in P. Pettit & J. McDowell (eds.) *Subject, Thought and Context*. Oxford: Clarendon Press.

[6] —— (1998) *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press.

---

[18]Williamson [9].

[7] Pascal, L. & Emilie, T. (2014) 'Against Metaphysical Disjunctivism', in J. Dutant, D. Fassio & A. Meylan (eds.) *Liber Amicorum Pascal Engel* Geneva, Switzerland.

[8] Overgaard, S. (2013) 'Motivating Disjunctivism', in *Husserl Studies* 29/1.

Sturgeon, S. (2006) 'Reflective Disjunctivism', in *Aristotelian Society Supplementary* 80/1.

[9] Williamson, T. (2008) 'A State of Mind', in E. Sosa et al. (eds.) *Epistemology: An Anthology*. Oxford: Oxford University Press.

Modal Reduction Made Easy: Reducing Metaphysical Modality to Ideal Conceivability*

Eugene Chua
University of Cambridge

1 Introduction

In metaphysical discourse, philosophers often resort to conceivability or imagination claims in the course of making m-modal claims. For example, Kripke gives the following argument for essentialism of biological origin.

> How could a person originating from different parents, from a totally different sperm and egg, be this very woman? One can imagine, given the woman, that various things in her life could have changed.... But what is harder to *imagine* is her being born of different parents. It seems to me that anything coming from a different origin would not be this object.[1]

The question is: what *explains* the fact of the matter that, if I am born of X and Y, then I am *necessarily* born of X and Y? What facts can we appeal to? For example, claims about *physical* modality (p-modality) – i.e. empirical claims such as 'it is impossible for water at 30 degrees Celsius to melt typical diamonds' – are explained by the physical facts in *this* world,[2] while claims about *logical* modality (l-modality) – i.e. claims about theorems and axioms of logic like 'given $((A \rightarrow B) \wedge B)$, B necessarily follows' – are explained by the theorems and axioms of the various logico-mathematical systems. But *what*, and *where*, are the facts that explain *metaphysical* modality (m-modality) claims such as Kripke's?[3] The ensuing debate has largely divided into two: primitivists about

---

[1]Kripke [4] p. 113. Emphasis added.

[2]I take, but do not argue here for, a Humean/best systems view of laws of nature, taking them to be non-modal summaries of facts under a Mill–Ramsey–Lewis 'best system', and whatever is p-possible is just whatever does not contradict the laws of nature, while the laws of nature themselves are p-necessary.

[3]Kit Fine argues that essence can be distinguished from necessity, that "the notion of essence which is of central importance to the metaphysics of identity is not to be understood in modal terms

m-modality argue that m-modal facts are *fundamental*, i.e. while there is no further fact that explains an m-modal fact, m-modal facts can be expressed in terms of possible worlds (which may be constructed out of abstract resources such as sentences, propositions, etc.). However, since we make use of m-modal facts in describing possible worlds, m-modal facts are arguably *conceptually prior* to the apparatus of possible worlds[4] and thus cannot be explained by possible worlds. Contrariwise, reductionists about m-modality claim that m-modal facts are *non-fundamental*: they can be explanatorily reduced to *non*-m-modal facts. One way of conceiving of these non-m-modal facts is, according to Lewis, to reduce m-modal facts to facts about possible worlds. Under Lewis's modal realism, possible worlds are concrete, spatiotemporally extended and, thus, non-modal entities. However, both approaches have been deeply dissatisfying so far. On one hand, the primitivist has much to account for in simply assuming modality to be fundamental, and primitivism should be seen as a last resort after we have exhausted all means of explaining m-modality. On the other, reductionist accounts such as Lewis's modal realism postulate an infinitude of spatiotemporally and causally isolated possible worlds, which we can nevertheless come to have knowledge of. It is hard to see how we can come to know of such concrete worlds,[5] and also hard to accept that this infinitude of worlds is spatiotemporally extended.[6] Lewis argues that the theoretical benefits – the capability to reduce modality to non-m-modal facts – are worth the ontological costs,[7] but, as I shall argue, we have other means of reducing m-modality and thus no need to take up Lewis's gambit.

---

or even to be regarded as extensionally equivalent to a modal notion" (Fine [2]). Not wanting to derail the argument, I simply take it here in this essay (following Kripke, who made the biological origins argument) that F is an essential property iff $\forall x\,(F(x) \rightarrow \Box F(x))$.

[4]For example, ersatz modal realism characterises an ersatz world as possible iff it would be possible for all the members of the set to be true together. As Sider comments, "[T]his characterisation of an ersatz possible world makes use of the notion of possibility." "Therefore, a possible-worlds analysis of modality that makes use of ersatz possible worlds in this way is circular." (Sider [8] p. 10) So we cannot explain modality by simply appealing to ersatz modal realism.

[5]Lewis suggests that we can have knowledge of mathematical objects through some rational faculty, and that we can likewise have knowledge of possible worlds through some such similar faculty (Lewis [5] p. 109). However, the status of mathematical objects as *concrete* objects is dubious at best, and if we take mathematical objects to be abstract then there is a disanalogy between Lewis's concrete possible worlds and abstract mathematical objects that demands elaboration.

[6]Thus the 'incredulous stare' given by many philosophers to modal realism.

[7]Lewis [5] p. 4.

My view is this: we like to think that what is possible is what is conceivable, but we hesitate because such conceivability claims are subjective and limited by our cognitive abilities. Thus, we require rational reasoning to minimise that error. The intermingling of these two facts allow us to talk about something we call 'm-modality', but there are no such modal facts to be found. Why not settle for facts about conceivability and rationality then? Why not claim: the fact of the matter will be decided simply if everyone involved in the debate, in conceiving of such a statement, either (i) *cannot conceive of that statement otherwise* – hence said claim is *necessary* – or (ii) someone provides a *rational defeater* to that statement by providing an alternative conceivability claim, hence making the claim merely *possible*; that is, someone *can* conceive otherwise. Together they explain the uses of m-modality we have.

There has already been work done in this direction. As Yablo writes, "conceiving involves the appearance of possibility."[8]. Chalmers further remarks that "[i]t is striking that many of these purposes are tied closely to the rational and the psychological: analysing the contents of thoughts and the semantics of language, giving an account of counterfactual thought, analysing rational inference".[9] What I am offering is simply a further step: facts about possibility are just explained by facts about conceivability and rationality – no more, no less – because there are no other facts to be found. In slogan form: m-modality is explained by conceivability, modulated by rationality. For example, some logical truths are often claimed to be m-necessary, in addition to being l-necessary: in that case, $(P \wedge \neg P)$ is m-necessary iff no one can conceive of $\neg(P \wedge \neg P)$. Of course, the metaphysically inclined will be quick to point out that conceivability is indeed a good *guide* to m-modality, but this in no way means that facts about m-modality can be reduced to facts about conceivability. M-modality is an important part of our intellectual enterprise – surely it is not as trivial as what *someone* can conceive. However, it seems like no other account can adequately answer the questions of what *other facts* there could be, and how we can come to know of them. In addition, I will show that facts about conceivability have to make m-modality trivial.

---

[8] Yablo [10] p. 5.

[9] Chalmers [1] p. 39.

2   Ability and Conceivability

2.1   Abilities

I begin by defining *abilities* as properties of individuals, the possession of which allows them to act in certain ways. Abilities relate to the domain of p-modality, and depend on the actual physical properties a subject has. Thus, a person *S*'s possession of an ability is a *physical* fact about an individual: if I have an ability to recite Shakespearean sonnets, it is p-possible for me to recite Shakespearean sonnets; it is not merely m-possible for me. There are different abilities – to write an essay, to play tennis, etc. Likewise, to conceive of *P* as the case (or otherwise) is also an ability. Thus:

(1) *P* is conceivable for a subject *S* iff *S* has an ability to conceive of *P*.

(2) *P* is inconceivable for a subject *S* iff *S* does not have the ability to conceive of *P*.

To conceive is, vaguely, to 'grasp' an idea: it shares similarities with imagining something, but imagining specifically requires perceptual information, while conceiving is something wider in scope. For example, to conceive of a 1000-sided polygon (1000-gon) is different from imagining one. I can grasp the concept of a 1000-gon and solve problems about it, but I cannot imagine how a 1000-gon would look in my mind: it would look just like a 1001-gon or a 1002-gon, and they probably all look like circles to me anyway. Thus, imagining is a form of conceiving, though not the other way round.

So far, we have been looking at facts about conceivability as facts about an arbitrary subject *S*'s ability to conceive of *P*. This makes conceivability out to be a somewhat subjective notion, since different people arguably conceive of different things. An easy objection is to simply question whether specifically *S*'s ability to conceive of *P* falls apart from the metaphysical necessity about *P*. For example, Pryor asks:

> Suppose I get amnesia, and forget my name. I think that Jim Pryor is somebody else. I hear bad things about Jim Pryor and so I decide to kill him. I think 'The world will be such a better place, with me in it but with Jim Pryor no longer existing.' Isn't that imagining myself existing without Jim Pryor's existing?[10]

---

[10]Pryor [7].

Pryor's example suggests that something can be metaphysically impossible, e.g. 'I can be alive while I am dead', involving a contradiction between being alive and being dead, even when it is conceivable by someone that it be the case. Firstly, it is unclear whether conceiving of *someone with the name Jim Pryor* is the same as conceiving of *oneself*. However, even granting him that fact, I believe this objection disappears once we distinguish between *ideal* and *prima facie* conceivability.

2.2   *Prima Facie* and Ideal Conceivability

On one hand, *P* is *prima facie* conceivable for a subject if, in Chalmers' words, "after some consideration, the subject finds that [*P*] passes the tests that are criterial for conceivability".[11] Thus: if *P* is *prima facie* conceivable for *S*, then *S* has the ability to conceive of *P*, i.e. can conceive of *P*. It is true that many things are *prima facie* conceivable for people even when the conceived claim is something typically known to be metaphysically impossible; there is no doubt a mismatch between what we take to be metaphysical modality and *prima facie* conceivability. *Prima facie* conceivability, after all, is limited to the speaker's cognitive and evidential limitations, and thus is dependent on who makes the claim. Thus we update our definitions:

(3) *P* is *prima facie* conceivable for a subject *S* iff *S* has an ability to conceive of *P*.

(4) *P* is *prima facie* inconceivable for a subject *S* iff *S* lacks the ability to conceive of *P*.

On the other hand, *P* is *ideally* conceivable if conceivable on 'ideal rational reflection'.[12] Note that *P* is stated to be *ideally* conceivable *tout court*, rather than for *any* subject – there is some objective import to a statement being ideally conceivable. Obviously, it has to be *prima facie* conceivable by *someone* but there is a further fact about *P*, viz. it conforms to ideal reasoning. People often conceive of inconsistent or incoherent things (e.g. geometrically impossible 3D shapes, killing oneself while staying alive) but eventually realise it only *seemed* conceivable; further reasoning dissolves such claims. The notion of ideal rational reflection is, as Chalmers suggests, best understood as 'unde-

---

[11]Chalmers [1] p. 2.

[12]Ibid.

featability by better reasoning',[13] where undefeatability should be understood not as 'possibly undefeatable' (to avoid any modal notions) but as actually following from or contradicting the rules of ideal reasoning. Thus:

(5) *P* is ideally conceivable iff, for some *S*, *P* is *prima facie* conceivable for a subject *S* and *P* conforms to ideal reasoning.

(6) *P* is ideally inconceivable iff either there exists no *S* such that *P* is prima facie conceivable for a subject *S*, or *P* fails to conform to ideal reasoning.

This account deserves more filling out. However, we address Pryor's complaint adequately: we can simply say that the subject in his anecdote made a *prima facie* conceivability claim (i.e. he *thought* he could conceive it), but it can be defeated by another better conceivability claim in accordance with ideal reasoning. When we reason about Pryor's subject's claim, another conceiver C might say that the subject's claim, "I can be alive while Pryor is dead", is *prima facie* inconceivable by C because in the subject's claim, both "Pryor" and "I" are used to refer to the subject, and it is a tautology that, if he is alive, then he is alive – so the subject cannot be conceived of as being alive while Pryor dies. The subject is misguided because he only knows that "Pryor" refers to someone, whoever did the bad things, but he does not know that "Pryor" refers to him. Once we follow through with the claim's conclusions, however, the subject's claim is *defeated*, and is thus not an ideal conceivability claim. In sum, if our reduction of facts about m-necessity to facts about conceivability makes use of the concept of ideal conceivability, then Pryor's complaint is not a problem at all.

We now need a way to characterise ideal reasoning in a way that does not presuppose m-modality, as that leads us down the path to circularity or primitivism. As such, we impose ideal reasoning on conceivability claims because there are certain rules and axioms we hold to be important to rational discourse. M-modal claims involve discussing what *could* have been, but there are no modal facts in this world that explain claims of this sort. There are only *prima facie* conceivability claims, disparate, subjective and dependent on each conceiver's abilities – we conceive of scenarios based on the limits of our experiences and knowledge. However, as we have seen with the Pryor example, *prima facie* conceivability claims do not cohere with the importance we ascribe to m-modality. Hence, we need something to sort the bad claims from the ones that are good. As Chalmers writes: "One needs justification that cannot be ra-

---

[13] Ibid. p. 3.

tionally defeated."[14] What better criterion for rational justification than the rules of rational thinking, the rules we *already* use on a day-to-day basis to judge mundane claims like whether your niece's answer to a maths question was correct?

One might argue that such rules are of a certain privileged sort which lend us *insight* into m-modal truths. I beg to differ – there is no reason why this is so. We know the rules of arithmetic, the rules of logic, some notions of mereology and causation, etc. They are true in our world. However, what else unites them, making them m-modal truths, rather than just reliable tools? I follow Sider's conventionalism about such rules of ideal reasoning.

> [T]he 'certain sorts' of propositions invoked by the Humean are not objectively distinguished, that no joint in reality encircles the class of logical and mathematical truths. So what *does* select this class? Something about *us* ....[15]

Nothing in the world brings coherence to the rules other than us, and what selects these rules as rational rules, over and above axioms and truths in their various fields of knowledge, is a fact about us as rational thinkers. Here, explanation bottoms out in facts about human psychology and rationality. For example, the basic laws of logic are such that conceiving of their negation necessarily involves the use of the laws themselves; it is perhaps one reason why we take mathematical and logical necessity to imply m-necessity. These laws are thus part of the rules of ideal reasoning. On the other hand, the laws of nature are not psychologically compelling – we do not argue about the laws of nature rationally, in an *a priori* fashion, but take them to be given empirically, to be 'whatever they actually are'. Thus, when we conceive of the billiard ball being hit in one way, we can always conceive of it moving in another way, and this is true even if determinism is true (i.e. even if it is physically necessary on some accounts that the ball moves only one way). Perhaps one day, when we have a complete science, such rules will become rationally given as well, but so far we do not have this.[16] As such, many empirical statements end up being merely contingent or possible – they do not explicitly go against any ideal rules of reasoning, i.e. contain no contradiction or fallacy, so they are ideally conceivable, and thus m-possible. I have tried to characterise ideal reasoning

---

[14]Ibid.

[15]Sider [9] p. 270.

[16]Some scientists are certainly beginning to show dogmatism with regards to some tenets of physics.

by listing a few kinds of rules, but regardless of what exactly the rules are, it suffices to say that they exist. As Chalmers recognised, "any such attempt would end up being open-ended and incomplete".[17]

2.3 Necessity and Conceivability

I previously claimed, with (i), whatever is m-necessary is whatever cannot be conceived as otherwise. In light of our explication, it is clear that facts about m-modality are not to be explicated in terms of *prima facie* conceivability, but *ideal* conceivability. So:

(7) *P* is *m-necessary* iff ¬*P* is not ideally conceivable.

(8) *P* is *m-possible* iff *P* is ideally conceivable.

With this, we can show how our account coheres with some basic axioms relating m-necessity and m-possibility in traditional modal logic:

(9) *P* is m-necessary iff it is not m-possible that ¬*P*.

If ¬*P* is not m-possible, then ¬*P* is not ideally conceivable. This is exactly the case when *P* is m-necessary under our account.

(10) *P* is m-possible iff it is not m-necessary that ¬*P*.

If ¬*P* is not m-necessary, then ¬¬*P* is ideally conceivable. In other words, *P* is ideally conceivable. Again, this is exactly the case when *P* is m-possible under our account.

To sum up, *P* is m-necessary iff we cannot conceive of ¬*P*, either because no one finds ¬*P* *prima facie* conceivable, or because ¬*P* violates ideal reasoning. *P* is m-possible iff it is *prima facie* conceivable to someone, and it conforms to ideal reasoning. Thus, two types of facts explain *P*'s m-necessity or m-possibility: whether its negation is *prima facie* conceivable by anyone, and whether it conforms to ideal reasoning. The former is a psychological fact about human beings, while the latter is a rational fact about human intellectual activity. Together they determine the m-modality of any claim.

The Humean point comes in here. *P*'s status as an m-necessary claim simply involve two facts: either ¬*P* is *prima facie* inconceivable for no one, or ¬*P* contradicts the rules of ideal reasoning. Note that I do not proclaim that these

---

[17]Chalmers [1] p. 3.

facts about us make *P true*. *P* is true by virtue of whatever makes it true in our world – that debate is left to the other fields to decide. As Sider writes, "The proposition that $2 + 2 = 4$ is made true by whatever makes mathematical truths true generally (facts about mathematical entities, perhaps). Its status as a *mathematical* truth is secured by whatever generally makes mathematical truths mathematical."[18] However, what makes *P m-necessary*, over and above its truth, is explained by two further physical, non-m-modal facts about human beings. If no human has the ability to conceive of ¬*P*, either because no one ever finds ¬*P prima facie* conceivable, or, even if someone does, ¬*P* contradicts the rules of ideal reasoning that we have – sensibly so, since typically, if ¬*P* is ideally inconceivable, then *P* is either an ideal rule of reasoning, or follows from the ideal rules of reasoning. So, under our account, facts about m-necessity reduce to at least one of two physical facts: one about human beings' ability to conceive of *P*, and another about whether what we conceive conforms to the rules of ideal reasoning.

M-modal facts are reducible to non-m-modal facts about the human ability to conceive of things, and non-modal facts about whether our *prima facie* conceivability claims are in line with the rules of ideal reasoning. Questions about possibility reduce to questions about ideal conceivability, and we are familiar with the epistemology of such facts. Facts about our abilities to conceive of things are physical facts about us, and facts about whether our conceivability claims follow the rules of ideal reasoning are logical facts. The realm of metaphysical modality is thus explained. One might object to the claim that this makes m-necessity trivial, since the rules of ideal reasoning in the convention might arguably change, but given how long some of these rules (e.g. logic, certain folk-metaphysical laws) have stuck with us, and how many of these rules are cross-cultural and stable across time, the importance of these rules may be due to something intrinsic to our psychology as human beings, and less subjective than we think. The notion of m-modality is thus demystified.

3   Objections

3.1   Ability and Modality: Purported Circularity

One main question that looms over my account is whether my reduction of m-modality to facts about the human ability to conceive of things implicitly

---

[18]Sider [9] p. 321.

involves m-modality again if we analyse the notion of 'ability'. If m-modality figures again when we try to explain the notion of an ability, then it would be a fatal blow to my account, which seeks to reduce m-modal facts to non-m-modal facts and explain them through non-m-modal facts. Luckily, I do not think that this is the case.

Firstly, I have characterised abilities as p-possible facts. That is, if someone *can* do something, or has the ability to do something, it is *p-possible* for someone to do that thing. While the orthodox debate has often claimed p-possibility to entail m-possibility, I explain it as such: if something is in fact p-possible, then when a conceiver conceives of it, the rules of ideal reasoning do not preclude its being conceivable, and thus it is ideally conceivable. If that is the case, it follows from my account that it is *m-possible*. So p-possibility *does* entail m-possibility, but by way of facts about ideal conceivability. P-modality itself is not the same as m-modality. P-modality relies on facts about the world, while m-modality relies on facts about us, and facts about ideal reasoning. There is some overlap, viz. the facts about us as human beings, but m-modality also relies on rules independent of the world, viz. ideal reasoning rules. So, even if we took abilities to be explained in terms of p-modality, there is no reason why it must entail an explanation in terms of m-modality. More likely, I insist that m-modal facts are explained in terms of abilities, which are then explained in terms of p-modal (non-m-modal) facts. Furthermore, if we take a Lewisian best systems approach to p-modality, the laws of nature just are the summaries of *non-modal facts* under a Mill–Ramsey–Lewis 'best system'– whatever regularities in our world past, present and future that best explain the world in accordance to the virtues of strength, simplicity and fit. Whatever is p-possible is just whatever does not contradict the laws of nature i.e. best system regularities, while the laws of nature themselves are p-necessary, i.e. the regularities themselves. So, there is arguably a way to analyse p-modality in terms of non-modal facts. Thus, using abilities need not lead to circularity.

Even if one does not buy that argument, it is unclear whether abilities are analysed in terms of m-modal facts. It is known that m-modal facts can be analysed in terms of modal logic. Some have argued that abilities are *restricted possibilities*. Maier writes:

> To have an ability is for it to be possible to *A* in some restricted sense of possibility . . . . [T]his is to be understood in terms of accessibility relations among possible worlds: someone has the ability to *A* just in case there is an accessible world where she *A*s. The task of giving a theory of ability is then simply the task of articu-

lating the relevant accessibility relation.[19]

However, Kenny showed that facts about abilities are incompatible with several theorems of even the most basic systems of modal logic. He gives two plausible theorems of modal logic:

(11) $A \rightarrow \Diamond A$

(12) $\Diamond(A \vee B) \rightarrow (\Diamond A \vee \Diamond B)$

The first is true in modal logics T or stronger, and it says that if A, then possibly A. If we take ability to P as possibly P, as Maier suggests, then, according to (11), someone has an ability to P just if he Ps. But this is wrong – I can accidentally score a ball into the hole at golf, without having the ability to do so. Kenny writes "A hopeless darts player may, once in a lifetime, hit the bull, but be unable to repeat the performance because he does not have the ability to hit the bull."[20] In all such cases, I do something, but I lack the ability to do it. So (11) is wrong if we take abilities to be restricted possibilities. One might argue that (11) is not true on K, the weakest modal logic. If that is the case, perhaps one might argue that K should be the modal logic of choice, in explaining abilities.

This may work for (11), but (12) is true even under K, yet untrue if we apply it to an analysis of abilities. If it is possible that A or B, then either possibly A or possibly B. However, applying it to abilities, we have the statement 'if someone has an ability to A or B, then someone either has an ability to A or an ability to B'. However, as Kenny writes, "[g]iven a pack of cards, I have the ability to pick out on request a card which is either black or red; but I don't have the ability to pick out a red card on request nor the ability to pick out a black card on request".[21] Thus, it seems that even the weakest modal logic, K, has trouble accommodating an analysis of abilities. If there is one at all, it is not clear what that analysis is.

Either path we take, we notice that abilities are resistant to analysis by m-modality, and capable of analysis by other non-modal means. If that is the case, the alleged circularity disappears.

---

[19]Maier [6].

[20]Kenny [3] p. 214.

[21]Ibid. p. 215.

3.2 The Modal Apparatus

Another challenge comes in the form of scepticism about applying my account to modal investigations. Here, I provide some examples with regards to certain key concepts of modal investigations – viz. possible worlds, *a posteriori* necessities and *de re* modalising – and show that most such tools can be left untouched at the surface level.

I. Possible-worlds can be explained as a *tool* for communicating and making precise our conceivability claims. The ersatzer was not wrong: a possible world is a set of propositions describing a scenario involving *P*, corresponding to the claims we make in a *prima facie* conceivability claim about *P*. Now, not all such possible worlds are candidates for m-modal talk, since there will be contradictory ones – useless for rational discourse. By applying ideal reasoning, we end up with at least logically consistent possible worlds including the claim *P*.[22] That is, only possible worlds that are ideally conceivable will pass ideal reasoning, and thus facts about them will be candidate facts for explaining m-modality.

II. My account of m-modality suggests modal knowledge is arguably *a priori*; *P*'s m-modality does not involve the fact that *P*, only what I can conceive about *P* and whether *P* coheres with ideal reasoning, which is free from empirical evidence and thus *a priori*. This suggests that all m-necessities, insofar as they are dependent only on ideal reasoning and my ability to conceive of it, are *a priori*. However, what about Kripkean statements of *a posteriori* m-necessity, such as statements that (given that gold actually has the chemical structure bearing 79 protons) gold m-necessarily has the chemical structure bearing 79 protons? It seems that such statements, if they exist, pose a huge problem to our account which seems to suggest that all modal facts are explained by *a priori* facts.

However, recall the distinction at the end of Section 2 between truth and necessity. What makes *P* true is whatever makes it true based on its subject matter, e.g. mathematical/logical/scientific truths, but what makes it m-necessary is

---

[22]Logical consistency is often claimed to involve hidden modal talk, where logical consistency is defined in terms of whether it is *possible* for statements to be true together. However, it seems at best talk about l-possibility, and not m-possibility. It is unwise to simply conflate them together; my account explains the common conflation by this. Our talk about m-modality is explained by conceivability and rationality, and the rules of rationality involves the basic laws of logic. So in ideal conceivability, l-modality sneaks into our facts. But the modality that we depend on for the use of consistency remains the realm of logical modality, not metaphysical modality.

explained by conceivability and ideal reasoning. In this case, we can think of statements of *a posteriori* m-necessity (as they are usually conceived) as involving a putative schematic claim (E): things have essential properties, without which that thing ceases to be *that* thing. Many philosophers would argue that essences are important to our conceptual framework – perhaps they are so important as to be psychologically binding.[23] In chemistry, we have a specific version of it: we think of elements as having the essential property of having a chemical structure. Such claims, being psychologically binding – that is, people actually *prima facie* conceive of such rules – and not clearly violating any ideal reasoning, become ideally conceivable and thus *a priori* m-necessary. However, conceiving of this claim with regards to an element, e.g. gold, requires both the schematic rule (E) and an input into (E) of an actual empirical fact about that gold's chemical structure, i.e. having 79 protons. With the addition of an empirical fact, the claim becomes *a posteriori* m-necessary. However, the m-necessity of that statement is derived from the application of an *a priori* m-necessary claim (E) to an *a posteriori* non-m-modal fact in our conceivability claim. We can then account for *a posteriori* m-necessities in our theory: while they seem to be *a posteriori m-necessary*, what makes them *m-necessary* does not come from an *a posteriori* source but an *a priori* one, viz. the claim (E). What makes the claim m-necessary is still the fact that it is ideally conceivable – thus, such cases do not defeat our thesis that what makes a claim m-necessary is explained by purely ideal reasoning and conceivability, both of which are *a priori*.

III. *De re* modalising suggests that things have genuine modal properties. In our account, modal claims are not based on any modal facts or modal properties of objects; rather, they are based on our abilities to conceive of those objects with certain properties, subject to ideal reasoning. If $P(x)$ is 'x is P', then for x to m-necessarily be P is to say that I cannot ideally conceive of $\neg P(x)$. There is no reference to modal properties here. For example, Kripke argues that (EO) we could not have had different parents, and thus our having the parents we do is *m-necessary* for our existence. There are, however, as I have questioned in the beginning, seemingly no m-modal facts fit to explain such a claim. One simple way is to think of (EO) as either an ideal reasoning rule or something psychologically binding – it is thus ideally inconceivable that

---

[23]It is arguable that essences are surely metaphysical. But *prima facie* I see no strong reason to think so. Sider concurs "We should not regard non-modal essence as being metaphysically basic…fundamental theories need essence no more than they need modality." It seems that essences are more important relative to human discourse than anything else. I avoid debate about this in this essay.

(EO) is false, and thus (EO) is *m-necessary*. It is unclear whether this is indeed a rational rule, or something that no one can conceive of otherwise, but both seems likely (depending on who you ask). Either way our account does not obviously preclude explanation of (EO) in terms of ideal conceivability facts.

As such, it is clear that we can account for some of the modal apparatus we often use in philosophising, i.e. possible worlds and *a posteriori* necessities. If that is the case, then nothing on the surface level has been lost – we simply press on recognising that m-modality is in fact short-hand talk for facts about conceivability and rationality.

4   Conclusion

At the end of Section 1, I promised a non-trivial account of how facts about m-modality are to be reduced or explained in terms of facts about conceivability. Through the use of the concept of ideal conceivability, we see that m-modal facts can be explained in terms of facts about conceivability and facts about rational reasoning – both of which are non-m-modal. By understanding conceivability as an ability, we can either reduce it to p-modal facts and so to non-modal facts (according to the best systems view), or at least resist the view that abilities must themselves be explained by m-modal facts. Such an account respects the importance of m-modalising; under my view, m-modal facts are the facts we come to when we reason about the world, attenuated by the rules of ideal reasoning. They are thus our *cognitive best* as human beings, and that is all there is to m-modality – what we take to be possible, given our abilities to conceive, and the rules of ideal reasoning. If these claims match the facts in our world, they are p-modal facts; otherwise, they remain m-modal facts to be explained by conceivability and rationality. We m-modalise in order to explore whether something is p-possible. That, I believe, is what is vital to m-modality. Like a good Humean, then, I have provided an account of m-modality that makes m-modalising epistemologically and ontologically easy: epistemologically, the facts of conceivability and rationality are commonly accessible facts, unlike the queer possible worlds of Lewis, or the unexplanatory ersatz possible worlds of primitivists. Ontologically, they are also non-m-modal facts of a common kind – facts about human psychology and rationality. My account can also accommodate common modal apparatus, e.g. possible worlds, *de re* modalising and *a posteriori* necessities. Given these results, I conclude that my account of explaining and reducing m-modal facts to facts about ideal conceivability is a very plausible one. There is nothing more we need of m-modality that cannot be explained by ideal conceivability.

References

[1] Chalmers, D.J. (2002) *Does Conceivability Entail Possibility?* [Online] Available at `http://consc.net/papers/conceivability.html`

[2] Fine, K. (1993) *Essence and Modality* [Online] Available at `http://www.academia.edu/8908669/essence_and_modality`.

[3] Kenny, A. (1976) 'Human Abilities and Dynamic Modalities', in J. Manninen & T. Raimo (eds.), *Essays on Explanation and Understanding: Studies in the Foundations of Humanities and Social Sciences*. Dordrecht, Holland: D. Reidel Pub.

[4] Kripke, S. (1980) *Naming and Necessity*. Cambridge, Ma: Harvard University Press.

[5] Lewis, D.K. (1986) *On the Plurality of Worlds*. Oxford, UK: Blackwell.

[6] Maier, J. (2014) 'Abilities', in E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* [Online] Available at `http://plato.stanford.edu/archives/fall2014/entries/abilities/`

[7] Pryor, J. (2016) *Necessity, Possibility, and Conceivability* [Online] Available at `http://www.jimpryor.net/teaching/courses/intro/notes/necessity.html`

[8] Sider, T. (2003) 'Reductive Theories of Modality', in M.J. Loux & D.W. Zimmerman (eds.), *The Oxford Handbook of Metaphysics*. Oxford: Oxford University Press.

[9] Sider, T. (2011) *Writing the Book of the World*. Oxford: Clarendon Press.

[10] Yablo, S. (1993) 'Is Conceiving a Guide to Possibility?', in *Philosophy and Phenomenological Research* 53/1.

Repacking the Package Argument: Diagnosing and fixing George Klosko's Argument for Political Obligation

George Coiley
University of Sheffield

George Klosko's 'package' argument attempts to use the obligation one has to contribute to presumptive public goods (PPGs) to create obligation to also contribute to discretionary public goods (DPGs). This paper first argues that the package argument is unsuccessful. The second half then tentatively suggests how it can be repaired.

Specifically, Section 1 sets out Klosko's position; describing his conception of PPGs and explaining their significance to his package argument. It also introduces 'unassailability', a characteristic of PPGs. Section 2 demonstrates that Klosko's package argument fails by showing that, even when his conditions are met, obligation to DPGs is not necessarily present. Section 3 diagnoses this failure and shows that the package argument fails for two reasons. First, unassailability makes the package argument too strong. Second, PPGs fail to transmit obligation onto DPGs. Finally, Section 4 puts forward a repair for the package argument. It is suggested that consent can be created for a PPG-like good through the use of normative consent theory. When these PPG-like goods are integrated into the package argument, the above problems are solved and obligation to a larger state is justified.

1   The nature of PPGs and Klosko's package argument

The Principle of Fairness states that 'we are not to gain from cooperative efforts of others without doing our fair share'.[1] Since we benefit from living in a political state, Klosko thinks this can be used to found political obligation. However, the limiting argument responds that political obligation cannot be created just by the receipt of a benefit because consent is a precondition of obligation.[2] In answer, Klosko presents PPGs, benefits that fulfil the following

---

[1]Rawls [4] p. 112.

[2]Nozick [2] pp. 90–96.

conditions:

  (i) Public (non-excludable schemes, so universally received, regardless of contribution)

  (ii) Presumptively beneficial ('indispensable' for an acceptable life in the given community)

  (iii) Net beneficial (worth the recipient's effort in providing it)

  (iv) Fair (the burdens of providing the good and the good itself are fairly distributed)[3]

Confronted with a PPG, the limiting argument is defeated. Briefly, this is because the nature of PPGs is such that they create hypothetical consent. Therefore, the receipt of benefit from PPGs creates an obligation to take on some of the burden of supplying the benefit.

Klosko emphasises the indispensability of PPGs to answer his detractors and unambiguously defeat the limiting argument.[4] However, it is important to note here that if (ii) is true then it follows that (iii) must also be true, since the nature of an indispensable good means that it is worth *any* effort expended providing it. The result of this is that (iii) is superfluous because, if the PPG is presumptively beneficial, it *must* also be net beneficial. This constitutes a specific quality of Klosko's PPGs that is subsequently referred to as *unassailability* in this essay.

Only a few goods fulfil the conditions required to constitute a PPG, so PPGs only justify obligation to a very limited state. Klosko is not satisfied with this, so expands his argument with the intention of creating obligation to contribute to the burden of providing goods that are public, yet not indispensable: DPGs.[5] He proposes that DPGs be added to a package that includes at least one PPG. Since the individual is already obligated to the PPG, she remains obligated despite the addition of DPG(s). The conditions Klosko sets out for such a package are that it is public, fair, net beneficial and contains at least one PPG.[6]

---

[3]Kloskco [5] pp. 247–250.

[4]Detractors include Simmons and Nozick, both of whom diminish the power of Klosko's argument by presenting examples that are not PPGs in Klosko's strict sense.

[5]Klosko [5] p. 253.

[6]Ibid. p. 255.

2 Demonstrating the package argument's failure

This section demonstrates that Klosko's package argument is unsuccessful by providing examples where obligation to the additional DPGs is not present, despite the package fulfilling the conditions just given. It also clarifies Klosko's fairness condition.

Example 1. In Country 1 there used to be a dragon problem. Crops were regularly torched, children plucked up while playing and the population lived under a shadow of constant fear. Now, though, the government-run National Hero Service employs competent slayers who ensure the population's safety. The NHS is a PPG, so each of the population has an obligation to bear some of the burden of running it. Coincidentally, the population of Country 1 happens to be predominantly Christian and a law exists that states that every citizen must decorate their house with lights over the Christmas period. Since no-one is able to walk down the street without noticing them, these lights constitute a DPG. Most of the population enjoy the displays, but there are a few atheists and people who follow other faiths who find the law tiresome or even offensive. Here, it seems clear that the non-Christian minority has no moral obligation to contribute the Christmas light DPG.

However, the package described in Example 1 unambiguously fulfils the conditions of being public and including a PPG. The unassailability of the PPG constituted by the NHS means the package is also net beneficial: although a minority do not approve of the decorations, dragons are far worse than Christmas lights. The fairness condition is the only one not clearly fulfilled. Indeed, it is clear that the Christmas light DPG imposes a disproportionate burden on the non-Christians.

Therefore, it is important to clarify what Klosko means by fair distribution of burdens. Citing David Lyons, he claims that a fair distribution of burdens only demands that each participant is an 'integrated element' in a given scheme.[7] The consequence of this is that burdens can be shared unequally, yet the scheme remains fair in his terms. It is the circumstances of each individual that decide the extent of the burden for which the PoF creates an obligation. So in Klosko's example, soldiers, who are much more likely to be killed than their superiors, remain obligated despite their heavier burden, so long as there are good reasons for their being assigned their particular task.[8] This concep-

---

[7]Klosko [5] p. 251.

[8]Ibid. pp. 250–251.

tion of fairness is attractive, as common schemes frequently demand differing levels of burden from their contributors and we do not want this to invalidate obligation to such schemes.

So the fact that non-Christians have a heavier burden when contributing to the DPG does not mean the scheme is not fair. Furthermore, compared to the burden that the dragon slayers carry, it is insignificant. Such a comparison is permissible because Klosko's argument does not impose obligation to contribute to a public scheme separately, but as a package. This means that there is no distinction between helpful burdens (risking one's life fighting dragons) and unhelpful burdens (putting unwanted Christmas lights up).

A final counter-point might be that there is no *good* reason non-Christians should put up decorations, but this fails to get off the ground. The good reasons that someone is assigned the role of dragon slayer might be that they are courageous and handy with a sword. As long as these reasons exist, the dragon slayer is obligated to contribute to the scheme according to the principle of fairness. In the case of the Christmas light DPG, the good reason for allocating the duty to everyone is simply that everyone has a relatively equal capacity for putting up lights; the fact that non-Christians dislike the duty more is irrelevant. So the package described in Example 1 does fulfil the fairness condition.

Example 2. The citizens of Country 2 are under constant threat of outside attack from neighbouring countries whose sole objective is to destroy them. The survival of their society is entirely due to an army that is run on a conscription basis. This army constitutes a PPG. Unrelatedly, the citizens of Country 2 enjoy bloodthirsty entertainment. No-one can explain why, but it is accepted fact, so once a week a citizen is selected randomly and hung live on television. This has no effect on the war effort, so is entirely dispensable and constitutes a DPG. But again it is clear that no-one can have a moral obligation to be hung for the entertainment of onlookers.

However, once again the package unambiguously fulfils the conditions of being public and including a PPG. Due to the unassailability of the PPG, the package is net beneficial: without the PPG of the conscripted army the entire population would be slaughtered. It is also fair, as the selection of those to be hung is done by lottery, so the distribution of burden is at least as fair as for being conscripted into the army.

There is clearly something wrong here and the next section explains what. For now, let us accept that on Klosko's terms; all necessary conditions are fulfilled in examples 1 and 2, yet the packages both contain a DPG that fails to create obligation.

3   The impotence and corrosiveness of unassailability: diagnosing the problems

It has just been established that a package including at least one PPG and DPG can fulfil all conditions set out by Klosko, yet it is conceivable that contributions to the DPG still fail to be obligatory. This section begins by pointing out that unassailability is devastating to the usefulness of Klosko's package argument. It is then argued that unassailability is not helpful anyway, since PPGs fail to transfer obligation onto DPGs.

First, look again at the conditions that a package must fulfil:

(a) Public

(b) Contains at least one PPG

(c) Net beneficial

(d) Fair

A PPG is unassailable: its indispensability entails that any package in which it is included must have a net benefit. In other words, the inclusion of at least one PPG in any given package ensures that the package is net beneficial since any additional DPGs cannot conceivably bring a burden that outweighs the benefit provided by the PPG. Crucially, this means that if (b) is true then (c) is also true.[9] This relationship makes the package argument *technically* too strong since, as long as PPGs are unassailable, (c) is obsolete. This explains why Klosko's conditions are fulfilled even in examples where the DPG is clearly immoral, as demonstrated in Example 2 above.

The other more fundamental problem is that PPGs cannot be used as a source of obligation for DPGs because the relevant moral circumstances are not transferable. To show this, let's revisit the case for PPGs. PPGs provide benefits that are indispensable and this obligates the recipients of the benefit to take on some of the burden of providing the benefit, due to the Principle of Fairness. This argument is convincing, especially in the light of David Estlund's normative consent theory, which claims that non-consent is null in certain moral situations, even if the person whose consent has been nullified has not been given

---

[9]It also follows that if (c) is false then (b) cannot be true. However, this is not a transitive relationship: (c) can be true while (b) is false. This is because one can conceive a state that distributes goods that are net beneficial without providing PPGs (whether the distribution of these goods creates political obligation is beside the point).

the opportunity to consent.[10] Estlund gives an example of a flight attendant giving orders after a crash. In this situation, a surviving passenger, Joe, has a moral obligation to follow particular orders, for example to fetch bandages, even if he refuses to consent.[11] This obligation tracks the obligation due to PPGs; when public benefits are indispensable, non-consent to obligation to contribute is null due to the moral prerogative entailed by the Principle of Fairness.

But if the attendant demands a back rub from Joe, clearly he would not be under a moral obligation to comply and non-consent would be morally permissible.[12] Imagine further that the attendant insisted on a backrub, reasoning that because Joe had followed his previous command regarding the bandage, Joe was obligated to his every command. This reasoning is wrong because non-consent is only null in particular moral circumstances, so one incidence of it cannot be used to arbitrarily transfer obligation to morally non-equivalent circumstances.

The back rub scenario is analogous to Klosko's package argument. It is arbitrary to claim that the presence of a PPG transfers obligation onto DPGs originating from the same source, since the moral obligation a person has to contribute to PPGs is not automatically transmitted to DPGs. Therefore, while the presence of a PPG ensures that a given package is net beneficial, it is a mistake to say that this net benefit is sufficient to create obligation to contribute to all DPGs within the package. Klosko is begging the question by claiming that the presence of PPGs creates obligation to additional DPGs, without explaining how PPGs transfer moral obligation onto DPGs.

4   Repairing the package argument?

As shown in section III, normative consent theory has a near identical structure to hypothetical consent theory, which raises the possibility of conceiving of a PPG-like good that is based on normative consent yet still coheres with the rest of Klosko's package argument. This section explores that possibility, arguing that both the obsolescence of the net benefit condition and the seeming failure of PPGs to transfer moral obligation on to DPGs can be solved

---

[10]Estlund [1] pp. 356–361.

[11]Ibid. p. 357.

[12]This example follows a discussion with Dr Ashley Taylor, University of Sheffield (4/11/15).

by abandoning unassailable PPGs and replacing them with broadly conceived presumptive goods (BCPGs). First, BCPGs are explained and it is shown that they still create obligation, despite lacking unassailability. Second, it is argued that BCPGs give potency to the net benefit condition and are able to transfer obligation onto DPGs.

4.1

BCPGs are distinct from PPGs because they are founded in normative consent rather than hypothetical consent. As such, they are constituted by goods that have a high normative value in whichever political state is being discussed. Before assessing them it is worth noting that BCPGs have an intuitive appeal. PPGs, although conceivable, are really quite rare. Much more common in developed societies are goods that only just fail to be strictly indispensable for almost all citizens, which are captured by BCPGs: a stable economy, an efficient national health service, a fire service, an education system etc.

But what impact does substituting PPGs for BCPGs have on the creation of obligation? The vast majority of the population will use a given BCPG and all these people are obliged to contribute to it due to the Principle of Fairness. But, because BCPGs are not indispensable, obligation stemming from the Principle of Fairness is not general since at least a few people will never use a given BCPG. So the limiting argument is encountered again; the mere imposition of a benefit upon someone (the availability of a health service, for instance) does not create an obligation to the cooperative scheme that supplied it, because no consent was given to receive the benefit in the first place.[13] Therefore obligation to contribute to BCPGs in these cases must be justified.

This can be done by positing situations on close possible worlds that nullify non-consent to receiving BCPGs. If someone uses the BCPG in a close possible world, it is plausible that this is sufficient for acquiring an obligation to contribute to it from the Principle of Fairness, even if they do not require the good in the actual world. To elucidate this idea, consider the example given by Simmons intended to refute Klosko's argument for PPGs. A rural village is suffering from a drought. All the members of the community apart from Jane join forces and dig a trench from a nearby river to supply the village. Jane decides to dig a well in her garden instead of contributing to the scheme and in this way does not need to make use of the common trench of water. However,

---

[13]Simmons [3] p. 34.

in a close possible world Jane does not have a spring under her garden so would be obliged to take advantage of the BCPG constituted by the trench. So Jane is obliged to contribute regardless of the fact she has no need for the trench herself. This solution has an intuitive appeal; for instance healthy people pay taxes to maintain hospitals, with the awareness that it could have been them who got sick. [14] If this argument is accepted then BCPGs, despite lacking the unassailability of PPGs, are still able to generate obligation.

4.2

This section assesses the impact BCPGs have on the package argument. It is first quickly shown that BCPGs revitalise the package argument by allowing the net benefit condition to become potent. It is then argued that BCPGs can also create obligation to non-arbitrary DPGs.

BCPGs are not unassailable. Consequently, the net benefit condition (c) achieves significance that was impossible when the net benefit of a package was guaranteed by the unassailability of PPGs. Because c) is no longer obsolete a given package's benefit can be usefully assessed, rather than being overwhelmed by included PPGs.

The problem of transferring obligation onto DPGs is not solved so easily because obligation to DPGs is still not assured when Klosko's conditions are fulfilled. For example, if the aforementioned National Hero Service is reconceived as a BCPG, it remains the case that the package provided by Country 1 is of net benefit to its citizens; lack of dragons still trumps the nuisance of decorating. So according to the package argument, there should be obligation to decorate one's house, yet there is not. This can be explained by revisiting Estlund's plane crash. The flight attendant has moral authority derived from the normative force of the situation. Joe has an obligation to follow certain commands, even if he knows they are made in error.[15] This is the characteristic of authority: commands are followed due to the status of the commander, rather than the status of the command itself. But crucially the range of the at-

---

[14] An alternative way to create obligation is by showing that in normative terms Jane has acted immorally. Returning to the site of the plane crash, Joe would clearly be acting immorally if he failed to fetch a bandage for the attendant simply because he did not need one himself. Jane is guilty of the same type of immorality by refusing to contribute to the water trench. Therefore, Jane's non-consent is null when it comes to contributing to the BCPG constituted by the water trench, since morally she ought to contribute.

[15] Estlund [1] p. 357.

tendant's authority is limited, as shown by the back rub example. In the case of country 1, the range of the state's moral authority does not reach far enough to create obligation to put Christmas lights up.

For a solution to this problem it is useful to look at the counter-examples that have been offered to Klosko, both in this paper and by his other critics. From Christmas decorations to public address systems, these often share a similar formula: take a benefit that is of such importance there is no doubt that it generates obligation and package it with another quite arbitrary DPG.[16] It appears that arbitrary DPGs often fall outside of the state's moral authority.

However, it is plausible that BCPGs are able to transmit obligation onto *related* DPGs that are within the range of the state's moral authority. For example, if a well-functioning domestic economy is decided to constitute a BCPG, it conceivably follows that a common contribution to road upkeep could be morally obligatory, since a domestic economy cannot function well without sufficient transport infrastructure. With this in mind, let's allow the package argument an assumption: that it need not have to create obligation to arbitrary DPGs.[17] Working on this assumption, the reformed package argument is successful, since if its conditions are fulfilled, moral obligation to contribute to BCPGs and related DPGs is assured.[18]

5   Conclusion

This paper argued that Klosko's package argument for obligation to DPGs fails since PPGs do not seem to transmit obligation to DPGs and unassailability makes the net benefit condition of the package argument obsolete. It was then argued that these problems can be solved by substituting PPGs for BCPGs, consent for receiving the benefits of which is grounded in normative consent theory and an appreciation of close possible worlds. In this way, obligation to a large state, stemming from the Principle of Fairness, can be explained.

---

[16]Nozick [2] pp. 90–96.

[17]Another assumption that might be considered necessary is that it is operating outside extreme circumstances, like dragons, since in these situations the net benefit condition is still very powerful. However, it is actually intuitive that extreme circumstances can lead to greater obligations. For instance, in times of total war, conscription is viewed as a legitimate and obligating policy.

[18]It was pointed out to me at the Sheffield Undergraduate Philosophy Conference by Professor Fiona Macpherson of the University of Glasgow that under this assumption PPGs too would be able to transfer some obligation to related DPGs. This is true, but the state would still be limited in size and the problem of unassailability would remain.

The conception of consent argued for here gives legitimacy to what before was just a gut feeling; that not contributing to the public goods constituted by BCPGs is immoral in some way. For this reason, although the reformed package argument set out here is not flawless, it does represent a theory of political obligation that has quite powerful intuitive appeal.[19]

References

[1] Estlund, D. (2005) 'Political Authority and the Tyranny of Non-consent', in *Philosophical Issues* 15.

[2] Nozick, R. (2003) *Anarchy, State, and Utopia*. Padstow: Blackwell Publishing.

[3] Simmons, J. (2001) *Justification and Legitimacy: Essays on Rights and Obligations*. Cambridge: Cambridge University Press.

[4] Rawls, J. (1971) *A Theory of Justice*. Cambridge: Harvard University Press.

[5] Klosko, G. (1987) 'Presumptive Benefit, Fairness and Political Obligation', in *Philosophy and Public Affairs* 16/3.

---

[19] One problem is that BCPGs cannot create obligation to contribute to non-related DPGs, which might be demanded as a prerequisite for a successful theory of political obligation. As well as this, the limiting argument is still potent if a situation is conceived where in no close possible worlds an individual would take advantage of a certain public benefit. (I actually think this is implausible. For example, imagine an entirely self-sufficient hermit living in a dome. The possible world where he lacked the resources to build said dome is not that far away).

Journal of the British Undergraduate Philosophy Society