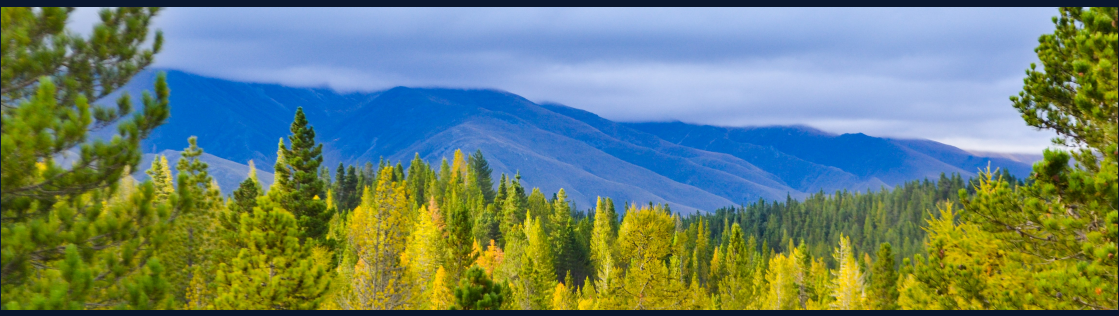
Issue 12 (2) Annual Conference 2018

British Journal of Undergraduate Philosophy



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

Journal of the British Undergraduate Philosophy Society

British Journal of Undergraduate Philosophy

British Journal of Undergraduate Philosophy

The Journal of the British Undergraduate Philosophy Society

Issue 12 (2) Annual Conference 2018

Editor in Chief:	Alexander Raubo
Commissioning Editor:	Edmund Smith
Editorial Officer:	Nathan Oseroff
Manuscript Editors:	Nathan Oseroff Edmund Smith
Peer Reviewers:	Edmund Smith Thomas Alexander Kersley Conor Thompson-Clark Benedetta Delfino Sophie Osiecki Alexander Raubo
President of BUPS:	Edmund Smith

www.bups.org.uk

British Journal of Undergraduate Philosophy

www.bups.org.uk

All rights reserved.
©2018 British Undergraduate Philosophy Society, and the authors.

No part of the *British Journal of Undergraduate Philosophy* may be reproduced or used in any form or by any means (graphic, electronic, or mechanical, including photocopying, taping, recording or by any other information storage and retrieval system) without written permission from the British Undergraduate Philosophy Society, and the authors.

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

Editorial

Writing is a canonical discomfort. Writing philosophy perhaps particularly so. At moments of discomfort it is easy to become self-conscious, to see nothing but what is feared, to resort to an *argumentum ad misericordiam*: 'If you do not like my text, then I am ruined. You do not want me ruined. Therefore, you should like my text'. A version of such an argument crops up frequently in academia, whether it is unsuccessful applicants entreating admissions officers or struggling students pleading for a re-marking. The argument is a bad one since the material conditional does not hold in most cases: your philosophy paper or your application gets unfavourably assessed, yet you are not ruined. Beaten up, shaken, grieving perhaps, but not destroyed. *Was mich nicht umbringt macht mich stärker* – what doesn't kill me makes me stronger – as Nietzsche remarks in the *Twilight of the Idols*. The next piece of work might be better, if only there is a will to improve.

Yet the conventions – of form, of style, of argument – may retain their daunting appearance. In the first instance, unfamiliarity with what the conventions are might be the biggest obstacle. But working on the assumption that they are static and consciously adhered to will quickly make the task even more of a discomfort. Thus the second instance is one of uncertainty about the possibility for effortful compliance. An adequate training seems to be the only way one can simultaneously manage the wilderness of one's own philosophical persuasions and the demands of orderly (and social) convention.

There is the possibility, of course, for a shortcut to virtue. Where conventions are not strongly enforced, one may resort to cynicism. Conventions become mere obstacles to a good life. But simply negating conventions would, in the end, obliterate the possibilities for communication. Since communication is at least one thing that academia, societies, and journals like BJUP aim to facilitate, conventions must be enforced by us. So some discomfort remains. But as per our educational aims, we hope that the discomfort – in writing papers and drafting presentations – will be an instructive and fruitful one.

It is in this spirit of mutual instruction that I am pleased to present the papers selected for this issue of BJUP. Since the publication of the journal nearly coincides with our annual conference to be held at the London School of Economics on the 14–15th of April, I also would like to extend a warm welcome to all the authors who will be presenting and discussing their work. I hope you will find the experience as edifying as in speakers have in past years.

In committee news, this issue marks the last for our fastidious and ever reliable Editorial Officer, Nathan Oseroff. Nathan has been a part of BUPS for nearly four years, two as Editorial Officer. His absence as an active committee member will be deeply

felt. However, we hope that he will continue engaging with the society and the journal more informally when he joins the ranks of our esteemed *emeriti*.

Finally, some acknowledgements. Many thanks to Edmund Smith and Sophie Osiecki who have worked tirelessly with organising the conference and producing the journal. Thanks also to Benedetta Delfino and Conor Thompson-Clarke for their help peer-reviewing. I very much hope you enjoy this issue.

AR

Contents

iii. Editorial

1. Have *Homo sapiens* ever lived in a 'state of nature'?
Dave Knowles

12. Hobbes and Descartes: 'Idea' and The Idea of God
Cody Myers

17. Equality of Welfare and Priority of People
Michael Ratcliff

23. Peoples as a unit of consideration – a Defence for Rawls' Laws of Peoples
Bon Bong

36. Intention: A Missing Element in Frankfurt's Principle of Alternate Possibilities
Shane Neagle

44. Hegel and Belinsky on the theory of 'Reconciliation with Reality'
Colette Olive

Have *Homo sapiens* ever lived in a 'state of nature'?

Dave Knowles
The Open University

Introduction: A Gap

Thomas Hobbes famously depicted what was later to be called the 'state of nature' as 'solitary, poor, nasty, brutish and short' [12, 103]. That sounds like an actual historical entity, but a more widely accepted definition is to be found elsewhere, for example in *A Dictionary of Philosophy* [11]. This defines a state of nature as 'the condition in which people either did live before, or would live *without, making or acceding to a social contract*' (italics added). It is, therefore, a state that may be thought of as either an alleged historical fact or an hypothetical claim about what would be or would have been the case, given certain conditions that may or may not occur or have occurred.'

This longer definition confirms that it is generally accepted that a state of nature is either historical fact or hypothetical. So my first supplementary question is – have *Homo sapiens* ever lived in an actual historical state of nature?

In the long definition above the state of nature is to live '*without a social contract*'. This lack of a social contract is a clear enough statement of principle to be used in testing three types of society.

1. Baseline Zero – a state of nature, (primitive individuals with no social contract or society).
2. Society One – a prehistoric Stone Age group whose social contract status we will assess below.
3. Society Two – a modern hunter-gatherer society with a social contract. Some anthropologists call these either bands (for groups of less than 40) or tribes (for larger groups).
4. Society Three – a nation state with a social contract. This category also includes simpler states that anthropologists sometimes call Chiefdoms, as well as modern states (i.e. the state Hobbes lived in).

In *Leviathan* Hobbes is discussing Zero and Three, as if there is no One or Two, or if there is it is inconsequential. My contention is that this is an over-simplification. I am therefore introducing Societies One and Two because my argument is that, although it is difficult to be certain about the details of life in prehistoric times, the evidence (in A-D below) is sufficient to support the existence of societies with social contracts.

What follows next then are brief summaries of four approaches (A) to (D) used by anthropologists to see into our prehistoric past.

(A) Triangulation – a guide to our evolutionary ancestors:

Homo sapiens share a common (extinct) ancestor with Chimpanzees, which lived 6 million years ago. Chimps live in large groups of up to 30 or 40 individuals some of whom are kin and others non-kin [18, 15]. Modern *Homo sapiens* hunter-gatherers frequently live in groups of 20 to 60 individuals some of whom are kin and others who are non-kin [15, 80]. Using this kind of information evolutionary biologists and anthropologists triangulate between current species back to their extinct ancestors, to establish a rough guide to the extinct species characteristics. On this basis they would say that our predecessors of 6 million years ago probably lived in some kind of extended social groups, because chimps and humans both do so. In which case, their estimate by this approach would be that all our predecessors over the last 6 million years probably lived in social groups that were a mixture of kin and non-kin. Using this method, as a first estimate, it seems possible that humans and their predecessors have lived in Society One for 6 million years.

(B) Archaeological examination of fossilised skulls or parts thereof:

1. We *Homo sapiens* have been around since 250,000 to 200,000 years ago [23, 63], and we have a brain size of 1450cc. (Rather than say from 250,000 to 200,000 every time it comes up [opinions vary], I shall use 200,000 as the more conservative figure.) Neanderthals were also around from about 200,000 years ago, but died out 30,000 years ago.

2. Before us, the species *Homo erectus* was extant back to at least 1,500,000 years ago, and using fossil skulls archaeologists have confirmed it had a brain of about 1100cc [25, 116].

3. The name *Homo habilis*, is often used for the various species that lived roughly between 1,500,000 to 3,000,000 years ago, it had a brain of about 750cc.

4. The skulls are smaller and smaller going back 4,000,000 years to Australopithecus (500cc). Beyond this, at 6,000,000 years ago, the size of the brain of our common ancestor with chimps is uncertain. Modern chimpanzees have a brain size of about 400cc.

There are variants of these species names, but those quoted above will suffice for our purposes. The larger the brain, the larger the neocortex, which governs cognitive processing. (In addition the ratio of the volume of the neocortex to overall brain size in humans is 1:30, double the ratio of 1:60 in other species – and thus doubling the volume of the neocortex.) A large neocortex can keep track of favours returned and favours not returned, the debts and credits of a social group. Anthropologists use neocortex

size as an indicator of brains that developed by being social. Using this method they conclude that *Homo sapiens* and *Homo erectus* were social, but *Homo habilis* was much less so [16, ch. 8]. As a second estimate this approach suggests that humans and their immediate predecessors have lived in Society One for the last *1.5 million years*.

(C) Exposure of former Palaeolithic 'floors':

Stone Age archaeological sites reveal middens of food waste, bones and caches of stone bladed tools and weapons when excavated. These are so large in number that either the *Homo sapiens* who created them occupied these sites for hundreds of years, which they could have done, or they lived in social groups, or both, [15, 255-258]. Based on this kind of material and the size of the exposed Palaeolithic 'floors' it seems probable that Stone Age *Homo sapiens* lived in groups of between 20 to 50 individuals [15, 300]. Although group size seems to have varied, as it does in modern hunter-gatherers, to suit environmental conditions and animal migrations.

Many (cruder) stone tools and blades made by *Homo erectus* have also been found. We can therefore tentatively, bearing in mind the possibility of extremely long occupation of the sites investigated, conclude that humans and their immediate predecessors may have lived in Society One for at least 200,000 years and possibly for as long as *1,500,000 years*.

(D) Survival of serious injuries:

Another indication of the reality of Society One is that fossil skeletons of *Homo sapiens* and Neanderthals have been found showing survival into 'old age' despite signs of earlier very serious injuries [22, ch. 6]. They could not have done so without the care and support of their fellow group members, thus this estimate suggests *200,000 years* of Society One.

Summarising the above, it is rare indeed these days to find an anthropologist who is not of the opinion that *Homo sapiens* have lived in extended (kin and non-kin) social groups for the whole 200,000 years of their existence.

Approach (A) indicates there have been social relationships suggestive of social contracts during all of the last *6,000,000 years*, approach (B) indicates that there have been such relationships during 'only' *1,500,000 years*, and approaches (C) and (D) suggest from *200,000* up to *1,500,000 years*. Taking a rough average of these results, it seems reasonable to say that their have been social relationships suggestive of social contracts for at least the last 200,000 years and probably for the last 1,500,000 years.

We can therefore initially conclude that the first actual historical state of nature was probably to be found over 1.5 million years ago with *Homo habilis*. In which case *Homo sapiens* have never lived in an actual historical state of nature (Baseline Zero).

The state of nature can only be hypothetical for *Homo sapiens*. (The state of nature

may well be historical for *Homo habilis*, but they are long gone, and furthermore, that is not the question.)

Before we dismiss the subject, what is it that caused Hobbes to use his famous phrase or Locke to ponder in *Two Treatises of Government* 'what estate all men are naturally in' [14, 118]? The hypothetical state of nature is not like John Rawls' veil of ignorance [19, 118], namely a pure thought experiment. The state of nature is brought to their minds by what these philosophers think they see. Hence the confusing double definition at the start of this paper that includes both fact and hypothesis. Hence Hobbes in *Leviathan* says 'they live so now' amongst 'the savage people in many places of America', [12, 104]. They are seeing something, so if what they see can't be accurately described as the state of nature (Baseline Zero) then it must be Society Two, before which came Society One.

What society did people live in before the first appearance of the nation state? Based on the evidence in (A) to (D) above the answer is a less sophisticated social contract. And before that? Probably an even less sophisticated one. And before that? We were no longer *Homo sapiens*, so a still less sophisticated *Homo erectus* society. And before … well, you get the idea, my argument is that there was never a time when anything that could be called 'we' didn't live in social groups. So my proposal is that Hobbes and Locke, and Rousseau for that matter, with his no 'need of each other's assistance' [21, 32], were taking as their inspiration primitive societies that looked (to them) as if they had no social contract (Baseline Zero). But in fact they were primitive societies with social contracts (and thus versions of Society Two).

Therefore my second supplementary question is as follows. Are we more likely to be led into error or towards clarity by using a hypothetical state of nature that could be described as jumping over, in a single bound, 200,000 years of human development?

There are both advantages and disadvantages to the hypothetical approach. An advantage is that the simplification involved provides us with a baseline from which to judge whatever actual societies follow the state of nature. A disadvantage is that it can delude us that we *Homo sapiens* once really did have a choice whether or not to enter society. When in reality, if such a choice was ever available (which is highly unlikely), the 'decision' was taken at least 1 million years before we existed.

To help us decide if the provision of a simple (hypothetical) baseline is worth this risk of delusion we need to try and see what we are missing. A first step then would be to take a closer look, if we can, at what went on during that gap of 200,000 years, and if possible identify what we mean by a 'social contract'?

Where to start?

We know that the nation state is a product of human agriculture and industrialisation, which resulted in larger and larger static societies. We also know that this process commenced about 10,000 years ago in the fertile crescent (Iraq), the lower Nile valley (Egypt) and China in particular [9, 27]. So we are going to need to go back at least 10,000 years.

If we went back 60,000 years (some claim 90,000) we would be beyond the limit of grave goods being found with buried human remain, usually assumed to be for use in the afterlife [13, 98]. To ensure that we concentrate on the evidence for social interactions and steer clear of the possible confusing factor of religion, leaving that for later consideration, we should therefore go back to at least 100,000 years ago.

The other advantage of 100,000 years ago (again here, some say 120,000) is that at this time all the *Homo sapiens* in the world were still living in Africa [22, ch. 2]. Anthropologists favour 100,000 years ago as the time we *Homo sapiens* probably started to come out of Africa and spread around the world. So any social contract we derive will be a basic one that applied to all humans the world over, because they will be from a time when we were all still living together in Africa. And as things turned out were about to burst out into the rest of the world – and then diverge and evolve into many different cultures.

In view of the above 100,000 years ago in the middle Palaeolithic or Middle Stone Age seems to have many advantages as a place to locate our enquiry. In addition to the above, also because we *Homo sapiens* have spent 95% (or 99%, depending on when you start counting) of our existence living in small hunter-gatherer groups, and only 5% (or 1%) living in nation states (Panter-Brick et al, p. 4).

What were *Homo sapiens* like 100,000 years ago, nice or nasty, easy to get along with or not, trustworthy or not? Trusting only within their small group? Less or more cruel than the Romans? How did they treat people from other groups? Did they eat them – were they cannibals? Did they sacrifice even their own children to the Gods?

Well, as mentioned above, we can answer that one, because 100,000 years ago they don't appear to have had any Gods. So it seems unlikely that there were sacrifices being made of valuable food or lives to propitiate the Gods. As to the rest of what they were like, God knows – if you'll forgive the pun, we really don't know much.

A Primitive Social Contract (Society One)

We know one thing though. In order to have lived in groups for at least 200,000 and maybe as much as 1,500,000 years we must have been willing to co-operate with other

members of the group we lived in, because if this were not the case then clearly there would have been no group. The fact that there were groups that cohered for generation after generation confirms that there was 'a society with a social contract' of some kind. There can be no co-operative if there are no individuals able and willing to co-operate.

No co-operation would equate to no teamwork, but what was the level of such Stone Age co-operation? Chimpanzees have a set of relationships between the individuals that make up their groups that looks like co-operation – sometimes [10, 222].

We can look at ourselves as we are now, co-operating but sometimes falling out with each other, and compare this behaviour to chimpanzee groups, who frequently fall out with each other, but sometimes co-operate. In doing so we can reasonably suppose (by triangulation) that *Homo sapiens* of 100,000 years ago probably behaved more towards our end of this spectrum than chimps seem to exhibit today.

That's all we can say I think, but it's an important starting point.

Despite this co-operation we can assume that there will also have been a pecking order within groups of *Homo sapiens*, because such rankings are found in present chimp and present human groups, as well as the chickens that the term is derived from. There is no reason to think that there won't have been feuds, simmering anger, likes and dislikes, insults and slights between some of the individuals who made up the team [5, 252].

At the top of the pecking order there will have been an Alpha Male who will probably have demanded pack co-operation. But there are limits to his ability to enforce it. For example no matter how overbearing a male lion is, lions do not hunt in packs the way wolves and wild dogs do. Living organisms have developed such that group life comes more 'naturally' to some than to others. Alternatively our willingness to co-operate may have been an extension of our feelings for kin, based on smaller nuclear family group. Either way, or another way, we can leave the mechanism of co-operation's first appearance to the Evolutionary Biologists [5, 312]. All we need to know for now is that to hold a group together co-operation must have existed. There can be no co-operative if there are no co-operators.

Co-operation in Action

To gather food and then share it is one level of co-operation. But *Homo sapiens* don't seem to have just gathered, they hunted as well, for small and middle sized game, but also it would seem, from their first appearance in the fossil records 250,000 to 200,000 years ago, for big game as well [5, 146]. Hunting potentially aggressive big game with only stone bladed spear heads fastened to wooden shafts, or a point formed on the end

of the wood itself, is an enterprise that is extremely dangerous. (Bows and arrows were not invented until about 20,000 years ago [13, 107-108]).

Despite this danger caches of the bones of large animals, even ones as aggressive as rhinoceroses have been found at certain sites [3, 146]. It is astounding that prehistoric humans could prey on such creatures. It is unbelievable that they could do so without co-operation. Hunting creatures so much larger, faster and aggressive than oneself requires formidable levels of co-operation. Co-operation, planning, communications and *trust* in ones fellow group members.

Chimpanzees (bbc.co.uk/earth/story/20150728) co-operate to hunt creatures that are smaller than themselves, colobus monkeys being their favourite prey. The human step up from small and middle sized game, must have been rather like chimps stepping up to trap, kill and eat a large silverback gorilla. Hard to imagine, but that's what *Homo sapiens* seem to have been doing from their first appearance on the planet, and in doing so, in due course became the planet's top predator (nothing to be proud of necessarily – just a fact).

In attempting to drive such animals into the waiting arms and spears of other hunters, or over a cliff, or into a swamp for 'easy' killing, each hunter must have been dependent (sometimes for their life) on the hunters on each side standing firm. That kind of co-operation means that you must stand firm and trust that others will also stand firm. To fail to stand firm would be an offence against ones fellow group members, to fail to stand firm could be said to be to fail *to do your duty*.

Fairness and Selfishness

What human groups primarily undertake, is support for each other, for some group members to sometimes make sacrifices (of food for example) on behalf of others. And for those others to in turn make similar sacrifices when for whatever different circumstances the boot is on the other foot and the need is in the other belly. Therefore we can be reasonably sure that Middle Stone Age individuals will have lived within social groups in order to have better food, shelter and defence, and thus a better chance of survival than they would as individuals alone.

Humans have big brains, but small teeth. We can run long distances, but have slow short distance speeds. We can make weapons and form teams, but have no claws. So improved survival chances are very difficult to reach without an accommodation with the other humans in the team. An accommodation made to produce sufficient *mutual assistance* to achieve *individual satisfaction* of wants.

This accommodation asks of us that we co-operate and be unselfish towards others on our team. If we humans were to make our living alone, as many big cats do (and as

the 'state of nature' assumes that we once did), the concept of selfishness would have no meaning because selfishness, to look after oneself first, second, any left self again, is necessary to survive. What is thus entirely permissible for the individual, living alone in a state of nature, has to be moderated when living as a member of a group.

Thus it is only after we became group dwellers that selfishness is frowned upon. That is when we are first called upon to 'think of others', rather than pursuing as our sole one and only focus the single minded gratification of our own needs and desires. To be unselfish is to be 'fair' towards others in our social group. And as alluded to above, fairness to others needs constant monitoring, careful assessment and re-assessment, because modifications in our responses to others are required dependent upon how others behave towards us.

I am not suggesting that 100,000 years ago our predecessors understood these lines of reasoning. I am arguing that, as we see with chimps nowadays, a leading adult will sometimes smack a youngster (bbc.co.uk/earth/story/20150728) who is trying to grab too much food. There is no need even for words, the message gets across, don't be selfish, don't be greedy, know your place. The message is likely to be twofold, it is also that I am bigger and older than you, and so if I want to I can be greedy, but you can't – in other words the Alpha response. Play nicely little one. This is an idea that is easy enough to get across, even with a non-verbal or limited vocabulary. Moderate your looking-solely-after-self mechanism or as we say now 'do as you would be done by'.

Reciprocal Altruism

The process of evolution (often portrayed as remorseless selfishness) can explain and engender co-operation via 'reciprocal altruism' [24].

Let's call the choices we have out in a state of nature as either to be selfish or to co-operate. We see it in the wolf pack gathered around the carcass once the hunt has succeeded – should I observe the pecking order, or fight for a higher place within it? One can convert this longstanding situation into a modern computer programme, as a game, with real human beings trying different behaviour pattern strategies [1].

If you are selfish and your opponent gives way (co-operates), then you take all you want and your opponent, especially when times are hard, gets nothing. When set up as a computer game you get five points and your opponent gets zero, hooray, you've won! Not so fast, this game is different to any other game you've ever played – it never ends. In the next round of the game, in the wild at the next hunt a few days later, your partner/opponent last time is still on the team, he/she is hungry and knows your face, he/she is human and has a good memory for faces.

This time there's no co-operation. In fact you fight each other almost to exhaustion.

Now neither of you eats much, because of your mutual opposition. Let's give you one point each, you ate a little more than the one who lost last time, but at the price of injury, exhaustion, pain, wasted time while others of the pack ate well, or worse, neighbouring packs ate well.

Next time, at the next round of the game, you both decide to go for co-operation, partly because you've now tried out each other's strength and realise you're almost equal, and partly because you're still too hungry from the last round to waste time fighting again. Your co-operation is successful, it is a good hunt, you both eat well.

You didn't get as much as when you failed to co-operate, and got away with it, but you're still well ahead of the game. Well ahead of the time you fought yourselves to a standstill. In this round you both get three points, which while not as good as five, is a damn sight better that one or zero.

This analysis of co-operation in evolutionary terms explains how reciprocal altruism can arise from instinctive selfishness, and thus be evolutionary stable, or pretty close to it. By following a simple strategy of co-operating on the first round followed by doing on the next round whatever one's partner/opponent did on the last, tit for tat, most of the time this will stack up more points than other behaviour patterns [8, 212].

It thus seems that tit for tat could be an initial unspoken contractual agreement, that then becomes the central principle of the *social contract* we've been using to define Society One and Two. You scratch my back and I'll scratch yours – an eye for an eye, the simplest contract in the world. And still in worldwide use to this day. No need to sign on the dotted line, just act in such a way that you are seen to be 'firm but fair', so that it is understood that you will return like for like, that you 'will do as you are done by'.

Reciprocity, *quid pro quo*, is the golden rule common to all religions and cultures – we all share it. A hell of a coincidence? Not if it's been bred into us via endless cycles of reciprocal altruism, generation after generation after generation.

Conclusion

To the question, have any *Homo sapiens* ever lived in a state of nature, we previously gave a preliminary 'no', so this is now confirmed. No because instead of finding Hobbes' 'savages' in a state of nature we have found social contracts based on reciprocity. There is thus sufficient evidence to show that humans have co-operated with each other for thousands upon thousands of years, and they have done so without a Leviathan to frighten them into it.

Which leaves us with the hypothetical version – as a baseline. We can see that this

version does have a limited use, as long as we keep in mind that even hypothetically it is more accurate to apply it to our long lost predecessors than directly to ourselves. With this proviso it can function as a base from which to measure firstly the various primitive societies (versions of Society One and Two) found at different modern and historical locations. And secondly a baseline for the complexities of modern nation states (versions of Society Three). It's use as a base is worthwhile as long as we maintain awareness that the 'state of nature' is probably a better summary of the life situation of a big cat hunting alone than of any hominid of the last 1.5 million years.

What we seem to have missed by Hobbes et al jumping over 200,000 years of human development is that the basis of all human social contracts is reciprocity. The golden rule version of reciprocity is subject to some criticism by various writers. Perhaps though it is time to re-evaluate it in a prehistoric context as the first principle of the social contract, rather than the whole story?

If this is the case then work will be required on the details – what other parameters might have sat alongside reciprocity in prehistoric times? Can they plus reciprocity be the basis of an original social contract?

As previously mentioned,'To fail to stand firm would be an offence against a fellow group member, to fail to stand firm would be to fail to do your duty'. An extension of this would be 'To fail to do your duty and stand firm would be an offence, and by the same logic it is also an offence to kill, injure, or steal from a fellow group member'. When do such offences become criminal offences? Because when they do this is no longer just a simpler society than a nation state, it is a society that probably contains the origins of morality.

References

[1] Axelrod, R. (1984) *The Evolution of Co-operation*, London, Penguin Books

[2] bbc.co.uk/earth/story/20150728

[3] Berkoff and Pierce, (2009) *Wild Justice*, Chigago, University of Chicago Press

[4] Boehm, C. (1999) *Hierarchy in the Forest*, Cambridge Mass, First Harvard University Press

[5] Boehm, C. (2012) *Moral Origins*, New York, Basic Books

[6] Brody, H. (2001) *The Other Side of Eden*, London, Faber and Faber

[7] Champion et al, (1984) *Prehistoric Europe*, London, Academic Press

[8] Dawkins, R. ([1976] 1989) *The Selfish Gene*, Oxford, Oxford University Press

[9] Dawkins, R. (2005) *The Ancestor's Tale*, London, Phoenix Orion Books

[10] de Waal, F. (1982) *Chimpanzee Politics*, New York, Harper & Row.

[11] Flew, A. (Ed.) (1979) *A Dictionary of Philosophy*, London, Macmillan

[12] Hobbes, T. ([1651] 2017), *Leviathan*, London, Penguin Classics

[13] Jordan, P. (1999), *Early Man*, Stroud, Sutton Publishing

[14] Locke, J. ([1690] 1964) *Two Treatises of Government*, London, Everyman Paperbacks

[15] Lee and De Vore, (1968) *Man the Hunter*, New Brunswick, Aldine Transaction

[16] Maslin, M. (2017) *The Cradle of Humanity*, Oxford, Oxford University Press

[17] Panter-Brick et al, (2001) *Hunter-Gatherers: An Interdisciplinary Perspective Cambridge*, Cambridge University Press

[18] Pond et al, (Ed. J. Silvertown), (2008) *99% Ape*, London, Natural History Museum

[19] Rawls, J. ([1971] 1999) *A Theory of Justice*, Oxford, Oxford University Press

[20] Ridley, M. (1997) *The Origins of Virtue*, London, Penguin Books

[21] Rousseau, J. J. ([1757] 2013) *A Discourse on Equality*, Aziloth Books.

[22] Stringer, C. (2011) *The Origin of Our Species*, London, Penguin Books

[23] Stringer & Gamble, (1993) *In Search of the Neanderthals*, London, Thames and Hudson

[24] Trivers, R. L. (1971) 'The evolution of reciprocal altruism', *Quarterly Review of Biology* 46

[25] Vialou, D. (1998) *Our Prehistoric Past*, London, Thames and Hudson

Hobbes and Descartes: 'Idea' and The Idea of God

Cody Myers
University of Michigan

Descartes and Hobbes have different definitions of 'idea,' which lead them to disagree whether there is an 'idea of God'. Descartes claims there is such an innate idea, whereas Hobbes denies there is any idea of God. I first consider Hobbes's and Descartes' definitions of 'idea,' after which I show how these definitions relate to an idea of God, concluding with objections to, and implications of, Descartes' account. I show that Descartes' argument for the idea of God need not preclude an infinite regress, given his definition of 'cause,' nor does his account of truth and falsity cohere with such an idea, as Descartes can only obtain a negative idea of God. Ultimately, I conclude in favor of Hobbes's claim that we can have no idea of God.

Definitions of 'Idea'

Descartes defines an idea as "everything that is the form of some perception" [1, 110]. The inference made from a chain of reasoning may be an idea, as well as an image of some (possibly) sensed object, since both are "immediately perceived by the mind" [1, 106]. For instance, I am aware that I reason, or that I feel warmth, or seem to see a 'tree', etc. An idea is not limited to the mental picture I have of something. This entails a distinction between different "components [of ideas] … called volitions … emotions … [and] judgments" [2, 16]. Therefore, with regard to affects, for example, an idea is an image of something, with the added component of fear (or any other emotion). However, there may also be ideas which are more strictly mental images of something.

Hobbes, however, has a different definition of 'idea'. An idea for Hobbes is "a representation or appearance of some quality … of a body without us" [3, 13]. That is, an idea is an *image*, either imagined or sensed, and there are no components apart from this image to be added to and differentiated from the mental image. Consequently, Hobbes does not view affects–such as fear–as being anything other than the image of what one fears "combined with the effect that such an idea produces in the heart" [1, 107]. That is, affects are simply images plus the physical, bodily effects caused by that image; Descartes, however, argues that imagining or perceiving an image and "simultaneously fearing it is different from merely seeing it" [1, 107]. Therefore, Descartes' 'idea' may be a sort of compound of different components, whereas Hobbes's idea is an imagined or sensed image of some external object, without additional mental aspects.

The Idea of God

The two different notions of idea, then, lead to disagreement over whether there is an *idea of God*. Descartes claims the idea of God is "real and positive" [2, 28], whereas Hobbes claims "there is no idea in us of God" [1, 106]. Because Descartes understands an idea to be whatever is "immediately perceived by the mind" [1, 106], he argues, contrary to Hobbes, that it's possible to have an idea of God "by means of reasoning" [1, 109]. This reasoning has first to do with Descartes' account and definition of causes and effects.

Descartes reasons that, because "there is at least as much in a complete efficient cause as in its effect" [2, 18], there must be some first cause in which all things are contained "either formally or eminently" [2, 18]. Because it's impossible for something to come from nothing, a cause must have everything in it that its effect has, otherwise it would violate this necessary truth. Descartes also says that the causal chain "can't go back to infinity … [but that] we must eventually arrive at a primary idea" [2, 19]. The primary idea is God, and it is from God that all other causes are effected, and thus God must contain all things.

Descartes argues that this idea of God cannot have come from ourselves, because "I who am finite would not have the idea of infinite substance in me unless it came from a substance that really was infinite" [2, 21]. That is, since we are finite beings, that which is infinite must be outside of us, since a cause must have "at least as much in [itself] …as in its effect" [2, 18]. Descartes concludes that there must be an infinite substance, otherwise we can't conceive of ourselves as finite, "deficient and imperfect … [without] something more perfect to use as a standard of comparison…" [2, 22]. The infinite substance, God, is that by which we conceive of ourselves as limited, according to Descartes.

Objections to Descartes' Idea of God

So far, there may be two different objections to Descartes' account of the idea of God. First, if Descartes claims that a cause must have "*at least as much*" [2, 18, emphasis added] in itself as in its effect, then his claim that there must be a primary cause is not a necessary implication. He says there may be no infinite causal chain, but that "we must eventually arrive at a primary idea" [2, 19], this idea being God. However, if a cause may have "as much" [2, 18] in itself as in its effect, there doesn't appear to be any reason why this chain cannot extend to infinity, in which case there might be said to be no first or primary cause which is itself infinite, but a chain of causes and effects that is infinite. If a cause is permitted to have "as much … reality" [2, 18] as in its effect, it seems there's no reason to suppose a sort of causal hierarchy, wherein there are 'greater'

and 'lesser' causes and effects. It seems logically permissible, given Descartes' definition of 'cause', that all causes contain only as much as their effects, and vice versa. Therefore, that which is infinite might not be an original, perfect cause (God) from which finite things with "less reality" [2, 18] emanate. What is infinite might be that endless process of modification of 'cause' into 'effect,' where there is an equal exchange of 'reality' from one to the other, since a cause (and so all causes) might only contain exactly as much as its effect. What is infinite might be the process of one substance modifying itself, of which each of us are modifications. If this is so, it is granted there is something which is infinite; however, this conception of the 'infinite substance' would not entail God's being more real or more perfect than any of its allegedly finite modifications. If there exists an infinite substance which infinitely modifies itself, then that entails such modifications contain no less reality than anything else. It may be we are no less perfect than God, because we are modifications of God.

However, it may be argued Descartes conceives of perfection and infinitude *at least* not in the only way possible given his definition of 'cause'. Though a cause must contain at least as much as is in its effect, there is no reason to suppose that there is ever more than such an equal exchange of 'reality' between cause and effect. Since this is allowed by the definition, it does not appear Descartes can claim there must be no infinite regress of cause and effect.

Furthermore, one possible argument against Descartes' depiction is as follows: infinity is not the "standard of comparison" [2, 22] by which we measure our "deficient and imperfect" [2, 22] natures, but rather that it is the endless extension of finite boundaries from which we derive the notion of 'infinity' (as suggested by the prefix 'in-'). Hobbes claims as much in his objections to Descartes' third meditation, when he says that, by 'infinite,' he understands "something whose boundaries or extremities I cannot conceive or imagine without imagining still more extremities beyond these" [1, 109]. What is 'infinite' is our reason pushing conceivable limits indefinitely. This, however, is an idea of our own finitude, endlessly extended. And, though Descartes may claim that the awareness of this endless extension is an 'idea', it is not clear how he can claim it is a "real and positive" [2, 28] idea. Rather, this reasoning suggests that whatever we call by the name of 'God' is a *negative* idea, an idea of everything we are not, a 'not-us' idea, because of its being defined in comparison with our finitude. It appears, then, that our finitude is the really standard of comparison by which we derive 'infinity'. We are aware only of our own traits (finitude, limitedness, etc.), and the term 'God' only negates these terms, and gets its traits by the negation of our traits. If so, 'God' and all the traits attributed to 'God' are no more than negative terms. Hobbes argues that those names "called *Negative*; which are notes to signifie that a word is not the name of the thing in question …be not names of any thing" [3, 30, emphasis in original]. Among these negative terms he includes "infinite" (ibid.). Though Descartes, in the third meditation, says "I shouldn't think that, rather than having a

true idea of infinity, I grasp it merely as the absence of limits … [and that] there is more reality in an infinite than in a finite substance" [3, 21], he supports this with the claim that infinity is the standard by which we measure our finitude, which claim has already been shown to be dubious.

Descartes, however, claims that by "an act of understanding, and, by the indefinitely extending it… [one] form[s] an idea of the divine act of understanding… And [that] the same applies to the rest of God's attributes" [1, 110]. Because we perceive that we can endlessly extend the boundaries we necessarily conceive of, we may understand that there is something which exceeds our limits. But again, this seems to be a negative idea. We are only aware of our own understanding in this, and Descartes himself says "[w]hen it is asserted that God is inconceivable, this is understood with respect to a concept that adequately comprehends him" [1, 111]. But, how can it be said that an idea of X is an idea of X, if it doesn't capture all those things that make X an X? If what it is to be an X is to have traits a, b, and c, and our idea, say, captures only a and b, then it's not an idea of X, but of something else. Nor does it help to define a term by reference to things it is not, i.e., not finite, not limited in understanding, etc. So, it is unclear how Descartes can maintain we have an idea of God which is "real and positive" [2, 28].

Setting aside the nature of the idea of God, I think Descartes' claim that this idea is innate may be problematic. He claims, in response to Hobbes' objections, that "when we assert that some idea is innate in us, we do not have in mind that we always notice it… but only that we have in ourselves the power to elicit the idea" [1, 111]. This power, though, seems to be something like a faculty, if we consider that the idea of God is no idea, but really only an endless operation of the reasoning by which we extend boundaries.

Idea of God Incompatible with Account of Falsity

If, as Descartes says, our ideas can neither be true nor false, "[i]nsofar as the ideas of things are just modifications of thought" [2, 18], and if error consists in "judgments dots assuming that the ideas in me have a similarity or conformity to things outside" [2, 16], then the 'idea of God' seems to destabilise Descartes' account of truth and falsity laid out in the fourth meditation. Descartes says that falsity and error consist in "participat[ing] in nothingness or the nonentity" [2, 28], and that "error is a lack" [2, 28]. This participation in nothingness occurs when the will, which is not "restricted in any way" [2, 29], is not kept within "the same bounds [of the understanding]" [2, 31], but is extended beyond "to that which [one doesn't] understand" [2, 31]. It's unclear, then, how Descartes can claim that the idea of God is "real and positive" [2, 28] if such an idea is also only an awareness of our own limitations.

The judgment that there is a God may not only be a "formal falsehood" [2, 20], but may also be a "material falsehood" [2, 20]. An idea may be materially false "when it presents what is not a thing as though it were a thing" [2, 20]. Descartes uses the example of heat and cold to illustrate this: "[T]he ideas that I have of coldness and heat are so unclear… that I can't tell from them whether coldness is just the absence of heat, or heat just the absence of coldness, or both are real qualities, or neither is" [2, 20]. However, Descartes claims that "the idea of God cannot be said to be materially false… since [it] contains more subjective reality than any other idea" [2, 22]. This, though, is supported again by his claim that the infinite is the standard by which we measure our own finitude. But, if we derive some incomprehensible 'idea' of 'infinity' by an endless extension of our own finitude, it seems more reasonable to assume that 'infinity' is something like a negative term that tells us nothing beyond our own limitedness. Furthermore, if the idea of God is a negative idea, and the term 'God' is a negative term, then it follows that the idea of God is materially false. If, to posit the existence of a thing corresponding to a negative term is to "[present] what is not a thing as though it were a thing" [2, 20], then any 'idea of God' is a materially false one.

Conclusion

As says Hobbes, "If there is no idea of God (and it has not been proved that there is one), this [Descartes'] entire inquiry falls apart" [1, 107]. Though Descartes wants to maintain that such an innate idea exists, I don't think the arguments he gives in support of his claim stand up to scrutiny. In conclusion, I don't believe Descartes' argument necessarily excludes the possibility of an infinite regress with regard to causes, nor do I think it's possible for an idea of God to be "real and positive" [2, 18], given the account of falsity that he offers. So, I believe Hobbes is correct to claim that there is no idea of God at all.

References

[1] Descartes, Rene. (2006). *Meditations, Objections, and Replies*. Ed. Roger Ariew and Donald A. Cress. Indianapolis, IN: Hackett Pub.

[2] Descartes, Rene. (2001). *Meditations on First Philosophy*. Ed. Ronald Rubin. Claremont, Ca.: Arete.

[3] Hobbes, Thomas. (1991). *Leviathan*. Ed. Richard Tuck. Cambridge: Cambridge University Press.

Equality of Welfare and Priority of People

Michael Ratcliff
University College London

I believe that equality of wellbeing[1] is instrumentally good, insofar as it is achieved by prioritising the wellbeing of those worse off in absolute terms. As such, I take a Prioritarian view, in opposition to telic egalitarianism (TE), which maintains that equality has intrinsic value. I will begin by outlining some differences in approach between these two theories, before giving an account of TE. I will then examine the Levelling-Down Objection (LDO) to TE and consider whether a pluralist form of TE avoids this concern, before using the Person-affecting View (PAV) to demonstrate that it does not. After considering and resolving a potential issue with the PAV, I will outline Prioritarianism, and in doing so demonstrate how it avoids the LDO. This will lead me to conclude that Prioritarianism is a more convincing theory than TE, meaning that what really matters is that some people are badly-off in *absolute* terms, rather than that they're worse off *than other people*.

Differing Approaches

Two important differences in approach between TE and Prioritarianism are the distinctions between comparative/non-comparative justice, and equality as an intrinsic/instrumental good. TE takes the former side of both distinctions, whilst Prioritarianism holds the latter. As such, TE holds that unjust treatment of a person depends on how they are treated *with respect to* other people [1, 207]: inequality is unjust because people have *different* levels of wellbeing. Contrastingly, Prioritarianism takes a non-comparative approach. Unjust treatment of an individual purely relates to facts about that individual, and not how their treatment compares to other people. In this way, inequality itself isn't the concern, although by helping those worst-off in *absolute* terms, this would tend to produce a largely equal society [1, 207]. As such, Prioritarianism takes equality to be an *instrumental* good: working towards equality is good as a means to prioritise benefitting the worse off. Contrastingly, TE takes equality to be an *intrinsic* good: equality is good in-and-of itself.

[1]For the purposes of this essay, 'equality' will be used to refer purely to equality of wellbeing.

Telic Egalitarianism: Explanation and Justification

With these distinctions in mind, TE's core claim becomes clear, and can be summarised by what Parfit calls the Principle of Utility:

> "It is bad in itself if some people are worse off than others" [1, 204].

With the exception of inequality that comes about from some fault or choice by the worse off person. As such, TE asserts that the ideal society would be totally equal; or at the very least, would only allow for the slight inequality that can arise from poor decision-making. Furthermore, due to the intrinsic badness of inequality, the scope of redistributive action needed to ensure a just society is incredibly broad. Given that *all* inequality is bad, inequality derived from good luck wouldn't be the only sort that needs redressing; rather, inequality arising from differences in "natural endowments" (intelligence, athletic prowess, etc) would also need eliminating [1, 208]. As such, inequality arising from smart decisions made by intelligent people would also be unjust under TE - although this may be impossible to totally eliminate, and if less intelligent people were in some way given extra help to balance out these endowments, then any resulting inequality from decisions would be justified.

Though this system may seem unreasonable at first, its intuitive appeal becomes clear once applied to examples like the Divided World scenario. Let humanity be divided in two, with no communication or awareness between the two sides. Now, let one half have wellbeing of 200, and the other half 400. This scenario seems worse than one where everyone has 300. However, it's difficult to see how this could be because of the effects of the inequality, rather than just some intrinsic injustice present in the inequality. As the two societies are divided, the better off half has no social or political superiority over the worse off half. Therefore, it seems the only justification for scenario 2 over scenario 1 is that it is more equal.

The Objection: First Attempt and Response

The most potent response to TE is the LDO, which aims to contradict the claim to TE's intuitive appeal. The objection points out that, if inequality is intrinsically bad, then any increase in equality must be in some way good, even if this is achieved by inflicting harm. This is best illustrated by returning to the Divided World. Let some natural disaster sweep across the world, lowering the wellbeing of both the privileged and less privileged halves to 50. Everyone's wellbeing has as such been depleted, but in such a way as everyone is now equally well off. The Telic Egalitarian seems constrained to admit that there is some morally good component to the natural disaster, as it has achieved the irreducible moral good of equality. However, to claim that inflicting a

mass harm on the entire world is a morally good thing seems counterintuitive to the point of preposterousness.

Though the Telic Egalitarian must 'bite the bullet' to some degree here, they may still respond with a pluralist account of TE). This would mean combining the principle of equality with some lexically inferior principle of utility, such as "It is in itself better if people are better off" [1, 205]. Although this principle must presumably carry less weight than the principle of equality (otherwise the theory would be a form of utilitarianism rather than egalitarianism), such an approach may allow TE to escape the LDO. Pluralist TE could assert that, whilst the natural disaster does have some morally good component, it is on balance still bad, as the overall loss of utility was so great as to outweigh the increase in equality. This kind of approach is of course imprecise, as it seems unclear how much utility would be needed to outweigh a particular increase in equality. Nevertheless, it seems to do a better job than conventional TE of explaining our intuitions, insofar as it deems wellbeing splits of 300/300 better than 400/200, which is now better than 50/50.

The Objection: Second Attempt, Problem, and Solution

Thus, in order for the LDO) to be effective, it must demonstrate that there is "*nothing* good" about achieving equality through harms. This may be achieved by appealing to the PAV, which asserts that "One situation *cannot* be worse (or better) *in any respect* if there is *no one* for whom it is *worse* (or better) *in any respect*" [2, 256]. In the Divided World, the natural disaster fails to improve the lot of anyone in any way: as the two sides were totally disconnected, achieving equality failed to give the less privileged side any greater socio-political clout, nor did it improve the material welfare of anyone. Thus, according to the PAV, it has no good component. The PAV also explains why a Divided World where both sides started at 100 before one upgraded to 110 has no morally bad component, whereas TE is constrained to condemn such a move. Such an approach seems itself intuitive, as well as explaining our intuitive response to certain scenarios. Morally good and bad actions seem to require a subject: giving money to someone in need is good because someone is being helped, whilst hitting a stranger is wrong because someone innocent is being harmed. But, if TE is correct, it must be possible for a morally good action to have no recipient of this goodness. Although the natural disaster itself has subjects (namely everyone), the moral *goodness* of the action has no recipient, as everyone ultimately receives a harm.

However, the PAV seems to entail concerning conclusions in other areas. In particular, the view that there are no impersonal Goods or Harms (actions which are good or bad without being good or bad for any specific person) may prevent us from criticising what appear to be clearly immoral actions. For example, let there be two friends,

who decide to make a joint bet on a horse, each putting forward 5 pounds. However, the friend responsible for physically placing the bet (Friend A) decides to secretly bet on a different horse instead. This horse in fact wins, and Friend A keeps the money for herself. Friend A feels no guilt over this, and Friend B never finds out: as far as he is concerned, he placed a bet on a losing horse. Furthermore, Friend B hasn't lost anything due to Friend A's dishonesty: either way, Friend B would lose the 5 pounds. The PAV seems to assert that there is nothing wrong with Friend A's behaviour, as no-one is worse off as a result of her actions. Nevertheless, this response clashes with our instinct to condemn such dishonesty as immoral. Thus, this example poses a clear problem for the PAV, which aims to replicate and explain our intuitions.

However, I believe this concern can be avoided. We may accept that some actions can be harmful to a person without decreasing their welfare. The dishonest friend example is intended to serve as a counterexample to the PAV's contention that there are no impersonal Goods or Harms; in essence, Friend A seems to have committed some morally bad act without harming Friend B. Thus, if it can be shown that Friend B has in fact been harmed, then the PAV will survive this criticism. If this is the case, then the PAV will be able to criticise TE – which is concerned purely with welfare, and so cannot assess harm via any other method than welfare – but avoid condoning other seemingly-impersonal Harms.

It can be argued that the two friends have entered into a kind of unspoken contract, which hinges on the notions of respect and honesty. This can be seen in their betting interaction, but applies more generally to their friendship. Friend B agreed to participate in the bet on the understanding that he would receive a share of the winnings if his £5 was used to acquire these winnings. Similarly, two friends agree to spend time with each other on the understanding that they appreciate each other's company, and don't have some ulterior motive. Someone who breaks this hypothetical pact is considered to be a 'bad friend': by going against the spirit of the friendship, they are taken to be doing something wrong. If we are to take friendship as a good thing, then to damage the friendship constitutes a harm. It also seems plausible to assert that Friend B is the recipient of this harm: it is his time, money, and trust that are being exploited. As such, regardless of Friend B's awareness of the changed bet, he has been disrespected through the breaking of the contract. Thus, it seems that the dishonest friend example in fact isn't a case of an impersonal Harm. Therefore, the PAV is able to criticise Friend A, and as such continue to successfully replicate our intuitions.

The Other Approach

A different approach to equality is Prioritarianism, which asserts that "Benefiting people matters more the worse off these people are" [1, 213]. As has been discussed above,

this is a non-comparative theory which takes equality to be an instrumental good. As such, Prioritarians don't advocate redistribution of welfare because inequality is in itself bad; rather, they do so because less well-off people have priority when it comes to maximising welfare. Any increase in equality that results from this is a by-product of the real moral good, prioritising the worse off.

Similarly to TE, Prioritarianism is at its strongest when given a pluralist formulation, wherein a lexically inferior principle of utility combines with the principle of priority given above. As such, a better version of Parfit's assertion would attribute weighted priority to the worse off – priority which can be outweighed by a large enough degree of utility. This alternative is provided by Roger Crisp [3, 752-3]. The alternative to this, absolute priority, would lead to the somewhat distressing scenario that a fractional improvement in the life of the worst-off person would take priority over a vast improvement to the lives of everyone else. It is also worth emphasising again – as was mentioned briefly in the section on the different approaches – that Prioritarianism is interested in those worse off in *absolute* terms, rather than those who are worse off compared to the majority [3, 751]. It doesn't matter that these people are worse off than other people; it simply matters that they are badly off.

Returning to the original Divided World example, Prioritarianism would produce the same intuitive outcome as TE, namely that both sides at 300 is preferable to the 400/200 split. However, its justification would be different. The importance of improving the lot of the less well-off takes priority over those who are better-off in absolute terms. Furthermore, Prioritarianism is able to dodge the LDO. As equality is an instrumental rather than intrinsic good, Prioritarians see no good component to an act that achieves equality through harm. In fact, if this act was to decrease the well-being of the worse off, the act would be entirely morally reprehensible. Thus, Prioritarianism produces the same intuitive appeal of TE, whilst also avoiding the counter-intuitiveness of the LDO.

To conclude, the comparative approach of TE does a good job of reproducing our intuitions on a surface level. However, in scenarios where equality conflicts with our interest in maximising welfare, this appeal collapses. Although a pluralist approach to TE may be enough to salvage the theory for some, the need to allow for at least some moral good in a scenario where no-one is benefitted remains at best bizarre, at least within the realm of welfare concerns. Contrastingly, the non-comparative Prioritarianism requires equality in scenarios where our intuitions expect this, but by modelling equality as an instrumental good it avoids the troubling conclusions of TE). Therefore, it doesn't matter that some people are worse off *than others*; instead, it simply matters that some people are worse off.

References

[1] Parfit, D. (1997). 'Equality and Priority', *Ratio X* 3, (Oxford: Blackwell Publishers), 202–221.

[2] Larry Temkin, L. (1993). *Inequality*, (New York: Oxford University Press).

[3] Roger Crisp, R. (2003). 'Equality, Priority, and Compassion', *Ethics* 113, 745–763.

Peoples as a unit of consideration – a Defence for Rawls' Laws of Peoples

Bon Bong
University of Hong Kong

In the Rawls' Laws of Peoples (hereafter referred as *LP*), he considered a second Original Position in the same spirit in A Theory of Justice (hereafter referred as *LP*), this time globalised so to determine the principles of justice that ought to govern the interaction between societies. The participants in *LP*, however, are unlike those who in the in *TJ*. They are representatives of "peoples", not individuals. Like in *TJ*, they were placed behind a Veil of Ignorance that obscured their knowledge of the socio-economic status of their own peoples that they will have to live with after the Veil-lifting, but that is essentially where similarities between *TJ* and *LP* end.

This choice to take "peoples" as the ultimate unit of consideration and representation in this second deliberative session became the centre of academic criticism.[1] It was argued that this construction essentially allowed states to take precedence over individuals in terms of consideration, which potentially allowed the justification of gross acts of injustice. It was also argued that the framework *LP* was completely unnecessary, as the framework in *TJ* was already sufficient in allowing us to determine all the principles that should govern human society, be it "domestic" or "globalised".

I shall give a more thorough treatment in the next section, but before that, I shall introduce some new terminologies to relieve ourselves of ambiguity and terminological inaccuracy. Let us call the deliberative process outlined in *TJ* the Original Congress of Individuals, or the Domestic Original Congress, and the conclusions they (would) have produced therein the Original Position. Let us call that outlined in *LP* the Original Congress of Peoples, or the Global Original Congress. Call the conclusions produced therein the Laws of Peoples. This saves us from the clumsiness of using the term "Original Position"to mean both the deliberative process and the conclusions they produce in both the deliberative sessions of *TJ* and *LP*.

[1] Unlike *TJ*, *LP* not only dimmed in comparison to *TJ* in terms of appraisal, it was almost universally poorly received. Reidy summarised the landscape of scholarly response to *LP* as "difficult in the secondary literature to find positive or even sympathetic treatments of Rawls'work in *LP*". The general scholarly sentiment expressed toward Rawls'analysis and conclusions in *LP* can be characterised by polite, respectful rejection or dismissal. It is a widely shared opinion that Rawls work in *LP* was thoroughly or nearly thoroughly wrongheaded.

This essay aims to focus on one issue: the choice of taking "peoples" as the unit of consideration and deliberation. I would argue that Rawls was reasonable and ultimately justified, and specifically, why "peoples" as the unit of consideration and representation in any Original Congress-like thought experiment that hopes to determine the principles that ought to govern all humanity is appropriate, and is far more philosophically reasonable. To do this, I shall first consider some general objections towards Rawls'choice of using "peoples", especially those raised by Pogge. In this section, I will also discuss Reidy's defence for Rawls' theory in *LP*. I will then, following Reidy's analysis, analyse Rawls' general philosophical strategic spirit to analyse Rawls' constructivist approach. Thirdly I will discuss the Cosmopolitans'proposed version of the Global Original Congress, and establish why it goes against Rawls' methodology. In particular, I will demonstrate how and why the Cosmopolitan's conception of a globalised Domestic Original Congress requires us to make unreasonable modifications to what the Veil obscures. For the sake of argumentative brevity, let us simply call the Cosmopolitans'issue with the unit of consideration and representation in the Global Original Congress, "the Issue of Representation".

The Cosmopolitans' Objection – the Issue of Representation

Cosmopolitanism can be summarised in their central belief that when it comes to international matters, especially on matters of socio-economic distribution, human beings, not other forms of human association, should be the ultimate consideration unit. Pogge summarised three essential common elements shared by all Cosmopolitan positions. First: *individualism*: the ultimate units of concern are human beings or persons – rather than say, family lines, tribes, ethnic, cultural, or religious communities, nations, or states. The latter may be units of concern only indirectly, in virtue of their individual members or citizens. Second, *universality*: the status of ultimate unit of concern attaches to every living human being equally, not merely to some subset. Third, *generality*: this special status has global force. Persons are ultimate units of concern for everyone – not only for their compatriots, fellow religionists, or such like [3]. It is therefore no surprise that Cosmopolitans *LP* unacceptable.

There are three major strands of objections towards the *LP*. The first claims the various ideas of a second round of Original Congress to determine the laws that ought to govern peoples in *TJ* (1st edition) and in *LP* are at worst contradictory, at best unnecessary. Pogge, in particular, observed that there were already issues of internal tension in the first edition of *TJ* where Rawls tried to extend his results of the Domestic Original Congress to a global level. Considering the 1st edition of the *TJ* with *LP* together, Pogge believes three views of international justice are embedded therein: the first comes from the first edition of *TJ* which has been taken out of subsequent revisions. There, Rawls spoke of a second session of the original position as featuring

persons from the various societies who make a rational choice of principles so as best to protect their interests. Pogge calls this reading R1. The second view, which is presented in the same page where the first view was introduced, was a conception of a second session as featuring "representatives of states who are to make a rational choice to protect their interests". Pogge calls this reading R2. A third reading suggests that Rawls endorses the traditional (pre-World War II) principles of international law [4]. Pogge tried to show in *Realizing Rawls* that no two of these three views are compatible. Rawls has subsequently resolved the tension by clearly and consistently endorsing the second view R2. However, Pogge criticises Rawls for doing so without justification why he favoured R2 over R1 [4]. Furthermore, Pogge himself noted that if we were to consider the entire Earth as a community, it fits the presumptions as outlined in *TJ*'s Original Congress of Individuals far better. Cosmopolitans in following Rawls'argument in *TJ*, argues that since the entirely world is quite literally the "closed society" in which individuals "enter by birth and exit by death" presumed in the Original Congress of Individuals in *TJ*. Furthermore, there is clearly a global Basic Structure that exerts a profound influence on the lives of individuals. Surely then if the Original Position derived from the Original Congress of Individuals in *TJ* was valid, then any construction of a Global Original Congress thought experiment that does not take humans as the representation unit must be inconsistent. Let us call this version of an Original Congress with global jurisdiction operating on the broad principles of *TJ* the Poggesian Original Congress (POC for short).

The second argument begins by first asserting that all constituting institutions of the Basic Structure are subject to scrutiny and revision at the Original Congress. There is no reason at all why nations, states, or peoples, as institutions or concepts warrant special exemption.

Their existential specificities must be understood as revisable, since they, as part of the Basic Structure, directly contribute to the allocation of socio-economic and political privileges, and exert a profound influence over the lives of individuals. By this logic, since the nation, the state, and the people, are all subject to revision, and since the participants of this POC are likely to recognise that institutions such as the nation, the state, and the people, are all going to contribute to the rise of grave socio-economic injustices, they will therefore radically reform and restructure them (as Pogge has suggested in *Cosmopolitanism and Sovereignty* where he advocates for some framework of vertical sovereignty [3]), if not their outright abolition.

The third argument, advanced by Pogge and Buchannan in particular, is that the only world in which the theory in *LP* is relevant, is one in which domestic societies are completely closed, where individuals have no interaction with those belonging to another society, and that these closed societies are maximally socio-economically independent from each other. Buchannan argues that the world in which the paradigm in *LP* is applicable, is and must essentially be a Westphalian world, which is no longer exists [1].

Pogge agrees and argues that not only does the high degree of international interconnectivity render Rawls'paradigm in *LP* inapplicable in today's world, but that there has never been a world in modern history or recent history where the paradigm was *ever* applicable.

Defences for Rawls are thin, but attempts might still be found. The first takes on the Issue of Representation by noting that the thesis is not just a positive assertion about the right to representation of the individual. It also asserts that other social entities and institutions have and should have no claim to moral consideration. The first defence asserts that they do in fact possess such claims, and that such claims are irreducible. For example, Reidy argues that the individual's claims to justice are claims necessarily directed towards other individuals, and towards other socio-economic and political constructions, associations, and institutions – one's family, one's friends, one's nation, one's government, etcetera [9]. This echoes and reflects the difficulty in resolving the problem in jurisprudence of whether one ever owes any duties or rights to oneself [2]. Humans are indeed self-originating sources of valid claims to justice, but the existence of a claim is predicated upon the existence of a structure to which the claim is directed to. "Nature permits these capacities [of humans claiming justice], but it does not deliver them" [8]. Reidy argues that claims of justice can only be grounded and manifested, essentially caused to exist, only when human beings form associations, socio-economic associations with other human beings. And since it is overwhelmingly the case that human beings are incapable of surviving in complete and utter isolation on their own, human beings, are, on the whole, not self-sufficient entities. This is different when it comes to the matter of peoples. Reidy argues that peoples can both come to be and persist over time as corporate moral agents from any cooperation with other peoples. Therefore, it is entirely reasonable that the claims of justice of peoples be directed to nobody external, but instead to itself, or the individuals constituting the body. For example, Reidy argues that a people may reasonably adopt the internal securing of domestic justice for its members as its only fundamental end.

Reidy offers a second defence for Rawls. Reidy remarks that individual participants of the Original Congress of Individuals are thought to participate involuntarily in basic socio-economic institutions because no single individual is or can ever be truly self-sufficient to the point that he can survive on his own. Social and economic cooperation is almost absolutely necessary for all individuals under any usual circumstance [8]. This is not the case for peoples. Almost by virtue of definition, a people is already considered to be self-sufficient. If that reeks of abuse of formalism, one can accept that at least in principle, all recognisably people-like peoples are self-sufficient, or at the very least, potentially capable of being so. This is to say, even if we were to assume that all peoples are now tightly interwoven into the fabric of the global Basic Structure, it is not impossible for a people to rescind that connection and withdraw into itself. It will probably suffer, but it is not impossible [9].

These are not fool proof solutions to the Issue of Representation. Although they have probably succeeded in establishing why if it were legitimate to think peoples were to have any claims to justice, their claims to justice must necessarily be structurally different from the claims to justice of the individual, the question why is it at all justified to think peoples have any claims to justice remains unanswered. Reidy's argument tells us why if peoples have moral claims, their moral claims are structurally different from those of the individual – but he has not produced reasoning as to why they should have moral claims at all.

The third defence is one external to and independent of the theoretical framework of the Original Congress thought experiment. It points to the spirit governing Rawls'philosophical methodology. Rawls began with his theory-building in *LP* with a seemingly tangential discussion on a concept that he calls as "realistic utopia". His methodology in *LP* aims to build a theoretical framework of the principles that governs the peoples of the world in a manner that would allow them to approach a realistic utopia. This ideal world is "realistic" because it could exist under the laws of physics and general limitations of reality; it is a "utopia"because it "joins reasonableness and justice with conditions enabling citizens to realise their fundamental interests" [6, 7]. The way to understand this is that Rawls is saying he is not trying to investigate the characteristics of normative moral principles in a vacuum, but to consider them within the limits of what is realistic possible and permitted to happen by the realistic constrains of human society. This Rawlsian spirit of philosophical investigation, emphasised again in *Justice as Fairness*,[2] was correctly highlighted by Reidy. Reidy himself emphasised that "the fundamental task of philosophy is to chart the course we must collectively follow if the empirical, contingent, and local forces of history and politics that make us what we are are themselves to remain bounded by or confined within the content of a common human reason, itself also an empirical contingent historical achievement". Rawls' elaboration of the idea of a "realistic utopia" is therefore a manifestation of a belief of "approaching fundamental issues in normative political philosophy from a point of view that takes person not as they are under existing conditions, but as they might be, given the empirical limits of human psychology, biology, and the like, under conditions that might reasonably be hope for" [9].

I believe this is the correct approach to properly counter the Cosmopolitans'challenges, yet it remains to be shown that acknowledging the above spirit of doing philosophy

[2] "We view political philosophy as realistically utopian: that is, as probing the limits of practicable political possibility. Our hope for the future of our society rests on the belief that the social world allows at least a decent political order, so that reasonably just, though not perfect, democratic regime is possible. So we ask: what would a just democratic society be like under reasonably favourable but still possible historical conditions, conditions allowed by the laws and tendencies of the social world?" [5, 4],

justifies Rawls'configuration of his construction in *LP*. I shall elaborate more on this in the next section.

Critique of the Poggesian Original Congress

The questioned posed by the Issue of Representation, can be transformed from "why represent peoples?" into "why assume peoples at all?". For Rawls, the existence of Peoples is not a debatable matter for the Congress, but is in fact a fundamental aspect of reality, almost like the laws of physics. In other words, the participants at the Congress of Peoples have no right nor power to determine whether Peoples will or will not be a major component of the Global Basic Structure after the dissolution of the Congress. As Pogge has put it, "This institution [the state, peoples] would simply be taken for granted. In Rawls's sketch the mere existence of the modern states system today supports the precedence of domestic over global principles of justice and tends to reduce the agenda of the global original position to dealing between governments. However, if the endorsement of this institution is to have any force, then it must have been subjected to moral examination just like other institutions. Otherwise Rawls would be begging a crucial question? that justice may fail to require this system in its present form. [...] Rawls cannot defend this presupposition by asserting that states, or the traditions and communal life they may reflect, have in their own right a claim to exist" [9]. As an alternative, Cosmopolitans argue that the version of Original Congress that should be used to determine the principles of justice on a global scope, is precisely the one operating on the broad principles and mechanisms outlined in *TJ*, albeit with global jurisdiction over all humanity. This is the Poggesian Original Congress (POC) which briefly discussed above.

I believe why Rawls cannot accept a POC as the model of the thought experiment, is because however utopian it is, it is unrealistic. It is unrealistic because it neglects the anthropological limits and the structures of humanity that cannot be reasonably expected to be done away. And the concept of peoples as an anthropological structure, in Rawls'view, is precisely one of those concepts that cannot be done away.

It may be argued that the POC, whilst using only individuals as the unit of representation in the Original Congress, it does not affect peoples as social and anthropological structure at all. I will demonstrate this is ill-founded, and a POC that can ultimately produce principles of justice operating only through the consideration of individuals as the only representation unit, must be configured in such a way that the participants themselves not only not know to which people they belong to, but also that they must be ignorant of the very concept of "peoples" as well. In other words, a POC that aims to be faithful in constructivist set-up to the Original Congress as outlined in *TJ* except for its global jurisdiction, must have a Veil of Ignorance that is thick enough to

obscure even the concept of "peoples".

Let us consider the following three versions of POC. All three of them are characterised by the fact that participants participate in the capacity of individual human beings, representing themselves. Unlike the Global Original Congress of *LP*, they represent neither peoples, states, nor nations. All three versions of the POC have global jurisdiction. The principles outlined therein are to cover all human beings connected to the global Basic Structure, whose existence we have assumed for the sake of argument.[3] The Veil of Ignorance obscures their knowledge in the manner as outlined in *TJ* in all three versions of the POC – the difference then between the three, is what the Veil obscures on top of that. In the first version of the POC, the Veil obscures nothing beyond those in *TJ*. Call this P1. The second obscures the participants'knowledge of which peoples they belong to. They know that they belong to a people, and that they are aware of the existence of the concept of "a people", but they do not know which people they belong to. Call this version of the POC P2. Finally, the third version of the POC is the one in which the Veil is the thickest. In this version, on top of the censorship commitments in P2, the knowledge of the existence of "peoples" as a concept is also obscured. Call this P3. I shall argue that the only POC in which the participants will truly take human beings as the ultimate and sole unit of moral consideration, and the only legitimate source of claim to justice, is P3 – in both P1 and P2, the participants will produce principles that are either similar or stronger to those in *LP*.

Let us consider P1, where participants of the Congress know the existence of peoples, and indeed which peoples they belong to. This Congress is no good. Here, participants know not only which people they belong to, but also their various specificities. They also presumably identify quite strongly towards their own people. Since they know they are going to be most associated to the people they belong to, as opposed to those other peoples that they do not belong, when faced with the problem of organising the global Basic Structure, they are likely to organise it in two levels – one for their own peoples, and one for the relationship between different peoples. This is phenomenon is due to three reasons: first, their very knowledge of which peoples they belong to restricts and diminishes their empathy of members belonging to other groups. Second, having arranged themselves into smaller units along the lines of peoples, they will realise that the socio-economic structures exclusively or almost exclusively accessible only to their own peoples are likely going to impact their lives the most, as opposed to other institutions established elsewhere. They recognise the fact that their membership entails and is entailed by the fact that they are going to be most socio-economically associated with their own peoples before any other. Third, given

[3] So completely isolated communities such as the Amazonian uncontacted tribes and the Sentinelese people are excluded out of our consideration.

the above reasoning, but also in recognition of the fact that socio-economic decisions of one peoples may or may not impact members of another, they are likely going to come up with principles to govern the interaction of peoples, in which peoples and their overarching sovereign institution will be the unit of consideration and organisation. Their reasoning is clearly not acceptable to Cosmopolitans.

P2 resolves the problem in P1 by obscuring the participants' knowledge of their membership of a particular people, but does not obscure their knowledge of the concept of a peoples. However, this is still no good. In fact, P2 is structurally identical to that the Global Original Congress in *LP*, except without Rawls' limitation to liberal and decent peoples. Why will P2 participants not reason the principles of justice as Cosmopolitans want? Three reasons might be given: first, because the participants are aware of the existence of peoples, and that they will belong to one of them after the dissolution of the Congress, participants here will take into consideration of the fact they will be associated with a people after the dissolution. Presumably they may anticipate that these peoples have different socio-economic strengths, and that the people they belong to will be one of the most impactful socio-economic structures they will be associated with. Second, their identity and personal and emotional connection to their people constitutes a force that makes these individuals internalise the moral acts of justice and injustice inflicted upon their people (however such an act inflicted on a people might be perceived or constituted) as one inflicted upon themselves. The individuals internalise this injustice, but also externalise it and project it onto the construct of peoples. This projection of the sense of demand for justice onto their greater community of peoples to which individuals belong, turn peoples into a source of claim to justice that is more than just a sum of individual claims. Justice owed to a people in this sense is not reducible to the private individual. Third, given that they all know they belong to some people but not the socio-economic specificities of their peoples which might or might constitute as major sources of socio-economic advantages and disadvantages of their existence in the post-Veil-lifting world, they are likely going to reason that their good-will directed towards other peoples may not be entirely and unconditionally reciprocated. This coupled with the fact that they are likely going to have diminished levels of sympathy for individuals not belonging to their people, makes it reasonable that they will construct principles of justice using peoples as a unit of consideration and claim to justice.

P2 with further modifications as to which type of people is exactly allowed in the Global Original Congress would be the Global Original Congress in *LP* which is precisely what Cosmopolitans reject. It should be clear now that the POC Cosmopolitans are really looking for is one in which not only are the participants' knowledge of their membership is obscured, but one that the very concept of peoples is censored. That is P3.

The Veil of Ignorance and what lies behind it

We have seen that a Cosmopolitan's Original Congress requires the participant's unknowledge of the existence of the concept of peoples. This requires one to place the concept of peoples itself behind the Veil of Ignorance. This is problematic and a misapplication of the Veil.

The Veil of Ignorance serves to help us conceive the configuration of the "realistic utopia" and epitomises Rawls'methodology in his investigation of normative political philosophy – that is to join the limits of reality with the construction of a political order that allows the maximisation of one's freedom to self-actualise. The role the Veil plays in the Original Congress thought experiment, is to rid its participants, and us as external observers of the procession of the experiment, of the arbitrary beliefs that obscure our reason, and the variable physical traits of the human being that fertilises arbitrary beliefs. The Veil atomises the participants of the Original Congress, takes away their individual traits, and distillates them till only what essentially makes them human remains.

But the Veil does more than that. It deprives the participants of the Original Congress not only the knowledge of the themselves in the post-Veil-lifting world, but also the knowledge about certain external environmental configurations of that world. This has been consistently asserted in both *TJ* and *LP*. In the *TJ*, the participants "do not know about the particular circumstances of their own society", "they do not know its economic or political situation, or the level of civilisation and culture it has been able to achieve", though "they understand political affairs and the principles of economic theory"and "they know the basis of social organisation and the laws of human psychology" [[7], 118–119]. In *LP*, the parties "do not know, for example, the size of the territory, or the population, or the relative strength of the people whose fundamental interests they represent", and that "they do not know the extent of their natural resources, or the level of their economic development, or other such information" [6, 32–33]. The purpose of this obscuration is to ensure the Original Congress thought experiment functions as it should be: "a natural guide to intuition", which "must be interpreted so that one can at any time adopt its perspective" [7, 120].

Unfortunately, Rawls has not justified why he chose not to censor the concept of peoples. I believe we may reason out this ourselves by examining the general characteristics of the Veil (particularly that in *TJ*) and the information that Rawls has explicitly outlined that it hides, and therefore derive what other information we the Veil may reasonably censor.

From *TJ*, we can know that the Veil of the Original Congress of Individuals obscures the following: one's class position or social status, one's distribution of natural assets and abilities, one's intelligence and strength, etcetera. Furthermore, the participant

does not know "his conception of the good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism"– so individuals are obscured from their own philosophical and moral metrics used to assess what is good (their moralities and religious beliefs) they will possess as beings after Veil-lifting; they are deprived knowledge of their own dreams and life plans.[4] Observe that the effect of the Veil is that it essentialises the human being. It atomises all participants in the Original Congress by severing them from their human connections, and it renders them all isomorphic by subtracting all the qualities that differentiate them. Because "the differences among the parties are unknown to them, and everyone is equally rational and similarly situated, and each is convinced by the same arguments", the individuals are all the same with each other, and "therefore, we can view the agreement in the original Position from the standpoint of one person selected at random" [7, 120]. The participant of the Original Congress is a being with all the essential features to make it human, but none of the characteristics to make it any one particular human.

What then cannot be placed behind the Veil without seriously jeopardising the thought experiment? The knowledge of the limits of the laws of physics and reality is an obvious example. The Veil simply cannot censor the participants of the Original Congress from the most basic knowledge of physics.[5] The justification is straightforward – obscuration of such information renders the participants so free to conceptualise a world that is so detached from our own reality that they might not be able to generate principles of justice relevant in the post-Veil-lifting world. This allows us to extend the unobscurable class of qualities. In particular, the various features of the human condition. It cannot be reasonably justified that the participant behind the Veil knows not of the emotions of joy, sadness, grief, desire, envy, pride, and fury, or the sensations of hunger, lust, and ambition, the longing for human connection in all its conceivable forms, friendship, family, romantic and sexual relationship, and religious association. The Veil obscures their configurations in which they desire, need, abhor, or care about these aspects of humanity, but it does not take away their knowledge of such conditions, because it fundamentally begs one to imagine someone to be a human if one is not to have conception of such primal and basic qualities that define what is humanity. Furthermore, Rawls has explicitly allowed the participants to know "political affairs", "the principles of economic theory", the "basis of social organisation" and "the laws of

[4] For example, their desire to be a master of the piano, their desired version of their physical self, career plans, etcetera.

[5] These include, for example, the intuitive understanding of the mechanics of gravity, the general biological functions of the human being (say for example that natural human reproduction is to be conducted through sexual intercourse), and the knowledge of what is food and the consumption of which materials are necessary for basic, fundamental human survival.

human psychology" in the Original Position. If the participants knew them not, then the participants would simply be operating on pure, unfounded imagination. But implicit in Rawls'Veil in the *TJ*, is the assumption that the participants know what "politics" is, what "economy" is, and what "social organisation" and "human psychology" are. Indeed, behind the participant whose reasoning came up with the Principle of Fair Equality of Opportunity, which states that "?offices and positions'should be open to any individual", is the fact that he must know what a conception of an "office" and a "position" in the socio-economic and political sense. If one does know the existence of a concept, then one certainly cannot conceptualise it; and if one does not know a concept, then one certainly cannot conceptualise the various possible configurations and variations that the concept can manifest in the real world. Knowledge of the existence of the very concepts of the socio-economic structures the whole thought experiment is used to determine must be left known to the participant of the Original Congress. It would be self-defeating to obscure the existence of the concept of "political office" if the point of the experiment is to determine what principles of justice should govern political offices.

Having established the characteristics of what the Veil may or may not obscure, we come to the issue of whether it is legitimate to place the concept of "peoples" itself behind the Veil. To this, I answer in the negative. I argue that the concept of "peoples" belongs to the same class of concepts such as "race", "family", "friendship", "gender", and that the participants' knowledge of such concepts are necessary for the maintenance of an understanding of society reasonably close enough to our own for them to generate principles of justice applicable to a world similar as ours. In particular, I believe general knowledge of such concepts and the relationships they entail are necessary for these participants to be reasonably "human". Therefore, to take away the participant's knowledge of these concepts would be to decrease their humanity, and render them further away from our own world.

We identify three commonalities between concepts such as "race", "family", "friendship", "gender", and "peoples". First of all, these concepts are almost universal and natural, in the sense that various human societies have arrived at the construction of these concepts despite a complete lack of centralised coordination, and that it has historically been and continues to be a widespread manner in which human societies organise, classify, and structuralise their own societies, though the manner in which they do so might change drastically over time and space. Nevertheless, the rough constant is the existence of these concepts and the associated general underlying paradigm. The same might be said of the concept of "peoples". Second, it is very likely that these concepts will remain relevant in the foreseeable future of humanity, albeit their substantive meaning might radically change. Thirdly, an overwhelming number of humans define themselves and indeed others in terms of these concepts – so much so that I believe it is an alienable part of the human condition. Notwithstanding how might

philosophers view them, humans are likely going to continue using these concepts to organise, relate, alienate, privilege, oppress, and to generally action upon their own fellow human beings for the foreseeable future. Indeed, any society that is recognisably human is likely going to develop lines of partition and union according to such lines drawn by these concepts. If that reasoning stands, then it is not justifiable to place the concept of "peoples"the Veil, and Rawls' outline of the thought experiment in *LP* must be understood as largely correct.

Conclusions

The nature of these Original Congress thought experiments is to determine the fundamental principles that any reasonable and rational individual would agree ought to govern our most fundamental socio-economical structures and institutions. The beauty of Rawls's idea of an Original Congress is that it bridges the normative good and the constraints of reality, giving us a picture of what would have been reasonable to do as beings naturally motivated to do what is moral in the face of certain facts of reality. The Cosmopolitan's argument fails to appreciate the more unmovable components of our world and aims to philosophise in a vacuum. This methodology potentially runs against the entire spirit that underpins these thought experiments. Having recognised that, *LP* would cease to be as unattractive as it purports to be.

References

[1] Buchanan, A. (2000). "Rawls's Law Of Peoples: Rules For A Vanished Westphalian World."*Ethics* 110 (4): 697–721.

[2] Hart, H. L. A. (1955). "Are There Any Natural Rights?" *The Philosophical Review* 64 (2): 175–191.

[3] Pogge, T. W. (1992). "Cosmopolitanism And Sovereignty."*Ethics*103 (1): 48–75.

[4] Pogge, T. W. (1994). "An Egalitarian Law Of Peoples."*Philosophy And Public Affairs*23 (3): 195–224.

[5] Rawls, J. (2001). *Justice as Fairness, A Restatement*. Cambridge, Mass.: Harvard University Press.

[6] Rawls, J. (2003). *The Law of Peoples*. 1st ed. Cambridge, Mass.: Harvard University Press.

[7] Rawls, J. (2005). *A Theory of Justice*. 2nd ed. Cambridge, Mass.: Harvard University Press.

[8] Reidy, D. (2002). "Re-Reading Rawls: A Review-Essay."*Soundings: An Interdisciplinary Journal*85 (1/2): 173–202.

[9] Reidy, D. (2004). "Rawls On International Justice: A Defense." *Political Theory*32 (3): 291–319.

Intention: A Missing Element in Frankfurt's Principle of Alternate Possibilities

Shane Neagle
KU Leuven

Introduction

In Harry G. Frankfurt's article titled "Alternate Possibilities and Moral Responsibility", Frankfurt exhibits a thought experiment which changed the course of moral responsibility. Prior to Frankfurt, ethicists generally agreed that one is morally responsible for one's actions so long as one could have done otherwise. As will be explicated throughout this paper, Frankfurt disagreed with such a principle and put forth his own, claiming a person is not morally responsible for an action so long as he performed the action *only because* he could not have done otherwise [2, 838]. Since then, Frankfurt's principle has been identified as problematic by authors such as Philip Gosselin, who has constructively amended Frankfurt's incomplete principle by implementing his own. In this paper I argue that despite contributions to the evolution of identifying moral responsibility, both Frankfurt and Gosselin's principles have severe problems and require reconstruction. I will demonstrate these problems and proceed with the necessary reconstruction to arrive at a new principle of moral responsibility:

> A person is morally responsible for what he has done, so long as he could have done otherwise. If he could not have done otherwise, he is morally responsible if:
>
> 1. He is responsible for the fact that he could not, or,
> 2. The absence of the ability to have done otherwise was not caused by another moral agent and the result of the action performed by the moral agent in question was caused by the intention of that moral agent.

In section one, I explain Frankfurt's thought experiment which leads him to his own principle of moral responsibility. In section two, I discuss the plausibility of Frankfurt's thought experiment while defending it against a critique of hard determinism. However, Frankfurt's principle continues to be problematic. In section three I depict a principle put forward by Phillip Gosselin which attempts to add an element missed in Frankfurt's principle. Despite such progress in the endeavor of accounting for moral responsibility, a severe obstacle remains. In section four I articulate a thought experiment of my own which reveals the absence of an important element in both Frankfurt

and Gosselin. Finally, in section 5, I portray the lacking element, restate our new principle, and explain how our further developed principle intuitively captures every case of moral responsibility discussed in this paper.

Frankfurt's Principle and Thought Experiment

Frankfurt seeks to demonstrate his revised principle of alternate possibilities through a thought experiment involving Black and Jones. Black wants Jones to perform a specific action, which we will call 'action x'. Black is willing to manipulate the circumstances to ensure Jones carries out action *x*, but despite Black's desire, Black much prefers Jones to conduct action *x* on Jones' own accord. So, if Black identifies that Jones will not perform action *x* (Frankfurt says Black is an "excellent judge of such things"), then Black will intervene by whatever means necessary [1], ensuring that Jones carries out action *x*. Thus, in terms of Jones performing action *x*, we have two possibilities:

> Jones 1: Jones can perform action *x* on his own accord, without any interference from Black.
>
> Jones 2: Black can manipulate the circumstances to ensure Jones carries out action *x*.

Therefore, according to Frankfurt, when it comes to Jones performing action *x*, he does not possess the ability to do otherwise. He only has a mere choice in how it becomes the case that he performs action x- either on his own accord, or through the intervention of Black. Nonetheless, according to the original principle of alternate possibilities, which states that a person is morally responsible only if he could have done otherwise, neither Jones 1 nor Jones 2 is morally responsible for performing action *x*, which Frankfurt appropriately identifies as problematic. Thus, in identifying and attempting to solve this problem, Frankfurt put forth his own revised principle (hereby referred to as 'Frankfurt's revised principle'): "a person is not morally responsible for what he has done if he did it only because he could not have done otherwise" [2, 838].

Now let us envision our two possibilities involving Jones 1 and Jones 2 given Frankfurt's revised principle. With Jones 2, Black took control of the circumstances and- through whatever means necessary- ensured Jones 2 conducted action *x*. Since the only reason Jones 2 performed action *x* was *because* he could not have done otherwise, Jones 2 is not morally responsible, which seems intuitive. With Jones 1, Black never interfered, and Jones 1 preformed action *x* on his own accord. Therefore, Jones

[1] The means used by Black are not important here. Frankfurt leaves this open.

1 did not perform action *x* only because he could not have done otherwise, so Jones 1 is morally responsible, which also seems intuitive.

At this point, it seems as though Frankfurt has not only identified a problem with the original principle of alternate possibilities, but he has amended it successfully. However, despite being a negative principle, Frankfurt's revised principle remains problematic.

The Plausibility of Frankfurt's Though Experiment

One potential issue rests in Frankfurt's thought experiment, which he uses to demonstrate the necessity of his principle. More specifically, it deals with the way in which Black will be able to manipulate the circumstances to ensure Jones performs action *x* in the case of Jones 2. It's important here to examine how Frankfurt precisely describes such an occurrence could transpire. Frankfurt says,

> [Black] waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something *other* than what [Black] wants him to do [2, 835].

We can now see that Frankfurt leaves it unclear as to how Jones' decision will- or can-be made clear to Black. Instead, Frankfurt forces his readers to accept that Black is an 'excellent judge of such things'. This poses a problem. How plausible is it that Black can identify Jones' mental decision, before Jones reveals his decision through action? It seems as if we are presented with two possibilities.

One possibility could be that Black has access to the future. Perhaps Black possesses a clairvoyant faculty. It is plausible to say that such a possibility exists in a possible world, but let us assume that this possibility does not exist in the world of Black and Jones.

Another possibility, which seems much more realistic, is an idea posed by David Blumenfeld. Roughly, the idea claims that if we are to accept Frankfurt's words that something can be clear to Black (regarding Jones performing an action) in an infallible manner, then we are forced to accept that Black must bear an ability to identify the *causes* of things [1, 340]. This seems very plausible, but despite its attractiveness, there is an extremely important implication which must be discussed.

If we accept the concept that Black can identify the actions of Jones through the causes of Jones actions, then it seems to follow that the actions of Jones are *causally determined*. The most important aspect of this hard determinist perspective is that if Jones' actions are causally determined, then he must be excused of moral responsibility. I disagree with this hard determinist outlook, and I argue that it is not necessary to excuse

moral responsibility even if we accept that Black can identify the *causes* of Jones' actions.

Black could very well identify a causal chain which ends in Jones' action, and begins with Jones' free will. In this type of free-will initiated causal chain, moral responsibility is not excused. Hence, Black can be an excellent judge of the causes of actions, while simultaneously, Jones can be morally responsible for his actions. Therefore, this 'hard determinist' critique of Frankfurt's thought experiment has no impact on Frankfurt's revised principle. However, there are further arguments which expose important elements that Frankfurt does not account for. Philip Gosselin articulates such a case.

Gosselin's Account for the Situation

Gosselin argues that a critical component missing in Frankfurt's revised principle is the responsibility one has for one's situation. Gosselin illustrates this through an example of an intoxicated driver who cannot control his vehicle due to his intoxication. The intoxicated driver–in terms of not controlling his vehicle–couldn't have done otherwise due to his intoxication level, and therefore caused an accident *only because* he could not have done otherwise. Therefore, according to Frankfurt's revised principle, the driver is not morally responsible. Of course any rational person would argue the intoxicated driver is responsible, and with this mentality, Gosselin proposes his own principle for moral responsibility, which he calls 'S-PAP':

> A person is morally responsible for what he has done only if (1) he could have done otherwise or (2) he is morally responsible for the fact that he could not [3, 92].

Through the S-PAP, Gosselin strives to account for the moral responsibility one has for one's situation. There are many situations–including Gosselin's thought experiment with the intoxicated driver–in which this principle aligns with our intuitions regarding moral responsibility in a much stronger manner than Frankfurt's revised principle. Yet, despite a valuable contribution to moral responsibility, a vital feature remains to be seen.

Smith and Rob

In this section I will first construct a thought experiment of my own which demonstrates a major problem with the principles reviewed thus far, and I will then proceed to explain the demonstration.

Let us imagine a scenario in which a man, named Smith, desires the death of another man, named Rob. Smith has kidnapped and restrained Rob to a chair in his

basement, making it impossible for Rob to escape. In front of Rob is a mechanism–manufactured by Smith–where, when a button is pressed, the trigger of a gun will be pulled, resulting in the fatal shooting of Rob. However, after this entire situation has been established, Smith wishes to wait one day before executing Rob to ensure that the Rob held captive by Smith is the correct Rob, and not Rob's twin brother. So one day transpires, and Smith receives confirmation that he has the correct Rob. He then decides to enter the basement and press the button. However, Smith lives in an old house with faulty wood floors. On Smith's way to the stairway, Smith walked directly above the mechanism where Rob was held, when the wood floors collapsed. As gravitational forces acted upon Smith and pulled him to the basement floor, his elbow initiated the button, entirely unintended, which resulted in the death of Rob. It would seem intuitive to consider Smith as morally responsible for the death of Rob. However, if we examine this case under the principles of both Frankfurt and Gosselin, Smith is not morally responsible.

Recall Frankfurt's revised principle. In the previously described example, Smith, being acted upon by gravitational forces due to a reasonably unforeseen event, initiated the button unintentionally, and therefore pressed the button only because he could not have done otherwise. Smith is therefore excused of moral responsibility. Clearly, this can't be so. Further, even Gosselin's principle does not account for Smith's responsibility. We cannot consider both the collapse of Smith's floor and gravitational forces to be elements that Smith is morally responsible for. And after-all, these are the conditions that made it so Smith *could not do otherwise* when his elbow pressed the button. Considering this, it seems as though an important dimension is unaccounted for in both Frankfurt's revised principle and in Gosselin's advanced principle.

Intention

The element missed by both Frankfurt and Gosselin is the intention of the moral agent. I will now restate our new principle and then discuss the previously mentioned cases to determine if this principle encompasses that which Frankfurt and Gosselin have missed.

> A person is morally responsible for what he has done if he could have done otherwise. If he could not have done otherwise, he is morally responsible if:
>
> 1. He is responsible for the fact that he could not, or
> 2. The absence of the ability to have done otherwise was not caused by another moral agent and the result of the action performed by the moral agent in question was caused by the intention of the

moral agent.

Let us turn back to Jones 1, who performed action *x* on his own accord. If Jones 1 conducted action *x* on his own accord, then it follows that he could have done otherwise. He could have chosen not to perform action *x*, which would have caused Black to intervene and force Jones 1 to conduct action *x*. But Jones 1 did not do this. Therefore, Jones 1 could have done otherwise, so he is morally responsible, which aligns with our moral intuitions.

Recall that Jones 2 only performed action *x* due to the intervention of Black. Jones 2 couldn't have done otherwise. Since Jones 2 is *not* responsible for the fact that he could not do otherwise, section A does not apply. Further, section B provides us with a conjunctive proposition which will only hold if both propositions hold. The first proposition states "the absence of the ability to have done otherwise was not caused by another moral agent". In the case of Jones 2, Black, a moral agent, caused the absence of such an ability[2]. Thus, section B does not apply, which means Jones 2 is not morally responsible, which also aligns with our intuitions.

The case of Smith and Rob also deserves some analysis. The difference between Smith and Jones 2 is seen in section B where the reason Smith could not have done otherwise was not due to another moral agent- it was instead due to the natural law of gravity. Provided this, we must go one step further, to examine Smith's intention. Smith intended to kill Rob, and Smith had fabricated the circumstances to actualize his intention. Due to such an intention, the result of Smith's falling through a collapsed floor and pressing the button with his elbow- despite lacking the ability to do otherwise in this specific act- was the actualization of Smith's original intention: to kill Rob. If Smith had never fully intended to kill Rob, and Smith never manipulated the circumstances to commit such an evil act, then Rob would not have been killed. Therefore, according to this new principle, Smith maintains morally responsibility, which once again, is parallel to our moral intuitions.

The fact that the moral responsibility of the agents used in our previously mentioned examples is reflective of our moral intuitions will not suffice to consider our new principle as satisfactory. We must explore different situations to see if the satisfactory of our new principle remains.

We shall return to our example involving Smith and Rob. This time, Smith2 discovers that Rob2's identical twin brother is held captive by Smith2. Upon realizing this,

[2] The reason why the absence of an 'ability to do otherwise' caused by another moral agent can excuse moral responsibility of the moral agent in question, is because the moral responsibility is transferred to the moral agent responsible for the absence of an ability to do otherwise. This transfer can of course not happen when an entity other than a moral agent is responsible for the absence of the ability to do otherwise.

Smith2 runs to his basement to release Rob2's twin brother, but on his way, the wood floors collapse. Smith2 falls to the basement, with his elbow pushing the button, resulting in the death of Rob2's twin brother. Is Smith2 morally responsible for the death of Rob2's twin? Following our new principle, the question is this: was the death of Rob2's twin brother caused by Smith2's intention to kill Rob2? The answer, put simply, is yes. Had Smith2 never intended to kill Rob2 and had he never took the measures that he did in an effort to actualize his intention, then Rob2's twin brother never would have been killed. Thus, Smith2 is responsible.

Prior to concluding, we will examine one more variation to our thought experiment involving Smith and Rob. Smith3 and Rob3 are friends, and they are practical jokesters. They want to trick a friend into thinking that Rob3 is held captive in the same way described in the previous example involving Smith and Rob. In the same manner as before, Smith3 falls through the collapsed floor, his elbow presses the button, and Rob3 is executed. Is Smith3 morally responsible? It seems as though we are intuitively mixed here. Did Smith3's intention- to create a very realistic joke using a real weapon in a given situation which could possibly result in the death of Rob3- cause the death of Rob3? The answer, once again, is yes. Therefore, even in this case, Smith3 is morally responsible.

Conclusion

In this paper I have shown a major problem with Frankfurt's revised principle. I have discussed a notable effort by Gosselin who contributed to the evolution of the principle of alternate possibilities. Further, I have explained why- as it stands today- the principle remains incomplete. I therefore attempted to reestablish a principle of moral responsibility that accounts for the missing element of one's intention, which must be considered in cases where a non-moral agent is responsible for the absence of an ability to do otherwise. Lastly, I would like to emphasize that in the previously mentioned examples, it does not seem as though Smith, Smith2, and Smith3 all share an equal level of responsibility, despite maintaining moral responsibility. It follows then, that if we accept the consideration of intention in assigning moral responsibility, we should also recognize that there are different levels of moral responsibility.

References

[1] Blumenfeld, D. (1971). "The Principle of Alternate Possibilities." *Journal of Philosophy*, 68(11): 339-345.

[2] Frankfurt, H.G. (1969). "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy*, 66(23): 829-839.

[3] Gosselin, P. (1987). "The Principle of Alternate Possibilities." *Canadian Journal of Philosophy*, 17(1): 91-104.

Hegel and Belinsky on the theory of 'Reconciliation with Reality'

Colette Olive
King's College London

Introduction

This paper is concerned with the Hegelian notion of 'Reconciliation with Reality' and whether it is a desirable feature of Hegel's philosophy. My contention is that there is value in the idea and that, when understood correctly, it can be defended against a number of objections. The idea of a reconciliation of reality was a key preoccupation for the Russian Hegelians of the 19th century and was to be heavily critiqued by Belinsky, the prominent Russian philosopher. Early 19th century was riddled with poverty and political corruption, and relied heavily on the serfs for inexpensive labour. Consequently, for many philosophers at this time a key preoccupation was making sense of the terrible world they lived in. Belinsky popularised the conservative interpretation of the *Doppelsatz* in Russian which unfortunately was based on key misunderstandings about Hegelian philosophy. The *Doppelsatz* is the foundation of the idea of reconciliation with reality and is formulated in the preface to the *Philosophy of Right* as "What is rational is actual; and what is actual is rational" . I intend to first give an exposition of the reconciliation with reality and look at what the key aims of the project are. Then, I will present Belinsky's critique charitably and identify the different strands of his argument, concluding that two of these rests on an incorrect reading of Hegel. Having defended Hegel against most of Belinsky's critique but acknowledging one of his objections is strong, I will suggest that the reconciliation with reality can play the role that Belinsky wanted it to play all along i.e. make sense of a difficult world and will conclude that it is an attractive and useful theory.

Exposition of 'reconciliation with reality'

I now want to explicate what is meant by reconciliation or *Versöhnung* with reality and how it fits into the Hegelian picture. Reconciliation with reality, for Hegel, is to "recognise reason as the rose in the cross of the present and thereby to delight in the present" [7, 22], essentially to affirm the inherent rationality of the present. Hegel believes that "the rational becomes actual…by entering into external existence" [7, 22] and consequently what is actual will be rational. He does not want to focus on "how it (the state) ought to be, but rather at showing how the state, as the ethical

universe, should be recognised" [7, 23] indicating he is more concerned with affirming reality as it is and not making normative claims about how it ought to be. This doctrine is known as the *Doppelsatz* or sometimes 'the double dictum' and is the claim that "What is rational is actual; and what is actual is rational" [7, 20] . To understand and affirm this claim is to reconcile oneself with reality. The *Doppelsatz* is traditionally understood either in a conservative or a progressive way. The former interprets it as meaning 'what is, is good' and the latter as 'what is when properly realised is good' [11, 3]. To be reconciled is to affirm that social institutions such as state and family aim to be, and often already are, rational.

The aim of reconciliation is to reconcile oneself with the modern social world and the state thereby avoiding subjective alienation. The question then arises, why ought we to reconcile ourselves with the idea that the actual is rational? The main reason to affirm reality as rational seems to be as a way of avoiding alienation. Indeed, reconciliation is a key part of overcoming alienation according to Inwood [8, 37]. More precisely, the process of reconciliation is one in which we are no longer alienated from the social world and the state but become at home in the world (*zuhausesein*). Reconciling with social reality thus means affirming that the social world is a home which means "it makes it possible for people to actualize themselves as individuals and as social members" [4, 114]. Alienation is thus the opposite i.e. not being able to actualise oneself as an individual or social member in the world. When we are zuhause in the social world, we are more realised than in our alienated state. We are also more connected with reality and others. It is thus better for us to be reconciled than to not be.

Hegel's account of history is integral for understanding the *Doppelsatz* and rests on the doctrine that "reason governs the world" [6, 31] and that consequently, history and the unfolding of the universe is a rational process. This follows naturally from Hegel's metaphysics, his absolute idealism, given that reason is imbued in nature and so it must also be imbued in history. History is essentially teleological for Hegel and is governed by laws that allow it to reach its end. Since the essence of Geist is freedom [?, 261], Hegel concludes that the end of history is the self-awareness of Geist and freedom thus history progresses towards this goal in accordance with reason. This means that history is necessarily progressive since it is always directed towards the self-actualisation of Geist. The other reason history is necessarily progressive is because history is inherently dialectical. The Hegelian dialectic posits that everything can be analysed in terms of thesis, antithesis and synthesis. When the thesis' internal contradictions are realised by its negation, a synthesis is created by the two. Thus, the grand narrative of history moves from thesis to antithesis and then synthesis. Reality and our social world play a role in this narrative. They are instantiations of the self-realisation of Geist and are thus rational which is at the heart of the *Doppelsatz*.

Another helpful way of understanding the Hegelian project of reconciliation is by seeing Hegelian social philosophy as a "social theodicy". Traditionally conceived, a theod-

icy is a way of solving the problem of evil; the existence of suffering and a benevolent god. The social theodicy is a way of solving the problem of evil in society and the existence of rational institutions. Essentially, it justifies "the ways of society to its members" [5, 176]. Hegel does not shy away from the existence of evil and suffering in a society and describes history as a "slaughter-bench" [6, 27] but an important aspect of reconciliation is to acknowledge the existence of evils and to understand them as necessary for social progress. By reconciling ourselves with reality we complete the task of the social theodicy by making sense of the flaws in the social world.

Belinsky's critique of Hegel

It seems to me that Belinsky's critique, given in a letter to Botkin, has three distinct strands which I will untangle here. Belinsky takes a conservative reading of the *Doppelsatz* which I believe is what leads to his complaints. Consequently, he finds affirming the conservative *Doppelsatz* unappealing for three reasons. Firstly, by affirming that reality is rational we are seemingly forced to justify and accept suffering that exists in the world rather than seeking to remedy it. This is also closely related to the rational progress of history. As well as this, Belinsky rejects the doctrine because he does not see how it can solve the social and political problems of his time which is connected to his belief that reconciliation with reality requires one to accept rather than try to change reality. Finally, it seems that reconciling with the rational involves accepting the relegation of the value of the individual.

Belinsky's first objection deals with the idea that reconciling oneself with reality means that we simply accept and indeed can justify suffering in the world. This is a natural consequence of the conservative reading of the *Doppelsatz*, since if we interpret it as the idea that the real is rational, it looks at first blush like we are committed to saying that the state of the world as it is, is rational. If we affirm that the world is rational, how can we make sense of suffering? The progress of history or the development of Geist are rational processes so instances of suffering and evil, as part of that process, must also be rational. Belinsky voices this concern with reference to the dialectic method and says "disharmony is said to be a condition of harmony, that may be profitable and pleasant for some?but not for those whose fates are destined to express the idea of disharmony" [2, 159]. The complaint is that Hegel asks us to affirm a reality filled with suffering because it is part of the rational self-realisation of spirit. But Belinsky does not think that reality, as it is, is rational or good because of the presence of suffering and the suffering of those that get caught in the crossfire of history. For him, affirming reality means affirming this suffering and this to him is abhorrent.

The second objection Belinsky levels at Hegel is that his theory of reconciliation doesn't solve any political or social problems but is "mere abstraction" [2, 162]. Belinsky doesn't

elaborate on this objection very much but I believe we can supplement in with some thoughts from Shkolnikov such that we can see the complaint more charitably. He notes a rupture in Belinsky's thought where Belinsky became concerned with "real socio-political problems replacing his former preoccupation with "abstract" philosophical questions" [10, 181]. Belinsky opposes the idea of the individual/personality with the universal and accuses Hegel as being preoccupied with the abstract and the universal rather than the individual, which is the epicentre of suffering in the world. He believed that the universal was "a swindler and executioner of the poor personality" [10, 181]. To be concerned with the individual is to want to address socio-political problems and suffering in the world but according to Belinsky, Hegel has done the opposite. For Belinsky, socialism did what Hegelianism couldn't by addressing the concerns of the individual and seeking to remedy them. Belinsky believed only socialism could solve the socio-political problems that plagues the individual, particularly in Russia, and that pure abstraction is not sufficient to do this.

The other reason I think Belinsky comes to this conclusion is that his conservative interpretation of the *Doppelsatz* commits him to the idea that the status quo is rational which seems to go against ideas about change and progress. As a revolutionary, this would have been deeply unattractive to Belinsky. He felt that reality needed to be radically altered, and indeed it really did, in order to meet the needs of the individual. He felt that Hegel thought reality provided everything the individual needed through rational social institutions and that Hegel was opposed to change.

What I believe is the strongest objection given by Belinsky is the complaint that reconciling reality with the rational would involve relegating the value of the individual. Affirming that reality is rational means accepting that reality is an instance of Geist self-actualising itself and doing this means affirming that history is always progressing as the Spirit moves toward self-understanding. Belinsky expresses his discontent with this idea when he says "The subject for Igor Fedorovitch (Hegel) is not an end in himself but the momentary expression of the universal" [2, 159]. Hegel seems to view the individual as the self-actualisation of the spirit and consequently a 'momentary expression' of the grand narrative of history. The individual is a means to the progress of Geist and as such, in Belinsky's words, Geist becomes "the moloch of the subject" [2, 159]. The progress of history can exact sacrifice from individuals and fundamentally takes precedent over their own lives. In Hegel's story, man is no longer the measure of all things whereas Belinsky seems to believe something akin to this. He claims that "the fate of a subject, an individual, a personality is more important than the fate of the world" [2, 160] which is diametrically opposed to the Hegelian view.

The reason for placing the emphasis on the individual rather than the grand narrative is rooted in his view that concrete reality is of greater importance than the abstract ideal [9, 4]. Through Belinsky's intellectual development, he gradually became disenchanted with the idea of abstract ideals because they didn't seem to capture reality, in-

deed so often accident seemed to prove stronger than reason and he bitterly remarked that he had "been loyal" to Hegel in "tolerating Russian reality" [2, 162]. This connects to Belinsky's emphasis on the individual because he believed the individual was more concrete than abstract ideas. We ought to be concerned with the real, concrete individuals suffering now, rather than the development of abstract ideas. For Belinsky, truly affirming reality will be focusing on the individual and the Hegelian reconciliation does very much the opposite. The Hegelian notion of reconciling oneself with reality, according to Belinsky, is in some sense realising that we are of very little value and are dispensable to the progress of history and Geist.

A defence of Hegel against such objections

A central worry we might have about Belinsky's objections is that (1) and (2) seem to rest on a misreading of the *Doppelsatz*. As discussed in the exposition, we can generally distinguish between conservative and progressive readings of the *Doppelsatz*. The conservative reading, however, is problematic and appears to misinterpret a lot of Hegel's intentions regarding the *Doppelsatz*. Indeed, it neglects the important idea that "The mere fact that a state exists, on Hegel's view, does not entail that it is either rational or, in Hegel's technical sense, 'actual.'" [12, 234]. Existence and actuality are distinct ideas and the conservative reading conflates them. Thus, by saying that the rational is actual, Hegel is not simply endorsing the status quo. We will see below that in fact Hegel was under no illusions about the state of reality and did not think the world was perfect as it was.

Belinsky's first worry is that by reconciling ourselves with reality we are forcing ourselves to accept evil and suffering in the world. His understanding is that reason becomes a Moloch, exacting suffering in the world for the sake of the Geist's progress and more worryingly becoming justified in doing so. But again, this relies on the conservative understanding of what it is for reality to be rational. Hegel is not saying that we should resign ourselves to the existence of suffering in the world and explain it away with the claim that suffering is necessary for the progress of history. What we must do, as part of reconciliation, is acknowledge suffering. Hardimon refers to this as the "melancholy" [4, 121] in reconciliation. He stresses that we must not see reconciliation as a "state of perfect harmony" [4, 110] which contains no conflict, because this is not how reality is. Reconciling oneself with reality is of course seeing the rose in the cross but delighting in this requires we acknowledge not just the rose but also the cross. It is an important feature of reconciliation that we recognise all that is wrong with the world rather than sweeping it under the carpet. As well as recognising, part of reconciliation is accepting the existence of suffering. Again, this is not to be confused with resigning ourselves to the status quo, rather it is about understanding the true state of the world but affirming that our social institutions aspire to, and indeed often succeed

in, becoming rational. We can at once be "genuinely reconciled and recognize that the fundamental features of modern social world?will inevitably exhibit defects and imperfections" [4, 108]. This means that Belinsky's worry that we must accept and resign ourselves to suffering in the world can be dissolved by understanding the melancholic elements of reconciliation correctly.

With regards to Belinsky's second objection we can make a similar case. Hegelian philosophy is not an instruction to accept the status quo and do nothing about the problems facing modern society. This is of course what the conservative reading seems to say. If we adopt the progressive reading and understand that what is actual is spirit and that spirit is rational, then we do not need to say that the status quo is perfectly fine and needs no improvements. Belinsky adopted socialism instead of Hegelianism because he felt it was the only way of effectively addressing social and political problems. Unfortunately, Young Hegelianism and Marxism had not reached Russia by the time of Belinsky's death in 1848 and he had not worked out on his own that the two were completely compatible. The marriage of Hegelianism and socialism was possible and would have alleviated Belinsky's worries about the social utility of Hegelianism.

It is clear to me that the first two objections given my Belinsky rest on the misreading that Hegel was blindly endorsing the status quo rather than accepting and seeking to remedy suffering in the world. But, as discussed in the exposition, Hegel was under no illusions about the state of the world and was clear that it was not perfect in the *Philosophy of Right* which includes discussions of war, divorce and poverty. Part of the Hegelian project of social theodicy involves justifying and explaining the existence of these phenomena. Hardimon suggests Hegel can do this by saying that such things are necessary for dialectical progression for example war is necessary for peace etc. I think this misses Belinsky's point though as it is precisely the idea of justifying suffering that he objects to, the idea of disharmony being necessary since it involves the necessary suffering of some. What I think it stronger is the notion of Melancholy previously discussed which demonstrates that acknowledging the bad in the world is the only way to reconciliation. Belinsky's final objection about the state of the individual does seem to me a legitimate worry for Belinsky to have and I will discuss the weight of this objection in the next section.

The value of reconciliation

Belinsky believes he has found a number of problems with Hegel that render the notion of reconciliation with reality as deeply unattractive. I want to now evaluate the concept of reconciliation and determine whether Belinsky's objections carry any weight. Whilst it seems clear to me that his reading of the *Doppelsatz* is too conservative to be accurate, it seems one of his objections does carry some force. It is a feature

of Hegelian philosophy that in some senses the self is relegated from an end in itself to a means. The unfolding of history and consequent self-realisation of Geist does take precedent over the needs of the individual and many have sacrificed their lives in the name of the progression of society/history. We might even see the Master/Slave dialectic as a relegation of the value of the individual since it looks like they can no longer self-conscious themselves but rely on the recognition of others for self-consciousness. These are just a few of the facets of Hegelian philosophy in which the individual might be relegated from an end in themselves to something instrumental. Belinsky rejects this on the grounds of his own conception of the individual and it seems to me there is no way of making the two views compatible. Thus, this might seem to some a drawback of the notion of reconciliation depending on whether one wants the individual to have a Kantian status as an end not a means.

Whilst Belinsky may have highlighted a potentially unattractive feature of Hegel's project of reconciliation (again, depending on one's conception of the individual), I see a great deal of value in the project of reconciliation and in fact I think it does the job Belinsky wanted it to do. Reconciliation provides us with an excellent way of making sense of the world. By placing an emphasis of affirmation, we have a more positive *Weltanschauung* but that this does not entail we stop trying to improve the present. There seems to me something intuitively appealing about the idea of affirming aspects of reality that are rational and ethical and trying to improve the aspects that are not good. This stems from the progressive interpretation of the *Doppelsatz*, which is my preferred reading. We can actually see the project of reconciliation as an incredibly progressive one. I particularly like Hardimon's notion of melancholy when thinking about reconciliation as I think it is in fact a necessary part of progress. If we do not reflect on the negative features of the world there is no way we can remedy them. We must be acutely aware of suffering in the world such that we might reduce it. What Belinsky was looking for was a way of understanding the world around him such that he could work to improve it. Indeed, reconciliation does exactly this by focusing both on the 'rose' and the 'cross' of the present. This to me seems a balanced way of thinking about the world.

Conclusion

I see the project of reconciliation as a pragmatic and realistic way of understanding the world. Contra to Belinsky's counterargument, it does not involve merely accepting suffering in the world and doing nothing about it but acknowledges the existence of problems in the social world. Equally, it acknowledges that a lot of social institutions such as the family, the state etc are in fact rational and beneficial to society when properly realised. I see great value in the idea of a balanced outlook that in Hardimon's words espouses "realism and sobriety" when thinking about the social world and real-

ity. I do, however, agree with Belinsky regarding the status of the individual. I believe there is a relegation of the value of the individual and that one becomes a 'cog' in the machine of historical progress and social order rather than an end in oneself. Whilst this is perhaps a concession for many people, I think it does not render the theory devoid of use. The theory's attractions lie in its realistic and pragmatic way of understanding reality and its ability to connect us to the world thereby avoiding alienation. To my mind, it is thus a valuable theory.

References

[1] Beiser, F. (2005). *Hegel*. London: Routledge.

[2] Belinsky, V.G. (1948). *Selected Philosophical Works*. Moscow: Foreign Languages Publishing House.

[3] Dallmayr, F.R. (1993). *G.W.F. Hegel: Modernity and Politics*. London: Sage Publications.

[4] Hardimon, M.O. (1994). *Hegel's Social Philosophy: The Project of Reconciliation*. Cambridge: Cambridge University Press.

[5] Hardimon, M.O. (1992). "The Project of Reconciliation ." *Philosophy & Public Affairs* (Wiley) 21(2): 165-195.

[6] Hegel, G.W.F. (1953). *Die Vernunft in der Geschichte*. Translated by Robert S Hartman. New York: The Liberal Arts Press.

[7] Hegel, G.W.F. (1991). *Elements of The Philosophy of Right*. Translated by H. B. Nisbit. Cambridge: Cambridge University Press.

[8] Inwood, M. (1993). A Hegel Dictionary, *Philosophical Quarterly* 43(173):583-583.

[9] Plekhatov. (1987). *Belinski and Rational Reality*. Accessed March 22, 2018. https://www.marxists.org/archive/plekhanov/1897/belinski/part4.htm.

[10] Shkolnikov, Vadim. (2013). "The philosophical cap of Yegor Fjodorovič or becoming Belinskij." *Studies in East European Thought* (Springer) 65(3):175-187.

[11] Stern, Robert. (2006). "Hegel's *Doppelsatz* : A Neutral Reading." *Journal of the History of Philosophy* 44(2): 235-266.

[12] Westphal, K.R. (1993). The Basic Context and Structure of Hegel's Philosophy of Right, in F.C. Beiser (ed.), *The Cambridge Campanion to Hegel*, Cambridge: Cambridge University Press.

Durham University

MA Philosophy
Durham University

MA Philosophy

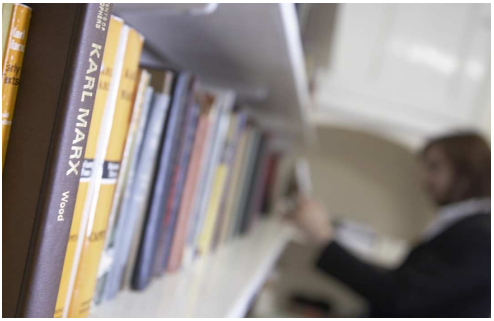
Course structure

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

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

Modules offered are subject to change.

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



Career Opportunities

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

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

Entry requirements

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

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

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

How to apply

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

Enquiries

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

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

Durham University and Durham University logo are registered Trade Marks of the University of Durham. Unless otherwise stated, all material in this publication is copyright of the University of Durham. The University makes every effort to ensure that the information contained here is accurate. Please note that the University's website is the most up to date source of information and we strongly recommend that you always visit the website before making any commitments.

Issue 12 (2)
Annual Conference 2018

Have Homo sapiens ever lived in a 'state of nature'?
Dave Knowles

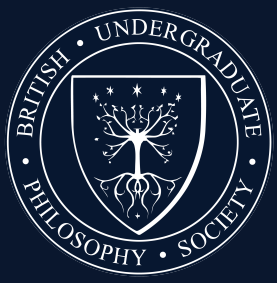
Hobbes and Descartes: 'Idea' and The Idea of God
Cody Myers

Equality of Welfare and Priority of People
Michael Ratcliff

Peoples as a unit of consideration – a Defence for Rawls' Laws of Peoples
Bon Bong

Intention: A Missing Element in Frankfurt's Principle of Alternate Possibilities
Shane Neagle

Hegel and Belinsky on the theory of 'Reconciliation with Reality'
Colette Olive



Journal of the British Undergraduate Philosophy Society