Issue 9 (2) Autumn Conference 2015 10th Anniversary Edition

British Journal of Undergraduate Philosophy

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

Journal of the British Undergraduate Philosophy Society

British Journal of Undergraduate Philosophy

The Journal of the British Undergraduate Philosophy Society

Issue 9(2) Autumn 2015

Editor-in-chief:	Geoff Keeling
Assistant editors:	Farbod Akhlaghi-Ghaffarokh Bede Hager-Suart Nathan Oseroff
Manuscript editors:	Andy Bates James Hart Ella Langham Conor Thompson-Clarke
Commissioning editor:	Achilleas Sarantaris
Assistant commissioning editor:	Gabriel Chiu
President of BUPS:	Farbod Akhlaghi-Ghaffarokh

www.bups.org/bjup-online

British Journal of Undergraduate Philosophy

www.bups.org/bjup-online

All rights reserved.
© 2015 British Undergraduate Philosophy Society, and the authors.

Cover image by Wikimedia Commons user 'Matthew Pron'
(`https://commons.wikimedia.org/wiki/User:MatP`) used under the terms of the GNU Free Documentation License. Photographs of Jennifer Saul, Jerry Fodor, Huw Price, Mary Margaret McCabe, and John Worrall used with permission.

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.

Editorial

> I love deadlines. I like the whooshing sound they make as they fly by.
>
> —— Douglas Adams

There is some possible world, or some set of possible worlds, where this all went calamitously wrong. Possible worlds where the Journal, through some stroke of misfortune, received no submissions. Possible worlds where the Commissioning Team cracked under the pressure of a countably infinite set of submissions. Possible worlds where, heaven forbid, I decided to forego my duties as Editor in order to find out exactly what the turtle-necks mean by 'poststructuralism'. But, as the dice fell, or as the Good Lord would have it, none of these conditions obtained.

I am immensely proud to announce that this edition marks the 10th Anniversary of *The British Journal of Undergraduate Philosophy* (*BJUP*). The first edition of the *BJUP* was published in December 2005, under the editorship of Robert Charleston. His aim, in founding the Journal, was to 'catch the attention of people who want to push themselves in our subject; who want to read and write extra work; to relish in the day-to-day experience of being a philosopher'.

Meanwhile, in December 2005, the prospect of me ever having a hand in something of this calibre was slim. Still fumbling through my tent-like blazer, and particularly fond of my Total 90s, I hadn't the faintest idea what philosophers do. And to be honest, I still don't. But I am enormously conscious of the great succession of coincidences that conspired to put me in this seat. And I could not be more proud to be listed alongside individuals who gave so many human-hours in the name of allowing undergraduate philosophers to flourish. So immense thanks go to Robert Charleston for setting up this enterprise. And to all the Editors, Presidents, Assistant Editors, Commissioning Editors, Manuscript Editors, Finance Officers and Conference Co-ordinators that kept this Journal alive over the last decade.

If any edition were deserving of the accolade '10th Anniversary Edition', it would be this one. The ordeal began with 122 submissions in the commissioning inbox. That's quite a lot, given that each one must be peer-reviewed, thought about, debated, discussed, committed to the flames, fished out of the fireplace, lost, found, critiqued and cried-over before receiving the final verdict. In fact it's an awful lot. But I would not have missed this opportunity for the world. The call for papers, in true fulfilment of Charleston's aim, travelled east to Lahore, west to Los Angeles, south to Cape Town and north to St Andrews.

This decade has seen the *BJUP* grow into a truly international enterprise. And a great deal of thanks is owed to those, from so many countries, who took the plunge and submitted their most outstanding piece of work.

From there, things took a slight tumble for the worst. We didn't have nearly enough dedicated postgraduate reviewers to make the workload bearable. So I cannot thank the Commissioning Editor, Achilleas Sarantaris, and his assistant Gabriel Chiu enough, for badgering this tremendously generous fleet of philosophers to take the requisite time from their postgraduate studies to critique our papers. Likewise, without these wonderful individuals, who between them have expertise on everything from Anaximander to axiomatic set theory, the excellent standards of the *BJUP* could not be maintained.

Once the insurmountable task of peer-reviewing was over, we arrived at what might plausibly be called 'Black Tuesday', or 'The Night of Editorial Damnation'. Farbod Akhlaghi-Ghaffarokh, our esteemed leader and beloved President, sat down with me and scrutinised each paper in turn, critically assessing the feedback provided by reviewers and arguing over the relative merits and demerits of each submission. This took a long time. In fact, we didn't complete the process until 2:30 a.m., which given that we'd started as night was falling, was quite the achievement. To make matters worse, I had a driving test in the morning. Remarkably, I passed, despite being so pumped on Red Bull that the prospect of an untimely loo-break during the dreaded thirty-eight minutes was beginning to look like a genuine possibility.

But somehow we pulled through. And I could not have done it without Farbod, who had to endure screams of undying terror, as I stared into the blue-abyss of Skype and shuddered as it stared back at me. And so we pressed on. Papers were distributed to the Assistant Editors, Bede Hager-Suart and Nathan Oseroff. Due to what might be described as seriously extenuating circumstances, the wonderful Nate was incapacitated and a terrible load befell Bede. Needless to say, the Journal was compiled and we owe everything to Bede for his exemplary contribution.

Last, and most certainly not least, we have our manuscript editors, who had the good fortune of meticulously scourging the spelling, punctuation and grammar of the entire Journal. A huge debt is owed to the manuscript editors, in particular Andy Bates, James Hart, Ella Langham, and Conor Thompson-Clarke for their immense contribution in the editorial process.

And now we move to more sombre territory. Farbod Akhlaghi-Ghaffarokh, our noble leader, who has dedicated so much time, whilst battling dissertations, his final year at the University of Reading, his MLitt at the University of St Andrews and the associated phenomenology of Geoff Keeling abandoning

all hope and reason, sinking into despair and staring blankly into the nothingness, is leaving us.

Farbod's term as President of the British Undergraduate Philosophy Society (BUPS) has been remarkable. I have never met anyone more willing to increase the objective probability of their contracting coronary heart disease to ensure that conferences go smoothly and journals are published on time. Neither have I met anyone of Farbod's age so competent in as many areas of philosophy as he is. He's not just good. He's really good. Or, in his own unforgettable words, there's independent good reason to think and no good reason not to think that the necessary and jointly sufficient conditions such that Farbod is an exemplary philosopher-in-training and President of BUPS have obtained.

I first met Farbod in November 2014, when we talked until 4:30 a.m., despite having a conference to attend three hours later. Needless to say, we were both very tired in the morning. But Farbod, with a level of endurance matched only by the tardigrade, stood up and presented a wonderful paper on the meaning and reference of the first-person term 'I'. He preceded to blast through the day, asking cutting but charitable questions to each and every speaker on topics ranging from Kant to the philosophy of time.

I once said to my pocket-anthropologist Matej Križnar, "I think he's the best undergraduate philosopher I know." To which Matej replied, "I think he's the best undergraduate I know." All things considered, including his rather suave elbow-patched tweed jacket, I have no doubt that Farbod will go on to become a remarkable professional academic philosopher ('in the relevant sense') and wish him all the best in his postgraduate studies. At the same time, it gives me great pleasure to welcome our new President, Eugene Chua. These are big shoes to fill, but I am more than confident that Eugene will make an excellent successor to Farbod. I look forward to sharing many laughs, tears and stress-induced nervous ticks over the coming months.

At the same time, Rajeev Dass, whose aptitude for remote-planning is matched only by that of Sauron, is leaving us. Among other commendable abilities, such as ruthless haggling, spreadsheets, poster-making, food-consumption-estimation and salesmanship, Rajeev has an unmatched ability to smile a Colon-D smile no matter how extreme the hardship. But Rajeev's head is not in the clouds, which is not a jab at his height, but a sincere complement about Rajeev's ability to remain level headed and optimistic no matter what the circumstances. He will be dearly missed by all of us, and I wish him the best of luck for the future.

And that just about concludes everything I had to say. I must add that I am extremely grateful to the Department of Philosophy, Logic and Scientific

Method at the London School of Economics, first for being a bastion of decent philosophy, but more importantly for funding a good-portion of our conference. In particular, my thanks go to Dr Peter Dennis for his role in the organisational process. I am also grateful to the University of Durham for their funding.

On a personal note, I must thank Farbod, for endless discussions over the last few months on everything from ontological pluralism to ontological monism, and for his wonderful understanding and friendship. I must thank my long-suffering partner Yousuf, for providing more moral support, hugs and pep-talks than anyone else. And last to my housemates, Matej, Achilleas, Ziga and Peter, for putting up with the immense stress generated through my undying commitment to this Journal. Like Sisyphus, we climb.

GK

Acknowledgements

Alister Gray

Angelo Cei

Anna Ciborowsky

Charlotte Withers

Chloé de Canson

Chris Atkinson

Christopher Burr

Christopher Ivins

Dan Wooley

Daniel Hill

David Olbrich

Douglas Giles

Jake Wojtowicz

Ji Yong Lee

Julia Tarnawska

Julie McGlynn

Kai Kualulaau

Leonidas Oikonomou

Michael Plant

Noemi Magnani

Oising Foggarty Graveson

Patrick Hassan

Peter West

Ryan Nefdt

Sam Spurrell

Sloveig Aasen

Steven French

Contents

Biopower, Subjectivity, Liberation

Carolina Iribarren
Cornell University

> The fear of losing the self, and suspending with it the boundary between oneself and other life, the aversion to death and destruction, is twinned with a promise of joy which has threatened civilization at every moment.
>
> —— Horkheimer & Adorno, *Dialectic of Enlightenment*

> For when you say "this is that", "'this' means 'that"'…the very form of your proposition, the 'is' ['est'] affiliated with *trying-to-say*, essentialises … substantialises … immobilises.
>
> —— Jacques Derrida, *Dissemination*

Biopower is, plain and simple, power of life: a form of power that "defines itself in relation to a biological field – which it takes control of and vests itself in."[1] The implied opposition to death in both the term and its definition is not accidental.[2] According to Michel Foucault, it is the end of the pre-modern sovereign right to kill that marks, precisely, the beginning of "*a biopolitics of the population*".[3] Biopower is, then, a mode of power "whose highest function [is] perhaps no longer to kill, but to invest life through and through".[4] Even though biopower functions on and through life, it is important to note, as Giorgio Agamben does, that "a power whose aim is essentially to make live [can still] exert an unconditional power of death".[5] At first glance, an aporia – biopower's domain over death – is perfectly consistent with its investment in life. Insofar as biopower "is coextensive and continuous with life" and "linked with a production of truth – the truth of the individual himself", life, that is, the locus of biopower, needs to be apprehended as a category itself, a category whose meaning is derived negatively from and thus is dependent on its op-

[1] Achille Mbembe [7] p. 166.

[2] As Slavoj Žižek [7] puts it, "death is, quite logically, the ultimate traumatic point of biopolitics, the politics of the administration of life".

[3] Michel Foucault [15] p. 139.

[4] Ibid.

[5] Agamben [4] p. 84.

posing term (death) within the dyadic binary of which it is part.[6] Within the biopolitical framework, life, in fact, operates as the ultimate category, that is, as the final but also the best and most extreme example of its kind – a category that presents and defines itself against mortality but that in practice shares a relation of contiguity, not opposition, with death.

Under the logic of biopower, however, 'ultimate' carries yet another meaning: basic and fundamental. Whereas previous forms of power organised and legitimated themselves primarily via naturalised identity categories, such as gender, race, and so on, biopower concentrates on the naturalisation, totalisation, and eternalisation of life itself. To accomplish this, biopower operates through two poles: 'the disciplines of the body' and 'the regulations of the population'.[7] The former refers to the techniques of individualisation that "bring the individual to bind himself to his own identity and consciousness and, at the same time, to an external power"; the latter, to the integration of "the care of the natural life of individuals into [the] very centre" of the biopolitical machine.[8] As Agamben rightly points out, these two poles – technologies of the self and political techniques – are but two faces of the same power; it is through these two modalities, both predicated and centred on the sanctity of life, that biopower, specifically, articulates itself.

It is exactly this "tricky combination in the same political structures of individualization techniques and of totalization procedures" that generates biopolitics' horrifying efficacy.[9] As both an individualizing and a totalizing form of power, biopower attaches itself to everyday life, "transforming an essentially political body into an essentially biological body, whose birth and death, health and illness"[10] become the target and site of state administration and regulation; at the same time, biopower applies itself to the individual – the body within the body (of the population) – "mark[ing] him by his own individuality, attach[ing] him to his own identity, impos[ing] a law of truth on him that he must recognize and other have to recognize in him".[11] Leaving behind the Hobbesian body politic, the biopolitical state adopts what Foucault terms

[6] Foucault [12] p. 333.

[7] Ibid.

[8] Agamben [7] p. 137.

[9] Foucault [12] p. 332.

[10] Agamben [4] p. 84.

[11] Foucault [12] p. 331.

'pastoral power',[12] which manages to integrate the individual into the matrix of the population, "penetrating into a threshold in which it is no longer possible to establish [identitarian] caesuras." Stripped of its qualifiers, life becomes, all at once, sanctified (think of Agamben's *homo sacer*) and abstracted to the point of total dehumanization (Arendt's[13] 'nothing but human' being). The paradox of life's lethal quality, epitomized by the now maxim "to make live and let die"[14] is by no means accidental: it is not unknown that biopower makes living dead out of us.[15]

Identity categories do not, of course, disappear under biopolitical regimes – as the recently coined term 'identitarian politics' demonstrates.[16] Biopower still deploys those 'old' social demarcations; in fact, many of its administrative practices depend and are justified through these categories, which tacitly – though at times very explicitly, as in the case of criminals and the insane – designate who is a viable subject and who is not: "In order to be actualized effectively, life requires a series of mediations constituted precisely by these categories."[17] Nevertheless, biopower's production and application of identity works to subjectivitely subdue rather than to objectively totalize and essentialize the subject: "apparatuses aim to create – through a series of practices, discourses, and bodies of knowledge – docile, yet free, bodies that assume their identity and their 'freedom' as subjects in the process of their desubjectification."[18] This means that, as "a technology aimed at permitting the exercise of biopower",[19] identity serves, as before, to parcel the population in order to make its administration more effective and less painstaking. Though still entangled in regimes of interpellation and normativity, identity nevertheless

---

[12] See Foucault's [14], Lectures 5–7 for a more indepth explanation and analysis of pastoral power.

[13] Arendt [7] p. 92.

[14] Foucault [13] p. 241.

[15] See Alexander Weheliye's description of the Muselmänner in *Habeas Viscus: Racializing Assemblages, Biopolitics, and Black Feminist Theories of the Human* for a description of the "phlegmatic but still living corpses" paradigmatic of the biopolitical space.

[16] See Adorno [1] p. 149 for an analysis of what arguing under the banner of identity entails ("The supposition of identity is indeed the ideological element of pure thought, all the way down to formal logic; but hidden in it is also the truth moment of ideology, the pledge that there should be no contradiction, no antagonism") and Gayatri Chakravorty Spivak's 'Can the Subaltern Speak?' for one instantiation of the discourse of identitarian politics.

[17] Esposito [9] p. 55.

[18] Agamben [5] p. 19.

[19] Mbembe [7] p. 166.

ceases to mark the threshold of power/knowledge. It is the unatomizable category of (bare) life that becomes, in biopolitics, the language through which the totalizing structures and procedures of power dictate who must live and who may die.[20]

When the "political 'double bind', constituted by individualization and the simultaneous totalization of structures of modern power"[21] is apprehended in its plenitude, the notions of substance and essence suddenly, albeit inadvertently, become very attractive. Now that our very life has become power's most lethal weapon, liberation seems available only in the form of some substantial content, some secret treasure that remains, within the self "as a primary condition of existence".[22] Such reasoning is apparent in Agamben's emancipatory *profanation*. Although the object of this liberatory practice is never defined, its substantive quality is nevertheless more than implied: "[W]hat we are dealing with here is the liberation of *that which remains captured and separated by means of apparatuses*, in order to bring it back to a possible common use."[23] It is also present in Hardt and Negri's *Empire*, where 'postmodern liberation' is posited on 'a properly ontological power',"a radical counterpower, ontologically grounded … on the actual activity of the multitude, its creation, production, and power – a materialist teleology".[24] Another, though less glaring, instance of this accidental call for essentialism can be found in the work of Foucault himself. As Butler points out, Foucault inevitably falls back on the body as the prediscursive material upon which biopower inscribes itself in all of its might, but which could be salvaged via "a tactical reversal of the various [biopolitical] mechanisms": "The rallying point for the counterattack against the deployment of sexuality ought not to be sex-desire, but bodies and pleasures."[25] What is most striking about all of these examples is that their emancipatory turns, predicated on pre-served substance(s), occur after lengthy discussions of the (re)production and therefore implication and complicity of subjects by and in political regimes – a blind turn that takes the form of a 'I

[20] This formulation is a reversal of Achille Mbembe [7] ("who may live and who must die") seems more in line with the Foucauldian definition of biopower as "organized around the management of life rather than the menace of death" (Foucault [15] p. 147).

[21] Foucault [12] p. 336.

[22] Nietzsche [19] §I:10.

[23] Agamben [5] p. 17.

[24] Hardt and Negri [17] p. 61.

[25] Ibid. p. 66.

know well, but all the same.'

What these gestures toward alternate possibilities risk, however, is more than their critical impetus. By alluding to liberation in the form of substantive recuperation, do these questions not re-inscribe and further consolidate the very source of biopolitical legitimacy, which is also biopower's locus of operation, namely, life? Do they not rely on the presupposition that there is, and must be, such a substratum that underlies all forms of qualified life – that is, a metaphysical substance or ontological condition that pre-exists biopolitical subjectivation? And, in doing so, are they not further advancing biopower's move to naturalize the very constructed outlines of life? We must not forget:

> [T]he most important political categories of modernity are not be interpreted in their absoluteness, that is, for what they declare themselves to be, and not exclusively on the basis of their historical configuration, but rather as the linguistic and institutional forms adopted by the [biopolitical] logic in order to safeguard life from the risks that derive from its own collective configuration and conflagration.[26]

Life, as biopower's preferred discursive resource, becomes the regime of truth of the biopolitical society. At this point, it is worth recalling Foucault's insistence on the complicity and inseparability of power and knowledge, the latter understood as the product of epistemic regimes that come about as a result of contingent historical conditions.[27] Insofar as "truth isn't outside power, or lacking in power",[28] biopower's articulation of life constitutes (one of) the constraints through which the politics of truth – "the types of discourses which [power] accepts and makes function as true; the mechanisms and instances which enable one to distinguish true and false statements, the means by which each is sanctioned; the techniques and procedures accorded value in the acquisition of true; the status of those who are charged with saying what counts as true" – are produced and, as such, must be heavily questioned.

The problematics of truth inevitably lead to the problematics of ontology. Biopower exploits our thirst to know who we *truly* are by producing this truth, a truth that revolves around the question of life as not only the condition of being alive but also as status (lifestyle). Once essentialism is on the

[26] Foucault [15] p. 157.

[27] Foucault [11] Preface.

[28] Foucault [16] p. 72.

table, there is the added danger of revalidating the ontological status of the subject as a given, preexisting, non-arbitrary category that does, in fact, exist behind the doing.[29] Our stance here should be unequivocal. Given the conflation of the subject of the enunciation and the subject of the enunciated that is paradigmatic of biopower[30] – mainly, due to the necessary (re)production of a biopolitical subject[31] – it is not only futile but utterly irrelevant to focus on the notion of human agency, when agency is taken to be a precondition or necessarily correlative to self-realization and self-identity.Emancipatory discourses that resort to the notion of agency not only turn a blind eye to the impossibility of subjective agency – precisely due to the coincidence of subject and object *within* the very body of every individual – but also implicitly and complicitly reinvites and reinvents the positing and self-subsisting 'I' of idealist readings of Descartes – exemplified by Fichte in its positive ontological condition and by Spinoza in its negative – the same presentist view of the subject[32] that has been much to blame for the tragedy of hegemonic structures that rely, by necessity, on essentialized, naturalized, and totalized ontological and epistemological categories. Though, as discussed above, these categories are neither the site or the object of biopolitics, their deployment is tactically and effectively in alignment with the interests of biopower, for they facilitate the securing (in its dual meaning: protecting and constraining) of life. Identity categories are, among others, "the forms that [have] made an essentially normalizing power [i.e. biopower] acceptable".[33]

That we need to scrutinize biopower, alongside its regime of truth and privileged category of operation – life – goes without saying. But, in doing so, we must not forget Foucault's addendum to the lifeblood of liberatory discourses: "Where there is power, there is resistance, *and yet, or rather consequently, this*

[29] A reversal, of course, of Nietzsche's [19] §I:13 formulation: "[T]here is no 'being' behind the doing, effecting, becoming; 'the doer' is simply fabricated into the doing – the doing is everything".

[30] This conflation occurs via the "*technologies of the self* by which processes of subjectivization bring the individual to bind himself to his own identity and consciousness and, at the time, to an external power" (Agamben [12] p. 137).

[31] "The term 'apparatus' designates that in which, and through which, one realizes a pure activity of governance devoid of any foundation in being. This is the reason why apparatuses must always imply a process of subjectification, that is to say, they must produce their subject." (Agamben [5] p. 11)

[32] This is Butler's term.

[33] Foucault [15] p. 144.

resistance is never in a position of exteriority in relation to power."[34] It is when we deny our implication in power relations of domination by fixing a comfortable distance that we blindly perpetuate them. Even Foucault, when discussing the specificity of biopower, makes a precarious distinction between 'individual' and 'subject': "[Biopower] is a form of power that makes individual subjects."[35] Again, rather than making such a conceptual separation, which only ends up reifying the notion of the core self, we must question the status of the individual-subject within and as a product of the regime of ipseity that "forces the individual back on himself and ties him to his own identity in a constraining way" precisely by acting as the guarantee of continuity and identity. If a refusal of the mechanisms of subjectivization and individualization is possible, such rejection should resist taking the form of a narrative of recuperation, according to which "individuals" fight against "abstractions, of economic and ideological state violence, which ignore who we are individually"[36] precisely in order to return and recover our "individual" essences. Not only does such a project presuppose the existence of the metaphysical self-positing, self-present 'I' – ignoring thereby the historical conditions that both motivate and make possible our conceptualizations – but also it betrays its own cause by remaining within the circular logic of identity production. While the 'subject' in this context could be equated to the first of the "two meanings to the word 'subject'", that is, "subject to someone else by control and dependence"; while the 'individual' coincides with the second definition: "tied to his own identity by conscience or self-knowledge" – both forms of self-formation, as Foucault himself explains, suggest "a form of power that subjugates and makes subject to".[37] In other words, anything that gestures – be it from or by the 'individual' – at determining who one is or, rather, that one is at all, must be heavily interrogated if a critique of (bio)power is to carry any weight.

By no means am I gesturing to some possibility for subversion or hope of any kind located in the interstices between the multiple processes of subjectification, which are being replicated, *ad nauseam*, by the schizophrenic logic of late capitalism. As Agamben writes:

> The boundless growth of apparatuses in our time corresponds to the equally extreme proliferation in processes of subjectification

[34] Ibid. p. 95. Emphasis added.

[35] Foucault [12] p. 331.

[36] Ibid. pp. 330–331.

[37] Ibid.

> . . . a dissemination that pushes to the extreme the masquerade that has always accompanied every personal identity.[38]

What I am suggesting, however, is that the notion of agency must be interrogated and, if need be, tossed aside if the "necessary error of identity"[39] is to be shown, once and for all, unnecessary and detrimental to, if not the opposition, at least the understanding of the current economy of power relations. If "today there is not even a single instant in which the life of individuals is not modeled, contaminated, or controlled by some [biopolitical] apparatus",[40] it is difficult, to put it mildly, to see how and from where subjective agency would arise. Moreover, if "apparatuses are not a mere accident in which humans are caught by chance, but rather are rooted in the very process of 'humanization' that made 'humans' out of the animals we classify under the rubric of *Homo sapiens*",[41] then human, individual, and subjective agencies can be nothing other than fantasies predicated on the logic of the Enlightenment – as a result of which "humanity had to inflict terrible injuries on itself before the self – the identical, purpose-directed, masculine character of human beings – was created"[42] – and on the fictitious possibility of a prediscursive form of subjectivity, of being – as if we could operate outside the symbolic, without the very imaginaries that constitute our social milieu – and therefore an uninhabitable condition.

If, according to Foucault, "power is tolerable only on condition that it mask a substantial part of itself [as] [i]ts success is proportional to its ability to hide its own mechanisms"[43] – a sentiment that Catherine Malabou echoes when she says that biopower always and necessarily hides itself behind the 'ideological *mask*' of bare life[44] – then, perhaps, theories of biopower should shift their focus and redirect their efforts from recuperation and liberation to unmasking. Though the distinction may seem subtle, if not altogether precarious, especially since *unmasking* appears to promise a hidden treasure just as much as emancipation and liberation might, the difference lies in the thing – or, rather,

---

[38] Agamben [5] pp. 14–5.

[39] Attributed to Gayatri Chakravorty Spivak in Butler [6] p. 229.

[40] Agamben [5] p. 15.

[41] Ibid. p. 16.

[42] Horkheimer and Adorno [2] p. 26.

[43] Foucault [15] p. 26.

[44] Malabou [18].

the no-thing – that awaits behind the mask. The notion of masking of course implicates that of hiding, but hiding does not necessarily imply an object; a hole or a lack is as much of a contender for concealment as a secret treasure. A hope that I am willing to indulge is, then, one that involves the unmasking of power only to arrive, not to any sort of substance (ontological or otherwise), but to the 'no-thing' which Roberto Esposito attaches to his elaboration of *communitas*. This kind of analysis would bring us face to face with the notion of

> [*N*]*ot* subjects. Or subjects of their own proper lack, of the lack of the proper. Subjects of a radical impropriety that coincides with an absolute contingency or just simply 'coincides', that falls together. Finite subjects, cut by a limit that cannot be interiorized because it constitutes precisely their'"outside'; their exteriority that they overlook and that enters into them in their common non- belonging.[45]

Rather than analyzing biopower from the point of view of subjective transcendence or substantialized immanence, this analysis would insist on the constitutive lack that lies at the heart of subjectivity. Perhaps the rallying point for the counterattack against the deployment of the individual ought not to be the materiality that feeds biopower and the biopolitical machine, but "that void, that distance, that extraneousness that constitutes [us] as being missing from [our]selves", forcing us to reconsider the facticity of life all over.[46]

References

[1] Adorno, T. (1973) *Negative Dialectics*. London: A & C Black.

[2] Adorno, T. & Horkenheimer, M. (2002) *Dialectic of Enlightenment*. Stanford: Stanford University Press.

[3] —— (2009) *What Is an Apparatus?*. Stanford: Stanford University Press.

[4] Agamben, G. (1999) *Remnants of Auschwitz*. New York: Zone Books.

[5] —— (2009) *What is an Apparatus?* Stanford: Stanford University Press.

[45] Esposito [10] p. 7.

[46] Ibid.

[6] Butler, J. (2014) *Bodies that Matter: On the Discursive Limits of 'Sex'* New York: Routledge.

[7] Campbell, T. & Sitze, A. (2013) *Biopolitics: A Reader*. Durham: Duke University Press.

[8] Derrida, J. (1972) *Dissemination*. Chicago: The University of Chicago Press.

[9] Esposito, R. (2008) *Bíos: Biopolitics and Philosophy*. Minneapolis: University of Minnesota Press.

[10] —— (2009) *Communitas: The Origin and Destiny of Community*. Stanford: Stanford University Press.

[11] Foucault, M. (1970) *The Order of Things: An Archaeology of the Human Sciences*. New York: Pantheon Books.

[12] —— (2000) *Power, Volume 3*. New York: The New Press.

[13] —— (2004) *Society Must Be Defended: Lectures at the Collège de France, 1975-76*. London: Penguin Books.

[14] —— (2007) *Security, Territory, Population*. New York: Palgrave Macmillan.

[15] (2012) *The History of Sexuality*. New York: Knopf Doubleday Publishing Group.

[16] —— & Rabinow (1984) *The Foucault Reader* New York: Pantheon Books.

[17] Hardt, M. & Negri, A. (2000) *Empire*. Harvard: Harvard University Press.

[18] Malabou, C. (2014) 'The King's Two (Biopolitical) Bodies', in *Representations* 127/1: 98–106.

[19] (1998) Nietzeche, F. *On the Genealogy of Morality*. Indianapolis: Hackett Publishing.

[20] Weheliye, A. (2014) *Habeas Viscus: Racializing Assemblages, Biopolitics, and Black Feminist Theories of the Human*. Durham: Duke University Press.

Does Our Temporal Experience Favour The A-theory of Time?*

Xintong Wei
Universities of St Andrews and Stirling

1 Introduction

In this essay, I will argue that our temporal experience does not favour the A-theory of time. Specifically, I will defend the B-theory against two arguments arising from the experience of the present: the argument from occurrence and the temporal knowledge argument. Let me first explain A-theory and B-theory of time. McTaggart[1] famously distinguishes two ways in which positions in time can be ordered. The A-series of time is the one in which positions are ordered according to their A-properties such as 'being one day past', 'being near future', 'being present' and so on. The B-series of time is the one in which positions are ordered in accordance with B-relations such as 'one hour earlier than', 'two days later than', 'simultaneous with' and so on. Philosophers who think that the A-series of time is not real and the B-series is all there is to time are B-theorists. The B-theory of time claims that the only temporal properties exemplified by events are the relations of precedence and simultaneity. There are no tensed facts or tensed properties as being past, present or future. A-theoretical statements are meaningful and can be true, but their truth conditions are to be expressed in tenseless terms. All tensed statements can be reduced to facts about a manifold of equally real past, present, and future events or objects and no time is ontologically privileged. The A-theory denies this reduction. The A-theorist of time contends that there are genuine tensed facts and properties that are not in any way reducible to facts about B-relations. (McTaggart himself is neither an A-theorist nor a B-theorist of time since he believes that A-series is self-contradictory and B-series cannot exist without A-series.)

The B-theory of time has been objected to on the ground that its austere on-

*Delivered at the BJUPS Autumn Conference, 21–22 November 2015 at the London School of Economics.

[1] McTaggart [12].

tology cannot account for various aspects of our experience of time including the irreducible temporal nature of our psychological attitudes, the experience of the passage of time and the presence of experience. In this paper, I focus on the third aspect of our experience of time and two arguments that arise from the presence of experience. We have the impression from daily experience that events are occurring at the moment. The present is experientially privileged since we can only experience what occurs in the present. It seems that such experience favours A-theory of time, which commits us to the ontology of presentness. In Section 2 and 3, I will set up the background of this debate and introduce the problem of the presence of experience. In Section 4, I will evaluate the argument from occurrence, according to which our temporal experiences have the property of 'occurring' in addition to the much-discussed property of 'being present'. The temporal knowledge argument will be discussed in Section 5. This argument is a temporal version of the knowledge argument in philosophy of mind (in which colourblind Mary recovers colour vision and thereby learns a new non-physical fact). It concerns the puzzle of how I seem to learn a new fact when I know that an event is present even if I always know about the corresponding B-fact that the event occurs at *t*.

2 The A- and B-Theory of Time and the Reduction of Tense

There are two issues at the centre of the debate in metaphysics of time. The first concerns the ontological status of the past and future. According to *eternalism*, reality is a four-dimensional spatiotemporal manifold of objects. Past and future times are no less real than present times just as distant places are not ontologically less significant for being spatially distant. Past and future objects are as real as the currently existing ones. According to *presentism*, only present objects are real. The 'growing block universe' theory – a position in between eternalism and presentism – holds that only the past and the present are real but the future is not.

The problem of ontology is closely related to the second issue of the debate, namely, the status of tense. Tensed sentences are those that presuppose a certain privileged position within the whole of time and are often marked by 'now' 'was' and 'will' such as:

> It is *now* snowing.
>
> I *was* reading *Tractatus* yesterday.
>
> She *will* go to the concert.

Tenseless sentences are those that are not tensed such as:

The concert is later than the meeting.

His book is published in 2014.

Following McTaggart, tensed sentences are called 'A-sentences' while tenseless sentences are called 'B-sentences'. The dispute is whether tense is an ultimate feature of reality. The majority of eternalists are reductionists. They deny the irreducible existence of tense, for all times are ontologically on a par. Tensed statements can be reduced by means of tenseless analysis of the truth conditions of tensed sentences. The combination of eternalism and the reduction of tense is called the 'B'-theory of time. Anti-reductionism about tense, on the other hand, is often called the 'A'-theory of time and A-theorists are those who 'take tense seriously'. The B-theorists believe that we can give an account of our tensed language without appealing to the reality of tense. For example, how am I going to construe the truth conditions of 'it is now snowing'? One natural suggestion is to index utterances to a time. Temporal indexicals such as 'past' and 'present' refer to their time of utterance directly. Whether these indexical sentences are true depend on the context in which they are uttered. So we have:

For all t, 'it is now snowing' uttered at t is true iff it snows at t.

To put such analysis for tensed sentences in schemata we have:

Present tense

(A1) For all t and t^*, 't is now' uttered at t^* is true iff $t = t^*$

(A2) For all t, 'e is now' uttered at t is true iff e occurs at t

Past tense

(B1) For all t and t^*, 't is in the past' uttered at t^* is true iff t is earlier than t^*

(B2) For all t, 'e is in the past' uttered at t is true iff e occurs earlier than t

Future tense

(C1) For all t and t^*, 't is in the future' uttered at t^* is true iff t is later than t^*

(C2) For all t, 'e is in the future' uttered at t is true iff e occurs later than t

The A-theorist denies such reduction and argues that tense is part of the ontology. There are tensed facts such as 'it is now snowing' and 'Wittgenstein

was born in Austria'. One reason against reduction of tense is from the presence of experience. How could the present moment appear to be ontologically privileged if the property of being present does not exist? The existence of tensed facts seems to be the best explanation for the experiential data. The A-theorist Craig thinks that the reality of tense is obviously given to us through experience:

> Indeed, I should say that the belief in the reality of tense and temporal becoming enjoys such powerful epistemic status for us that not only can we be said to know that tense and temporal becoming are real, but also that this belief constitutes a defeater-defeater which overwhelms the objections brought against it… We have already observed that the experience of tense is universal among mankind.[2]

Craig is arguing in the passage that the A-theory of time is true since the belief that some events and some times are present is basic and universal. And its veridicality is given to us in experience. Such belief on the basis of daily experience constitutes an intrinsic defeater-defeater against the B-theory which denies the irreducible reality of tense. A belief is an intrinsic defeater-defeater if it, all on its own, overwhelms any potential defeater brought against it. But what exactly is the presence of experience? And what do those experiential data show?

3 The Temporal Limit of Our Conscious Experience

Our experiences can be associated with temporal properties in two ways.[3] On the one hand, conscious experiences can represent temporal properties of events and objects that we perceive. For example, one might experience the lightning as occurring before the thunder. This is the temporal structure of the content of our conscious experience, namely, the subject represents these two events as such that the time at which the lightning occurs is earlier than the time at which the thunder occurs. This representation is false as the sound travels much slower than the light, giving the incorrect impression that the lighting occurs before the thunder. This should be distinguished from the temporal structure of our conscious experiences themselves, which also have temporal

[2] Craig [5] p. 138.

[3] See Dennett [7], Mellor [13].

properties. The temporal location of my experience of the lighting is earlier than the temporal location of my experience of the thunder. Although my representation of the temporal structure of the object of experiences is false, it seems that my judgments concerning the temporal structure of the conscious experiences themselves wouldn't go wrong. In normal circumstances, I am able to distinguish various experiences and am aware of the temporal order of my experiences themselves.

Mellor is the first philosopher who formulates the *problem of the presence of experience*. In his *Real Time II*, he observes four features of conscious experiences:

(1) '. . . it seems a striking fact that events can only be experiences while they are present.'

(2) '. . . all the experiences I am having now must be located in the present. . . not include events that are now past or future.'

(3) 'Being present must itself be an aspect of experience, something of which we are directly aware.'

(4) 'our perception of the presence of our experience, like that of many if not all its other features, seems to be infallible.'[4]

Although Mellor is not explicit, I think he applies the above distinction and characterizes the phenomenon of the presence of experience in two ways.

Firstly, we might say that the temporal structure of the experience itself has the property of being present, which corresponds to the third and fourth point Mellor makes. We have the impression that our conscious experience is always in the present and such perception seems to be infallible. It seems intuitive that we are experiencing the present in the sense that in so far as we are conscious, we are aware that the conscious experience is present when I have it. All my experiences from different times such as having Christmas dinner, writing this paper and watching a movie later had, have or will have the property of being present. In so far as I am conscious, whatever I might be experiencing, it seems that my conscious experience is always located in the present in objective time. Let us call this characterization of the phenomenon the *presentness of consciousness*.

Secondly, we might say that, in relation to the temporal structure of the con-

[4] Mellor [13] pp. 43–44.

tent of our conscious experiences, the events are *presented* to have the property of being present, which is captured in Mellor's first two points. Let us deal with the second feature first and try to address Mellor's question of why all the experiences I am having now must be in the present and why they couldn't include events that are now past or future. Unlike our spatial experience, in which we can perceive objects standing in various relations to us, up or down, near or far, an event in the distant past or in the future could not be a part of our conscious experience. But this fact can be easily explained away. Since perception is a causal process and backward causation is rare, the event of which I am having an experience is very unlikely to be located in the future. There also are empirical explanations of why we cannot perceive the distant past (when one is not admiring a star). First, it seems that light, sound, etc. would have to travel *slower* than they do for us to perceive events that occurred a few minutes ago. Since they travel fast enough between the perceived object and the perceiver that there isn't a significant time lag in the case of the nearby objects that we ordinarily perceive, we only perceive nearby events that are in the very near past or in the specious present. Second, in order to be efficient agents we have to store incoming information in our memory bank once they are absorbed to make room for more updated information.[5] If I could directly perceive events that occurred minutes ago, say, I was drinking a cup of tea and at the same time directly perceive that I am writing a paper, I would get confused about where I am and what I am doing. Thus the second feature could be explained empirically and is not really relevant to the debate of metaphysics of time.

Now let us look at the first feature. Mellor is not claiming that the events themselves are present, but rather that they are presented to be present. If there is indeed an objective present, the events cannot be located in the present when I perceive them because there is a time lag between its occurrence and my perception of the event. The content of conscious experience is believed to have the property of being present. But such belief is false. Although things change much more slowly compared to the speeds of light and sound, so we can capture what is happening with only a short time lag (unless we are observing the distant galaxies), there is always a small interval between the time of the event and that of my experience. A-theorists might contend that, although I perceive events and times through my perceptual and cognitive 'lens' such that the time at which I am aware of them is slightly later than the time at which they actually occur, I can somehow infer that the event has the property of presentness a while ago. Let us call this kind of characterization associated

[5] Butterfield [2].

with the temporal properties of the object of experiences the *presentness of content*.

Mellor then goes on to summarize what he thinks is the problem of the presence of experience that B-theorists must be able to explain away:

(5) 'whoever I am, and whenever I believe my experience to be present, that now-belief will be true.'[6]

The problem is why it is necessary that our experiences characterized in terms of the presentness of consciousness should be in the present in so far as they are accompanied by A-beliefs. That is, my experience is always believed to be located in the present and that now-belief is always true.

Since the problem concerns the A-beliefs or A-judgments about the temporal properties of our experiences, a natural B-theoretical move is to employ the token-reflexive analysis of their truth conditions as sketched in Section 2. Let us take my belief that my experience of writing a paper is present as an example:

(E) My experience of writing a paper is present.

Recall that:

(A2) For all t, 'e is now' uttered at t is true iff e occurs at t.

We have the truth conditions of (E):

(T_E) For all t, my belief that 'my experience of writing a paper is present', had or asserted at t (the time I have this belief), is true iff the experience occurs at t.

Since my conscious experiences and my now-belief about them are simultaneous, (E) is true. This now-belief ascribes presentness to my experience of writing a paper (which occurs at the B-time at which I have this belief). Unless we have a mistaken now-belief about a past conscious experience, the now-belief is always true because the temporal locations of our conscious experiences will automatically make the now-beliefs about them true. This is Mellor's account of the presence of experience formulated in (5).

However, there are two ways to counter Mellor's argument. First, one might question whether (5) is the real problem of the presence of experience on the basis of (1)–(4) (assume that all parties accept (1)–(4) as basic experiential

[6] Mellor [13] p. 44.

data). Second, one might question whether Mellor's B-theoretical solution is adequate to account for (5). I shall deal with the first one in Section 4 and the second one in Section 5.

4 Argument From the Occurrence

Balashov makes a distinction between the easy and the hard problem of the presence of experience.[7] What we have in (5) is only the easy problem of the presence of experience, which concerns the A-beliefs or A-judgments about the temporal properties of our experiences. B-theorists insist that the problem is solved in so far as we are able to account for the tensed semantics without appealing to any tensed facts. The reason why I attribute the A-property of being present to my experiences is the B-fact that my experiences occur at the same time as the attributions in question and the general fact that our A-beliefs about experiences tend to be veridical. There is nothing more to be added, i.e. the reality of tense, than the fact that my experience of writing an essay is believed by me as present. Balashov thinks this reply fails to capture all there is to the presence of experience and leaves out an important feature he called occurrence. The hard problem involves a more radical sense of presence, in which our conscious experiences are "known to be occurring, or present, as opposed to not occurring, or absent".[8] It is the problem of why a particular temporary experience I have, i.e. writing an essay, as opposed to, say, having Christmas dinner, is occurring *simpliciter*. The basic strategy of this form of A-theoretic argument (which is also employed by other arguments discussed in this paper) is to appeal to a certain feature of our experience of time and argue that such a feature is best explained by the A-theory of time. More specifically, this form of argument seems to indicate that we can directly perceive certain temporal properties of our conscious experiences on the basis of which we form the basic and universal belief about the reality of tense. The suggestion is that (5) should be reformulated as:

(5*) Whoever I am, and whenever I believe my experience to be occurring *simpliciter*, that now-belief will be true.

The token-reflexive analysis of tensed beliefs *fails* to accommodate the occurrence of my experience because it is not the case that, for all t, 'the experience

[7] Balashov [1].

[8] Ibid. p. 269.

that I am writing a paper is occurring *simpliciter*' uttered at *t* (the time I have this belief) is true iff the experience and the utterance are simultaneous. Because it is possible that the experience and the utterance are simultaneous while the experience is not occurring *simpliciter*, i.e. the above sentence would be true of past utterances simultaneous with past experiences. The veridicality of (5*) then, the A-theorist contends, is given to us in experience of A-ontology.

But what is this veridicality given to us in experience that A-theorists often appeal to? Why do they take it for granted? I will try to work out the argument in support of the claim. The proponent of such argument may be talking about either the temporal properties of the content of experience or the experience itself, or both. We can spell out the parallel argument for both of them. Firstly we have:

Presentness of content

(P1) *S* has an apparent presentation of an event, *e*, as occurring *simpliciter*.

(P2) If there are no special considerations and if *S* has an apparent presentation of *e* occurring, then *e* is occurring *simpliciter*.

(P3) In the absence of special considerations, *e* is occurring *simpliciter*.

(P4) *e* is occurring *simpliciter* iff reality is irreducibly tensed.

(P5) Therefore: in the absence of special considerations, reality is irreducibly tensed.

By special considerations I mean (a) the subject is cognitively incapable; (b) the occurrence of *e* is impossible; and (c) there are relevant background facts which rule out the occurrence of *e*. The subject should have the cognitive and perceptual capacity to distinguish the temporal structure of events. The event should not be self-contradictory or metaphysically impossible, i.e. I am stepping up and down the stairs. And there should not be any background facts that are inconsistent with (and thus rule out) *e*'s occurring *simpliciter*, i.e. the B-fact that the supernova occurs a long time before I have the perception of the event.

One difficulty with this version of the argument from occurrence with respect to presentness of content is that it is hardly effective in establishing the irreducible reality of tense since there is always a special consideration in place. The events we are experiencing are certain to be in the past when we perceive them. However, this argument might be effective if we make a distinction between mental events and physical events. Mental events such as my decision

to read David Lewis's paper and my experience of them are arguably simultaneous and are located in the present. If the apparent presentation as occurring *simpliciter* was of mental events, then the special consideration (c) could be ruled out in many cases and the argument would be effective. But whether such a distinction can be drawn is a matter of debate. Many have argued that this distinction is only semantic and all events are physical whether or not one describes them in psychological terms, though some reject the claim that mental events can be explained purely in physical terms.[9]

To avoid this controversy, we can run the argument with respect to the presentness of conscious experience.

Presentness of consciousness experience

(Q1) *S* has an apparent presentation of her conscious experience, *c*, as occurring *simpliciter*.

(Q2) If there is no special consideration and if *S* has an apparent presentation of *c* occurring, then *c* is occurring *simpliciter*.

(Q3) In the absence of special considerations, *c* is occurring *simpliciter*.

(Q4) *c* is occurring *simpliciter* iff reality is irreducibly tensed.

(Q5) Therefore: in the absence of special considerations, reality is irreducibly tensed.

Given that the special considerations are the same as above, it seems at least, *prima facie*, plausible to conceive of a situation in which special considerations are absent. The special consideration (c), which is present in the argument of presentness of content, is absent in this argument because the apparent presentation of my conscious experience is always in the present. It then follows from the argument that the event I perceive is occurring *simpliciter*, which is, according to Craig, a defeater-defeater against B-theoretical denial of tensed-ontology.

But is occurrence a unique phenomenon? To begin with, is the concept of occurring *simpliciter* even intelligible? An initial objection pointed out by Callender is that the concept of occurrence is ambiguous.[10] What is occurring *simpliciter*? We are not more enlightened by Balashov's short answer that such

[9] Macdonald [11], Nagel [14], Davidson [6].

[10] Callender [3].

experience is '*sui generis*'. But I think there are two reasons why occurrence *simpliciter* is, *prima facie*, intelligible and distinctive. First, what is unique about occurrence *simpliciter* is that it is something in addition to occurring (being present) when they are. Suppose that I am now having a headache. This experience is not merely known to be present at the time I form this A-belief. It is occurring, as opposed to, not occurring which is the sheer absence of certain experience. The fact that I am now having certain experience A not another experience B can only be explained by the fact that A is occurring while B is not. Second, this radical sense of presence seems to be implied by many A-theorists' accounts of temporal experiences. To start with, in the passage quoted above, Craig says that we 'know that tense and temporal becoming are real'. Schlesinger also observes that the present is "palpably real".[11]

And it is argued that this unique property is best explained by the A-theory of time that there are tensed facts such that the experiences in question are indeed occurring in the present. These remarks all seem to point out that 'presentness' isn't just a referent or index to a certain B-theoretic time. It is something real and whatever is in the present is occurring.

However, the above example of having a headache does not really tell us what is the 'something' in addition to being present. The reformulation of (5*) is meaningful only if 'presentness' is occurring *simpliciter*, which is distinct from being a mere temporal indexical. If occurrence and presentness so understood are indeed two distinct phenomena, then either (a) there is an event that is occurring but not present; or (b) there is an event that is present but not occurring. Since the A-theorist couldn't assert (a) given her account of occurrence, the A-theorist might wish to assert (b). This could be achieved if the 'present' is used as a temporal indexical while the 'occurring' is used in a non-indexical way. For any event e, 'e is present' is true iff e occurs at the time at which the now-assertion is made. For any event e^*, 'e^* is occurring' is true iff the event is really happening. An event can therefore be present but not occurring. But this characterization of occurrence seems to me to presume the truth of tensed metaphysics. It is simply another way of claiming the ontologically privileged status of the present: the occurrence of an event is ontologically real. For what could this 'presence as opposed to sheer absence of certain experience' be, other than existence itself? If the claim is about the existence of an event at a particular time as opposed to all other times, then it presupposes A-ontology. We have no reason to assert (b) because the argument is artificially constructed

[11] Craig [4] p. 427.

on the basis of A-metaphysics. The only way to assert (b) is to read the theory into the data. Therefore, if we do not presuppose A-metaphysics, occurrence is not really a distinctive phenomenon from presentness.

(Q1) is also problematic. Do I really have an apparent presentation of *c* being occurring? Is being present, either *qua* occurrence or presentness, a phenomenal or sensible property? There are two arguments against the claim that presentness is a phenomenal or sensible property. First of all, as I have shown in our discussion of the presentness of content, the idea that we are sensing the property of being present regarding events seems problematic, for they are in the past when we perceive them.[12] Objects look the same to us even if they are in the past. Surely, in this argument, we are not talking about perceiving events being present but perceiving the property of the conscious experience itself. But one might wonder why one and the same property of being present is perceivable when it comes to the experience itself but not the content of the experience? This puzzle is solved if the property of being present is not something that can be perceived but a feature of 'the perceiving' itself, i.e. as an indexical. It is not really a phenomenal property. Furthermore, an objection similar to Mellor's can be constructed if we accept the possibility of time travel (although some versions of the A-theory are incompatible with the possibility of time-travel, i.e. presentism, some A-theorists think time-travel and the A-theory are compatible.[13] Suppose that I travel back to the age of dinosaurs, I will have an apparent presentation that my conscious experiences have the property of being present. I will have the apparent presentation that it is occurring just as the one I have that my current experience of writing the paper is occurring. Conscious experiences appear to be at present even if they are in the past (in the objective time of the universe). It is true that this particular instance of (Q1) may not establish the reality of tense because there is special consideration, i.e. (c) there is a background fact which rules out the possibility of my conscious experience being present. But dialectically, what matters is that we find an example where we have apparent presentation of conscious experience being present but which is in fact in the past. This example shows that being present is not a phenomenal property.

Second, phenomenal or sensible properties distinguish some experiences from others. But the property of being present doesn't distinguish one experience

[12] Mellor [13].

[13] For further discussion, see Keller & Nelson [10].

from the other since every experience is present when it occurs.[14] Furthermore, consider the following plausible and defensible criterion for a sensible-property:

(Cr) *F* is a sensible-property for a modality iff it is not the case that there are possible pairs of individuals *x* and *y* such that (a) *x* and *y* are indiscernible for that modality and (b) *x* is *F* and *y* is not-*F*.

The property of being orange is a sensible-property for vision because under normal lighting conditions etc., there aren't possible objects such that they are visually indiscernible but one is orange and the other is not orange. However, the property of being present is clearly not a sensible-property for any modality. It is possible that I heard middle C played twice with an interval in between. My hearing of the first middle C earlier on is indiscernible from my hearing of the second middle C for the modality of hearing, but the first event is not present now while the second event is. The phenomenon of dynamic recalibration is also worth noticing.[15] Research from cognitive science has shown that our brain can adjust its expectations to achieve subjective simultaneity when it is receiving signals from different modalities at different speeds. The brain assumes that sound, sight and touch should be simultaneous when a person, say, knocks on something. If one of the signals arrives later, the brain will recalibrate the expected time of the signals to better simultaneity. For example, in one experiment, when a subject pressed on a button, a beep 250 ms later appeared to be 'pulled' slightly toward the act. Motor-sensory recalibration can also lead to other kinds of illusion such as illusory reversal of action.[16] Dynamic recalibration shows that, in many cases, instead of being a genuine sensible property, the property of being present for a modality might result from such brain adjustment.

5 The Temporal Knowledge Argument

I have shown that our experience of presentness does not support (5*) because occurrence *simpliciter* is not a distinct phenomenon and our seeming presentation of an event or conscious experience as occurring does not lead to a tensed ontology. But the A-theorist might contend that Mellor's B-theoretical solu-

[14] Hestevold [9].

[15] Eagleman [8].

[16] Stetson et al. [19].

tion is not an adequate account for (5). What is missed out is the cognitive significance of now-beliefs that cannot be captured by tenseless truth conditions. Recall that my now-belief about my experiences is always true when it obtains because of the B-fact that now-belief is always formed simultaneously with present experiences. However, by this account, the truth conditions of '*e* is now' believed at *t* are the same as the truth conditions of '*e* is *t*' believed anytime. If this is so, then the tensed 'now is 12 noon on 4th March 2015' uttered at 12 noon on 4th March 2015 is logically equivalent to the tenseless '12 noon on 4th March 2015 is 12 noon on 4th March 2015' because it is impossible for the former to be true while the later is false and vice versa given that they have the same truth conditions. But this cannot be right because the tensed belief seems to be cognitively significant whereas the tenseless belief is tautological. I always know that '12 noon on 4th March 2015 is 12 noon on 4th March 2015' but not that 'now is 12 noon on 4th March 2015'. I seem to learn a new fact when I know that 'now is 12 noon on 4th March 2015', and my now-beliefs play an important and active role in affecting my actions.

Perry presents a temporal version of Jackson's knowledge argument in philosophy of mind that captures the above thoughts.[17] Suppose that Perry knows all the tenseless facts about the past, present and future events hosted by the Royal Institute of Philosophy from 2000 to 2016. For example, he knows that 2015's Annual Lecture will be given by Amartya Sen on 12 March 2015 in Beveridge Hall, at 5.30 p.m. However, it seems that Perry comes to know something new when he knows that today is 11 March 2015. He infers that the Annual Lecture is tomorrow. And he might thus decide to fly to London and attend the lecture. Knowing that today is 11 March 2015, Perry acquires a new bit of knowledge, a fact that he did not know until he had a look at the calendar. The temporal knowledge argument can thus be formulated as the following:

(W1) Perry knows all the B-facts about the Annual Lecture from 2000 to 2016.

(W2) Perry comes to know a new fact that the Annual Lecture is tomorrow when he knows that today is 11 March 2015.

(W3) Therefore: not all facts are B-facts.

The key premise in the argument is (W2), which B-theorists would deny. Why does it seem that Perry acquires a new fact? Perry has the belief that 2015's Annual Lecture is on 12 March 2015. And after looking at the calendar, Perry forms a tensed belief that 2015's Annual Lecture is tomorrow. According to

[17] Perry [17].

(C2), the latter belief is true iff the lecture occurs one day later than the time at which the belief is formed. Since Perry forms his belief on 11 March 2015, the truth-maker of both beliefs, according to the B-theory, is the B-fact that 2015's Annual Lecture is on 12 March 2015. However, the latter belief leads to very different actions, subsequent beliefs, desires and emotions. For example, Perry immediately books a flight to London. He believes that he will have a pleasant stay in London. He cannot wait to have a chat with his colleagues and friends after the lecture and is too exited to get sleep. If Perry does not learn a new fact after looking at the calendar, what explains the change of his behavior? So we have a subsidiary argument in support of (W2):

(i) The A-belief '*e* is today' believed at *t* and the B-belief that '*e* is at *t*' believed at anytime lead to very different actions, subsequent beliefs, desires and emotions.

(ii) If two beliefs lead to different actions, etc., then they must have different content and thus different truth conditions.

(iii) By (i) and (ii), the A-belief and the B-belief do not have the same truth conditions.

(iv) The best explanation for (iii) is that the A-belief is made true by a tensed fact.

(v) Therefore: when one comes to know that '*e* is today', one knows a new tensed fact that makes this A-belief true.

The above argument in support of (W2) relies on the traditional Fregean view of belief, which contains three parts: (1) belief is a relation between a believer and a proposition; (2) propositions are objects of our belief and belief is the apprehension of propositions; and (3) propositions are identical to the meaning of that-clauses. But the Fregean view has been challenged on the ground that it cannot accommodate indexical belief in general. Consider the following two sentences:

> You are in front of a wild bear.
>
> I am in front of a wild bear.

On Frege's account, since 'you' and 'I' have different senses, these two sentences have different meanings and thus express different propositions. It is thus possible for someone to believe that 'you are in front of a wild bear' without believing that 'I am in front of a wild bear'. The truth values of these two beliefs might differ because they do not have the same meaning. If the second is believed by me, and the first by my friend referring to me, the two beliefs will have the same truth values and thus on Frege's view, the same meaning.

But these two beliefs clearly have different cognitive significance and seem to have different meanings.

The Kaplan–Perry theory of belief claims that belief involves not only a believer and a proposition believed, but also the belief state one is in. The belief state is a way of believing, the *character* of one's belief, which is fixed by the conventional rules of use of indexicals. The linguistic function of the temporal indexical accounts for the cognitive difference between the two beliefs. On this view, when Perry forms his A-belief that the lecture is tomorrow, the proposition believed is the B-proposition that the lecture is on 12 March 2015. He is also in a certain belief state such that anytime he was asserting the above A-belief he would have been in the same belief state – believing that the lecture starts *tomorrow*. On this view, the cognitive significance is to be explained by the belief state one is in, not the content of the belief. Suppose that the belief that 'the lecture is tomorrow' is formed in 1990 instead of 2015. Although the proposition believed is a different B-proposition, the young Perry was nevertheless in the same belief state as the old Perry and was disposed to behave in similar ways. The content of the that-clause in a belief is separate from the character of the belief. Cognitive significance of a belief is tied with the character, not the propositional content. Sameness of character leads to the same action. Therefore, (ii) is false because beliefs with the same propositional content but different characters do not lead to the same actions. Apprehending the same B-proposition via tensed sentence would lead to actions different from doing so via tenseless sentence. But neither does it imply a tensed ontology nor the irreducibility of tensed beliefs.

One objection says that a tenseless ontology cannot account for our experience, in this case the characters of A-beliefs. As Craig put it, the characters of beliefs with tenseless propositional content must be really tensed, otherwise "one comes away with no more cognitive significance" than that of the Fregean view of belief.[18] He thinks that tensed facts are necessary for the cognitive significance of A-belief, even if those facts are non-propositional, i.e. the facts that are represented by tensed characters.

But the characters of A-belief need not be tensed. In fact, it does not really make sense to say that the character of certain belief is tensed. Because a character is the conventional rule that governs the use of an indexical in a given context. A rule itself cannot be tensed, otherwise it can only work for a specific moment and then remains useless for the rest of time. Furthermore, why

[18] Craig [4] p. 257.

must they have any ontological implication in order to explain the cognitive significance? Think about a case involving the personal indexical 'I'. We need not posit 'I-facts' to explain the I-character of the belief that 'I am about to be attacked by a bear' that is responsible for subsequent desires and actions of the person who has that belief. I screamed and I ran away as fast as I could (but should you encounter a bear in the wild, do stay calm and walk away slowly). Anyone who has such belief and has no knowledge of what to do when encountering a bear in the wild might be in the same psychological state as I am and behave in this silly way. The I-character of the belief does have a cognitive significance for the belief '*X* is about to be attacked by a bear' (suppose that '*X*' refers to me). But we need not appeal to objective selfhood to explain such character. This is just the way we use the language. Similarly, the characters of A-beliefs can explain for the cognitive significance without themselves being tensed or positing any tensed fact.

But how can we use these linguistic rules correctly? Is the individuation of the characters of tensed beliefs possible without positing a tensed ontology? The A-theorist might appeal to externalism about mental content and contend that there cannot be characters that are cut off from the actual world. Externalism claims that to have certain types of intentional mental states, one should be related to the external world in the right way. One simple example would be our perceptual experience. To have a mental state of seeing a table, one must be causally related to the table, i.e. actually perceiving a table in front of him. The person's having this mental state is individuated by some fact about the external world he is in. Let us first explain the individuation of the character of belief that has the personal indexical 'I'. Rather than claiming that characters are cut off from the world, we think that the conventional use of the indexical arises from the constant interaction between the speaker and the external world. I realize that my experiences are different from that of others because I can see and hear things that others cannot, i.e. I can hear my inner voice reciting a poem but you can't. Similarly, for other people, there are experiences that are unique to them and are under the marker of 'I'. I observe how other people use the indexical to refer to themselves. When you said "I met Ishiguro at the Buried Giant book launch tonight", I wouldn't think that 'I' referred to me, for I stayed at home writing all the night. I would correctly understand, through past experience of thousands of instances like this, that you were talking about yourself. By such training, I grasp the rule that 'I' refers to the person who utters the word. It is not the property of selfhood that enables me to correctly apply the linguistic rule.

To grasp the rule for temporal indexicals is just like that for personal indexicals. Temporal perspective is no more a mystery than the fact that our con-

scious experience is subjective and is from a particular point of view. I have suggested that A-belief is reducible to B-belief but has distinct character. Take the A-belief 'the Annual Lecture is tomorrow' as an example. Let us call the character of this belief the A-character, which is the B-proposition 'the Annual Lecture is on 12 March' believed via the tensed sentence. I realize that my current perceptual experience differs from both memory and anticipation. Our temporal perspective is limited in the way that we only have conscious temporal experiences when we are conscious. That is, I can only perceive the world from the perspective of one of my temporal parts which occupies a specific spatio-temporal region in the four-dimensional world. (The A-theorists, especially the presentists, tend to reject four-dimensionalism, the view that a person persists through time in virtue of having a temporal part at each moment of his existence. But the view is growing in popularity and fits naturally with the B-theory of time.) I remember that in the past, say at t_0, I had a similar experience that the conscious experience through the perspective from which I was interacting with the surrounding environment was different from memory and anticipation. These experiences I have at the time I am conscious will be under the marker of 'now'. By such training, I grasp the rule that 'now' refers to the time at which I have the thought. Similarly, Perry also grasps that it is from his unique temporal perspective he is interacting with the surrounding environment. He was informed about the date by looking at the calendar. He came to be in certain belief-state on 11 March such that the B-belief he knows all the time has the A-character which prompts him to book the flight to London and so on. His acceptance of the B-belief via the tensed sentence is caused by some events or objects he perceived, i.e. the calendar, and by his self-referential awareness of his conscious experience, not by A-facts.

6 Conclusion

In this paper, I distinguished two ways in which our experiences can be associated with temporal properties. I presented the phenomenon of the presence of experience in terms of the presentness of consciousness and the presentness of content. I introduced the problem of the presence of experience as formulated by Mellor, whose solution can be countered by A-theorists in two ways. One way is to disagree with his formulation of the problem in the first place. A-theorists might reformulate the problem in terms of occurring *simpliciter*. The other way is to argue that Mellor's solution is inadequate to account for the phenomenon because we seem to learn a new fact when we form a true now-belief. I contended that the argument from occurrence fails because (1) occurrence is not distinctive from presentness thus we need not reformulate

the problem in terms of occurrence; (2) more generally, the property of being present is not a sensible property. I also contended that the temporal knowledge argument fails because B-facts are all there are and we can explain the difference between the cognitive significance of A-beliefs and B-beliefs without appealing to A-facts. Therefore, Mellor's solution to the problem of the presence of experience remains plausible. The presence of experience does not favour the A-theory of time.

References

[1] Balashov, Y. (2005) 'Times of Our Lives: Negotiating the Presence of Experience', in *American Philosophical Quarterly* 42/2: 296–309.

[2] Butterfield, J. (1984) 'Seeing the Present', in *Mind* 93/370: 161–361.

[3] Callender, C. (2008) 'The Common Now', in *Philosophical Issues* 18/1: 339–361.

[4] Craig, W.L. (1996) 'The New B-Theory's Tu Quoque Argument', in *Syntheses* 107/2: 249–269.

[5] —— (2001) 'Wishing It Were Now Some Other Time', in *Philosophy and Phenomenological Research* 62/1: 159–166.

[6] Davidson, D. (1970) 'Mental Events', in L. Foster & J.W. Swanson (eds.) *Experience and Theory*. London: Duckworth.

[7] Dennett, D.C. (1991) *Consciousness Explained*. Boston: Little, Brown.

[8] Eagleman, D.M. (2008) 'Human Time Perception and Its Illusions', in *Current Opinion in Neurobiology* 18/2: 131–136.

[9] Hestevold, H. (1990) 'Passage and the Presence of Experience', in *Philosophy and Phenomenological Research* 50/3: 537.

[10] Keller, S. & Nelson, M. (2001) 'Presentists Should Believe in Time-Travel', in *Australasian Journal of Philosophy* 79: 333–345.

[11] Macdonald, C.A. (1989) *Mind–Body Identity Theories*. London: Routledge.

[12] McTaggart, J.E. (1908) 'The Unreality of Time', in *Mind* 17/4: 457–474.

[13] Mellor, D. (1998) *Real Time II*. London: Routledge.

[14] Nagel, T. (1965) 'Physicalism', in *Philosophical Review* 74: 339–356.

[15] Perry, J. (1979) 'The Problem of the Essential Indexical', in *Noûs* 13/1: 3–21.

[16] —— (1997) 'Indexicals and Demonstratives', in B. Hales & C. Wright (eds.) *A Companion to Philosophy of Language*. Oxford: Blackwell.

[17] —— (2001) 'Time Consciousness and the Knowledge Argument', in L.N. Oaklander (ed.) *The Importance of Time*. Dordrecht: Kluwer Academic Publishers.

[18] Schlesinger, G.N. (1991) 'E pur si muove', in *The Philosophical Quarterly* 41: 427–441.

[19] Stetson, C. et al. (2006) 'Motor-Sensory Recalibration Leads to an Illusory Reversal of Action and Sensation', in *Neuron* 51/5: 651–659.

The Rule of Law and The Right-to-Have-Rights: An Arendtian Defence of Bingham's Substantive Definition of The Rule of Law*

Saw Watts
University of Bristol

1 Introduction

The rule of law is the implicit principle of legality that is present in all legal systems to a greater or lesser degree. The modern use of the concept was introduced by Dicey, although a precise definition is the subject of great contention. This paper shall focus on one aspect of the greater debate and consider the extent that the rule of law must protect human rights. Specifically, this paper attempts to defend the substantive claim that human rights must be protected by the rule of law; as such, this paper may be understood as an interpretation of the substantive position, which is then defended against formalist objections.

The first section shall seek to explore the foundations of this particular debate through a more general comparison between two competing definitions of the rule of law. In addition to providing some context for this paper, the section introduces the debate on whether or not the rule of law *should* protect human rights. This section roots the question in the dichotomy between legal positivism and legal naturalism; whilst positivists assert that laws may be judged by the conditions of their creation, naturalists judge laws against their conformation to some independent moral standard. The formal/substantive debate reflects this more fundamental disagreement; such that formalist arguments reflect the positivist claim that laws should only be judged based on the ways they were created.

After exploring the formalist argument, I suggest that a response may be found through a critical consideration of human dignity. I introduce the Arendtian account of human rights and human dignity, and suggest that, to the extent that her account is plausible, the rule of law must protect human rights. Thus,

*Delivered at the BJUPS Autumn Conference, 21–22 November 2015 at the London School of Economics.

my contribution is the application of the Arendtian understanding of human rights to the formalist/substantive debate; although its value as a solution is conditional on its plausibility, I argue that it is not inconsistent with either formalist or substantive positions. I conclude by claiming that, on at least one plausible account of human rights, they must be protected by the rule of law.

2 Defining the Rule of Law

Defining "the rule of law" is a matter of great contention, to the point where "the ubiquity of usage of the expression, is matched only by the multiplicity of its meanings".[1] Certain features may be identified without controversy. Bingham suggests that "the core of the existing principle is…that all persons and authorities within the state whether public or private, should be bound by and entitled to the benefit of the laws publically made, taking effect (generally) in the future and publicly administered in the courts".[2] So understood, the rule of law articulates an implicit principle of legality, that no man should be above the law.

Bingham develops his definition with eight further principles for the rule of law. Unlike the essential definition given above, these eight principles are more controversial:

(B1) The law must be accessible, and so far as possible intelligible, clear and predictable.

(B2) Questions of legal right and liability should ordinarily be resolved by application of the law and not the exercise of discretion.

(B3) The laws of the land should apply equally to all, save to the extent that objective differences justify discretion.

(B4) Ministers and public officers at all levels must exercise the powers conferred on them in good faith, fairly, for the purpose for which the powers were conferred, without exceeding the limits of such powers and not unreasonably.

(B5) The law must afford adequate protection of fundamental human rights.

[1] Loughlin [9] p. 2.

[2] Bingham [4] p. 8.

(B6) Means must be provided for resolving, without prohibitive cost or inordinate delay, *bona fide* civil disputes which the parties themselves are unable to resolve.

(B7) Adjudicative procedures provided by the state should be fair.

(B8) The rule of law requires compliance by the state with its obligations in international law as in national law.[3]

This paper shall focus on the controversy surrounding B5, or the principle that the law must protect fundamental human rights.

Raz offers an alternative definition; for Raz, the rule of law has two aspects. Firstly, the law has authority over people and they obey it; secondly, the law should be able to effectively guide people.[4] These two aspects reflect the fundamental intuition that "the law must be capable of guiding the behaviour of its subjects".[5] From this general intuition, Raz goes on to deduce the following principles:

(R1) All laws should be prospective, open, and clear.

(R2) Laws should be relatively stable.

(R3) The making of particular laws (particular legal orders) should be guided by open, stable, clear, and general rules.

(R4) The independence of the judiciary must be guaranteed.

(R5) The principles of natural justice must be observed.

(R6) The courts should have review powers over the implementation of the other principles.

(R7) The courts should be easily accessible.

(R8) The discretion of the crime-preventing agencies should not be allowed to pervert the law.

The above principles are plainly different from those offered by Bingham. They may be subdivided into two groups. (R1)–(R3) describe how the law should conform to standards that allow it properly guide action; (R4)–(R8)

[3] Bingham [4] Chapter 3–10.

[4] Raz [12] p. 213.

[5] Raz [12] p. 214.

show how various legal mechanisms allow adequate guidance in cases of distorted enforcement etc. ensuring adequate conformity to the rule of law and properly dealing with cases of deviation from it.[6]

How can we explain this difference? Raz and Bingham offer radically different explanations of what the rule of law includes; this, in turn, reflects substantially different understandings of what the rule of law is and the purpose that it is meant to serve.

Bingham implies that the rule of law should reflect "that fundamental truth . . . 'wherever law ends, tyranny begins'".[7] On Bingham's definition, the rule of law reflects a commitment to meeting some moral standard. For example, he assents to the existence of a "constitutional principle of the rule of law".[8] Bingham treats adherence to the rule of law as a moral virtue in itself, evidenced by his provision for the protection of human rights. In this regard, we may identify his position as a legal naturalist; his definition suggests that there is an independent standard against which positive law should be measured, one that includes the preservation of human rights. The rule of law epitomizes "an ideal worth striving for",[9] to the extent that conformity to it creates a just society.

Comparatively, Raz advocates a view of legal positivism; unlike Bingham who views the rule of law as moral, Raz takes a formal instrumentalist stance. The purpose of the law is not to create a just society; its primary function is to guide behaviour. The rule of law is nothing more than a principle which states that laws should be created in the correct legal manner, such that they are capable of guiding the actions of the individuals to which they apply. Unlike Bingham, the rule of law is only one virtue that a legal system may possess. Thus, the virtue of the rule of law is not the virtue of a legal system, or to be judged by its ability to reflect some constitutional principle; it "has a specific virtue that is morally neutral in being neutral as to the end to which the instrument is put".[10] The value of conforming to the rule of law is restricted to formal terms, or its ability to guide behaviour.

The rule of law is like other instruments, to be judged against its ability to per-

[6] Raz [12] p. 218.

[7] Locke quoted in Bingham [4] p. 8.

[8] Bingham [4] p. 67.

[9] Bingham [4] p. 174.

[10] Raz [12] p. 226.

form the function for which it was intended; it is "the virtue of efficiency; the virtue of the instrument as an instrument"[11]. It is not to be judged by its purpose as such but rather by its capacity to fulfil its function. Failure to comply with (R1)–(R8) would make it impossible for the law to guide human action; adherence to the rule of law may still be a virtue, but only in an instrumental sense.[12]

We can now see the distinction between the two previous definitions of the rule of law. While Bingham suggests that the virtue of the rule of law is in its creation of a just society, Raz gives a formal instrumental definition. This disagreement between Raz and Bingham, formal and substantive, has far reaching implications; disagreement about the purpose and exact nature of the rule of law can only limit its overall value as a practical concept.

However, such fundamental disagreement does not justify discounting the rule of law as an important legal and political principle. Two things can be said in its defence. Firstly, it is universally recognized that the rule of law is a normative value; it indicates how a nation's legal system should function. Thus, conforming to the rule of law is an aspiration that may or may not be satisfied completely. To the extent that a better defined concept is better applied, precisifcation of the definition of the rule of law makes its application more feasible. As Dworkin suggests, "governments and peoples agree that there is an important value named by democracy or liberty or the rule of law, though they disagree, perhaps profoundly, about exactly how that value should be identified".[13]

Secondly, adherence to the rule of law is linked to the preservation of human dignity. This is the focus of later sections, and recognition of this emphasizes the importance of the rule of law. Whilst this appears inconsistent with Raz's instrumental definition, this paper shall attempt to show that human dignity, as expressed by the protection of human rights (B5), is not only something that should be protected by the rule of law but is also a necessary condition for the rule of law's ability to guide behaviour.

---

[11] Raz [12] p. 226.

[12] Loughlin [9] p. 13.

[13] Dworkin [6] p. 5.

3 The Protection of Human Rights

Given the fundamental disagreements described in the previous section, the protection of human rights offers a useful point of comparison. Its inclusion in Bingham's definition, and exclusion in Raz's, highlights its potential to reflect the fundamental beliefs of each particular conception of the rule of the law. By considering their arguments we may access these fundamental beliefs and achieve reconciliation.

Arguably, Dicey did not include the protection of human rights in his original definition of the rule of law; this, along with differences in its supposed purpose has led to serious disagreement. Despite this controversy, the rule of law is often linked with the protection of human rights. Perhaps most explicitly, the Universal Declaration of Human Rights states "[I]t is essential, if man is not to be compelled to have recourse, as a last resort, to rebellion against tyranny and oppression, that human rights should be protected by the rule of law."[14]

This common link between the protection of human rights and the rule of law suggests that arguments to the contrary merit close consideration. This paper will first outline Raz's specific argument against this connection and then introduce Bingham's response.

3.1 Formal Definitions: Raz's Argument

Raz, in line with his formal definition, argues that the rule of law does not protect human rights. He argues that it is quite possible for a state to conform to the rule of law and yet still carry out systematic human rights abuses. Such a state "will be an immeasurably worse legal system, but it will excel in one respect: in its conformity to the rule of law".[15]

Many historical examples support Raz's claim. For example, the terrible human rights abuses committed by Nazi Germany may be seen to conform to the rule of law. In this case, the law expressed the will of the German nation, manifested in the office of the Führer; the law existed in a purely statutory sense, complying with the formal definition of the rule of law. This legal system, complying with this definition of the rule of law, was nonetheless able to

[14] [13] Preamble.

[15] Raz [12] p. 211.

commit human rights atrocities on a horrendous scale.[16]

Raz is not condoning human rights abuses; he merely suggests that the rule of law "is compatible with gross violations of human rights".[17] Conforming to the rule of law is only one of the virtues a legal system may possess and, as such, a legal system may conform to the rule of law and fail in all other regards.

Such examples show the implications of a formal position; It is useful to introduce Raz's distinction between "the rule of law" and "the rule of good law". Whilst Raz follows the formal tradition and separates the two, substantive positions (such as Bingham's) often collapse the distinction and view the rule of law as a moral virtue in itself. Substantive positions use the rule of law as a basis for further rights against which the content of laws may be judged, formalists are only concerned with whether the formal principles are met.[18]

Raz justifies this separation as follows. One of the primary virtues of the formal conception of the rule of law is "that there are so many values it does not serve"; the law is morally neutral and is judged by its ability to perform the function of guiding behaviour to whatever end.[19] If we deny this and assert that the rule of law should include other values and produce "good" laws, then the concept lacks any independent function. The constitution of a just society, the nature of rights, and the extent of governmental action are all important issues to be considered in their own right; there is no need to cloak these individual questions and their conclusions in the rule of law.[20]

Thus, the scope of the rule of law is limited to the formal mechanisms of the law, its processes and features. It follows that an actively immoral regime may totally comply with the rule of law, so long as the required formal principles are upheld. The fact that conformity to the rule of law may allow 'bad' or immoral laws to function effectively does not challenge this position; "that a sharp knife can be used for harm does not show that a being sharp is a good making-characteristic for knives" – the fact that the legal system can produce immoral laws does not necessarily mean that the legal system is not function-

[16] Zorkin [15] p. 47.

[17] Raz [12] p. 221.

[18] Craig [5] p. 1.

[19] Raz [12] p. 219.

[20] Craig [5] p. 2.

ing correctly.[21]

3.2 Substantive Definition: Bingham's Response

Bingham defends a substantive position, asserting that the law must afford adequate protection of fundamental human rights (B5). His argument is given along naturalist lines; "a state which savagely represses or persecutes sections of its people cannot in my view be regarded as observing the rule of law, even if . . . [the repression/persecution] is the subject of detailed laws duly enacted and scrupulously observed".[22] Since the purpose of the rule of law is the creation of a just society, an unjust society does not conform to the rule of law.

Conforming to the formal mechanisms of the rule of law is insufficient, "law is the essence of freedom"[23] and a state that denies freedom does not conform to the rule of law. To the detriment of his argument, Bingham seems to rely on the fact that current legal-political bodies often incorporate the respect for human rights within their definition of the rule of law and in practice. Although a link is officially recognized, the fact of its recognition does not justify his claims.

While Bingham's substantive position could be strengthened through close analysis of this official recognition, the required level of detail is beyond the scope of this paper. Instead, this paper shall recognize and attempt to address several problems with Bingham's account in *The Rule of Law*. Solutions to these problems will strengthen Bingham's account both in its own right, and as a counter proposal to the formal understanding advocated by Raz.

Firstly, Bingham recognizes a common problem in creating a general principle for the protection of human rights. Noting the cultural relativity of certain rights and freedoms, he states that "there is no universal consensus" on which rights are fundamental. Whilst versions of the rule of law exist in many societies, there is no inter-societal agreement on which rights are fundamental; for example, "in some developing countries a higher premium is put on economic growth than on the protection of individual rights".[24] Thus, it is very difficult

[21] Raz [12] p. 225.

[22] Bingham [4] p. 67.

[23] Zorkin [15] p. 47.

[24] Bingham [4] p. 68.

to justify how and why the rule of law should protect human rights if those rights are not clearly defined.

Secondly, Bingham does not offer a response to Raz on his own terms; he fails to demonstrate why we should include the protection of human rights within the rule of law. His argument does not provide a convincing reason for collapsing the distinction between "the rule of law" and the "the rule of good law", showing why protected human rights are integral to the functioning of a legal system. Such a response is as much a defense of legal naturalism as a defense of the substantive definition of the rule of law; to defend the claim that the rule of law must protect human rights, we must establish a connection between the rule of law as a legal principle, and the rule of law as a means for a just society.

3.3 Common Elements: The Preservation of Dignity in the Formal and Substantive Definitions

The goal is to offer a defense of Bingham's substantive position, justifying the preservation of human rights as a key principle of the rule of law. One approach would be to expand on the 'constitutional principle', developing the importance of this principle to the extent that we must consider it in all legal and political practice.[25] However, this would necessitate deviation from the main subject at hand and further emphasize the divisions present at the heart of the rule of law.

Dworkin indicates a better approach, stating that "disagreements about the character of the rule of law reflect even deeper disagreements about …the concept of human dignity".[26] The concept of personal dignity offers a shared normative principle that both Raz and Bingham seek to foster in their respective definitions.

Understood in this way, we must look at the intent of both positions to preserve human dignity; that is, if the rule of law fails, human dignity has been lost and a tyranny that exists above the law has triumphed. For Raz's position to make sense, "he must think that in some respect or to some degree his tyranny is better than it would be if it did not legislate openly and execute its

[25] Bingham [4] p. 67.

[26] Dworkin [6] p. 5.

laws faithfully".[27] Raz implies that a rule-following tyranny is better than one that is unrestricted; the benefit of the rule of law is the benefit of this former tyranny over the latter, or its ability to guide behaviour. Comparatively, Bingham thinks that a rule following tyranny is no better than one which fails to respect the rule of law.

This difference reflects different understandings of dignity. For Raz, dignity only demands a fair warning of when the government will interfere in people's behaviour; this account is consistent with formal positions, only demanding that "the law be clear and publicized, that it be capable of being complied with, and that institutions like independent courts exist with procedures well calculated to adjudicate whether the warning has indeed been ignored".[28] On this account, the rule of law can be defined as distinct from any other virtue in formal terms.

Bingham makes a further demand: "that all those subject to coercive dominion are equal before the law in a deep sense".[29] This account demands that the principles, which justify how some are treated, must be applied to all people equally. Thus, the benefit of the rule of law cannot be defined in purely formal terms; equality requires both "equal participation in law making and respect for a variety of political rights", such that the rule of law is not sufficient.[30]

At this stage, we can compare the two positions through their understanding of human dignity. A response to Raz's formalist position is required to defend the second, egalitarian, understanding of dignity; this understanding of dignity must be related back to human rights and, if possible, recognize that there is no universal agreement on the individual rights that may express this human dignity. We must show that fair warning of human rights abuses is insufficient to preserve human dignity, and show that such abuses fail to recognize human dignity. Thus, the aim of this paper is to give an account of human dignity with the following features:

- Human dignity is present in all people equally, to the extent that all people are equal before the law; that is, the account reflects Bingham's substantive position.

[27] Ibid. p. 6.

[28] Ibid. p. 9.

[29] Ibid.

[30] Dworkin [6] p. 10.

- Human dignity grounds human rights, such that the protection of human rights is in some way analogous to the protection of human dignity.
- The denial of an individual's dignity is incompatible with the rule of law being respected; that is, human dignity demands that the rule of law be respected. Thus, the account defends the substantive position.

If these conditions are met, then we can compellingly defend the substantive claim that the rule of law must protect human rights; later sections argue that the protection of human rights may be reduced to the protection of human dignity, and that the formal definition of the rule of law is not sufficient for this.

4 A Dignitarian Rule of Law: How does the Rule of Law Respect Human Dignity?

It is useful to briefly clarify the extent that dignity features in Raz's account. As we have seen, and as Dworkin notes, ordinary speakers give different definitions of the rule of law. Despite debates about the content of the rule of law, we must accept "that the rule of law is a value" and we must define it in such a way as to reveal what this value is.[31]

Raz recognizes this and elaborates the value of the rule of law; crucially, he notes, "observance to the rule of law is necessary if the law is to respect human dignity". This involves recognizing and respecting an individual's autonomy, "as individuals capable of planning and plotting their future".[32]

This entails that the law has the potential to violate human dignity, though this does not mean that the rule of law guarantees its protection. The most that is granted is that "a legal system which does in general observe the rule of law treats people as persons at least in the sense that it attempts to guide their behaviour through affecting the circumstances of their actions".[33] The rule of law is a negative value; the rule of law minimizes the danger created by the law itself, requiring that the law should be open, clear, and stable, and interpreted by an impartial judiciary. This is not sufficient to claim that the rule of law guarantees the protection of human dignity; for Raz, the rule of

[31] Dworkin [6] p. 6.

[32] Raz [12] p. 221.

[33] Ibid. p. 222.

law does not admit any responsibility for the content of the law. The most that we can claim on Raz's account is that "deliberate disregard for the rule of law violates human dignity" to the extent that the total absence of the rule of law is a total absence of protection.[34]

This is radically different to the normative claim that a substantive definition reflects; to defend a substantive version of the rule of law we must not only show that the rule of law guarantees the protection of human dignity, but why that should be the case.

How can this be achieved? We must provide an account of human dignity that necessarily follows from conforming to the rule of law, or is coextensive with it. We must show that human rights are inseparable from human dignity and that these rights, which ought to be protected by the rule of law, are held by all. In short, we need an essentialist dignitarian notion of human rights such that they are entailed by the rule of law. Only then can we convincingly respond to Raz's formal instrumentalism.

Thus, the remainder of this paper may be understood as a direct defense of a substantive rule of law; the following sections shall attempt to outline the above notion of human rights and human dignity, to the extent that they must be accounted for even with a formally defined rule of law.

5 Hannah Arendt: 'The Right-To-Have-Rights'

How then are we to understand human rights? Recent interpretations of Hannah Arendt have considered the importance of human dignity in her account of human rights; following Dworkin's *The Rule of Law*, we can apply this understanding of human rights/human dignity in defense of the substantive position. Although it is rarely mentioned, Arendt explicitly states that "human dignity needs a new guarantee which can only be found in a new political principle, in a new law on earth, whose validity this time must comprehend the whole of humanity while its power must remain strictly limited, rooted in and controlled by newly defined territorial entities".[35] This lends itself to a substantive interpretation of the rule of law, such that the rule of law positively guarantees the dignity of its subjects.

[34] Ibid. p. 221.

[35] Arendt [1] p. ix.

Her account also accommodates the lack of consensus on human rights, offering a response to the first of the problems faced by Bingham's substantive account.[36] She can defend an essentialist account of human rights that should be protected by the rule of law, even though these rights may vary in content and importance.

Writing in response to the human rights atrocities of the twentieth century, Arendt is critical of the relationship between politics and human rights. Her argument is too detailed to be explored here; it is sufficient to note that she explores the inevitable tension between political systems and human rights. We shall first explore the distinction between civil and (more fundamental) human rights; this forms the basis of her account of 'the right-to-have-right'. It is important to note that Arendt saw human rights as essentially political, rather than moral or legal.[37] Given Bingham's inherent legal naturalism, this political understanding may show how the rule of law is equivalent with the rule of good law.

5.1 'Human Rights' as Civil Rights

In direct criticism of the Universal Declaration of Human Rights, Arendt recognizes that the human rights which are often taken as universal depend on the political systems that create them. Such rights take their authority from mankind and were thought to be inalienable, to the extent that "no law was needed to protect them because they were the source of all other laws".[38] These rights were created politically, as "'rights' [that] only exist by virtue of legislation".[39] Such rights are limited to the political system that legitimized them, or the people that assent to their existence.

Thus, essentially civil rights (the rights of the citizen) were misidentified as the 'rights of man'; human rights became "equivalent to the rights of a people under the protection of a government".[40] The effects of this misunderstanding were the human rights abuses and totalitarianism of the twenthieth century. For example, in the Nazi denial of Jewish citizenship, the "loss of national

[36] cf. §3.2.

[37] Isaac [8] p. 61.

[38] Parekh [11] p. 44.

[39] Menke [10] p. 335.

[40] Parekh [11] p. 44.

rights was identical with the loss of human rights, that the former inevitably entailed the latter".[41] If the political system is the sole source and guardian of 'human rights', then the denial of membership to that political system is equivalent to the loss of these rights.

Her claim that 'the rights of man' are really the rights of citizens articulates Bingham's earlier concern. For example, many of the rights explored in the European Convention of Human Rights seem to reflect the current values and capabilities of the EU. When we talk about human rights, we actually talk about "the rights of each member of a political community and thus of rights which also can only pertain to the members of this political community".[42]

5.2 The Right-to-Have-Rights

Recognising this and alleviating Bingham's concern, Arendt suggests that the only fundamental right is the right-to-have-rights. Unlike the so-called 'rights of man', the only right we have in virtue of our humanity is this right-to-have-rights. The rights commonly understood as inalienable are, in fact, "the rights of man within a community".[43]the only truly fundamental right is the right-to-have-rights.

This is the sole human right, drawn from the misappropriation of civil rights as human. At its simplest, the right-to-have-rights may be understood as an individual's entitlement to civil rights. However, more fundamentally, the loss of this right-to-have-rights results in the deprivation of our humanity as we lose "a place in the world which makes opinions significant and actions effective".[44]

Whereas our civil rights define our position within society, this right-to-have-rights refers to our right to membership of that society; as such, it may be seen as the basis for our civil rights. It follows that a substantive definition of the rule of law may justifiably defend the possibility of 'human rights' (or civil rights in the Arendtian sense) through the protection of this right-to-have-rights. If we accept that (B5) refers to the protection of the sole human right, the-right-to-have-rights, then it follows that the rule of law may conceivably protect all other rights by implication. Thus we may claim:

---

[41] Arendt [2] p. 292.

[42] Menke [10] p. 336.

[43] Parekh [11] p. 45.

[44] Arendt [2] p. 296.

(B5*) The rule of law must afford adequate protection to the fundamental human right-to-have-rights.

Though Arendt's understanding of human rights is unorthodox, it does not mischaracterize Bingham's argument; a principle, which protects the right-to-have-rights, may protect other rights *in potentia*, whilst accounting for a lack of universal consensus on their particular content.

Subsequent sections develop this account, exploring how this right-to-have-rights is a reflection of our fundamental humanity. This can be applied to Raz's formalist argument, arguing that the rule of law must guarantee this fundamental human right.

6 The-Right-To-Have-Rights: An Expression of Human Dignity

We have already noted how the typically declared 'human rights' are contingent on the political systems that formed them; an individual is not endowed with these rights by virtue of their status as a human being, rather as a member of a given nation/polity.[45] The only properly human right is the right-to-have-rights, the denial of this being a loss of our humanity.

Justifying how the loss of this right results in the loss of our humanity serves two functions. Firstly, it further establishes the right-to-have-rights as the only fundamental human right; secondly, it allows us to respond to Raz by establishing that this right is an expression of our basic human dignity that must be protected by the rule of law.

The right-to-have-rights does not just refer to our entitlement to civil rights; it reflects our right "to live in circumstances where action and speech are meaningful". The importance of this is closely bound up with Arendt's understanding of the essential nature of humanity; these circumstances do not just refer to belonging in a society but are "equivalent to humanity itself".[46]

Essentially, humans are politico-linguistic beings. Human dignity it not a natural property, but a consequence of an individuals intersubjective, linguistic, existence within a political society.[47] This claim reflects Arendt's concept of

[45] Menke [10] p. 336.

[46] Parekh [11] p. 48.

[47] Menke [10] p. 340.

the '*vita activa*', which links the three conditions of labour, work, and action with the fundamental nature of our existence.[48] Action, accompanied by speech, "is the only activity that goes on directly between men without the intermediary of things or matter".[49] The essential role of speech and action is understood as self-disclosing; through speech and action we identify ourselves and attain humanity in the fullest sense.[50]

This action cannot occur in isolation, but depends on a plurality of individual beings; that is, the extent to which speech and action contribute to an individual's identity depends on intersubjective agreement.[51] Arendt recognises the relationship between equality within a political system, and the distinction of an individual within that system. Equality is based on the intersubjectivity of this political system, such that we can communicate and understand each other; through this we can disclose ourselves as distinct individuals and be constituted as such within a political society.[52]

Speech and action reveal the self and as such are fundamentally human capacities; that is, a person cannot refrain from this whilst still maintaining their humanity. For Arendt, this commits us to an understanding of humanity as essentially political-linguistic beings; it reflects the disposition of humanity and "a general characteristic of the human condition".[53] Our human dignity exists in this political-linguistic condition, because we are political-linguistic by nature.

It follows that the denial of our right to live in circumstances where action and speech are meaningful[54] is equivalent to the denial of our human dignity. The right-to-have-rights expresses our fundamental human nature as political-linguistic beings, identifying ourselves through speech and action in an intersubjective political system. Importantly this right, by virtue of our human nature, is political in recognising the importance of our membership in a political community; however, it is undeniably essentialist in its "insistence that human dignity must be maintained and that this can only be done through the

[48] Arendt [3].

[49] Ibid. p. 7.

[50] Parekh [11] p. 46.

[51] Ibid. p. 47.

[52] Ibid.

[53] Arendt , quoted in Menke [10] p. 340.

[54] See Parekh [?] p. 47.

guarantee of participation in a political community".[55]

This account of human rights is consistent with the first two conditions given in §3; the sole human right, or the right-to-have-rights, is grounded in human dignity and universally applicable to all people by virtue of their humanity. Thus, we may understand (B5*) as follows:

(B5*) The rule of law must afford adequate protection to the fundamental human right-to-have-rights; that is, the rule of law must respect the political-linguistic nature of an individual and protect their right to live in circumstances where action and speech are meaningful.

7 The Right-to-Have-Rights and The Rule of Law

It stands to be seen how the rule of law is meant to protect the right-to-have rights. Arendt does not mention the rule of law explicitly, and as such some may question the relevance of her account. However, she does offer a solution to the problem of protecting human rights in *The Origins of Totalitarianism*. She writes:

> We are not born equal; we become equal as members of a group on the strength of our decision to guarantee ourselves mutually equal rights. Our political life rests on the assumption that we can produce equality through organization, because man can act in and build a common world, together with his equals and only with his equals.[56]

Equality does not exist in nature, but only within the political community; if human rights are possible at all, it is only through political institutions.[57] In recognizing their political grounding, we find that rights are not automatic structures but depend on a decision to guarantee each other's mutually equal rights. As Parekh points out, the political realm is intersubjective, such that "without our political equals there is no possibility for genuine action and speech"; thus, the protection of the right-to-have-rights is essential for genuine politics. Without human rights, there could be no politics, only "coercion and

[55] Helis [7] p. 75.

[56] Arendt [2] p. 301.

[57] Parekh [11] p. 50.

confrontation".[58]

If human rights are possible at all they are, by the very nature of humanity, political. The rule of law exists as a means of safeguarding the right-to-have-rights and the equality of all within the law; this equality is produced through the structures laid out in principle by the rule of law. As such, the rule of law provisions for the generation and protection of future civil rights. It functions as a universal safeguard of human dignity, such that grants individuals the "right to belong to a genuine political community to resolve issues of rights".[59]

Given Arendt's account of human rights and human dignity, we can now apply this to the formal, substantive debate. We must show that the denial of an individual's dignity is incompatible with the rule of law being respected; that is, human dignity demands that the rule of law be respected.

However, we must first give reasons why Raz would accept this account as plausible; we must suggest why he should not reject Arendt's understanding. Arendt's account is coherent with Raz's formalist claims in several ways. Firstly, in giving a political account of human rights, Arendt presents a morally neutral account that nonetheless preserves the importance of these fundamental rights. Thus, this account can recognize the intuitive importance of human rights whilst remaining consistent with the morally neutral instrumentalist definition of the rule of law.

Secondly, Arendt "moves away from the problematic idea of natural or inalienable rights" found in traditional accounts.[60] Nonetheless, her account is essentialist in basing these rights in human dignity (via the right to membership in a political community). The importance of membership links her account to Raz's suggestion that the rule of law guides behaviour, such that the law itself is only applicable to members of that particular legal system.

This being the case, we may now defend the substantive definition. The validity of this response is conditional on the extent to which Arendt's account is plausible; if we accept this we can claim that, on at least one plausible account, the rule of law requires the protection of human rights.

---

[58] Ibid. pp. 59, 50.

[59] Helis [7] p. 75.

[60] Helis [7] p. 73.

8 The Protection of Human Rights Revisited: A Response to Raz

In §7 we applied Arendt's dignitarian account of human rights to modify B5; the rule of law must protect the fundamental right-to-have-rights as the only properly human right. We have constructed an Arendtian understanding of the substantive position that protects human rights. This final section shall offer a direct response to Raz's criticisms, drawing on the themes explored in the previous sections.

Recall Raz's argument against (B5); He argues on two fronts; firstly, he argues that the rule of law is a virtue in terms of its ability to perform its function, or guide behaviour. This being the case, he suggests that it is quite possible for a state to commit gross human rights abuses whilst exemplifying the rule of law. A response to Raz will deal with both of these points; it will show that the rule of law must respect human rights if the law is to guide behaviour, and it will show that a state cannot commit human rights abuses whilst conforming to the rule of law.

This response shall attempt to establish the following claims: firstly it will be argued that the rule of law should protect the right-to-have-rights. Secondly, it shall be argued that a statethat does respect the rule of law guarantees that human rights shall be respected. Both propositions draw heavily from Arendt's political understanding of human rights and human dignity; the result is that, if we accept Arendt's understanding of human rights, Raz's position fails.

8.1 Why The Rule of Law Should Protect Human Rights

Accepting the Arendtian account, it follows that the rule of law should protect human rights. A state which disregards the rule of law fails to properly respect human dignity. The rule of law functions as a universal protection for the-right-to-have-rights; if rights are to exist at all, then they depend on the rule of law as a provisional means for their protection. This point can be brought out by showing that a state that does not conform to the rule of law does not respect human dignity; or, by asserting that the rule of law is a necessary condition for the protection of human dignity.

This is not controversial, even under formalist definitions of the rule of law.[61]

[61] cf. §4.

Raz defines a respect for human dignity as "treating humans as persons capable of planning and plotting their future".[62] this is not inconsistent with Arendt's notion of the right-to-have-rights. The right to live in circumstances where action and speech are meaningful necessarily requires that people are capable of action; this action is implicitly directed towards the future.[63] Furthermore, Raz gives two ways in which violation of the rule of law violates human dignity:

(a) It may lead to uncertainty when the law fails to articulate the course of future developments.

(b) It may lead to frustrated/disappointed expectations when the perceived stability of a given law is destroyed through retroactive law-making or improper enforcement, etc.[64]

The result of (a) and (b) is that the law fails to treat individuals as having the right to define their own future, limiting its ability to guide behaviour. This is in no way inconsistent with the self-identifying role of human dignity within a society. For example, if (a) were to be the case, then an individual would no longer be able to structure any current action or speech against their future developments; if (b) were the case, then we would be unable to act or speak of our future with any certainty. To the extent that we identify ourselves through our hypothetical future action, Raz's violations of human dignity only serve to reaffirm the claim that flouting the rule of law directly contravenes human dignity.

At this point a criticism may be raised; although we have shown that the rule of law is necessary for the protection of human dignity, we have not shown that it is in any way sufficient for it. We have not defended the claim that conforming to the rule of law entails the protection of human rights (such that human rights abuses cannot occur whilst conforming to the rule of law).

[62] Raz [12] p. 221.

[63] Parekh [11] p. 48.

[64] Raz [12] p. 222.

8.2 The Right-To-Have-Rights as a Necessary Condition to Guide Behaviour

A response must show that if the essence of the rule of law is to guide behaviour, then the rule of law must protect the-right-to-have-rights. If "the law must be capable of being obeyed", then we must show individuals cannot obey if their dignity is denied.[65]

The solution lies in recognizing that a legal system only has authority over its subjects; the source or justification of this authority are not relevant, only that a legal system can only guide the behaviour of subjects within that particular system. If we understand the sole human right as the right to membership in a community in which our speech/actions are meaningful, it follows that the denial of this human right denies the authority of the rule of law to guide behaviour.

For example consider (R5), that "the principles of natural justice should be observed".[66] this refers to guiding principles such as open and fair hearings, absence of bias, etc. If this were to be fully applied, then individuals would have membership to a legal community such that they can speak for themselves and act either as claimants or defendants. Crucially, if membership is obstructed, then the principle fails to be observed; the legal system must respect human dignity in order for its members to act and speak within that community.[67] There is no way that the rule of law can effectively guide behaviour if this right is denied, as the rule of law depends on the membership of individuals whose actions are meaningful.

9 Conclusion

On at least one plausible view of human rights, they must be protected by the rule of law. Drawing on Arendt's account of human rights, this paper shows that the rule of law should protect the-right-to-have rights, and that conforming to the rule of law guarantees that this right-to-have-rights is protected. This account is plausible in its recognition of the morally neutral, political, nature of human rights; if we accept it then it follows that a legal system must respect

[65] Raz [12] p. 213.

[66] Ibid. p. 217.

[67] Waldron [14].

human dignity, epitomised in the right to have rights. Thus, our understanding of the rule of law, as the implicit principle of legality, must accept this responsibility.

References

[1] Arendt, H. (1958) *The Origins of Totalitarianism*. New York: World Publishing Committee.

[2] —— (1971) *The Origins of Totalitarianism*. New York: Meridian Books.

[3] —— *The Human Condition*. Chicago: University of Chicago Press.

[4] Bingham, T. (2010) *The Rule of Law*. St Ives: Penguin Books.

[5] Craig, P. (1997) 'Formal and Substantive Conceptions of the Rule of Law: An Analytical Framework', in *Public Law* 21: 467–487.

[6] Dworkin, R. (2012) 'The Rule of Law' (Keynote Speech). European Commission For Democracy Through Law (Venice Commission), Conference on 'The Rule of Law as a Practical Concept'.

[7] Helis, J. (2008) 'Hannah Arendt and Human Dignity: Theoretical Foundations and Constitutional Protection of Human Rights', in *Journal of Politics and Law* 1: 73–78.

[8] Isaac, J.C. (1996) 'A New Guarantee on Earth: Hannah Arendt on Human Dignity and the Politics of Human Rights', in *The American Political Science Review* 90: 61–73.

[9] Loughlin, M. (2009) The Rule of Law in European Jurisprudence. European Commission For Democracy Through Law (Venice Commission). Study 512/2009

[10] Menke, C. (2014) 'Dignity as The Right to Have Rights: Human Dignity in Hannah Arendt', in M. Düwell et al. (eds.), *The Cambridge Handbook of Human Dignity: Interdisciplinary Perspectives*. St Ives: Cambridge University Press 332–343.

[11] Parekh, S. (2010) 'A Meaningful Place in the World: Hannah Arendt on the Nature of Human Rights', in *Journal of Human Rights* 3: 41–52.

[12] Raz, J. (2009) *The Authority of Law: Essays on Law and Morality* (2nd edn.). Oxford: Oxford University Press.

[13] UN General Assembly (1948) Universal Declaration of Human Rights [Online], 217A (III). Available at `http://www.un.org/en/documents/udhr/index.shtml#a3` [Accessed 6 April 2015]

[14] Waldron, K. (2011) *The Rule of Law and Human Dignity*. Sir David Williams Lecture. Cambridge: Centre for Public Law.

[15] Zorkin, V.D. (2007) 'The Rule of Law and Legal Awareness', in *The World Rule of Law Movement and Russian Legal Reform* 46–62.

What is *A Priori* Induction? Roy Harrod's Thought Experiment and Why it Fails*

Sam Harman
University of Oxford

Introduction

Can we ever escape the radical implications, for our worldly knowledge, of Hume's famous argument against induction? In this paper, I revisit Roy Harrod's ingenious attempt to do so. Induction, he replies to Hume, can be justified *a priori* – thereby surmounting the circularity of an empirical grounding that has plagued post-Humean inductionism.[1] Harrod sets out to achieve this bold response by an imaginative and remarkably compelling thought experiment, which has hitherto invited insufficient and largely misdirected discussion. In Section One, I outline the experiment, and analyse the mathematical objections that constitute the current literature's overwhelming thrust. I begin Section Two, however, with the recognition that these objections fail to address the crux of the experiment's intuitive force and, consequently, the fundamental issues at stake. I then provide a new and proper account of why this intuitive force must be rejected: first, by showing that it is ultimately fallacious; second, by showing that, *even if* well founded, it would be a futile reply to Hume.

1 Harrod's Experiment

Harrod's thought experiment occupies a specific role in his project to defend induction against Humean scepticism.[2] Much of this project depends on Harrod's resurrection of the 'principle of experience', which states that empirical propositions concerning the past can, under certain circumstances, constitute

*Delivered at the BJUPS Autumn Conference, 21–22 November 2015 at the London School of Economics.

[1] Black [2].

[2] Brown [5].

evidence in favour of empirical propositions concerning the future.[3] To see why and how Harrod must resurrect such a principle, we must be precise as to where he and Hume diverge. As Ayer recounts:

> [Harrod is] at one with Hume in holding that any judgement of probability which relates to a matter of fact requires to be supported by empirical evidence. But whereas Hume maintained that in cases where two empirical propositions did not stand in any relation of logical entailment it was only in virtue of some factual assumptions that one could count as evidence in favour of the other, *Harrod thinks himself able to show that, in certain crucial instances, the relation of 'being evidence in favour of' is a logical relation*.[4]

Thus, Harrod concedes to the Humean sceptic that any claim regarding prospective experience can be justified only by appealing to some probability that arises as a result of previous experience. He also concedes that this relation of probability, between previous and prospective experience, cannot itself be empirically grounded. Committed to these views, Harrod must revive the principle of experience but as an *a priori*, logical relation rather than an *a posteriori* one.[5] It is of this relation that Harrod considers his thought experiment proof.

1.1 Original Formulation

Consider a steady journey by a blindfolded person *S*, who has absolutely no knowledge about the world, along a uniform road that at some point ends at the edge of a cliff.[6] Suppose that at regular and frequent time intervals *S* makes the following extrapolative claim:

(C) The road will extend at least one-tenth the distance I have already travelled.

According to Harrod, it is mathematically necessary that *S* will be correct ten times for every once that he is wrong. That is, (C) will be true ten times

[3] Harrod [8].

[4] Ayer [1].

[5] Harrod [9].

[6] Harrod's version did not include a cliff at the end of the continuity. I have inserted this to emphasise the value of the reliability of our inductive claims – a normative position that predicates any purposeful inquiry as to the justification of that reliability. The precipice awaiting *S* removes any unwarranted benignity towards inductive error.

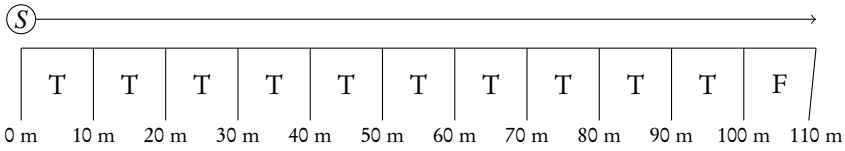


Figure 1: *S*'s blindfolded journey

for every once it is false. Consider Figure 1. Dividing the road into eleven equal stretches, it is clear that (C) is true anywhere along the first ten stretches (the first 100 m) and false anywhere along the final stretch (after 100 m).[7] Therefore, assuming *S* utters (C) at least once along every stretch, 10/11ths of *S*'s claims will be correct. This, Harrod claims, gives *S* a 'probability of being correct' of 10/11. Generalising:

> If we entertain the belief of continuance for at least $1/x$, it is quite certain we shall be right x times for every once that we are wrong ... a probability of being correct of $x/x + 1$.[8]

This result is profoundly promising for Harrod's project. Not only is the probability of $x/x + 1$ arrived at purely *a priori* – and thus appears to resurrect the principle of experience as the logical relation Harrod requires – but it is also encouragingly high.[9] Indeed, it further appears certain that the numerical probability will be higher when (C) is a more 'modest' extrapolation: that is, when x receives a higher value. If in our diagram *S* were to claim that the road ahead will extend for at least one-hundredth the distance he has already travelled, his probability of being correct would be over 99 per cent. In fact, our most fundamental real-world inductive beliefs, concerning the continuity of elements of the universe that have persisted for billions of years, are virtually certain *a priori* to be true when we extrapolate over a human time horizon. If Harrod's experiment succeeds, then I can be effectively positive that the sun will rise in the East every day until my death. This is an astonishingly robust reply to the Humean sceptic.

[7] The cardinal measurements in my example are irrelevant to the general conclusion, but serve as a useful means of illustration.

[8] Harrod [9].

[9] Kyburg [10].

Yet Harrod believes his result to imply an even more powerful foundation for our confidence in induction. Suppose *S* claims at ten metres along the road that it will extend ahead one more metre. Then the probability of this claim is 10/11, since it is mathematically similar to (C). Now suppose *S* claims at *one hundred* metres along the road that it will extend ahead one more metre. This is equivalent to an extrapolation of one-hundredth the distance already travelled, yielding a probability of 100/101. What this demonstrates, Harrod invites us to conclude, is that the probability of an extrapolation of absolute distance *increases* as *S* proceeds further along the road.[10] Every day I see the sun rise in the East, the probability that it will do so tomorrow *goes up*! It takes far less than a Humean sceptic to find this result counterintuitive. If Harrod is correct, then, he has not only restored the faith in induction that Humean scepticism eradicated; he has *strengthened* it.

1.2 A preliminary objection

It is undoubted that Harrod fails to establish his desired result via this first formulation of the experiment.[11] The flaw, being both simple and well expounded in the current literature,[12] is worth relaying but not extensively elaborating upon.[13] *S* is alleged to have a confidence corresponding to a probability of 10/11 that his utterance of (C) will be correct on any one occasion. This confidence is derivative of *S*'s knowing that he will be right ten times for every once that he is wrong. However, *S* knows that the instances where (C) is false are not spread out, but rather all occur at the end of the road. So *S* knows as he moves along the road that he is progressing through a solid block of true claims and approaching a solid block of false ones. Since *S* therefore knows that the proportion of *remaining* true claims to *remaining* false claims decreases with each successive claim he makes, he must know that the probability of (C) falls as he moves along the road. *S* cannot calculate by how much his chances of being correct decrease after each claim – nor can he estimate his chances of being correct in any single utterance of (C) at any single point on the road –

[10] Mathematically, an extrapolation of a fixed, absolute distance corresponds with successively more modest proportional extrapolations, as *S* covers a greater absolute distance.

[11] Millican [12].

[12] Harrod himself concedes the potency of the objection, and attributes its identification to J. O.Urmson

[13] Indeed, it is the current literature's absorption with this objection that has detracted from due attention towards the 'bigger picture', something this paper seeks to rectify.

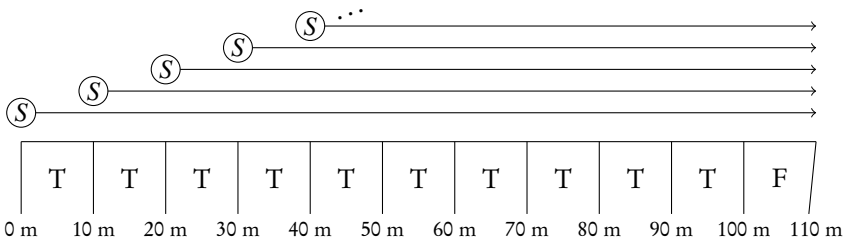


Figure 2: Harrod's 'square array' diagram adapted

because the calculation would require *S* to know the proportion of the road he had covered. Moreover, *S* is concerned only with the future probability of being correct in claiming (C) – a glance back at the road covered is little comfort if the next step leads him over the precipice. At any given point, therefore, the stock of claims that *S* has already seen fulfilled becomes irrelevant, and his ignorance of the proportion of the road covered leaves him unable to know whether the new *prospective* probability of (C) is 0.99 or 0. The promising position Harrod alleges *S* to occupy collapses.

1.3 The 'square array' modification

The flaw arises because Harrod slips into an external perspective of the probability of (C) rather than *S*'s perspective. Consequently, his solution seeks to consider the probability "from every point of view, namely from the point of view of every position on the line".[14] This leads Harrod to his 'square array' argument, by which he asserts that *averaging* the range of possible probabilities facing *S* at any point on the road can (i) discharge the truth of past claims in the grounding of *S*'s confidence in his extrapolations, yet (ii) retain this confidence to a degree sufficient to preserve the robustness of Harrod's reply to the Humean sceptic. Consider Figure 2, an adaption of Harrod's 'square array' diagram.[15] If *S* knew which one of the above positions he occupied at any given moment – that is, he knew the proportion of the road he had already travelled – he would be able to adjust his confidence in (C)'s being true accordingly. At 50 m, for instance, the probability of (C) is 5/6 after we have discounted pre-

[14] Harrod [8].

[15] Ibid.

vious and already verified claims. Now, *S* of course cannot know which of the above positions he occupies, and so is unable to calculate the exact probability of (C). But Harrod believes *S* can overcome this difficulty by recognizing that he is equally likely to occupy any one of these positions at any one moment, by virtue of the frequency and regularity of his claiming (C). *S* can then take the average of probabilities across all positions, including the infinitely many positions between those represented in the above diagram. This yields a probability of 100/121, and we are invited by Harrod to conclude that, given *S*'s ignorance of his actual position, *S* is justified in taking the probability of any single utterance of (C) to be 100/121.[16] Generally, *S* can have confidence in the truth of an extrapolation of $1/x$ corresponding to a probability of $x^2/(x+1)^2$.

Harrod believes this result to surmount the objection that previous fulfilments of (C) are irrelevant to its future probability, and that it would be irrational for *S* to successively maintain a confidence in (C)'s truth corresponding to 10/11 on the basis of previous fulfilments, given his awareness of (C)'s diminishing prospective probability. Crucially, it seems we can jettison any appeal to past instances of (C) in our calculation and yet leave *S*, despite his ignorance, in a similarly promising epistemic position as before. Though slightly more modest, $x^2/(x+1)^2$ remains an encouragingly high probability for inductive claims and still entails a rising probability of absolute extrapolations as the period of observed continuity lengthens.

1.4 Criticisms

The 'square array' refinement is designed to allow *S*'s confidence in the future truth of (C) to stand free from depending on previous cases of its being true. In fact, it merely commits a more convoluted guise of the fallacy that prompted its creation.[17] According to Harrod, *S* is entitled at any moment to regard the probability of (C) as 100/121 because this is the average probability across all possible positions that *S* could occupy at that moment. But this is clearly false: *S* knows at any moment that a certain portion of the road has been covered. Thus 100/121 no longer represents such an average, because it relies on positions that *S* has already occupied, and so cannot occupy now. *S* must therefore reject 100/121 as 'obsolete', though of course he cannot calculate the necessary adjustment of this figure, because that would require him to know

[16] Ibid.

[17] Blackburn [3].

what proportion of the road he had covered. All *S* knows is that, since his progression along the road diminishes the prospective probability of (C), the exclusion of earlier positions when making the calculation will yield a lower average probability. As originally objected, *S* has no grounds to pick out one value anywhere on the spectrum of possible probabilities. Consequently his confidence in the truth of his claim, derivative of any such value, evaporates once more.

Harrod's attempt to repair his result a second time is wholly unpersuasive:

> [*S*] is in complete ignorance as to what relation the length he has already travelled bears to [the] total length. Therefore, he cannot bring the length of travel already accomplished into any relation to the formula… What we have to do … is to decide on our formula at the outset, and then to adhere to it; we have to adhere to it because in the course of the journey we get no new relevant information.[18]

It is absurd that, in the absence of information required to calculate a correct result, one should instead favour a result they know to be incorrect. Yet this is exactly what Harrod has found himself requiring to defend: *S* is unable to accurately adjust his estimate of the probability of (C) from the initially correct 100/121, and so he must stick to that probability. That "we get no new relevant information" is clearly false: *S* does not know by how much the probability requires adjusting, but he knows that *some* adjustment is necessary. If *S* cannot accurately revise his initial value to take the distance covered into account, then he should not stand by it, but abandon it – for it is undoubtedly incorrect.

2 The bigger picture

We can now reject any justification for *S*'s being confident in the truth of an isolated claim. We can also reject the notion that progression along the road makes extrapolations of absolute distance more probable. What we cannot yet reject is the sheer thought that, considering the set of claims which *S* will make, a very high proportion will necessarily be true. This is a result that critics have generally failed to address, perhaps because they feel no need to go beyond disproving what Harrod himself professes to have shown. Yet, left

[18] Harrod [8].

standing, such a result may well preserve some of the intuitive force of Harrod's experiment. If *S* cannot have confidence in any single claim being true, he can surely have complete confidence in the vast majority of his claims being true – does this not remain at least somewhat promising for induction? I shall now extinguish this intuitive force: first, by showing that it is false to regard a very high proportion of the set of *S*'s claims as being necessarily true; second, by showing that *even if* such a result were established, it would be a worthless and misdirected reply to the Humean sceptic.

2.1 Continuities and continuance

All of our earlier objections take for granted that the road posited by Harrod does in fact extend a certain, though unknown, distance ahead of *S*. Yet the most devastating counterargument to Harrod exposes this most primitive of assumptions as rendering his entire experiment circular.

The fault is subtle but basic. The result that *S* is correct in claiming (C) ten times for every once he is wrong relies unquestionably on the assumption that *S* has the time to claim (C) *at all* – strictly speaking, that *S* has the time to claim (C) on at least 11 separate occasions.[19] But this is to assume that there is a non-zero length of time between *S* beginning his journey and ending it. In turn this is quite clearly to assume that there is a non-zero distance between the beginning of *S*'s journey and its end. That is to make the assumption that the road will extend for some distance into the future, which is precisely what Harrod's experiment attempts to establish.

Further to this circular premise about the road, Harrod makes an almost imperceptible assumption about *S* which threatens his result in a different and intriguing fashion. Harrod has been regarded by critics as generous to the Humean sceptic in his experiment's premise that *S* will arrive at a point at which he knows that the continuity ends. Yet Harrod is *advantaging* himself by assuming that *S* is thereby able to identify what constitutes the end of the 'continuity' he is travelling along.[20] Consider a modification of the original experiment, in which the cliff is replaced by a shallow stair, but also in which the 'continuity' consists in stretches of road punctuated at regular intervals by such stairs.[21] Now, consider *S* at the point of the first stair. It is at this point

[19] Braithwaite [4].

[20] Harrod [7].

[21] Against the objection that this is not properly referable to as a 'continuity', it must be urged

where Harrod implicitly considers *S* both correct and justified in regarding the continuity as having come to an end. That is, it is the point at which *S* moves from being correct in considering himself in a period of continuity, to being correct in considering himself in a period of discontinuity. But of course *S* is *not* correct in this latter consideration, for the real continuity is still continuing. Indeed, if the staired road extends a great distance further than its first plateau, then *S* will incorrectly consider himself in a period of discontinuity for a far greater distance than he will correctly consider himself in a period of continuity. Since Harrod's result relies on *S* manifesting such considerations in frequent claims, it follows that *S* will falsely claim that the continuity will *not* extend a further $1/x$ a great many times more than he will correctly claim the opposite.[22] So unless Harrod can impart upon *S* exactly the kind of *a posteriori* knowledge his project depends on jettisoning[23] – in this case the precise form of the continuity – it is as likely that a very *low* proportion of *S*'s claims will be true as it is that a very high proportion will be. Moreover, *S* will be utterly unable to determine which situation in fact occurs.

2.2 'Probability' and the value of induction

We have seen that Harrod fails to establish that a high proportion of *S*'s claims will necessarily be true. However, I submit finally that, even if he were to establish such a result, it would fail to give *S* the confidence which Harrod seeks to impart upon him, and consequently upon rational humans, in making inductive claims of this kind.

Harrod is aiming to resurrect the principle of experience – that empirical propositions concerning the past can, under certain circumstances, constitute evidence in favour of empirical propositions concerning the future – as a defence of induction against Humean scepticism. If the principle of experience is to be at all a substantive reply to the Humean sceptic, then it must be *valuable* to rational humans, in that it allows us to justifiably have confidence in the truth of inductive claims that actually matter to us. By the recognition of two crucial facts, Harrod's claim regarding the overall proportion of true claims can be shown – even if true – to have zero value as a defence against the scep-

---

that many of our inductive claims posit the extension of precisely so formed 'continuities' – the rising of the sun at regular intervals, the repeated beating of one's heart, and so on.

[22] Since *S* believes the continuity to have ended, he can be taken to believe that it will not extend.

[23] Levi [11].

tic. This is because it neither is one of, nor helps us to make, any of the kind of inductive claims that actually *matter* to us. The first fact is that the principle of experience has value to *S* only if he can be confident in the truth of his *imminent* claim – that is, only if some set of empirical propositions available to *S* can constitute evidence in favour of the *very next* empirical proposition concerning the future. The second is that no empirical proposition available to *S*, particularly that of the overall proportion of claims that are true, constitutes such evidence. If these two facts can be demonstrated, then the principle of experience is of no value whatsoever to *S*, and therefore provides no substantive reply to the Humean sceptic.

The first fact is neatly validated by Millican:

> What [*S*] wishes to know, regarding the series of future extrapolations that he is about to initiate, is not *what proportion of the extrapolations in that series will turn out to be true*, but simply *whether any at all of the extrapolations in that series will turn out to be true* … the series of extrapolations is not random but sequentially ordered. The future extrapolations in that series form an orderly queue, and the particular one-tenth extrapolation which [*S*] is just about to make is firmly at the head of that queue – if any other extrapolation in the series ultimately turns out to be true, then this one certainly will too.[24]

The point here is that *S*'s ability to use the principle of experience for any given future proposition is predicated on its being an effective tool for verifying his imminent proposition. Any future use of induction *S* might benefit from rides on whether his *next* step will lead him over the precipice. But that it will not is something which *S* cannot afford to merely *assume*. He must be justifiably confident that his next step will not lead him over the precipice. Equivalently, *S* must be justifiably confident that his imminent claim will be true. Thus, to give the principle of experience any value, Harrod's experiment must show it to establish this justification.

Does it? I do not think so. Harrod believes that the high proportion of claims that will be true can be used, via the principle of experience, to justify *S*'s confidence in the truth of his imminent claim. But this is a progression that stands on dubiously simplistic assumptions. Specifically, Harrod's flawed step is from a 'rate of truth' to a 'probability': 10/11 of *S*'s utterances of (C) will

[24] Millican [12], emphasis and underlining as in original.

be correct, hence the probability of *S* being correct in *any* given utterance of (C) is 10/11. Harrod holds this step to be "in accordance with the traditional notation", and indeed no critic has yet challenged it.[25] However, I propose that Harrod has fallen victim to the statistical 'reference class problem', of determining what constitutes the *class* that is relevant when calculating the probability applicable to a *particular case*.[26] The point is best made by way of two examples:

1. Consider a person *R* betting on a completely fair die with 11 sides, on which ten of have 'T' and one has 'F'. The die has been rolled many times under regular and consistent circumstances. 10/11 of these rolls have resulted in the die landing on 'T', and 1/11 in it landing on 'F'. *R* knows this, and so bets on the die landing on 'T'.

2. Consider a person *Q* boarding a flight on an airline. *Q* knows that, as a result of a new engine the airline began using ten years ago, every flight since then has been unsuccessful. But *Q* nonetheless boards the plane – for he knows that, of all flights in history, 10/11 have been successful.

It seems clear that *R* is justified in making the bet, while *Q* is not justified in boarding the plane. More precisely, the confidence that *R* has in the outcome of the next roll of the die, as a result of *all* previous rolls, is well grounded; while the confidence that *Q* has in the success of the next flight, as a result of *all* previous flights, is not. This is despite the fact that both *R* and *Q* are maximising the *sample* of cases, on whose proportion of outcomes they are basing the probability of the *next* case's outcome:

1. 10/11 rolls of the die produce 'T'. *R*'s imminent roll of the die is a case of a roll of the die. Therefore the chance that *R*'s imminent roll of the die will produce 'T' is 10/11.

2. 10/11 flights are successful. *Q*'s imminent journey on this airline is a case of a flight. Therefore the chance that *Q*'s imminent journey on this airline will be successful is 10/11.

In the die example, it seems good practice for *R* to maximise the size of his sample of cases relative to the population, such that the resulting probability is as reliable as possible. But this, importantly, is because *R* has no statistically good reason to employ only a sub-set of the above sample. Each case is identical

[25] Harrod [8].

[26] Williams [13].

and corresponds perfectly to the imminent case. *R* could of course obtain a radically different probability by employing as a sample only those rolls that, for instance, have produced 'F'. But this would be irrational, since it employs a criterion of sub-categorisation that is determined by pure chance. By using the sub-class formed by any such criterion, *R* unambiguously renders the resulting probability less reliable.

However, the essential difference, and the flaw in *Q*'s reasoning, is that in the flight example *Q* knows of a class division in his population of cases – namely flights on that airline in the past ten years – that, were he to use only that sub-class as his population, would (i) produce a probability that corresponded more accurately to his current case, and (ii) produce a starkly different and less promising probability of his imminent flight's success. By (i), I mean that *Q*'s imminent flight is a member of very many classes – all flights in history, all flights on that airline in the past ten years, all journeys in the history of transport in general, all journeys made by *Q* – and *Q* must select the class, on whose proportion of outcomes it is most appropriate to base the probability of the next case's outcome. *Q* knows that the airline's use of the new engine has causally and drastically affected the probability of the success of flights on that airline. So the sub-class *all flights on that airline in the past ten years* is more appropriate than the class *all flights in history*, because the criterion of sub-categorisation *directly* and *non-accidentally* affects the probability produced. We would say that *Q* is justified, in ascribing a probability to the success of his imminent flight, to the extent that he uses the most appropriate class as its basis. Moreover, we would say that *Q* is irrational in ascribing, to such success, a probability based on a class which is known to be radically inappropriate and unrepresentative in comparison to an alternative. *This* is what prevents *Q* from being justified in the way that *R* is.

Now, Harrod is making the same progression – from a proportion (10/11) of a population (all utterances of (C) by *S*) exhibiting a certain character (truth), to a probability (10/11) of an individual case exhibiting that character.[27] Therefore, if *S* is to be justified in ascribing that probability to his imminent utterance of (C), then Harrod's experiment must resemble the die example rather than the plane example, insofar as the two differ by the crucial property outlined above. But the experiment in fact resembles the *plane* example in this property, which is why *S* is not justified – in fact, *S* is irrational – in ascribing a probability of 10/11 to the truth of a given utterance of (C). For there

[27] Ayer [1].

is most certainly a sub-categorisation of *all utterances of (C)* whose criterion directly and non-accidentally affects the probability produced, in a way and to an extent that renders the resulting probability far more accurate for *S*'s *imminent* utterance of (C). That criterion is whether or not the utterance of (C) occurs when *S* is positioned in the last 1/11th of the road. The sub-class *utterances of (C) in the first 10/11ths of the road* will yield a probability of *S* being correct of 1, while the sub-class *utterance of (C) in the last 1/11th of the road* will yield a probability of *S* being correct of 0.

Although *S* of course does not know which of these sub-classes his imminent utterance of (C) is a member of, he undoubtedly knows that it is a member of *exactly one* of them. By virtue of this, *S must* be ascribing a probability, to the truth of his imminent utterance of (C), that is based on a class – *all utterances of (C)* – which is known to be radically inappropriate and unrepresentative in comparison to an alternative. The only difference between this and the plane example is that the *correct* alternative is inaccessible to *S*. But recall that the crucial condition, that is sufficient to render *S* irrational in such an ascription, is merely whether such a radically more appropriate and representative alternative is *known by S to exist*. And clearly, though *S* cannot identify which of the two sub-classes constitutes that alternative, he knows it to exist. *That* is what prevents *S* from being justified – indeed, renders him irrational – in ascribing a probability of 10/11 to the truth of his imminent utterance of (C), and in having the corresponding level of confidence in being correct in that claim. That the known alternative is inaccessible does not suddenly render the original class sufficiently appropriate, but rather prevents *S* from being justified in ascribing any meaningfully representative probability whatsoever.

Conclusion

In this paper, I have given Harrod's thought experiment the proper attention that its originality and importance demand. I exposed the bulk of previous critical assessment as a red herring to the deeper substance of its intuitive power – that is, the notion that *S* will be correct many more times than he will be wrong. In placing due focus on this particular result, we can address issues far more profound for the broader analysis of induction, and provide a truly complete critique of Harrod's remarkable contribution. I have shown that Harrod fails in his reply to Humean scepticism on two counts. The first is that the *a priori* necessity of *S*'s high degree of correctness, though at first glance compelling, is upon closer examination erroneous. Such necessity collapses unless Harrod makes both a circular assumption regarding the road and a self-defeating assumption regarding *S*. The second is that, *even if* the neces-

sity of *S*'s high rate of truth could be established, it would fail to justify the stance that Harrod wishes to take. I showed the principle of experience to be of value only if it can render *S* justifiably confident in the truth of his *imminent* claim. I then demonstrated that *S*'s high rate of truth is insufficient to warrant such confidence, and that Harrod's argument for its doing so commits a 'reference class' error. Since Harrod thereby comprehensively fails to resurrect the principle of experience, his defence of induction disintegrates, and his reply to Hume must be rejected.

References

[1] Ayer, A.J. & Macdonald, G. (1972) *Probability and Evidence*. London: Macmillan.

[2] Black, M. (1958) 'Self-supporting Inductive Arguments', in *The Journal of Philosophy* 55/17: 718–725.

[3] Blackburn, S. (1973) *Reason and Prediction*. Cambridge: Cambridge University Press.

[4] Braithwaite, R. (1958) review of H.R.F. 'Foundations of Inductive Logic', in *The Economic Journal* 68/269: 146–148.

[5] Brown, S., Collinson, D. & Wilkinson, R. (eds.) *Biographical Dictionary of Twenthieth-Century Philosophers*. New York: Routledge.

[6] Harrod, H.R.F. (1951) 'Induction and Probability', in *Philosophy* 26/96: 37–52.

[7] —— (1960) 'New Argument for Induction: Reply to Professor Popper', in *British Journal for the Philosophy of Science* 309–312.

[8] —— (1974) *Foundations of Inductive Logic*. London: Macmillan.

[9] —— (1960) 'The General Structure of Inductive Argument', in *Proceedings of the Aristotelian Society* 61: 41–56.

[10] Kyburg, H.E. (1964) 'Recent Work in Inductive Logic', in *American Philosophical Quarterly*1/4: 249–287.

[11] Levi, I. (1958) 'Foundations of Inductive Logic', in *The Journal of Philosophy* 55/5: 209–212.

[12] P.J.R. (1996) 'Hume, Induction, and Probability', PhD Dissertation, University of Leeds [Online] Available at `http://www.davidhume.org/papers/millican/1996PhD.pdf`

[13] Williams D.C. (1953) 'On The Direct Probability of Inductions', in *Mind* 62/248: 465–483.

Should States be Bound by Moral Side-Constraints?*

Dylan Keitz-Playford
University College London

1 Introduction

The question of how a State should proceed to achieve its moral goals within political philosophy is secondary only to whether or not States should exist. We can see this reflected in Robert Nozick's *Anarchy, State, and Utopia*, where the first two chapters deal with the existence of States, and the third the moral constraints of States. This paper seeks to defend Nozick's view that States' moral activities ought to be bound by side-constraints. This paper shall defend the claim against the following two arguments; one, the theoretical argument, made by Michael Neumann in his work 'Side Constraint Morality', that there is no distinction between side constraint and end-state maximizations; and two, against the practical argument, made by Daniel Nolan in his work 'Consequentialism and Side-Constraints', that absolute side-constraints do not provide correct moral answers in high-stake situations.

2 Definitions and Clarifications

A side-constraint[1] restricts a State (or individual) from "violat[ing] [certain given] moral constraints in the pursuit of [their] goals…"[2] This shall be compared to an end-state maximization which instead seeks to achieve a moral goal (say the minimization of violations of certain moral rules) by the weighted means that maximize that goal (i.e. the means that account for the highest quality maximization of the goal). This would allow for the "violat[ion] of [a

---

*Delivered at the BJUPS Autumn Conference, 21–22 November 2015 at the London School of Economics.

[1] Throughout this paper, we shall be using the terminology 'side-constraints' and 'end-state maximization' to differentiate between the two ways of understanding a State's moral goals as discussed in this paper. This corresponds to Nozick's 'side constraint view' and 'end-state maximizing view', Neumann's 'side constraint' and 'end state' and Nolan's 'side-constraints' and 'consequentialism', respectively.

[2] Nozick [3] p. 29.

moral rule] when doing so minimizes the total (weighted) amount of the violation... in the society".[3] To put the distinction into an example: side-constraints would hold that one must not violate constraint *C*, not even to achieve goal *G* (*G* being any particular goal that a State might wish to achieve). The end-state maximization would hold that the goal is the (weighted) minimization of violations of *C*, and in this way *C* can be potentially violated in an individual circumstance, if it minimizes the violations of *C* over the entire society. This paper does not necessarily accept any of Nozick's conclusions beyond the claim that States ought to be bound by side-constraints. Neither will it define what side-constraints a State should abide by. As a consequence, this paper will give no defence against the second argument Michael Neumann advances: that Nozick's side-constraint morality falls short of its Kantian origins. As that particular argument relates to the side-constraints that Nozick defends, it falls outside of the purview of this paper.

Given the above, when the need for examples arises, this paper shall use the following side-constraint: a State should not violate the reasonable protection of innocent citizens within its borders, namely from bodily harm. While what exactly 'reasonable' in this context means shall not be clarified, it is used to acknowledge the non-omniscience (and non-omnipresence) of a State, in order to avoid an example like the following:

> State A has the side constraint that it 'should not violate the protection of innocent citizens within its borders, namely from bodily harm'. State A determines that the only means to keep within this side-constraint is to imprison every one of its citizens. This is obviously morally wrong. Therefore this is an untenable sideconstraint.

However, while this paper will use the example above, the analysis will hold if one substitutes any other side-constraint in its place.

By the same token, while this paper advocates side-constraints as the proper moral bearings of a *State*, it will not answer the question of whether an *individual's* actions should also be bound by side-constraints. While these two questions have been conflated over time, it is easy to see that commonsense moral intuition holds the morals of a State to a different standard compared to individuals.[4] After all, if an individual makes it non-voluntary to pay for certain

[3] Ibid. p. 28.

[4] This is not, by any means, to accuse any of the authors mentioned in this paper of confusion in this regard. Indeed, they were all seemingly concerned with both the actions of States and

unrequested services (or rather, opting out comes with a punishment of some form) we call it theft and extortion. If a State does it, then we call it taxes.[5] And so, the morality of the State differs from the morality of individuals in fundamental ways, perhaps even in regards to the question of side-constraints. Fortunately, the analysis of sideconstraints can work independently of whether the criticism is aimed at an individual having side-constraints or a State having side-constraints (as we shall see when we begin to modify examples).

3 Overview of Nozick's Argument

Nozick's argument for side-constraints is fairly simple. For him sideconstraints, not end-state maximizations, recognize that we may not use another person as we please because "[i]ndividuals are inviolable"[6] And while Nozick admits that it is questionable whether "behavior toward a person [can] be constrained so that he is not to be used for any end except as he chooses" , it is obvious in cases where one "physically aggress[es] against" another that a person is being violated.[7] By contrast, end-state maximizations would allow for a person to be used as means, if it minimized the overall use of people as means.

Nozick argues that the "libertarian side constraint that prohibits aggression" is a necessary consequence of the side-constraint view (as well as how side-constraints flow from the Kantian idea underlying Nozick's particular conception of side-constraints). And in doing so, Nozick sets out a challenge: if we are to reject the libertarian side-constraint, we are left with three options: (1) we "must reject all side-constraints…", (2) we "must produce a different explanation of why there are moral side-constraints…" without necessarily "entail[ing] the libertarian side constraint", or (3) we "must accept the…idea about the separateness of individuals and yet claim that initiating aggression against another is compatible with this…idea".[8]

---

individuals. Nozick in particular makes it clear that he wishes to discuss both, in particular how the individual and the State reacts.

[5] While this may come off as a snide libertarian remark about the legitimacy of taxes, I mean it in no such way. After all, States are generally considered acceptable (and possibly necessary) theoretical institutions, and the power of taxation is frequently granted upon them. Extortion rackets have none of those qualities.

[6] Ibid. p. 31.

[7] Ibid. pp. 31–32.

[8] Ibid. pp. 33–34.

As arguing against (1) is the main thrust of this paper, we must proceed by either (2*) trying to discover an argument for side-constraints which does not rely on Nozick's Kantian roots, or (3*) somehow modifying the underlying basis of 'individuals are inviolable' without necessarily entailing any particular sideconstraint. We shall proceed with (2*), and it is here that this paper will make its most significant departure from Nozick's text. For while, it is true that Nozick's argument relies upon the idea that individuals are inviolable, it is not true that *all* side-constraints must necessarily rely upon this. Instead, side-constraints stem from moral maxims about activity. For example, one may create a side-constraint from the moral maxim 'one shall not steal' that may look something like this: the means used to achieve goal *G* must not include theft. Such a side-constraint does not necessarily require the idea that 'individuals are inviolable' and therefore does not necessarily arrive at the same conclusions as Nozick does.

This allows us to understand how side-constraints are formed. Side-constraints acknowledge moral maxims about activity (whether about what *must* be undertaken, or what *must* be avoided), and help create a set of morally permissible or morally impermissible actions. On the other hand, end-state maximizations do not tell us that certain activities are exclusively morally permissible or impermissible, but rather it tells us to strive toward a goal, allowing any tool that will bring us closest to the goal, while requiring we use only the tools that do.[9] While this argument does require the moral maxim to be truthful for the sideconstraint to work, by the same understanding the goal of the end-state maximization must be a moral goal.

To illustrate, let us presume that the moral maxim 'A State should reasonably protect innocent citizens within its borders from bodily harm' is truthful. The side-constraint that arises is 'A State should not violate the reasonable protection of innocent citizens within its borders, namely from bodily harm, while it pursues its goal'; while the end-state maximization is 'A State's goal should be the minimization of its weighted violations of the reasonable protection of innocent citizens within its borders, namely from bodily harm.' It becomes clear that we should prefer the side-constraint in this case: if the moral maxim is absolutely true, then it follows that it should not be violated.

[9] We can see how this argument is the same underlying one as Nozick's. The only difference is that Nozick has a particular side-constraint he wishes to defend; this argument can be applied to any and all side-constraints.

4 Neumann's Argument

While not essential to Nozick's argument we need to make note of two things, which he thought of as 'gimmicky', specifically for their potential ability to rewrite side-constraints into end-state maximizations.[10] These are (i) infinite weights and (ii) indexicals. An infinite weight is when an end-state maximization includes a clause that any violation of the goal has infinite negative weight. This means that the said violation is an impermissible action for that activity could in no way have the highest weighted path to achieve goal *G*. Indexicals are terms, such as 'me', 'he', 'she', 'they', where the language doesn't fix a specific person, but rather is speaker- and context-dependent. Therefore, if one were to "distinguish… between [one's]violating [side-constraints] and someone else's doing it") and if one gave their own violations infinite negative weight, then we end up with an end-state maximization that appears to be a side-constraint.[11] However, Nozick believes that "[a] careful statement delimiting 'constraint views' would exclude… transforming side-constraints into the form of an end-state view…".[12]

Michael Neumann notes an apparent problem with Nozick's distinction between side-constraints and end-state maximizations: they both seem to prescribe the same thing, even if one accounts and corrects for potential issues arising from the granting of infinite weight and indexicals.[13] On the technical end, one could remove indexicals and/or infinite weights and still have an end-state maximization that appears to be a side-constraint. This stems from the nontechnical question of 'Which transformations are gimmicky?'[14] Neumann argues that the linguistic reasoning for differentiation between side-constraints and end-state maximizations, (that side-constraints are constraints and end-state maximizations are goals) which makes certain transformations gimmicky and others non-gimmicky, seems arbitrary. After all, couldn't someone who argued for end-state maximizations believe that the maximizations "constrain[ed], or [prescribed] goals"? Or couldn't someone who argued for

[10] Ibid. p. 29.

[11] Ibid.

[12] Ibid.

[13] Along with Neumann's discussion of Nozick's apparent failure to match the bar that Kant sets, this paper will also ignore the strain throughout of a utilitarian reply to Nozick, for the utilitarian analysis does not encompass the entirety of his argument that end-state maximizations are not substantively different to side-constraints.

[14] Neumann [1] pp. 133–134.

side-constraints believe that the side-constraints "prescrib[e] goals. . . " and want to "minimize their violations" of the side-constraint?[15] And in such a manner, Neumann came up with the following principles used for an end-state maximization:

(A) Maximize *C*.

(B) Minimise violation of (A).

And if one wishes to see no violations of (A) occur then we must accept:

(A′) Minimise my current violation of (A).

The mixture of (A′) and (B) is the same as:

(D) Minimize violation of (A), with infinite negative weight given to my current violation of (A).[16]

Neumann states that (D) appears to be gimmicky; after all it gives infinite negative weight to an indexical. However, the use of the word " 'current" would be essential for the 'gimmicky' transformation of any 'real' side-constraint principle. . . ".[17] For if we are to attempt the task, then we must be allowed to deny our current violations.

Therefore Neumann advances two propositions involving side-constraints: (4) 'the linguistic boundaries between side constraint and end-state principles are extremely vague' and (5) 'many supposedly pure end-state and side constraint principles can be regarded both as setting goals and as imposing constraints'.[18] Which, when put together reveals an "essential similarity between the two approaches".[19]

However, while his analysis of the technical issue is correct (namely that if one wished to translate side-constraints into end-state maximizations then one needs to use indexicals and infinite negative weight), Neumann is wrong on an important point. Side-constraints allow for other goals, while end-state maximizations do not. If my goal (end-state maximizations) is to minimize viola-

[15] Ibid. p. 34.

[16] Modified this from Neumann [1] pp. 136–137. The only substantive difference is that he used "utility" instead of "*C*".

[17] Ibid.

[18] Ibid. p. 134.

[19] Ibid. p. 137.

tions of *C*, then I must minimize violations of *C* as a matter of moral priority. Comparatively, if my side-constraint is to not violate *C*, then I can seek any number of goals that would not have me violate *C*. And in this way, Neumann's psychological account of the moral actors is wrong; it doesn't matter if one feels as if side-constraints are a goal. Side-constraints are not goals to achieve, but rather form the edge of a path; if one wishes to follow the moral path, they must never stray from the side-constraints.

But, it could be argued, I am merely disguising the issue. Indexical end-state maximizations with infinite weighting are still essentially the same as side-constraints; after all they would always propose or bar the same activities, and in this case it's just a matter of perspective. One may see side-constraints as merely constraints and indexical end-state maximizations as goals, but if they necessarily form the same path, then aren't they the same? And in such a way, Neumann's psychological account holds.

However, if it weren't for infinite weighting, then there is a clear difference between end-state maximizations and side-constraints. After all, even indexical end-state maximizations that don't have infinite weighting could include violations in extreme situations (which cannot happen with side-constraints). And so, infinite weighting reveals that Neumann (and even Nozick to a certain extent) has the picture backwards. An indexical end-state maximization with infinite weighting is not turning a side-constraint into an end-state maximization, but rather it's the point where an end-state maximization becomes a side-constraint.

While that, again, seems to be disguising the issue, an analysis of how an indexical end-state maximization with infinite weighting would be fulfilled reveals it to be a side-constraint. Using the principles above let us presume that someone believes that their goal is (D) They must minimize violations of (A), while giving infinite negative weighting to their current violations of (A). In practice, the person would never violate (A) for they are restricted from violating (A). Therefore it is a guide of activity, telling that person what *constraints* there are, rather than what the goal is. And if we were to replace (A) with 'Libertarian Principle', we would have Nozick's side-constraint.

5 Nolan's Argument

Daniel Nolan, rather than arguing about the technical nature of side-constraints, denies that side-constraints are better than end-state maximizations, especially

when considering high-stakes situations.[20] In order to make his argument he focuses on absolute side-constraints rather than those with cutoffs. Cutoffs are conditions such as 'don't break a promise unless your child's life is at stake' where one may abandon sideconstraints and begin using end-state maximizations.[21]

In order to bring out his argument he uses an example similar to the following:

> Imagine there is the country of Cyrilstan. Its megalomaniacal dictator, Cyril, announces to the world that he has built nuclear devices on a large scale, with over a hundred devices ready for deployment and the most advanced security protecting hem. Also, in order to convince the world of his genius, he shows proof of the weapons and the advanced security. He announces that he plans to use every weapon within three days, targeting major populations centers, including parts of the neighbouring nation Astan. As luck would have it, in revealing his advanced security, experts recognize a weakness that will open up the nuclear stockpiles to coordinated airstrikes, disabling them and killing no one. However, the weakness requires using a device that only exists in Astan and in order to use the device, Astan will need to torture its innocent citizen Dudley. The torture would certainly cause him bodily harm, and Dudley (who is a bit of a disinterested person, uncaring for most affairs outside of his daily routine) would not consent to the torture. In addition, Astan has accepted the side-constraint 'a State should not violate the reasonable protection of innocent citizens within its borders, namely from bodily harm'.[22] What should Astan do?[23]

---

[20] While Nolan's paper is mostly geared toward answering Peter Vallentyne's paper 'Against Maximizing Act-Consequentialism', his answers are general enough to be applicable to this paper.

[21] Nolan [2] p. 7.

[22] We'll presume that the 'reasonable' gives no special insight to how Dudley should be treated.

[23] Nolan [2] p. 10–12 original example involved him acting as a nuclear inspector alongside Dudley, and the only way to stop Cyrilstan's attack was to knock Dudley over a railing (which Dudley would not consent to), therefore causing him serious (though not life threatening injury). It was also explicit that no one that Nolan or Dudley knew would be injured in Cyrilstan's attacks. Most changes to this example are due to the fact that this paper wishes to focus on States, rather than individuals; however the change about who gets injured is mostly for the sake of argumentative flow, as shall become apparent. None of these changes, however, alter the conclusions of the thought experiment.

Nolan would argue, presumably correctly, that the moral intuition is for Astan to cause harm to Dudley,[24] in order to overcome Cyrilstan's security.[25] Given this, side-constraints are not the absolute argument that proponents claim they are, and are inferior to end-state maximizations.

Nolan rebuts five arguments in defence of side-constraints. This paper shall focus on only one of them, in the section 'Moral Tragedy – No Right Answers'. In a moral tragedy there is "no permissible option to perform".[26] It is both impermissible to torture Dudley, and it is impermissible to let Cyrilstan launch the missiles; both would violate the side-constraint of reasonable protection of innocent citizens. So, the argument goes, we can rely on end-state maximization because Astan is left with no option. (It is important to note that Astan did not directly do anything to put itself in this position; therefore, it doesn't have a moral responsibility for the situation that might affect how we perceive Astan).

Nolan quickly rejects the Moral Tragedy argument, saying that "our intuition is pretty strong that it is permissible" to cause Dudley harm. For while "not hurting people is a good rule of thumb… it has exceptions, and this looks like an exception…". And a defender of side-constraints must believe either that Astan should torture Dudley which makes "the 'absolute side-constraint'…" appear less absolute than how it is presented, or that Astan should not torture Dudley, which is the seemingly immoral option.[27]

However, Nolan's argument requires us to believe that the permissibility of actions change based on circumstances. And while it may be true that certain activities that are immoral in isolation seem moral once the circumstances are understood, the action remains impermissible, and rather is the product of circumstances outside of the actor's control. (Imagine the impoverished person who stole food to save her family from starvation. The theft is still not permissible, but circumstances make it the most permissible of impermissible

---

[24] I will avoid saying that Nolan claims it's permissible to torture Dudley, to acknowledge the mutation of the argument; I would not wish to make claims on his behalf. And indeed, the argument seems to holds for any level of harm given to Dudley, with the possible exception of executing him.

[25] However said, I believe this is a slightly less clear case than Nolan's original example; perhaps because it involves the necessarily unequal power structure of the State versus an individual, or perhaps because torture is potentially more harmful than a broken bone. However this does not change the overall intuition, which would be to torture Dudley.

[26] Ibid. p. 18.

[27] Ibid.

options). And so it is with Dudley's torture. As torturing Dudley is not a permissible option in isolation, it remains impermissible even with Cyrilstan's threats. However, it appears the least impermissible of the two options. Therefore, while it is impermissible for Astan to torture Dudley, they still must torture him.

In response, a defender of his work might then argue as follows. What if, instead of Astan being at risk, Cyril explicitly mentions how Astan is safe; it is too close to Cyrilstan's borders. And furthering the complication, all other nations agree that if Astan does not torture Dudley, they will not retaliate against Cyrilstan nor Astan (therefore protecting Astan from direct attack, as well as any potential nuclear fallout). With this modification, the only citizen of Astan at risk of bodily harm is Dudley. By torturing Dudley, they violate their side-constraint, while if they do not, they commit the seemingly immoral act of not stopping Cyrilstan. There are two potential responses to this idea. (6) Astan ought to decide that its side-constraints are too important to ignore and not torture Dudley. It could be claimed it is well within the purview of a State to not bother with outside affairs; a State's only duty being to protect its innocent citizens within its borders. Outside of that duty, the State has no responsibility. However, while one might be able to develop the idea, it still goes against the intuition that, as Astan is the only nation with the capabilities to stop Cyrilstan, it should, especially when weighed against the torture of a single person. Therefore (6) is not a satisfactory response.

This leads us to the second response: (7) if the moral intuition against letting innocent people in other States die, when Astan is the only State capable of preventing it, is that strong then it must be adopted as a side-constraint (either as a side-constraint of its own, or somehow expanding the singular side-constraint to include it). Astan only has impermissible options that were forced upon it, and must decide which is the least impermissible. The main criticism of this response is its seeming ad hoc nature; it allows proponents of side-constraints to call any moral weakness of side-constraints a moral tragedy and revert to end-state maximizations. However, we commonly accept that there may be multiple moral restrictions on everyday behaviour; one may neither kill nor steal. So why should a side-constraint exist as a singular rule? Anytime there are multiple moral considerations there is potential for conflict; this is not unique to side-constraints. Therefore side-constraints can cover all morally permissible situations; it's only when a moral tragedy is occurring that side-constraint analysis falters (as with any moral system).

6 Conclusion

Side-constraints differ nontrivially from end-state maximizations, and attempts to write side-constraints into the form of end-state maximizations in fact achieve the reverse. Side-constraints also provide better answers for ordinary moral situations, while high-stake situations have no correct answers. After all, if we believe in a moral maxim, then we should follow it absolutely whenever possible. And when we cannot follow a moral maxim, then there must be a conflict, either with itself or a different moral maxim. Both in theory and in practice, side-constraints are consistent.

References

[1] Neumann, M. (1982) 'Side Constraint Morality', in *Canadian Journal of Philosophy* 12/1: 131–143.

[2] Nolan, D. (2009) 'Consequentialism and side-constraints', in *Journal of Moral Philosophy* 6/1: 5–22.

[3] Nozick, R. (1974) *Anarchy, State, and Utopia*. New York: Basic Books.

Defending Foundationalism from Michael Williams*

Matthew Laurence Kenney
University of Pittsburgh

Introduction

In his book *Groundless Belief*, Michael Williams presents a formal account of what he calls 'the no-foundations view' about justifications for our beliefs, which is a kind of constructivist coherentism whose essence the following quotation captures:

> If I had to offer an aphoristic statement of the essential nature of justification, I should probably say that being justified consists in doing or saying what your conscience and your society let you get away with.[1]

Towards defending that position, Williams weighs phenomenalism – what he sees as the only other option to his own view – and finds it wanting. This is a type of foundationalism that maintains that our beliefs are ultimately supported by a foundation comprising 'the given', i.e. something nonconceptual in perception that can justify our judgments thereof. Williams critiques phenomenalism via critiquing the idea of the given, maintaining that there are unsolvable logical inconsistencies in the notion of givenness. *In nuce*, Williams asserts that phenomenalism fails without givenness, that justification ipso facto has no foundation without phenomenalism, and therefore that justification amounts to a purely social phenomenon.

While I grant that phenomenalism requires givenness, I dispute his thesis here in several other ways: first, I maintain that Williams' arguments against givenness are not decisive on the matter, and second, that the possible failure of phenomenalism does not entail Williams' particular breed of coherentism. In what follows, I begin by sketching Williams' arguments against givenness, then dispute particular points, and close by briefly discussing some alternatives to phenomenalism and Williams' no-foundations view.

---

*Delivered at the BJUPS Autumn Conference, 21–22 November 2015 at the London School of Economics.

[1] Williams [8] p. 115.

Givenness

Williams' overall argument of *Groundless Belief* takes the following structure:

(1) If phenomenalism is the correct view of justification, then the given (more specifically, our apprehension of it) must "provide the ultimate check on empirical knowledge",[2] i.e. must serve as a foundation for empirical knowledge, since "the claim that empirical knowledge must rest on a foundation…[is] the fundamental idea behind the notion that there is a given element in experience".[3]

(2) The apprehension of the given is either propositional or non-propositional.

(3) If the apprehension of the given is propositional, then "the sort of awareness we have in our apprehension of the given is just another type of perceptual judgment and hence no longer contact with anything which is *merely given*".[4]

(4) If the apprehension of the given is non-propositional, then "the appeal to the given…cannot explicate the idea that knowledge rests on a perceptual foundation".[5]

(5) The apprehension of the given is either apprehension of a perceptual judgment and hence not *merely given*, or cannot explicate the idea that knowledge rests on a perceptual foundation. (3 & 4)

(6) If the apprehension of the given is not merely given, then it cannot provide a foundation for empirical knowledge.

(7) If the apprehension of the given cannot explicate the idea that knowledge rests on a foundation, then it cannot provide a foundation for empirical knowledge.

(8) The apprehension of the given cannot provide a foundation for empirical knowledge. (5–7)

(9) Phenomenalism is not the correct view of justification. (1 & 8)

---

[2] Ibid. p. 28.

[3] Ibid. p. 59.

[4] Ibid. p. 39. Williams gives additional arguments for this that will be discussed in Section 2.

[5] Ibid.

Simply put, givenness cannot be grasped propositionally and linguistically on the one hand or conceptually and ineffably on the other; consequently phenomenalism, requiring this very notion, fails to provide an account of how beliefs are justified. I will respond by taking each of the horns of Williams' Dilemma in turn, first considering non-propositional givenness, and then propositional.

Non-Propositional Games

Williams belabors (4) – that non-propositional knowledge (specifically sense-data here) cannot be a foundation for knowledge, maintaining that supposing the given to be nonpropositional simply makes it "unintelligible how awareness of the given could serve as a check on anything, how such knowledge could ever count in favor of one hypothesis rather than another," and that "the doctrine of the ineffability of the given is impossible to maintain. Necessarily, it breaks down when one attempts to give some characterization of the given whereby it is to be distinguished from other elements in experience."[6] Williams' reasoning behind this is that the doctrine of the given being non-propositional arises as a result of maintaining both that *particular things are given and that the given is always apprehended non-inferentially*:

> The idea that the given constitutes the sensuous element in experience strongly suggests that particular things are given, such things as colour patches, Lewis's qualia, and so on. But the theoretical goal which the concept of givenness is called upon to play requires that taking the given amount to acquiring non-inferential knowledge. That these two views must prove difficult to combine should be apparent.[7]

Williams is here pointing out the issue (first identified by Sellars[8]) when combining awareness *of*, for example, a certain sensation being a sensation of red on the one hand with knowledge *that* the sensation is red on the other. Simply put, the phenomenalist confuses *knowledge-of* with *knowledge-that* – the difference between simple awareness of a sense-datum and knowing that a fact about a sense-datum obtains. But once the description of the given has moved

[6] Ibid. pp. 32–33.

[7] Ibid. p. 34.

[8] Sellars [5] pp. 21–22.

from knowledge-of to knowledge-that, the discussion is no longer about *non*-propositional knowledge in the slightest. Unfortunately for the phenomenalist, there is no way for knowledge-of a sense-datum alone to serve in the role of the given.

In response, I wish to consider a possible alternative that Williams does not mention. In the same way that some phenomenalists (according to Williams) discreetly transitioned from knowledge-of to knowledge-that, Williams has discreetly transitioned from speaking of nonpropositional givenness to speaking specifically of the supposed givenness of things such as sensedata, qualia, etc. (although he focuses on sense-data), and has not supplied any reasons for doing so, even though one might want to speak of *objects* themselves being given instead of sense-data. Sense-datum theories were originally proposed as an explanation of variations in our perceptual experience of objects seemingly unchanged *an sich*: a table, for instance, can appear to be different colors under different conditions, and sense-data can supposedly account for these differences in perception. For my part, however, I suspect alongside other philosophers that sense-data are not required to explain the differences in perceptions of some static object. Take the following analysis from Austin, who writes of Ayer's bent stick version of the argument from illusion:

> Well now: does the stick 'look' bent to begin with? I think we can agree that it does, we have no better way of describing it. But of course it does *not* look *exactly* like a bent stick, a bent stick out of water – at most, it may be said to look rather like a bent stick partly immersed *in* water. After all, what in this case is supposed to be *delusive*? ...Does anyone suppose that if something is straight, then it jolly well has to *look* straight at all times and in all circumstances? Obviously no one seriously supposes this. So what mess are we supposed to get into here, what is the difficulty?[9]

I am inclined to pose the same questions to the sense-datum theorist. What Austin's argument suggests in a rather candid way is that there is nothing utterly absolute about the supposed 'illusions' of sense experience that *force* the philosopher to consider sense-data as necessarily existing. Proposing their existence is optional, and we might do just as well without it. So if we decide to drop sense-data and there is none thereby needed to explain the 'illusions' of experience, then there is no reason to suspect that we are *not*, in fact, acquainted with physical objects. This is not a possibility that Williams considers.

[9] Austin [1] p. 29.

So if talk of sense-data is optional, then we can bypass talk thereof and speak directly about physical objects. But the question still remains about whether talk of physical objects can sidestep the original issue of cross-breeding both propositional and non-propositional apprehension of the given. On the face of things, it seems that in talking about knowing, for instance, 'that conifers have needles instead of leaves', I may only be talking about propositional knowledge and not discussing any kind of non-propositional awareness at all. In reply, I propose that knowledge of physical objects (specifically the ones that require no mediation of sense-data in order to be true or false) can be expressed in knowledge-that language without any essential change to meaning, which (if correct) has a two-fold consequence: first, that the original objection posed to sense-data is not salient regarding physical objects, and second, that talk of objects does not need to have one foot in knowledge-that discourse in order to be given. In regards the first consequence, I can only say that there seems to be a certain amount of noteworthy difference between knowing 'of a red-colored sense-datum' and knowing 'that the sense-datum is red', but the difference between knowing 'of conifers having needles' and knowing 'that conifers have needles' seems inconsequential. Concerning the former case, the difference between a red-colored sense-datum and a sense-datum that is red is the difference between – as Williams' says – a "particular shade of color…[and] a particular which is thus coloured",[10] i.e. the difference between being presented with an independent bit of redness on the one hand and a thing (datum) which is doing the presenting of the redness on the other. This is an obvious difference in meaning, but not one that seems to arise when speaking of actual objects being perceived; when speaking about the greenness of the conifer needles I am by no means speaking of an independent bit of greenness as opposed to a neutral thing that carries green information – in both cases the talk is simply concerning green conifer needles. So while I may be able to speak about knowing 'of conifers having needles' instead of knowing 'that conifers have needles', I have difficulty explicating any kind of philosophically contentful difference between the two that is not purely verbal. I have neither the room in this paper nor the slightest hunch for beginning to explain why there might be such a difference between sense-data and physical objects concerning knowledge-of and knowledge-that, but if there is a difference, then articulating a knowledge-that statement about a physical object being given is no different than articulating a kind of non-propositional awareness about a more complicated topic.

[10] Williams [8] p. 35.

I anticipate the objection that rather than knowledge-of being expressible in knowledge-that for physical objects, the converse is true and therefore that knowledge about physical objects is always propositional, thereby banning physical objects from ever serving as non-propositional foundations.[11] This is a fair point, but if the same meaning is expressible in two different ways, then I doubt that there is a way to tell which expression is the 'right' or 'correct' one. Moreover, I maintain that it does not really matter: what is important here is that talk of some physical objects can fulfil the roles both of propositional knowledge and non-propositional knowledge without an essential change to the meaning. Whether knowledge concerning physical objects is *actually* one or the other seems unnecessarily metaphysical. To take a crude example, a mechanic's inner 'mechanicness' would not restrict him from completing the tasks of a janitor if he knew how to mop floors and mop them well. If one were to press this point by saying first that talk of physical objects is not *truly* non-propositional although it must be, and second that we must *know* that it is non-propositional in order for it to serve as a non-propositional foundation for knowledge, I can only reply that talk of physical objects has the capacity to behave and function non-propositionally, and I am not sure what else is even needed for a statement to be non-propositional. Simply put, if a sentence can be apprehended non-propositionally, then it can be known non-propositionally, and whether it is *actually* non-propositional is a discussion I would rather avoid.

In addition to this first consequence of talk of physical objects bypassing the difficulty present with talk of sense-data between knowledge-of and knowledge-that, the second *a fortiori* consequence is that talk of physical objects may not fall into the same trap present with talk of sense-data in regards establishing a foundation for empirical knowledge. While sense-data on a knowledge-of conception alone can in no way give a foundation to empirical knowledge, talk of physical objects does not seem to have this limitation, since it can satisfy the role of propositional knowledge without thereby becoming propositional (conversely for the post-analytic-neolinguistic-pragmatist types, it can satisfy the role of non-propositional language without thereby becoming non-propositional). Remember that the original problem for sense-datum theories (according to Williams) was their inability to maintain a non-propositional foundation for empirical knowledge without thereby becoming partially propositional as well. But if what I have been saying concerning talk about some physical objects is correct, then their ability to sidestep the issue of knowledge-

[11] Quine would, I believe, make such an objection. See Quine [3] p. 44.

of becoming knowledge-that grants them the ability to serve as the non-propositional apprehension of the given by being able to function propositionally without becoming propositional knowledge. If this is correct, then they are a candidate for non-propositional knowledge that Williams has not discussed. Non-propositional knowledge may be saved after all.

Propositional Givenness

(4) may be less certain than Williams thinks, but there are also things to be said about (3), which is really a critique of the notion of 'basic beliefs', specifically regarding the so-called incorrigible beliefs about the way things appear.[12] Throughout the book, he gives many arguments against basic beliefs in addition to (3), and I have attempted to respond to the most salient here in addition to the one mentioned in the formal argument outlined in the last section. In what follows, I will introduce an argument that Williams gives and respond to it, beginning with the one mentioned previously, and then continuing with three others. (4) maintains that since basic beliefs are a step removed from the given, we are no longer acquainted with the thing given but only acquainted with our perceptual judgment of the thing given. This strikes me as incorrect. Take this response from Ernest Sosa:

> Consider practices – e.g. broad policies or customs. Could not some person or group be justified in a practice because of its consequences: that is, could not the consequences of a practice make it a good practice? But among the consequences of a practice may surely be found, for example, a more just distribution of goods and less suffering than there would be under its alternatives. And neither the more just distribution nor the lower degree of suffering is a propositional attitude.[13]

This example shows an instance of a community or person who is justified in maintaining a certain belief because of events in the world. This does help to show that one can perhaps be acquainted with more than just the basic beliefs themselves. Consequently, while the apprehension of the given might be propositional, there is no reason to suppose that it thereby removes enough contact with experience to make any apprehension of the given impossible.

---

[12] Williams [8] pp. 69, 161–162. I accept this specification since there seem to be no other suitable candidates for basic beliefs.

[13] Sosa [6] p. 7.

Second, Williams maintains that a belief can be basic just in case it is *intrinsically credible*, by which he means that the belief must both require no further justification and cannot be given further justification: not only must it terminate the conversation of the present community, it must be impossible to ever give justifications for it, and since this can never be done, no beliefs are intrinsically credible and therefore not basic. Thus, no beliefs can serve as foundations to empirical knowledge.[14] Williams accepts that an actual regress is halted by the fact that communities currently accept some beliefs as terminating points in a discussion, but the potential regress is always present since more justifications *could* always be provided for a so-called basic belief: once, for instance, a person understands what makes basic beliefs basic, *that* becomes the new basic belief. But if further justification can always be given for a basic belief, then those beliefs cannot halt the threat of a potential regress, and they therefore cannot be intrinsically credible, and thus cannot be basic.

In response, I suspect that Williams is running together justification and explanation, where justification is (for our purposes) simply supplying good reasons for holding a belief and explanation is some further description that does not count as a *reason* for holding that belief – the difference between asking why and asking how. Granted, one might be able to ask for a causal ancestry of some incorrigible belief that 'this seems blue' and be given an answer, but that causal ancestry does not become the new basic belief. Nor does finding out 'why' certain beliefs are incorrigible and therefore foundational move the foundational role from the incorrigible belief to the explanatory belief, since no justification proper was ever given for the incorrigible belief, since that belief would still be basic for someone who held it even if they were unaware of why it was incorrigible.

This gives a response to Williams' example of a further 'justification' being given for a basic belief, but perhaps some other justifications for an incorrigible belief lie underneath that Williams has not considered. I doubt that this is possible: no accurate response can be given to the question of whether one has good reasons for believing that things are appearing a certain way, because the way things appear is beyond the control of the person who holds the belief about the way things appear, and providing justification of *anything* assumes some kind of autonomy (rightly or not) present in the individual. If, for instance, someone grabs my hand and thrusts it into a fire, and a passerby approaches me and asks why I did such a thing, I could only stare at him blankly:

[14] Williams [8] pp. 61–69.

I cannot give good reasons because it was beyond my control. Alternatively, if *I* decided to thrust my hand into a fire, and that same passerby approached and demanded reasons for it, then I could respond. So it seems straightforward that justification can only be given voluntary actions. Likewise, it seems rather straightforward that the way things appear is beyond the volition of the person, since the brain is hardwired to interpret experiences in a certain way. One might take a hallucinogen or through a sheer force of will command things to appear differently, but this only changes the input information with which the brain works, and does not give the person any control over the way things will seem. Even if it *did*, justification could still not be given for the way it seemed, but for why one *decided* to change the way things seem, which is a question of volition over action and not over appearances. It seems safe to say that appearances are not under direct control, and so no justification can be given for them. If this is correct, then these types of beliefs really are intrinsically credible because no more reasons can be given for them. Consequently, they can stop both the actual and potential regress of justifications, making them basic and therefore foundational.

I likewise think that this answers another objection that Williams musters against basic beliefs, that if our foundation is comprises such beliefs, then "the certainty of the given would seem to be threatened. For describing an activity is not an activity that brings with it a logical guarantee of success."[15] I have two brief responses. First, it is by no means obvious to me at least that the constituents of a foundation for knowledge must be completely certain. What makes a belief foundational is the fact that it requires no further justification, not that it could not be false or that it frequently turns out to be true. Second, even if certainty is required for phenomenalism, the fact that basic beliefs are incorrigible and cannot be given any further justification entails that they are certain. Of course, Richard Rorty famously maintains otherwise with the possibility of 'cerebroscopes' showing how things *really do* seem: if what the cerebroscope says is different than what the person reports, then how things seem to someone is not the end of the discussion.[16] But I suspect that there are three other distinct interpretations of a possible disparity between a report of an incorrigible belief and a cerebroscope result: first, the machine could be broken, second, the person could be lying, and third, mental states could turn out to be completely different than brain states and Descartes was right all along. So even if cerebroscopes were somehow invented, this by no means

[15] Ibid. p. 32.

[16] Rorty [4] p. 421.

completely eliminates the certainty of reports about the way things appear.

Williams' last argument I will address is that even if those basic beliefs are certain and no justification can be provided for them, they still may not be foundational since earlier information is needed in order to make these kinds of reports. Williams calls this the 'parasitism argument': epistemologically basic beliefs require prior knowledge that must be known in order for the latter knowledge to be known.[17] As Strawson points out in *Individuals*, one cannot speak of knowing about a certain sensation without knowing what a person is, and Rorty accurately observes that "before sensations can be talked of, certain pieces of behavior must be capable of being talked of as well".[18] But if there are necessary bits of knowledge that must be known before so-called basic beliefs can be known, then there is no reason to maintain that those basic beliefs really are basic at all, even if they are certain and terminating.

This talk of identifying certain things in experience like persons and behaviors pushes me to suspect that the parasitism argument is not so much a critique of basic beliefs as it is a sort of linguistic-Kantian critique about non-propositional knowledge. If so, this is easily dismissed by recognizing that although someone who points out a person in experience without the necessary language to identify a person only points out some ineffable patch of awareness, this person still has *some* kind of experience, which is all that the phenomenalist supporter of non-propositional knowledge requires, since that experience is itself a sort of ineffable givenness. But suppose that the parasitism argument applies to basic beliefs in addition to non-propositional knowledge. If it does, then it shows that I must have a prior knowledge of certain propositions before I can construct basic beliefs about my experience, and I have no wish to deny this. But all that follows from this is that there is a *temporal* priority of learning such things as language, what people are, etc. that is different from the *logical* priority of basic beliefs. This is evidenced by the fact that – presuming such knowledge to be necessary – these prior beliefs required to pick out behavior and persons would still be logically supported by the basic beliefs insofar as beliefs about persons, for instance, would still need to be logically upheld by basic beliefs thereof – and *that* can only arise from experience. Apart from that, there is no reason that the temporal order whereby different beliefs and propositions are learned must mirror the logical order of the justification of beliefs. In fact, the different ways that the edifice of knowledge is constructed

[17] Williams [8] p. 148.

[18] Rorty [4] p. 421.

can be entirely indexical to each individual without thereby altering the foundational structure of justification itself: there is no reason to assume that the order in which certain nodes on the logical chain of justification are discovered somehow changes the very order of that logical chain. Thus, even granting the parasitism argument does not really affect capacity of propositional apprehension of the given to be logically prior to anything else.

Alternatives

So far I have attempted to defend phenomenalism from Williams' critiques; insofar as it succeeds, Williams' arguments have not decided the fate of phenomenalism. At this point, however, I recognize that one might feel partial to Williams' position and desire to defend it, or one might join the ranks of the many who have critiqued all kinds of phenomenalisms on multiple fronts. Yet even if phenomenalism turns out to be completely unsalvageable and Williams' position redeemable, *the negation of phenomenalism does not entail Williams' view of justification*; for that to obtain, these would have to be the only options, but rather there are many other disjuncts available to the epistemologist when choosing a theory of the structure of justification. Williams has simply constructed a false dichotomy. With this in mind, I will devote this section to identifying and exploring other options available beyond just Williams' coherentism and phenomenalism. First, phenomenalism is only one form of foundationalism. One type of foundationalism (which I mentioned in passing before) is the Platonic Theory of Recollection. A consequence of this metaphysical theory is the epistemological point that our justification for our beliefs comes from remembering our soul's time in the world of the forms before incarnating here. This is best exemplified in the *Meno* dialogue when Socrates purportedly shows Meno's slave how his ability to calculate geometry arises from recalling his time in world of the forms.[19] While not explicitly stated this way, Meno's slave is justified in the conclusions he reaches about the lines and lengths of the geometric figures by his ability to remember experiences he had in his existence prior to this one. This is obviously a form of foundationalism since the experiences in the world of the forms terminate the justification process in an absolute way, and our recollection of the forms justifies our beliefs comprehensively in that there is not a justified belief which receives its ultimate justification from something other than a memory of a form. Thus, we have an instance of a foundationalism that bypasses the issues that Williams' raises

[19] Plato [7] pp. 77–78.

with the apprehension of the given, basic beliefs, etc., and we therefore have at least one alternative to both phenomenalism and constructivist coherentism that warrants the claim that Williams has thrust a false dichotomy upon us.

I mention Plato's Theory of Recollection in particular to show that foundationalism can encompass a wide range of views beyond the sense-datum theories of the first half of the twentieth century. I recognize, however, that this is not a popular view, so I will also bring to mind 'infinitism', expressed best in Peter Klein's work. The original Agrippan Trilemma proposes three possibilities: either our chains of beliefs stop arbitrarily, a belief reappears in the chain and makes it circular, or the chain continues infinitely. The foundationalist tries to show that some stopping points are not arbitrary and the coherentist that circular reasoning is sometimes appropriate, while infinitism maintains that chains of justification do continue infinitely, while we are still justified nonetheless. One recent supporter of infinitism is Peter Klein, who maintains that infinitism follows from two fundamental premises in epistemology: the principle of avoiding circularity, and the principle of avoiding arbitrariness.

(PAC) For all x, if a person, S, has a justification for x, then for all y, if y is in the evidential ancestry of x for S, then x is not in the evidential ancestry of y for S.

(PAA) For all x, if a person, S, has a justification for x, then there is some reason, r_1, available to S for x; and there is some reason, r_2, available to S for r_1; etc.[20]

(PAC) eliminates any type of circular coherentism, and (PAA) eliminates foundationalism. But the conjunction of both of these principles entails that justification extends infinitely. While I do not have the space to defend either of these, a detractor must at least admit that they are both very plausible, and that infinitism may therefore be something of a live option worth debating.

There are now two options in addition to phenomenalism and constructivist coherentism, and it should be clear that Williams has presented a false dichotomy. Many more options exist in addition to these, so one need not deliberate between the two positions that Williams suggests. While this has been a short section, I think it has succeeded in offering and describing two relatively plausible alternatives – each accepted by some contemporary philosophers – that may be adopted instead of either Williams' no-foundations view or phenomenalism.

[20] Klein [2] pp. 298–299.

Conclusion

This essay has essentially scrutinized premises of an implicit disjunctive argument in the background of Williams' *Groundless Belief*:

$$\text{Phenomenalism} \vee \text{No-foundations view}$$
$$\neg\text{Phenomenalism}$$
$$\therefore \text{No-Foundations view}$$

I first objected to the second premise by attempting to show that Williams' arguments against phenomenalism with either a propositional or non-propositional apprehension of the given fail to demonstrate that phenomenalism is logically inconsistent as a theory of justification. I then disputed the first premise by proposing alternative ways to view justification besides phenomenalism and the no-foundations view, which are themselves particularized forms of foundationalism and coherentism, respectively. Insofar as any of it succeeds, Williams' arguments do not entail his particular theory of justification.

References

[1] Austin, J.L. (1964) *Sense and Sensibilia*. New York: Oxford University Press.

[2] Klein, P. (1999) 'Human Knowledge and the Infinite Regress of Reasons', in *Philosophical Perspectives* 13/13: 297–325.

[3] Quine, W.V.O. (1961) 'Two Dogmas of Empiricism', in *From a Logical Point of View* (2nd ed.). Cambridge: Harvard University Press.

[4] Rorty, R. (1970) 'Incorrigibility as the Mark of the Mental', in *The Journal of Philosophy* 67/12: 399–424.

[5] Sellars, W. (1997) *Empiricism and the Philosophy of Mind*. Cambridge: Harvard University Press.

[6] Sosa, E. (1980) 'The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge', in *Midwest Studies in Philosophy* 5/1: 3–25.

[7] Plato (2002) 'Meno', in *Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo*, trans. G.M.A. Grube. Indianapolis: Hackett Publishing Company Inc.

[8] Williams, M. (1999) *Groundless Belief* (2nd ed.). Princeton: Princeton University Press.

Nozick, Kant, and Redistribution: They Can All Get Along

Deion Kathawa
University of Michigan

Introduction

Whether redistributive taxation systems can be reconciled with libertarianism is an interesting and (potentially) intractable question in political philosophy. Harvard philosopher Robert Nozick is well known for his opposition to any State other than one that takes the form of what he calls a 'minimal State': one that is very little more than a 'night watchman' and simply ensures that its citizens' liberties (i.e. negative rights) are protected. People are free to act however they choose so long as they do not harm anybody else – infringe on their basic rights. The State may only interfere when a person does in fact threaten another's rights. I will contend that Nozick's moral principles do not entail his conception of justice (i.e. no redistribution) because he is misappropriating Kant, and I will also present one additional argument against his prohibition of redistributive taxation in order to demonstrate that a state-run redistribution scheme is indeed possible, plausible, and workable within a libertarian framework.

For libertarians, the most basic right is the right to property. Nozick contends that any redistribution by the State is unjust because goods are not what he calls 'manna from heaven'; they do not suddenly appear out of nowhere. They are made by people through their labours; therefore, when the State – via taxes – collects and redistributes these goods, it perpetrates upon the taxed a sort of 'violence', infringing on their right to keep the property that that they produced by and through their efforts.

This paper will have three main parts and several subsections. It will first explore Nozick's moral commitments (self-ownership, the priority of liberty, the absolute right to property) and his conception of justice (historical theory of entitlement). Then, it will analyse Kant's categorical imperative as a basis for Nozick's rejection of redistribution, and explore an additional argument against the general libertarian's rejection of redistribution. Finally, it will respond to some possible objections. In so doing, libertarianism will be reconciled with a minimally redistributive taxation system.

1 Nozickian Libertarianism

1.1 Moral Commitments

Nozick's libertarianism is grounded in three separate, though interrelated, moral commitments: people's inherent right of self-ownership, the priority of liberty, and the absolute right to property. These elements constitute 'libertarianism' as such. First, Nozick's conception of self-ownership. This is the concept that, if we own anything, we first own ourselves. This is the first thing that we own, and it is because of and through this primordial ownership of self that we can ever come to own anything. Second, his prioritization of liberty. Nozick holds that people's right to do what they please, as long as what they choose to do doesn't harm another person, is utterly fundamental to their freedom as persons. The government cannot stop its citizens from acting and doing what they want; it must only secure their negative liberty: their right to be left alone. (Related is his conception of equality. Nozick believes that all people are *morally* equal, much like Locke, but that is where their equality ends. No one, simply in virtue of their equality, is entitled to anything except a chance at 'making their way' in the world.)

Third, the absolute right to property. Nozick holds that, because we own our bodies, or selves, absolutely, what we make out from our labours belongs to us in the same way that our bodies belong to us. Because our efforts are the only things that convert raw material into products and goods worthy to be sold on the market, one owns the fruit of one's labour as absolutely as one owns one's own body.

1.2 Justice as Entitlement

Nozick bases his theory on what he calls an 'historical' theory of entitlement. On this view, there are three ways in which a person can licitly come to own something.

First, through legitimate, initial acquisitions of property. If a person comes across property that is unowned, they may take it for themselves. This principle can be seen in action in North America in the sixteenth and seventeenth centuries when settlers staked claims to unoccupied land as they expanded farther and farther westward (though historical scholarship has revealed this theory of benign land attainment to be false). As long as there was no violence or coercion in the acquisitions, according to Nozick, they were just.

Second, though legitimate transfers of property. When one person owns some property, and they voluntarily transfer it to another person, a legitimate transfer has occurred. One person simply sold the land (or whatever it was) to another person. As long as there was no violence or coercion in the exchange, then it is within the person's right to offer up his property for sale to the other person; it is just.

Third, through rectification. Nozick understands that the history of the world is not one in which people have legitimately acquired initially unowned property or voluntarily transferred it on the market but, rather, was one of unjust, involuntary exchanges, in which those with better weapons typically forced a group of people to give up what was (to Nozick) rightfully theirs. Because of this reality, Nozick supports compensating transfers that would give to each person what they would have owned had there been no violent, coercive, and unjust transfers. But because he recognizes the challenges and limitations of this strategy (how can we know the amount of property each person would have had absent injustice?), he also supports equal amounts of property being initially doled out to members of society. In his mind, this is perhaps the best way to mimic what might have been the current allocation of property if all transfers throughout history had in fact been voluntary and just.

2 The Arguments

2.1 Kant-As- Nozick-Supporter

Kant formulated a well-known moral principle within his deontological ethical framework called the categorical imperative. There are three formulations of his imperative; the second states "So act that you use humanity, whether in your own person or in the person of any other, always at the same time, never merely as a means".[1] Nozick, in his *Anarchy, State, and Utopia*, explores whether something like redistribution is permissible. Because people often bear certain costs for a greater overall good (e.g. the pain of the dentist avoids greater suffering later), he reasons, why not similarly have some bear a certain social cost if it gives rise to a greater overall 'social good'? Nozick, in an effort to explain why this thinking is erroneous, writes "[T]here is no social entity with a good that undergoes some sacrifice for its own good. There are only individual people, different individual people, with their own individual lives.

[1] Kant [2] p. 496. Emphasis mine.

Using one of these people for the benefit of others, uses him and benefits the others. Nothing more."[2]

2.2 The Turn: Where Nozick Meets Redistribution

In a shallow sense, Nozick is correct: simple redistribution is merely the taking of one individual's resources and transferring them to someone else. What Nozick overlooks, however, is that Kant's second formulation of the categorical imperative stresses that we cannot use people 'merely as a means' to some an end. But redistribution does not necessarily use people as mere means to some end.

Nozick holds that "[t]axation of earnings from labour is on par with forced labour".[3] He believes that a taxation system turns the taxed into partial slaves because the government gives some people a right to the proceeds of another's labour, which is a part of their self-ownership; therefore, they now have partial ownership of this person through claims on their property. This has been popularised into the convenient quip'taxes as slavery'. Redistributive taxation allows for the creation of societal improvements and resources (highways, public parks, libraries, police and fire departments, the military, etc.) whose existences would be otherwise impossible for groups of individuals acting in concert (less government) to bring about. These societal features and institutions most definitely benefit even those taxed. To say that one is being 'partially owned' at first appears plausible but is a bit dishonest upon closer examination: the taxes are going toward improvements from which everyone benefits. No one is being used 'merely as a means' to some end, as Nozick contends.

Nozick's schema also implicitly assumes that people only want to keep all of their wealth and that they derive no benefit from having it redistributed to others. It is also almost certainly false that everyone who earns a living only desires to keep all that they earn only for themselves. There are some who wish to help others through charitable giving but for myriad reasons simply do not (e.g. they lack the proper motivation). In these cases, the State may act as a vehicle through which individuals are freed from the burden of having to choose what their Kantian 'higher self' would choose to do if that part of themselves was in total control. People will be 'forced' to give up some of their wealth, but it is actually what they truly desire to do in the first place.

---

[2] Nozick [3] pp. 23–33.

[3] Nozick [3] p. 169.

The State stands in as facilitator for the desires and actions of people's rational selves.

2.3 Loss of Autonomy Inheres within Vast Inequities

Nozick holds that all people have a right to liberty and property, or at least the ability to 'sell themselves' and work for someone who does have property, i.e. the ability to contract themselves out for a wage. He does not hold that vast inequalities are unjust. Perhaps charity and other private actions can ease them, but it is most definitely not the proper role of government to involve itself in rectifying these disparities. They are matters of charity and of brotherhood, not of State action.

But what happens when inequalities are so vast that they rob people of their autonomy and right to liberty and self-directedness? We can imagine a society with economic inequalities so vast that a subset of the population must contract themselves out as slaves (even in spite of the country's having formally outlawed slavery) because it is the only way that they can survive. Feminist philosopher Ann Cudd, in her book, *Analyzing Oppression*, writes that there are necessary background conditions for this type of institution to develop. One of them is the existence of a "desperately poor population that can be easily lured away from their traditional support systems". She then writes about a specific case: men who are "enticed by the promise of a salary, food, and shelter to leave their urban shantytown homes and families to go make charcoal from the Brazilian rainforests for sale to the steel mills".[4]

Cudd does not invoke a hypothetical thought experiment in her book: her analysis is rooted in the reality of the Brazilian economy. This system – one in which morally equal people with an inherent right to self-ownership and autonomy – exists, Cudd argues, because the country "suffers from a highly unequal distribution of income".[5] It is difficult to see how these people are free and self-directing. It is clear that the deplorable conditions in which these men find themselves are (1) brought about (or at least, made possible) by the income chasm that Cudd references and (2) severely diminishes (if not outright eliminates) these men's right to liberty.

Certain individuals are the ones who profit from this system. Indeed, Cudd

[4] Cudd [1] p. 114.

[5] Ibid.

writes again that "the ultimate beneficiaries here . . . [are] the Brazilians who own the forests and invest in the charcoal camps".[6] Everything in this real-life scenario follows Nozick's libertarian schema: a class of people owns land, invests in the productive use of that land, and inequalities result. But it does not stop there, as Nozick would prefer. Instead, the wealthy and powerful set up a system (or passively allow one to take root) wherein the desperate poor of the country are approached with an offer, and they 'freely' accept it. They are then enslaved against their will. In a word, they are preyed upon. They are no longer free and autonomous as they have a right to be. It seems that government must step in with some sort of redistributive system in order to rectify the economic inequities that are the driving force behind why Brazilian men are being enslaved, thereby preventing this gross violation of people's liberty.

3 Is Nozick Right After All?

3.1 Objections

Several objections can be raised in opposition to the above arguments. First: individual citizens can collectively group together and form associations in order to accomplish the things that (allegedly) only government can. In this way, individual autonomy is protected, only those who actually pay for the improvements are permitted to use them, and the coercive State is kept at bay. While this seems plausible, upon closer examination it fails for at least two reasons.

Reason One. Government's existence and subsequent actions are likely the most efficient way in which large-scale projects can be tackled. The collective action problems (particularly the 'free rider' problem) inherent in tasks that government currently handles are likely to make the projects attempted under the Nozickian 'associations' model vastly more expensive than their current iterations or even altogether insurmountable. Individuals and corporations, in Nozick's world, will compete with one another, stalling projects and engaging in underhanded tactics to secure contracts and clients – all at a time when unity and a shared purpose are paramount. Allowing government to execute certain large-scale project avoids all of this messiness while also aligning with libertarians' desire for productive, cost-effective, and labour-saving plans and

[6] Ibid.

measures.

Reason Two. Even if individuals could somehow effectively join together and accomplish 'Massive Task A' – in this case, let's call it having built a highway – they will then need to create a system in which those who did not contribute to its construction (either monetarily or labour-wise) are for that reason barred from using it. A private security force would need to arise to guard the road, and a stratified society would emerge, with the rich being permitted to use the road and the poor being unable to use the road. This is problematic enough as it is, but imagine another scenario: one which a moderately wealthy individual contributed to the building of, for example, one-fifth of the highway. Is he only allowed to use only one-fifth of the entire highway? Only his one-fifth of the road? What would that system look like, and how would it even be enforced?

If we accept that these challenges are ones that we do not wish to deal with for whatever reasons, and that allowing everyone to use the highway increases overall productivity and economic growth, thereby making everyone wealthier, what reason would a libertarian offer so as to remain obstinate and persist in banning those who did not help pay for the highway or who did not actually physically help to build it? The scenario of growth seems to be exactly what a libertarian would want, and Nozick's world appears to be unwieldy and unworkable.

Second: in the Brazil situation, individuals will not do anything to end this system because (1) they are profiting from the status quo or (2) they are actually incapable of tackling such a systemic problem. In light of this, the problem of slavery persists. But this is very obviously repulsive to libertarian sensibilities regarding autonomy and self-ownership and liberty. So what to do? Perhaps Nozick will allow for a public police force to exist as per his conception of the State as a mere 'night watchman' (though some have interpreted his view as not even allowing for such a police force – only for private security firms). If that is the case, then the police can break up the institution of illegal slavery in Brazil. This appears to be a workable solution, but there are two problems, one practical and realistic, the other hypothetical and speculative.

The practical problem: in Brazil, the police and the courts are either complicit in the slavery ring or are too weak to actually put a stop to it. So much for the mere 'night watchman' State which exists solely to protect its citizens' negative freedoms! *The hypothetical problem*: even if the State's police force was actually effective in breaking up the system (as they are not in Brazil) . . . then what? The slaving is dismantled, sure, but the inequality persists. There are still no jobs for these people. What prevents another exploitative system from springing up or from the slaving to return, but in a new form? It seems that the

most appropriate way to proceed is to attack the underlying reason that people are misled into these types of lives: economic insecurity. Providing people with a minimum level of economic security so that they can (1) avoid these types of systems and (2) actually have the effective (not merely formal) ability to be autonomous seems to be the solution. (After all, it is hard to be autonomous when you are constantly having to stave off starvation.)

3.2 Concluding Remarks

In the final analysis it seems that Nozick's libertarianism is amenable to some sort of redistribution. It is unlikely that this paper will have pushed Nozickian libertarianism into Rawlsian (or even Dworkinian) liberalism, but it surely is able incorporate a 'welfare' system which meets people's basic needs without becoming not-libertarianism. Nothing in Nozick's particular political philosophical framework, his reading of Kant, or libertarianism generally, rigidly precludes even minimal redistributive taxation; in fact, it may require it in order to function as Nozick intended.

References

[1] Cudd, A. (2006) 'Violence as a Force of Oppression', in *Analysing Oppression*. New York: Oxford University Press.

[2] Kant, I. (2013) 'Groundwork on the Metaphysics of Morals', in R. Shafer-Landau (ed.) *Ethical Theory: An Anthology* (2nd edn.). Chichester: Wiley-Blackwell.

[3] Nozick, R. (1974) *Anarchy, State, and Utopia*. New York: Basic Books.

[4] Swift, A. (2014) *Political Philosophy: A Beginners' Guide for Students And Politicians* (3rd edn.). Cambridge, UK: Polity.

Can We Maintain Neutrality Towards Islamic *Nikkah* Marriages in a Liberal Society?*

Hettie O'Brien
St Andrews

Focusing on the institution of marriage within a liberal society, I will look closely at two arguments that aim to justify non-neutrality towards marriage. I will argue that both fail to align this non-neutrality with the liberal values of autonomy and self-elected ends. Firstly, I will examine Christopher Bennett's paper, 'Liberalism, Autonomy and Conjugal Love', in which Bennett invokes autonomy as a justification for the state's promotion of conjugal marriage. I object to this argument because it overlooks the subjective nature of realising self-respect. Secondly, I will look at Thom Brooks' paper 'The Problem with Polygamy'. Brooks opposes polygamy and the minimal defence of polygamy advocated by Martha Nussbaum. I object that both arguments are incompatible with the values of self-directed ends and autonomy because they both assert one that form of marriage is better or worse for an agent independently of the choice that agent might make. My objections to both arguments intend to show the difficulties a liberal society might have in justifying promoting one form of marriage over another. This is inspired by the idea explored by John Rawls in both *Political Liberalism* and later, that the state should not involve itself in '*fully voluntary*' areas of the family[1] because doing so might endorse one partial conception of the good over others.[2]

Section 1

The UK is a liberal society. I take the term 'liberalism' to refer to the type of Kantian liberalism advocated by Rawls and delineated by Martha Nussbaum. Nussbaum claims such liberalism is opposed to three things: the turning of morally irrelevant differences into sources of social hierarchy, corporatism, and ideologically based politics. "*Liberalism so conceived is about the protection*

---

*Delivered at the BJUPS Autumn Conference, 21–22 November 2015 at the London School of Economics.

[1] Rawls [13] pp. 147, 161.

[2] i.e. Rawls's idea of a comprehensive doctrine he explores in *Political Liberalism*.

of spheres of choice."[3] Defined as such, a liberal society is concerned with autonomy because it is concerned with protecting individuals' spheres of choice. It is opposed to corporatism, wherein individuals are not permitted to elect their own ends because those ends are subsumed by a teleological societal end. I take the UK to be this sort of society: it values individuals' spheres of choice and is opposed to corporatism, and definitely to ideologically based politics.

Unofficial polygamous marriages exist in the UK. Nobody is sure of the exact figures: some believe it to be up to 20,000, while the government contests it is more likely in the region of 1,000.[4] Many of these polygamous marriages exist within Muslim communities and take place in unofficial *Nikkah* marriage ceremonies.[5] Monogamous marriages are the only legally recognised form of marriage in the UK. This means that polygamous spouses do not receive the same social security benefits or legal recognition as monogamous spouses.[6]

A recent documentary interviewing wives in polygamous marriages in the UK raised interesting questions surrounding the consensual nature of these marriages. In the case of Nibila Phillips, a Cambridge ex-PhD student who claimed that she'd rather be in a polygamous marriage, it seems clear that some women do espouse this lifestyle.[7] If people choose freely to enter into such social arrangements, it raises the question of what grounds society has for promotingone form of marriage over another. My argument is not intended as an advocacy of polygamy. Rather, I am using the case of polygamy in the UK as an illustration of the difficulty a liberal society might have in justifying its non-neutrality towards marriage.

I will now examine the two arguments of Bennett and Brooks that attempt to align a justification of non-neutrality towards marriage and the liberal values of self-elected ends and autonomy.

---

[3] Nussbaum [11] p. 58.

[4] Khan [8], Kennedy [7]

[5] Serck [15]. 'Unofficial' because they are not officially recognized by UK law.

[6] Fairbairn [5].

[7] I mention that she is a PhD student because her access to education is further evidence of her having made the decision to enter into a polygamous marriage from an informed position.

Section 2

Bennett argues that the state is justified in endorsing conjugal marriage over other forms of marriage because it protects the type of conjugal relationships which promote autonomy. He is concerned with a normative rather than descriptive version of marriage. No doubt some actual marriages impede rather than promote autonomy, so Bennett bases his argument on an idealised form that promotes autonomy.

Bennett has a bipartite understanding of autonomy. Someone may have the external freedom from coercion to pursue an option, but be prevented from availing themselves of that option because they are "*lonely and lacking in self-respect*".[8] This lack of internal freedom may inhibit them from taking charge of and valuing their own projects. Having the self-respect to value one's own projects is instrumental for the second part of autonomy, as if the agent does not value themselves and their ends, they cannot take advantage of the freedom they have to elect their own ends. The first premise of Bennett's argument is that a liberal society which values autonomy will be interested in promoting both autonomy and the self worth that is instrumental in utilising this autonomy.

Bennett's second premise is that intimate friendship is crucial to autonomy. He justifies this with reference to the views of Hegel. The view that intimate relationships enable recognition of the self is evinced in Hegel's quote that "[in love] I gain myself in another person".[9] Bennett's second premise draws on the Hegelian notion of recognition. This can be understood metaphorically: it is through the eyes of another that I myself am reflected. More literally, it is through the fact that others recognise us as particular persons in our own right that we have the very sense of being particular persons in our own right.

This applies to societal relationships generally. What distinguishes intimate relationships from those which are more impersonal is two things. Firstly, in intimate relationships we share the content of our lives with others. We share the trials and tribulations that occur throughout our lifetimes, and value "the things [in our intimate friends] that make [them] the particular person [they] are".[10] We care about and recognize the content of our intimate friends' lives as important because we recognize and value them as individuals. Secondly, the

[8] Bennett [1] p. 286.

[9] Williams [16] p. 212.

[10] Bennett [1] p. 289.

recognition delivered in intimate relationships is from people whom we chose to value. Intimate relationships are distinct from parental love, as we choose our intimate friends. The recognition in these sorts of relationship is reciprocal rather than undesired, and is therefore empowering because we value the people who value us, rather than having undesired attention bestowed on us that may feel more like we are "being violated [or] stalked".[11] To be recognized as a particular person worth valuing is crucial to developing a sense of self worth that, recalling Bennett's bipartite definition of autonomy, is instrumental to exercising autonomy.

Bennett must show why conjugal relationships secure autonomy better than intimate friendships in order to defend his claim that the state is justified in promoting them. The art critic John Berger writes that in conjugal relationships "the sight of the beloved has a *completeness* that no words, and no embrace, can match".[12] Unlike intimate friendships, conjugal relationships involve the singling out of one individual for that role coupled with a concern for the entirety of their life. It is this aspect of conjugal relationships that I believe Berger touches on and Bennett defends: conjugal relationships deliver a sense of one's identity and value greater than intimate friendships because they involve a complete concern for the single party as a unique individual. This confers upon each party a correlating set of responsibilities. In certain situations, for example the 'next of kin' section on forms, one would expect ones partner to be on hand to take responsibility for this. Such a close concern for the entirety of a person's life is, Bennett argues, something unique to their relationship.

At first glance, Bennett's argument for the promotion of monogamous marriage seems to be compatible with a liberal society that invokes a minimal conception of the good. On the premises that a liberal society will be concerned with protecting autonomy, and monogamous marriage is instrumental to autonomy, a liberal society is justified in promoting monogamous marriage. I will now present my two objections to this argument.

Firstly, if we challenge the empirical grounds of the claim that marriage promotes autonomy better than other relationships, a liberal society is not left with much to justify its promotion of one type of relationship over another. This brings us to the second, broader problem: in promoting one type of relationship over another, a liberal society overlooks the fact that humans achieve

[11] Bennet [1] p. 289.

[12] Berger [3] p. 1. Emphasis added.

self-worth in a multitude of different ways at different points in their lives. Notions of how best to achieve self-respect are highly subjective.

The idea that conjugal relationships promote autonomy better than intimate friendships is problematic. It is extremely common that a married couple have friends outside of their marriage with whom they share intimate details of their lives that they may not share with each other. This can be said for most conjugal relationships. Furthermore, if conjugal relationships secure the self-respect instrumental to autonomy better than intimate friendships, would this mean that a person would have greater autonomy in a functional conjugal relationship while being without friends than they would have if they had a group of intimate friends, but not a functional conjugal relationship? Bennett might answer 'yes', but I am inclined to disagree. If person *X* possesses both intimate friends and a conjugal partner, and as a result of these relationships has a sense of self-worth, I do not believe that person *X*'s sense of self worth would remain the same if they had no intimate friends. It's not intuitively plausible that *X*'s intimate friendships could be subtracted and their self-respect remain. A more accurate depiction of how intimate and conjugal relationships secure autonomy is one in which they are both of indispensable importance in a person's life. This idea is further supported when we consider how important a strong group of friends is to securing the sense of self-worth as an individual *prior* to a relationship.

There exist many people who have yet to develop the self-worth instrumental to exercising autonomy. Possessing a group of intimate friends who value their qualities and characteristics will benefit and enable their sense of self-worth. Essential to developing a sense of self-worth is the process of *figuring out* what it is about yourself and identity that you value. It may be more beneficial to have an intimate group who each value a different quality or characteristic in you, rather than one person who values the entirety of you. For a more vulnerable, less self-assured person, this type of exclusive affection can be inimical to their development: to be reflected only in the eyes of one other is to only gain one perspective on who it is that you are, whether as to be reflected in the eyes of a group who value you for different reasons permits you to define and decide what it is you value in yourself. It is unlikely that we would advise a person lacking in self-confidence and a sense of self-worth that entering a relationship would be their best course of action. We would be more likely to suggest that they invest time in themselves, gaining a stronger sense of their own identity and pursuing activities that promote mental wellbeing prior to entering a monogamous relationship.

Furthermore, Bennett's argument that monogamous relationships secure autonomy better than intimate friendship doesn't account for cases where in-

dividuals abstain from conjugal relationships. An obvious example of this is religious devotion: would Bennett's argument necessitate the conclusion that those who abstain from conjugal relationships do not have the same sense of self-worth as those who have functional conjugal relationships? A more likely conclusion would be that the realisation of self-worth can take different forms dependent on the individual. The problem with claiming that conjugal relationships secure self-worth better than friendships (or other forms of devotion) is that self-worth is something that is highly subjective.

Section 3

I take Brooks' argument, like that of Bennett, to be a justification for a liberal society to legislate in favour of and against certain forms of marriages. Brooks objects strongly to polygamy. He attempts to show that polygamy is structurally unequal not just in *practice* but in *theory*.[13] If polygamy is only structurally unequal in practice, polygamy is justifiable if it meets with certain conditions in practice. Martha Nussbaum's minimal defence of polygamy claims that polygamy is justifiable iff: (1) It is sex-equal insofar as it is open to both men and women and (2) it is consensual, and the consent of all spouses is granted equal importance.[14]

On this basis, polygamy would be theoretically justifiable in a liberal society if it met with these two conditions. What Brooks wishes to argue is that polygamy is theoretically unjustifiable on all accounts: that, no matter if it meets these conditions, the concept of polygamy is unjust and therefore not defensible. He argues that there are two reasons for this. Firstly, polygamous marriages are always asymmetrically equal. Secondly, polygamous marriages discriminate against non-heterosexuals.

Brooks' first reason is as follows. Regardless of whether it is a woman or a man who is the person with multiple partners in a polygamous marriage, there will always be an asymmetry of equality because polygamy "threatens the equality among polygamous partners".[15] This is because the ability to leave a polygamous marriage is asymmetrical. While the person who has multiple partners can divorce multiple partners, each of those partners may only divorce the per-

[13] Brooks [4] p. 118.

[14] Nussbaum [12] p. 97.

[15] Brooks [4] p. 116.

son at the helm of the relationship. In contrast to polyamory[16] wherein there are more equal opportunities to divorce because "all have an equal voice in how the relationship develops", polygamy as a concept is asymmetrically equal because the ability to divorce multiple members of the relationship belongs only to one person.[17]

The second reason Brooks gives is that polygamy discriminates against non-heterosexuals. This is because polygamy always takes the form of polygyny, wherein one man has multiple wives, or polyandry, wherein one woman has multiple husbands. There is no category for non-heterosexual polygamous marriages, and thus it discriminates against non-heterosexuals.

I will shortly address the objections I direct at Brooks' argument. I do not intend these objections to stand as an advocacy of polygamy. They are intended to demonstrate how objecting to polygamy on a conceptual rather than empirical basis is difficult to align with the liberal values of autonomy and self-elected ends.

The first objection I direct to Brooks' claim that polygamy is always asymmetrically equal is that this ignores the consensual nature of some polygamous relationships. In many monogamous marriages, there is an asymmetry of inequality. Often, one partner will possess more power in the form of psychological leverage than their spouse. In extreme cases, two people might enter into explicitly dominant and subdominant relationships founded on the principle of an asymmetrical power balance. A more frequent occurrence is that power dynamics within a marriage or relationship evolve over time, perhaps because one party feels greater affection or the need to compensate. The idea that healthy relationships should be formed on the basis of equal respect and treatment of one another is compatible with the sorts of relationship which seem to display an asymmetry of equality. As long as there was no further cause for concern within a relationship, such as signs of coercion or abusive behaviour, we would find it normal to say that the matter of power dynamics within relationships is the business of those in that relationship. The underlying point here is that people freely choose what sorts of relationships they enter in to: if a person enters into a polygamous marriage, they do so in the knowledge that they are marrying one person who will also marry others.

[16] Defined by Brooks [4] p. 117 as "a relationship of multiple men and women". These relationships "may take the form of one man with several female partners, but they may also take the form of two men and several female partners or even several women with no men".

[17] Brooks [4] p. 117.

My objection to Brooks' claim that polygamous marriages discriminate against non-heterosexuals is simply that it is erroneous. It is true that polygamous marriages *so far* do not include cases of non heterosexual polygamous marriages. But this does not mean that there is anything governing the reason for this. Nor does it mean that polygamy includes in its conceptual definition an exclusion of non-heterosexual polygamous marriages. It is more likely that this is the case because polygamy occurs more frequently in cultures that also discriminate against homosexuality. Brooks' objection that polygamous marriages discriminate against non-heterosexuals is a case of, metaphorically speaking, addressing the leaves on the tree rather than the trunk itself. Polygamy within certain cultures necessarily excludes non-heterosexual marriages because those cultures discriminate against all non-heterosexual marriages. That does not mean that polygamy definitively discriminates against non heterosexual polygamous marriages.

Section 4

Both arguments serve as a justification for societal intervention in marriage. They argue that one form of marriage is better or worse than others, and on this basis justify non-neutrality towards marriage. Both align the opinion that one form of marriage is better or worse than others with the values of autonomy and self-elected ends that a liberal society values. Bennett's argument justifies the promotion of conjugal marriage on the basis that it is better at securing autonomy. Autonomy is valued in liberal societies, and thus Bennett's argument remains in line with liberal values. Brooks' argument strongly objects to polygamy on the basis that it is inherently unequal, and so his argument is justifiable within any society concerned with the promotion of equality.

In my opinion, both arguments cannot be aligned with the liberal values of autonomy and self-determined ends. I objected to Bennett's argument that it overlooks the subjective nature of self-respect. The claim that conjugal relationships secure the self-respect instrumental to autonomy better than intimate friendships is narrow, and discounts the important role friendship plays in securing self-respect alongside conjugal relationships. Furthermore, it leads to the conclusion that people who abstain from conjugal relationships do not have the same type of self-respect instrumental to autonomy as those who have conjugal relationships. This is not intuitively plausible, and overlooks the fact that people determine their self-respect in a multitude of different ways.

I objected to Brooks' argument because, like Bennett's argument, it ignores that individuals are separate beings who will find value and self-respect in dif-

ferent types of relationship. Brooks' objection that polygamy threatens the equality of polygamous partners due to an inherent asymmetry overlooks the fact that people often chose freely to enter into relationships with an asymmetrical power balance.[18]

If people enter into sorts of relationships free from coercion and of their own choosing, a liberal society will need to justify its intervention in such matters by claiming to know what is best for those individuals. This would be abhorrent to the principles of a liberal society, what Isaiah Berlin termed a "monstrous impersonation" of "[e]quating what *X* would choose if he were something he is not, or at least not yet, with what *X* actually seeks and chooses".[19] So the question remains: how might a liberal society which values autonomy, equality, and the ability of individuals to direct their own ends justify legislating against polygamy? In the case of polygamous *Nikkah* marriages in the UK, individuals arguably chose to enter into polygamous marriages of their own free will. If the argument that monogamous marriages promote greater autonomy is rejected, and the fact that these individuals have entered into asymmetrically equal marriages of their own accord is acknowledged, what grounds would a society have in promoting one form of marriage over another and legislating against polygamy? In short, what justifies a society's non-neutrality to the issue of polygamous marriages?

I suggest that a liberal society's intervention in marriage, insofar as it promotes monogamous marriage and does not legally recognise polygamous marriage, is better justified through an empirical rather than conceptual framework. As Brooks notes, polygamous marriages are proven to result in a greater likelihood of women suffering from depression and low self-esteem than monogamous marriages.[20] Furthermore, polygamous marriages are more likely to result in dynamics of subservience, wherein women are primarily responsible for child-bearing and may be prevented from seeking employment thus creating situations where women are financially dependent upon a marriage and unable to leave.[21]

Crucially, a further discussion of the idea of autonomy is needed to ascertain whether women entering into polygamous marriages really are doing so of

[18] Ibid. p. 116.

[19] Berlin [2] p. 18.

[20] Brooks [4] p. 111, Al-Krenawi & Slonim-Nevo [10].

[21] Gher [6], Al-Krenawi & Lightman [9].

their own free will. To chose something freely, one must be free from coercion and obstacles to exercising choice. To declare that women within polygamous marriages have entered into those arrangements free from coercion, it would have to be determined whether and to what extent financial or social pressures played a factor in their decisions. It would not be deemed a case of decision free from coercion wherein a woman entered into a marriage because they had no financial independence and risked becoming ostracised from their community.

I hope to have illustrated some of the contentions underlying a society's justification of non-neutrality towards marriage with the liberal values of self-elected ends and autonomy. To find an idea of the good that is both impartial and palatable to all parties within a society is a monumental philosophical challenge and will continue to result in a protracted debate of the sort seen in Rawls' *Political Liberalism* . I have suggested that in the case of marriage, legislation in favour of and against certain forms of marriage are better justified on an empirical basis concerned with how types of marriage are practiced rather than theoretically conceived. Furthermore, a greater discussion is needed of whether women within polygamous *Nikkah* marriages have entered into those marriages free from coercion. This will alleviate some of the difficulty a liberal society will have in involving itself in those areas of human life which are, as Rawls puts it, 'fully voluntary'.

References

[1] Bennett, C. (2003) 'Liberalism, Autonomy and Conjugal Love", in *Res Publica* 9/3: 285–301.

[2] Berlin, I. (1958) *Two Concepts of Liberty*. Oxford: Oxford University Press.

[3] Berger, J. (1972) *Ways of Seeing*. London: BBC and Penguin.

[4] Brooks, T. (2009) 'The Problem with Polygamy', in *Philosophical Topics* 37/2: 109–122.

[5] Fairbairn, K. (2014) 'Polygamy' [Online] Available at `http://www.parliament.uk/briefing-papers/sn05051.pdf`

[6] Gher, J.M. (2008) 'Polygamy and Same-Sex Marriage – Allies or Adversaries within the Same-Sex Movement?', in *William and Mary Journal of Women and the Law* 14: 559–603.

[7] Kennedy, D. (2007) '1,000 men living legally with multiple wives despite fears over exploitation', in *The Times* [Online] Available at http://www.thetimes.co.uk/tto/life/families/article1757524.ece

[8] Khan, M. (2014) *The Men with Many Wives* [Documentary], *Channel 4* Available at http://www.channel4.com/programmes/the-men-with-many-wives

[9] Al-Krenawi, A. & Lightman, E.S. (2000) 'Learning Achievement, Social Adjustment and Family Conflict among Bedouin-Arab Children from Polygamous and Monogamous Families', in *in Journal of Social Psychology* 140/3: 345–356.

[10] Al-Krenawi, A. & Slonim-Nevo, V. (2008) 'The Psychosocial Profile of Bedouin Arab Women Living in Polygamous and Monogamous Marriages', in *Families in Society* 89/1: 139–149.

[11] Nussbaum, M. (2008) *Sex and Social Justice*. Oxford: Oxford University Press.

[12] —— (2008) *Liberty of Conscience: In Defense of America's Tradition of Religious Equality*. New York: Basic Books.

[13] Rawls, J. (1999) *The Law of Peoples*. Cambridge MA.: Harvard University Press.

[14] —— (2005) *Political Liberalism*. New York: Columbia University Press.

[15] Serck, L. (2012) 'Polygamy in Islam: The Women Victims of Multiple Marriages', *BBC News Berkshire*. Available at http://www.bbc.co.uk/news/uk-england-berkshire-18252958

[16] Williams, R. (1997) *Hegel's Ethics of Recognition*. Berkeley CA: University of California Press.

To mark the 10th Anniversary of the
British Journal of Undergraduate Philosophy,
the Editors have selected a set of their favourite
interviews from the last decade.

Interview: Jennifer Saul

University of Sheffield

What is philosophy? What do you think of as its role?

I'm terrible at answering this kind of question, because what I love most about philosophy is the way that it helps us to see puzzles and problems where we might not have noticed them before. That's what I'm personally always drawn to. One effect of that is that I'm never happy with any definition – and this is at least as true for definitions of 'philosophy' as anything else.

What was your first exposure to philosophy? How did things progress from then?

I was very lucky to have a teacher at my secondary school who had nearly finished a PhD in philosophy many years before, and who taught a philosophy class every year. What drew me to him and his class was actually his sense of humour. I remember being extremely angry at first when we were reading the *Republic*. Initially I couldn't get past my horror at the censorship. And then there was a moment when I remember everything suddenly shifted and I saw the point of really taking seriously ideas that I had previously dismissed completely. It was tremendously exciting, and from then on I was hooked. (My parents tell me, though, that I obsessed over tiny nuances of linguistic formulation in Sunday comics. For some reason they take this to have been a clear indicator of a future in philosophy.)

What's your advice to students interested in becoming philosophers, or struggling with philosophy at the moment?

With a lot of subjects it's very worrying to be confused – things are supposed to be clear and if they're not clear it's problem. Philosophy isn't like this – if you're not confused at least a lot of the time, I'd say you're doing it's wrong. Philosophy simply is hard and puzzling, and if you do it well you will often

feel puzzled. Once you embrace this, it can be wonderful. But before that it can be very disconcerting.

Your recent book *Lying, Misleading, and the Role of What is Said* links ethics with the philosophy of language. Could you give us an overview of what it is about?

In recent years there have been very intense debates in philosophy of language over the notion of what is said (and other related notions). My initial thought was to see if I could shed some light on these debates by considering an area in which ordinary people really seem to care about this notion – the distinction between lying and merely misleading. Famously, Bill Clinton carefully said "there is no sexual relationship" when asked about Monica Lewinsky, hoping to succeed in misleading his audience without lying. The reason this utterance wasn't a lie was because it didn't say anything false. Ordinary speakers readily recognize the notion of saying involved in this kind of judgment, and they assign it significance (though whether they think it gets Clinton off the hook *morally* varies). My thought was to see what notion of saying is involved in these ordinary judgments that we make so easily. As it turned out, none of the current notions in the literature can do the work needed for the lying/misleading distinction, so I ended up having to sketch one of my own.

Along the way, I became really interested in the moral status of the distinction: whether it actually is, as it seems to many of us, better to avoid lying by merely misleading. To my surprise, I became convinced that (except in very special circumstances) it isn't, leading to the conclusion that if you're going to deceive anyway, you may as well just go ahead and lie.

In the final chapter, I apply all this to interesting cases, including a (very) detailed look at key Clinton utterances and the wonderful Jesuit Doctrine of Mental Reservation. (This doctrine allows a priest to avoid lying when asked if he's a priest by saying 'no, I am not a priest' while silently adding mentally '*of Apollo*'.) This turned out to be the most enjoyable part of the book to write.

Do you feel that philosophy has historically missed a trick in becoming more specialised and less synoptic?

I think specialization is inevitable and not in itself bad. But I also think a lot of the most exciting philosophy comes from encounters between these specialisations. Really wonderful things happen when someone like Sally Haslanger, originally trained as an analytic metaphysician, turns her attention to issues of gender and race. Each of these specialized areas (metaphysics, gender, race) has tremendous riches to draw on that they wouldn't have if there weren't traditions of specialization. But bringing these areas together yields something new,

and often even more illuminating.

Speaking from my own experience: I trained as a very traditional philosopher of language, doing propositional attitude semantics. But spending time in a deeply pluralist department at Sheffield (talking to e.g. Peirce and Hegel scholars) has made me a much better philosopher. It's not just that it makes me think about different areas of philosophy – it's that it makes me think in new ways even about philosophy of language. I think my work is far more interesting than it would have been if I'd just been talking to specialists. (Though I also think I need my specialist training to do the work that I'm doing.)

Feminism is mostly thought about as a social/political movement, and feminist thinking as sociological or political. What do you think feminism and analytic philosophy have to offer one another?

I never really understand why people think there's any kind of a tension here. Key ideas of feminism are actually philosophical ideas – questioning the public/private distinction; interrogating the relationship between freedom and social conditioning; examining the role of biases in our language and our thought, and so on. And philosophy has already benefited enormously from feminist thought. A lot of now quite mainstream ideas in philosophy – like those underpinning social epistemology – have some of their most important roots in feminism.

As well as being involved with the *Feminist Philosophers* blog, you cowrote a report about women in philosophy in the UK. Could you summarise your findings?

What we found is that women are approximately 24% of professional philosophers in the UK. This is the sort of percentage we normally see in the sciences and engineering, not in the humanities, which tend to be near parity. This is also much like what we see in every country that's been studied: women, in every such country, are 17–30% of full-time working philosophers. Some studies show that the percentage of women in philosophy is even worse than in maths.

We also found another pattern common to every country studied: women and men begin in approximately equal numbers, and then the percentage of women drops at each point at which it's easy for women to leave the field.

As students of any gender, what can we do to improve the situation for women in philosophy?

I think a lot of the most important things to do for women in philosophy are also quite simply good things to do for philosophy. The top one of these is sim-

ply to engage in philosophical discussion as a way of advancing understanding, rather than attempting to *win*. The best way to do that is to make sure you hear from as many voices as possible, and that that you really *listen* to them – that you work together to refine and improve ideas, even those you don't agree with. Critical discussion is important, but it needn't be about destroying. It can be about finding weaknesses and then building up and improving. The gladiatorial style of philosophy is actually terrible for the subject.

A part of having these good, constructive discussions is to notice if someone's trying to speak, or if their point is being attributed to someone else. You can say 'hey, Jane looks like she wants to say something'; or 'isn't that what Betty was arguing?'.

And of course, you should always speak up if you see people being dismissive of women, or of others who are not well-represented in philosophy. I hear from lots of students who have been quite put off the subject by casually sexist or racist comments, sometimes from their fellow students. The others in the room are undoubtedly appalled, but it's very hard to take that extra step of saying something. And yet it can make a huge difference.

And again, all this simply makes the discussion a better one – better for everyone, and better for philosophy.

What would you like to see change in philosophy in the next five, ten, or twenty years?

I'd like philosophers to be a much more diverse bunch. Frankly, I'm sick of talking to room after room of exclusively white people. We're losing out on a lot of talent, and we're missing out on a lot of important perspectives. I recently saw a quote from Angela Davis that shocked me in its simple undeniability. She was pointing out that if we want to understand freedom we really need to be hearing from philosophers who come from the social groups that have most been *denied* freedom. And the same point will hold true over and over again, for so many areas. The philosophers who are most read and most taken seriously have come almost exclusively from a narrow and especially privileged subset of society. Why on earth should we think the experience of these people best situates them to reflect on all the issues that philosophers reflect on? So my hope would be to see a more diverse profession. And I'd be really excited to see what this more diverse profession comes up with. I think our understanding will advance immensely.

Your profile on the University of Sheffield's website lists your 'proudest accomplishment' as having been a consultant on a zombie movie script. Do you have a zombie survival plan?

Sadly no, that's not my field of expertise. The film's zombie-slayer was to be a woman in philosophy, which is why I was consulted. But I myself would undoubtedly be one of the first to go!

What are you working on at the moment?

Right now I'm working on politically manipulative speech. I'm especially interested in the case of dogwhistles – utterances whose apparent content is (at least to some people) very different from a further message that's communicated surreptitiously. One kind of dogwhistle exploits something akin to ambiguity to address two audiences. A classic case of this from the US abortion opponents declaring their opposition to Dred Scott, a famous pro-slavery Supreme Court decision. To most people, this will seem irrelevant and a little strange (since few defend slavery nowadays), but to their fellow anti-abortion activists it's a conventional signal of agreement. Another kind of dogwhistle works wholly outside of the audience's awareness. A classic case here would be a political advertisement with absolutely no linguistic content about race, but accompanied by images suggesting black criminality. These turn out to be very effective at activating pre-existing racial attitudes without audiences even realizing it. And if they become aware of the racial messages, the ads cease to be effective. Philosophers have so far not studied dogwhistles much at all, but I think there's a lot of rich and important material here.

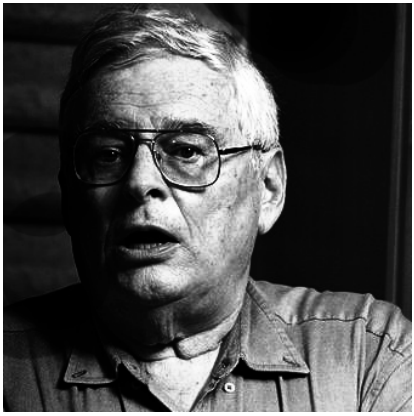
The *BJUP* team is extremely grateful to Professor Saul for taking the time to answer our questions in such detail.

Interview: Jerry Fodor

Rutgers University, New Jersey

What is philosophy and why should one 'philosophise'?

I think of philosophy as a kind of geography and as a service industry. Often one wants to know what else one will be committed to if one endorses a certain proposition. Answers to such questions can be distinctly unobvious and extremely hard to work out, to say nothing of distinctly embarrassing (as when you find that if you claim that *P* you must also claim not *P*, thereby contradicting oneself. There are well known reasons for avoiding self-contradiction if one can). There is no general agreement as to how this process of uncovering connections between propositions should proceed, though many useful models of successful attempts are to be found in mathematics and logic; hence the centrality of both in many philosophical discussions. But there isn't, as far as anyone knows, a general procedure for achieving for figuring out what follows from what. One uses one's experience, one's intuitions, and one's ingenuity, and one thrashes about as best one can. Interesting things sometimes turn up.



As to why you should bother doing philosophy, there isn't any reason that I can think of unless you find it interesting. In my experience, it keeps one up at night and the financial incentives are minimal. But 'what's-connected-to-what' questions do present themselves in all sorts of investigations other than the ones with which philosophers have traditionally been concerned, so it's just as well that someone should be trained in pursuing them. One way of getting such training is to work through some of the cases that have been examined in the history of philosophy. You won't find out what the good is, or what knowledge is, or what an object is, or whether abstract objects are real. But you will find out a lot about the consequences of endorsing one or another proposal, and why so many of the proposals that have been offered from time to time don't work. (In general, philosophy is better at finding out what doesn't work than what does). There is, by the way, nothing

against turning philosophy on itself; 'what is philosophy?' is a paradigm of a philosophical question. A flock of answers have been proposed over the years, to all of which philosophers have found persuasive counterexamples.

What is your first memory of philosophy? Whose works did you enjoy?

I don't at all remember when I first encountered philosophy. Probably it was when I first encountered Lego, which is also about taking things apart and then trying to put them back together. I suppose I've always been more comfortable thinking about the kinds of questions that one might reasonably hope to answer by doing thought-experiments and thinking up examples and counterexamples than about the kinds of questions that one can answer by doing experiments in laboratories. Socrates is the prototype. He was very good at, for example, showing people that what they say that they think virtue is, is in conflict with their own intuitions about which acts are virtuous. (Unsurprisingly, he was considerably less good at figuring out what, in fact, virtue is; this is a pattern that iterates throughout the history of philosophy).

Is there any advice you can give to undergraduates aspiring to be philosophers?

I think the best advice about learning how to do philosophy is to learn a lot of things that aren't philosophy. If you want to do philosophy, you have to know things to do it to. In that sense, philosophy is inherently a second-order discipline. I got into the philosophy of mind in large part by being exposed to empirical work in psychology; and I got into the philosophy of language in large part by being exposed to empirical work in linguistics. In both cases, it seemed that the questions the practitioners worked on were often more interesting than the questions that they said that they worked on when you asked them; their accounts of what they were doing were strikingly bad fits to their practices in doing it. Why, for example, do so many psychologists insist that they are interested in the prediction of a subject's behavior when it's perfectly clear that what they really care about is how the subject's mind works? Why, for that matter, are there so many philosophers who claim to be interested in "analyzing concepts" but are then quite content to leave it to somebody (psychologists maybe) to figure out what concepts are?

Another way to put this: I don't think there is any such thing as 'pure' philosophy, a discipline in which one starts with self-evident ("intuitively evident"), truths and arrives at true conclusions by employing only kinds of inference that are demonstrably sound. From Descartes to Wittgenstein (inclusive) that sort of view of the method of philosophy has been closely tied to the idea that the philosopher's goal is to refute skepticism: i.e. to refute claims like 'there are no such things as tables, chairs, or earthworms'; or claims that we have no

reason to believe that what appear to be other people aren't really just mindless machines, and so forth for a number of other theses which, quite rightly, strike the laity as either frivolous or plainly false (or both). I think, rather, that the (dis)confirmation of philosophical theories, much like the (dis)confirmation of empirical theories, is inevitably holistic, which is to say that it's inevitably sensitive to constraints like simplicity, face plausibility, coherence with other theories that are themselves prima facie plausible, etc. In particular, I don't think that there is, or even could be, a purely 'analytical' philosophy, the truths of which are all, in some sense, conceptual or semantically necessary. The relevant methodololgical maxim is not 'only say things that are conceptually or semantically necessary', but 'don't say things that are patently false'. Samuel Johnson was right: If someone tells you that rocks are just 'bundles of ideas', the right thing to do is suggest that he try kicking one. "Thus I refute Bishop Berkeley", Johnson is reputed to have said. A 'proof' that rocks are ideas would be a reductio of at least one of its premises (though which of its premises it refutes might easily turn out to be a question of genuine philosophical interest.)

What are your views on Connectionism and what place do they have (if any) in your theories?

I think the primary interest of Connectionism is that it has shown how hard it is for psychologists, neuroscientists and many in the AI community, to get it through their heads that Associationism isn't true. Associationism is the doctrine that learning complex consists of forming associative connections between its constituent concepts, and that thinking consists of the activation of such associations. So, to learn the concept BROWN SQUARE is to form an association between the concept BROWN and the concept SQUARE, and to learn the concept SQUARE is to form an 'associative connection between concepts *A* and *B* is one where there is a causal contingency between thinking one and thinking the other, the strength of which varies with the frequency that instances of one concept have regularly been followed by instances of the other in a in a certain creature's mind (it's in dispute whether there are other variables that also reliably affect the strength of associative connections; 'similarity' is often suggested, but for present purposes we can leave that moot). So, for example, for you to believe that cats have whiskers is for your mind to be so organized that thinking the concept CAT to reliably cause instances of thinking the concept WHISKERS. (Variants of this story available to accommodate psychologists of behavioristic bent, who don't talk of mental states, still less of causal relations among them. But here too we can skip the details). Given some version of association, the project of Connectionism is to simulate associationistic minds on computers.

I suppose this may sound like a parody, but it isn't. There really are people, not only in psychology, but also in artificial intelligence, in brain science, and even in linguistics, who really do hold that minds are networks of associatively connected concepts, and that thinking consists of the activation such paths. After all, the brain, suitably enlarged, does look rather like it consists of many, many things (neurons) that are directly or indirectly, connected , one to one another by axons, dendrites and the like. If that's what brains are, then perhaps what happens in learning is that the activations of certain neuron comes to increase the probability that some of the neurons to which it it is connected will become active too.

Earlier above, I deplored the regularity with which philosophers of mind have been content to proceed in ignorance of empirical findings about how minds work ('I am a philosopher, therefore do not bother me with data'). Connectionists are equally pig-headed only the other way around: 'I am a scientist, therefore do not bother me with arguments'. At least since the 18th century, the philosophical literature has offered what I take to be conclusive reasons for thinging that Associationism can't be true which, as far as I know, have never been seriously impugned. Thus, Kant argued that acquiring the belief that *A*s are *B*s can't be a matter of of instances of *A* coming to cause instances of *B* because, Kant pointed out, because it is one thing to think lead is heavy but it is something quite else to think lead and then have that cause you to think heavy. The differerence is that, in the former case, heaviness is predicated of lead, but in the second case it isn't. Accordingly, thoughts have truth values but assocations don't. (Frege repeated this line of argument a century later, but psychology didn't pay any attention to him either).

The moral, I suppose, is that academic disciplines shouldn't be mistaken for 'natural kinds'. The fact that theory *A* is taught in the Department of Psychology but theory *B* is taught in the Department of Philosophy (or vice versa) is not a good reason for thinking that they the truth of one is irrelevant to the truth of the other. Divisions between departments are there for the convenience of deans, not for the furtherance of research. The 'cognitive science' movement was, among other things, a self-conscious attempt to teach philosophers what they need to know about psychology and psychologists what they need to know about philosophy. It hasn't worked very well so far.

What was your motivation for writing LOT2? How have your views on LOT1 changed since you wrote LOT2?

This is a long story, and it's still unfolding. The latest version of the story is set forth in a new book with Zenon Pylyshyn (forthcoming from MIT Press in 2014) called *Minds Without Meanings*. The beginning was set forth in my *Lan-*

guage of Thought (hereinafter 'LOT1' (1975)) a soft-cover is available version from Harvard University Press of is available for nostalgia freaks.

The basic idea was to revive the sort of 'Representational' theory of mind familiar from the British Empiricists (Hume was perhaps the prototype); but LOT1 wanted to make a couple of fundamental revisions. First, the Empiricists had thought that mental objects (beliefs, for example) are relations between minds and 'mental representations' (which Empiricists often called 'Ideas'). LOT1 wanted to hold onto that: In the simplest kind of cases, to think about dogs is to entertain a mental object that refers to dogs (or, if you prefer, one that 'expresses the concept' DOG). However, though Empiricists had generally thought of Ideas as something like pictures, Berkey offered very convincing reasons for saying that they can't be; the question 'What's the alternative?' remained moot.

According to LOT1, by contrast, mental representations comprise a system of structured objects not very different from the sentences of a 'natural' language (e.g. English). To think that's a dog is to have in mind an object not unlike the English sentence "that's a dog" (there are important reasons for denying that mental representations could actually be sentences of natural languages, but for present purposes, we can skip them.) So far, then, the LOT1 story was: We think in things that are more like sentences than they are like pictures. Concepts are the constituents of thoughts, in much the way that words are the constituents of sentences. The constituents of the thought dogs bark are the concepts DOG and BARK so arranged that the second is predicated of the first. OK so far.

The second major revision of Empiricism that LOT1 proposed was to abandon the Empiricists' thesis that mental processes are associative. The suggestion was that, instead of thinking of learning as the formation of associations and thinking as their activation, LOT1 said we should think of them as computational prcesses defined over the structured mental representations that express thoughts. This was, in effect, to adopt the suggestion of Turing's that eventually gave rise to computers. Associationism having proved a dead end, Turing provided an alternative.

The significance of Turing's account of how the mind works (viz. by the manipulation of syntactically structured mental representations) really can't be over-emphasized. For just one example, we've seen that if mental processes are supposed to be associations, there seems to be no way to distinguish two quite different processes that a mind might undergo: on one hand, forming the belief that lead is heavy; on the other, forming a causal chain in which instances of the concept LEAD cause instances of the concept HEAVY. But if mental

processes are computations – operations defined over linguistically structured objects, we can think of the belief that lead is heavy as predicating heaviness of lead, much as the sentence 'lead is heavy' does. In effect, Turing's notion of mental processes as computations rather than association solved the problem that Kant and Frege had raised of distinguishing thoughts from mere causal sequences of concepts.

But if thoughts are structured arrays of concepts, where do the ideas come from? The Empiricists had an answer: All Ideas are associations of basic ideas, and basic ideas are traces of sensations. So the Empiricists' psychology was thought to support an Empiricist epistemology in which all one's concepts are definable in terms of sensations, and all one's thoughts and theories are ultimately thoughts and theories about one's sensations. Positivism followed.

One could easily imagine a computational psychology taking much the same path; and, if fact, almost all computational psychologists actually did, thereby raising a problem that still persists: Supposing that all concepts are definable in a vocabulary of of basic (e.g. sensory) concepts; how is it possible to learn any new concepts? On the one hand, sensory concepts aren't learned, they are just triggered by encounters with instances of the properties they express. Given the character of one's sensory mechanisms, 'learning' the concept RED is just an automatic consequence of encountering red things; and, of course, the structure of the sensory mechanisms are innate (for this sort of reason, many of the Empiricists took the innateness of basic concepts to be self-evident). And since, by assumption, all the nonbasic concepts are definable in terms of the basic ones, it seems that there can't be such thing as a kind of learning that increases the expressive power of the concepts one is born with. You can't 'learn a new' concept because, by assumption, every concept you can learn is equivalent, by definition, to some array of concepts that you already have. This conclusion was all the more embarrassing because several decades of inquiry, both in philosophy and in cognitive psychology, showed that Wittgenstein was right: practically no words have definitions, sensory or otherwise.

LOT1 was inclined to say 'well, so be it all concepts are innate'; in effect, that's what LOT1 did say. But that conclusion struck many philosophers and psychologists as preposterous on the face of it; from time to time it struck me that way too. However, it's one thing to disapprove of an argument's conclusion; it's quite another to show that the argument is unsound. What to do?

LOT2 reviewed this state of affairs and tried to put a bandage on the paradox by recruiting the notion of an innate 'attractor space' that had been developed in the Connectionist literature. In effect it offered a kind of concept Nativism that was at least slightly embarassing than the thesis that all concepts are innate.

But I'm now convinced that it was unsuccessful. I now think I know the way out; it's one of the things presented in the forthcoming book with Pylyshyn. But I don't think you are going to like it.

For the last hundred years or so , the tradition in theories of conceptual content has taken for granted that concepts have both 'senses' and 'extensions', with the former determining the latter. The picture is very intuitive: The concept DOG has a senses such that it applies (has in its extension) all and only dogs. If the concept DOG is definable, them maybe its sense is its definition; maybe 'domestic canine' would do? Likewise for every other concept that isn't sensory. Acceptance of some or other version of this this 'sense and extension' account of conceptual content (and hence of word meaning) has been practically universal in psychology, philosophy and linguistics. Indeed, there are lots of people philosophy who think that Frege had provided an argument that shows that this account of content must be right (see any standard introductory text in semantics.)

However, the 'sense and extension' account of conceptual content leads, by the kinds of considerations mentioned above, to the discomforting conclusion that all concepts are innate. *Minds Without Meanings* hopes to dispense with this paradox. Very roughly, it says that there aren't any such things as senses (that would explain why nobody has been able to explain what senses are; in particular, it's why definitions and the like are so hard to come by.) There are extensions, of course – it's quite true that DOG applies to dogs and to nothing else – but the extensions of concepts are determined not by senses but by causal relations between minds and the world. The claim is that this is sufficient apparatus for a theory of conceptual content that is adequate both to the purposes of a Naturalistic psychology and, as a by-product, to avoiding radical Nativism.

The suggestion is that 'meaning' is a useless concept; that we can do quite nicely without it, and that it has been the root cause of many, many failures in theories of mind and theories of language. It may well strike you as itself unbearably paradoxical; it may strike you that there sin't much to choose between 'all ideas are innate' and 'there is no such thing as meaning'. But do read the book.

The *BJUP* team is extremely grateful to Professor Fodor for taking the time to answer our questions in such detail.

Interview: Huw Price

University of Cambridge

How did you get into philosophy?

Well, as is true of quite a few philosophers, it was a little bit roundabout. As an undergraduate in the early 70s at ANU in Australia, I started thinking I wanted to be a physicist, in particular an astronomer. So I started off doing physics and pure and applied maths. For some reason I did philosophy one as an undergraduate, but then went on in my second year to do physics and maths and it was only at the end of my third year that I decided I wanted to give up the physics and concentrate on the maths that I had to find some way of filling up my degree in place of the physics units I would otherwise have done, and I found I could do a couple more philosophy units and make up a major that way. And it was at that stage, when I was a third year undergraduate, that I started getting interested in the subject but it still wasn't for another two or three years after that that I finally made the switch, I even started graduate work in maths at Oxford that would normally lead to a PhD or a DPhil but I left after the first year of doing an MSc and moved to Cambridge to do philosophy at that stage.

What area did you focus on when you moved to Cambridge?

A couple of things; the person I moved to Cambridge to work with was Hugh Mellor who was someone I met as an undergraduate in Australia when he was visiting ANU as a young lecturer from Cambridge. And so it seemed natural when I got to Cambridge to start thinking about the kinds of things he was interested in. So I started thinking about probability, but also about time, at that stage Hugh was writing his book *Real Time*. There's a sense in which almost everything I've done since then grows out of those two interests. I went on to write my thesis on probability defending a kind expressivist account of single case probability judgment, an account very different from Hughes. All the work I now characterised as expressivism or

pragmatism grew out of this kind of generalising and thinking about the more general issues which came up, arising out of that approach to a particular case, the case of probability judgment. So that's where my interests in pragmatism came from. And then on the other side I retained this interest in time, although I was always interested in doing it in relation to Physics, and so my interests in philosophy of time and philosophy of physics grew out from that side. I also do work on causation which is somewhere in the middle. Maybe that didn't develop until a bit later, probably that developed during the 80's when I was back in Sydney and talking to Peter Menzies a lot, he was a big influence at that point.

What do you think the role of philosophy is nowadays?

It's a question I find hard to answer in any general way. For myself I think some of the most interesting philosophical problems are problems that arise at the intersection between philosophy and other disciplines. Where, in a sense, philosophy is like a certain kind of method or a certain kind of focus and interests in the nature of the concepts which are specific to those disciplines. Something of that kind. So I suppose my interests in philosophy of physics could all be characterised in those ways. They arise from places in contemporary physics where the techniques and particular aptitudes that some philosophers have can, I think, be useful in understanding what the physics is telling us about the world. But I suppose that not everything I do would fall into that category. In the area of pragmatism there's a sense in which it's more internal. The problem space is more one that is created by philosophers of the past. So much of it is trying to suggest that some traditional ways in which philosophers have approached certain issues – what is causation for example – are in a sense misdirected because it's more fruitful to approach via certain questions; Why do creatures like us have a need for a concept of causation? What role does it play in our lives? So there it's more internal, more generated by a reaction against a certain way of thinking about causation within philosophy. Then again that's not a general characterisation, but it's a characterisation of two particular kinds of philosophy that I find interesting.

Do you think philosophy is quite apt to interface with some of the sciences? Would it take on the role of an 'overseer' in checking their work in a 'meta-science' kind of way?

I don't think I see philosophy as an overseer. I see it as something between an assistant and a kind of next destination. If you're coming from some particular discipline you're interested in the foundational issues within that discipline. It's very natural to appeal to some systems in philosophy or move in the direction of philosophy itself. So there's a sense in which it's on the same level, but it's

sort of out to one side or I suppose you could think of it as further back. From when I was an undergraduate I always very much liked that idea that you find in Quine – there's not a sort of 'first philosophy' which stands above this.

What do you think the role of metaphysics is in philosophy and in interfacing with other disciplines?

I think quite a lot of things that people do under the heading of 'Metaphysics', fall into the categories of the kinds of philosophy that I think would be better done in other ways. People think about questions like; what is causation? What are laws? What is the nature of something? – issues. At least a substantial class of those issues are unproductive and there are better ways of approaching the puzzles from which they arise. Often they arise because there seems to be some reason for thinking that the notion in question is difficult to make sense of and difficult to reconcile with the scientific world view. This is most obviously the case in the case of ethical judgment – what is the place of morality in the scientific world? Those are the kinds of cases where I think that focussing on questions like – Why do natural creatures like us have a need for concepts of this kind? Or a use for concepts of this kind? These questions are more productive than thinking about the 'metaphysical' issues. In a sense trying to deflate the metaphysical view and replace it with something else. Now it's probably fair to say that not everything that goes on under the heading of 'metaphysics' is of that kind. The thinking I do wearing my other hat – thinking about the nature of the quantum world, some people would certainly think of that as metaphysics. I don't know, but I think that's rather terminological. I'd rather regard it as something continuous with physics. They might come back and say that in calling it metaphysics they weren't denying that in any case.

I certainly think that the moves I take away from metaphysics aren't really in any sense trying to disagree with them and that's what Strawson called descriptive metaphysics – the mapping out of the space of concepts and understanding how things hang together. I don't think it's something I do, myself, very much. But it's not something I regard as in any sense a sort of disreputable or misguided activity. On the contrary I think it's probably very helpful as a precursor to the anthropological issues. We need to understand the relationship between various systems of concepts in order to help us think about the question of the genealogy of where and why those concepts came into use.

Would you disagree with the role of meta-ethics in trying to find a metaphysical underpinning to notions like 'the good'?

I do want to disagree with that program, but I don't see that as disagreeing with meta-ethics. It's just doing meta-ethics in a different way. Sort of starting the inquest again and thinking about the role of moral judgment and moral

concepts. That's still doing meta-ethics in that it's in some sense stepping back from ethics. Not doing first-order ethics. It certainly involves a certain amount of philosophical distance, but it's not using that distance to think about meta-physical questions, it's using that distance to think about Anthropological questions.

What advice would you give to undergraduates? Both those that aspire to be philosophers and those who are just trying to get through their undergraduate degrees?

To people in the second category my advice would be to try as hard as you can to find the bits of the subject that interest you. It's a very diverse subject and of course depending on where you are and where you study it's not all going to be available to you. In my own experience when I first came across it there were just certain little bits that I found really engaging and interesting. And that certainly wasn't true, by any means, for all the bits of philosophy I encountered. This is probably good advice for many fields but I think it's particularly true of philosophy. To people who are thinking about pursuing further study – possibly a career in philosophy – I think again, thinking about the bits that interest you is very important, but also thinking about people and thinking about where you might find interesting people, interesting teachers, at the next stage, when moving to graduate study. Most people who go onto graduate work do some kind of research masters, so I'd encourage them to look around at the various research masters programs that are available in the UK and US and other places. Keep in mind that philosophy is a big field. Even within what the English speaking world is considered strict philosophy there are lots of people doing things in very different ways. Doing a little bit of research can make a huge difference to your graduate experience.

You most recently published a book called *Expressivism, Pragmatism and Representationalism*; would you mind giving us an overview of what it is about and the area?

Maybe a good starting point for that in the sense that it is something that a significant number of undergraduates would have encountered, especially if they've done some meta-ethics, they would have encountered the distinction between (and there are different names for this distinction but the two common names are) cognitivism and non-cognitivism. The non-cognitivists, the early ones, were people who were then called emotivists – people like A.J. Ayer – these were people who thought that when we make moral claims we're not making genuine statements about the world we are doing something else, we are expressing our state of approval or disapproval or something like that. A more recent term for that kind of position about ethical claims, and other

kinds of normative claims, is expressivism. That's associated with the idea that we're expressing some kind of state that, at least on the surface, isn't to be thought of as a belief about the world, it's be thought of in some other way. If you have a view of that kind then you have a picture where there's a kind of boundary between the claims which should be treated that way, like the moral and perhaps the aesthetic ones, and claims that should be treated as regular, factual claims. It's a very interesting question what goes into each camp, as it were, each category. People who do meta-ethics courses may not come across the fact that there are all sorts of other topics where philosophers have found the non-cognitivist approach attractive as well. Some of them have nothing to do with ethics. People who hold that sort of view about conditionals, about laws, about probability judgments, about causal judgments – a very wide range of things. And indeed about truth itself, this is what's now a very popular view about truth that when we say *P* is true we're not, as it were, talking about some property of truth. The role of the predicate 'is true' is not to be seen as ascribing some real, substantial property but in some other way. So then it has at least a family resemblance to the views about ethics that we were talking about a moment ago.

So there's that bunch of issues and in most views of this kind you'll find what Rorty calls a 'bifurcation' between the genuinely factual cognitive language on one side and the non-cognitivist stuff on the other side. What I guess you could characterise my view as is as somebody who thinks that when you kind of push on that there's an important sense in which nothing ends up on the cognitive side. So there's really no interesting work or a theory of language which is theorising about language in the way that people took for granted that you would theorise about language on that cognitive side. So notions like really talking about the world or really representing the world or really being factual, all those notions turn out to have not very interesting theoretical content. And one of the reasons for thinking that is closely linked to the view about truth I've mentioned a moment ago. Truth isn't a substantial property and that naturally goes with thinking that this idea that there's this relation between words and the world that turns out to not be an interesting way of theorising about language either. So there's a number of labels for the resulting position. I tend to call it global expressivism. It's also called non-representationalism or global pragmatism.

In some sense this is a view I've been attracted to for a long time and in my first book, *Facts and the Function of Truth*, I was at least moving in that direction. But I think I've found a better way of saying things over the years and so the one way of characterising this view in Expressivism, pragmatism and representationalism is offering a defence of this global pragmatist view. The core

part of it are some lectures I gave in the Netherlands a few years ago. There are three lectures and then the book also has responses to those lectures by four people – Simon Blackburn and Robert Brandom and Paul Horwich and Michael Williams who in various ways are quite sympathetic to some of these moves, although people closer to you are often the harshest critics. And then there's also a long section in the end where I respond to those essays and add some new material. Some of it comparing my views to those of Wilfrid Sellars and that's, at least from my point of view, a very interesting connection in that it's something I wasn't aware of until quite recently.

One other thing I could say, just as a point of reference, is that some people might be familiar with Simon Blackburn's work and his program called Quasi-Realism. Another way of thinking about what I am doing is as a kind of generalised Quasi-Realism.

You are also well known for your views on time – would you mind telling us a bit about time, the arrow of time and how philosophy can interact with physics?

One way into that would be to think about the way in which I oppose or stand opposed to another view point about time. One of the debates that people will be familiar with, if they've done any sort of course in philosophy of time or time within a metaphysics course at all, they'll be familiar with a debate between two sides called A-theorists and B-theorists, and these are terms that go back to McTaggart. Roughly speaking, the A theorists are the people who believe that there is an objective distinction between past, present and future. B-theorists are people think that that distinction is in a sense just anthropocentric. They think that 'now' is just like 'here' we use it to mark where we stand but it is not a part of reality. What's in reality is the distinction between earlier and later. One way of thinking about my position initially is as opposing the A theory and defending the B theory. Defending what is often called the Block universe picture – in which all time is briefly real as against a picture in which the present moment is picked out in some way, either as the only moment that exists or as, you sometimes get it, as a kind leading edge of a growing reality, the picture called the growing block view. We think that the past exists and the past is continually expanding as the 'now' moves through time. So there are various versions of these A theory views and I oppose all of them I think they are all more or less incoherent. In some of my recent work one of the things I've done is to look at what could be meant by the idea that there's a 'flow of time' which is another thing I'm kind of opposed to. And I distinguish three elements of this kind of package that tend to be not properly distinguished. One is the idea that there's a distinguished present moment, which I mentioned before. The second is the idea that there is an objective direction to time, an

objective distinction between past and future. The third is the idea that there is something dynamic, some notion of a passage, which is essential to time. I think all of those notions turn out to be either incoherent or nearly so. Perhaps the one that has the best prospect of being made sense of is that there is a direction of time but even that, when you really look into it, it's very hard to see what could be meant by that claim. The picture you end up with actually, if you accept those arguments, is not McTaggarts' B-series, which would take for granted that there is an objective earlier/later than relation. It's what McTaggart called the C-series. Where you have the line, but the line doesn't have an objective direction on it, the relation is just between us. So you can say Kennedy's assassination was between the second world war and the election of Margaret Thatcher but you cannot say which one of those events came first. It was either war, assassination, Thatcher or Thatcher, assassination, war. But in reality neither of those depicts the right way. And this is one of the places I think it connects with Physics, because in physics the idea that there might be no objective direction of time came on the radar very precisely in the late 19th century in the work of Ludwig Boltzmann who's famous for work on thermodynamics and statistical mechanics. And what Boltzmann proposed was that the past and future are much like up and down, so if you live in the region of the universe where there is a thermo-dynamic gradient where the universe has been in low entropy and is approaching equilibrium, then in that region it would be natural to distinguish between the two directions. So creatures who live in that region would think of the direction in which they remember things as being the lowered direction and they would have expectations in the other direction. But Boltzmann suggests that's just like creatures on one side of the earth as seeing up in one direction and creatures on the other side of the earth seeing up in the other direction. But from the universe's point of view there's no real distinction between up and down.

That's one of the places where those rather metaphysical issues about time actually connect to the physics. Those are live debates because in physics, although it's probably true that most physicists believe in something like the block universe view, there are still examples around of very interesting and clever physicists who think there's something missing from that picture. There's a recent book by the physicist Lee Smolin, a very talented physicist but also a great writer, and he has this book defending the idea that there's something missing from the block universe picture and progress in physics depends on putting it back in.

That's my work on time, I guess, as it stands at the moment. The other thing I'm interested in in philosophy of physics is the foundations of quantum theory and the question of what quantum theory is actually telling us about the

world. There is this interesting possibility that hasn't been given the attention it deserves, which is that the answer to some of the quantum puzzles is that there's an element of retro-causality in the world, and I think that this is the most interesting sample of an area within contemporary physics where there is a deep roll for philosophical thinking, because I think there is an important possibility here, which may turn out to be the right possibility, the way the world actually is, which physicists are very, very ill equipped to think about, because it involves subtle thinking about bi-causation and questions about free will and agency, and questions about some of those questions that we were just talking about, about time. So then even the most insightful physicists who are interested in these issues about the meaning quantum theory, great, great thinkers like John Bell, haven't had the breadth of conceptual resources to tackle these issues in the way that they need to be tackled, in my view. So there's a real role here for bringing some clarity into this area.

One of my projects is working with a young physicist from San Jose, in California, a man called Ken Wharton who I started working with in the last three or four years and we now have a couple of papers. We have a plan to write a book trying to present these ideas in a semi-popular form, to make them more accessible to ordinary people as well as physicists and philosophers.

Do you think that the way philosophy interacts with physics now is how philosophy will begin to interact with other sciences and disciplines in the future?

I hope so, and I think that in the case of physics there has been a lot of progress over the last couple of decades. I know people in physics who were around before I was who remember what people were like in the field in the 60's and early 70's and they say that back then there was almost no interaction between physicists and philosophers. And they comment on the way that from how the 80's onwards there was a great growth of interaction, people going to common conferences and so on. I know in some ways it's been quite easy for me, I've had a lot of opportunities to go to conferences and talk to physicists. Cambridge has a very good tradition in this as do several other places in the UK and Europe and elsewhere. Now it's a very rich scene. Even if it's also true, and I think it's fair to say it is also true, that within physics and philosophy you'll find people who are a little bit negative about this activity. Physicists sometimes seem a little bit uncomfortable because their colleagues use terms like philosophy or metaphysics in a pejorative way. In philosophy you find instances of people being dismissive of some of the work in the philosophy of physics on the grounds that it isn't really philosophy, it's just physicists playing with numbers and stuff. There is a wonderful community at the intersection now, but to a slight degree, and maybe this is inevitable, I think it made itself

a community by distancing itself a little bit from the two communities.

I know a lot less about the intersection between philosophy and other disciplines but it's my impression that there are fascinating things going on there too. I know people who are interested in the intersection between philosophy and psychology for example, and I think that's very interesting. Some of my colleagues in Cambridge have contacts with linguists and with economics. On the whole I'm very optimistic about the continuing potential for philosophy to play a role in that sort of way. Because it's just the character of theoretical work that it tends to throw up these kinds of philosophical issues at the foundational level. So if you start thinking about the fundamental concepts of the discipline in effect you're going to be doing philosophy. That's good news for philosophers, good news especially for undergraduate philosophers, because it means that there's a sort of never ending supply of interesting problems to work on.

That reminds me of another piece of advice. That would be to take whatever opportunities you have to learn something significant about some other field as well. For example some branch in science. Another field which is clearly thriving that I haven't mentioned yet is philosophy of biology. I have many colleagues here, and used to have many colleagues in Australia when I was there, who work in those fields. What was characteristic of people like that and in other cases we talked about is that they often know a lot about the other discipline. So I think good advice for an undergraduate is to take opportunities as they come along to you, it depends what kind of programme you're in. In the UK people are often in single honours programs which makes it much more difficult, but if you do have opportunities to study something else as well then I'd recommend taking them, if your philosophical interests are in any sense inclined in that direction.

The *BJUP* team is extremely grateful to Professor Price for taking the time to answer our questions in such detail.

Interview: Mary Margaret McCabe

King's College London

What led you to choose Philosophy as a career?

Perhaps it was just that I couldn't stop — I started out as a classicist. When I was at school I studied Greek with the daughter of the wonderful Aristotle scholar and philosopher David Ross. We read two things together – Aeschylus' Agamemnon and Plato's Symposium. She showed me something about reading Plato that I never forgot: that he writes to be read with huge care, thought and reflection, and not just to tell us what to believe. When I was an undergraduate at Cambridge I had a supervisor, Denis O'Brien, who had the same view – that reading Plato just is doing philosophy, thinking about how thinking works; and then thinking about it all over again. He taught me how to see a page of Plato as though I had never seen it before, and to think about what might be going on there – so he showed me how these ancient texts could be absolutely fresh. Then I discovered that what completely captivated me about those texts were the arguments, the puzzles, the questions they were asking. The same, I found as I started teaching, was true of all sorts of parts of ancient philosophy – and I discovered at the same time just how much I loved the business of talking to others about the same puzzles and questions and texts – so perhaps it was teaching that really got me going: anything that I get right, I owe to those kinds of conversations with students and colleagues. The questions and the puzzles preoccupied me in ways that scholarship (I am afraid) never quite did… so I moved away from a Classics department in 1990 to KCL Philosophy Department – and that was, at that time and still, the most wonderful place to do philosophy, without the kind of 'philosophy as a competition' approach which I find so arid.

What do you think Philosophy is? What is Philosophy's role in the world?

So I think philosophy is just thinking about thinking, whatever form that seems to take, and however it is formalised. It is critical, highly demanding, precise, argumentative in the best way – and at its very best open-minded. This goes along, I think, with developing a kind of imagination that helps us think about how we are in the world, how we think about ourselves and others. All of this, I think, makes philosophy something that we do, not suffer; it makes it an activity, rather than a set of fixed views or beliefs – so I am more interested in philosophizing than in philosophies. If philosophy has a 'role in the world', it is to use that kind of critical care to make sure we think as best we can, and to make sure that our view of the world is rich and thoughtful. But I suppose

that brings responsibilities with it – failing to think clearly, especially in the public domain, is perhaps a kind of ethical lapse, rather like walking by the injured person on the other side of the street. I hope that doesn't sound just too horribly high-minded – but I am not convinced that there are firm lines of demarcation between the logical and the moral that makes it OK to see that an argument put forward in the public domain is horribly unsound – and not say so. I don't think, incidentally, that this means being aggressive or raging against others: indeed, I think that rage and aggression themselves damage how we think – and that too sounds horribly high-minded, doesn't it… Perhaps the best way to put it is that being civil to each other is an important aspect of how we do the subject.

What can we learn from Ancient Philosophy? What do you think ancient philosophy can teach us about ourselves as people?

One of the peculiar and difficult things about studying ancient philosophy is trying to get behind and away from the dominant assumptions we might bring to the interpretation of philosophy in the modern era. The ancients, at least in the Classical period, are not dominated, for example, by the problem of doubt – their accounts of knowledge are not, I think, driven by the question 'how can p be true?' but by the question 'how can p be false?'; equally not by the question 'how can we know p is true?', but rather by questions about explanation ('why did p come about?' etc.). Or – in ethics the dominant ancient worry is about the account we might give for virtue and goodness, rather than duty; and thinking about these thinkers gives us, I think, some kind of perspective on ourselves that closer thinkers don't allow (this is where the recent interest in 'virtue' ethics and epistemology comes form). At the same time, of course, they are somehow accessible to us all the same – the challenge is always to figure out how we negotiate between the strangeness of ancient philosophy and its familiarity – that negotiation, it seems to me, is something we can just keep right on doing.… And I completely love reading a text with others, sitting and working out what on earth is going on, together. Perhaps that is made more tractable when one is dealing with a text in a very difficult and distant language?

Who do you prefer: Plato or Aristotle? Why?

Plato – it seems to me that the utter brilliance of how he writes, and how reading what he writes engages us in philosophy, has no equal. But in my old age I am also coming to appreciate Aristotle much more than I did – as a student I found him just hugely daunting; and I found the received opinion at that time, that Aristotle writes what he does as a set of rather crabbed lecture notes, rather dry. I have since realized – better late than never – that Aristotle is

reading Plato too, in the rich ways that he should be read, and that Aristotle's detail can be as hair-raisingly exciting as Plato's. Jonathan Barnes says that he prefers Aristotle to Plato because Plato's arguments are such rubbish, and Aristotle's so good – I think the relation between them is far richer than this, and that thinking about it as a relation illuminates what otherwise seems to be just a contrast between two completely different ways of tackling some of the same questions.

What challenges have you faced as a woman in Academia? Do you think Academia is becoming more equal?

This is tricky. When I was an undergraduate in the '60s I thought we had all of this sorted out, that there was a very short distance to go before we would have something that felt like equality – or, rather, something that felt as if gender differences were just insignificant when it came to academic matters. I have since – and perhaps especially recently, come to be rather more dispirited about how much progress we have made. There is a fair amount of formal equality; but the figures for the progress of women, especially in some STEM subjects and in Philosophy, are dismal. Jenny Saul has done some excellent work on how this may be partly a matter of implicit bias; and there is a great deal being discussed in philosophy now about how to improve things – so perhaps it is not dismal after all. And perhaps some of the things that bother me may be just individual and not gendered at all. Perhaps my readers should decide. I suffer badly from 'Imposter Syndrome' – from expecting that just today I shall be found out to be a complete imposter and not belong anywhere in philosophy at all. I also find fierce argument rather disturbing; even although I can probably be rather fierce form time to time when I get excited. . . I find the 'philosophy competition' just makes me anxious, makes me think again and again about whether what I have to say is worth saying and then end up saying nothing. (This was at its worst when I as a graduate student; in King's again, I found that a very kindly environment helps this to stop). But as a consequence, I do try very very hard to make sure that seminars I run and lectures I give are not intimidating; that people see that I can be just as wrong and mistaken as anyone, and most of all that thinking about these questions is exciting and fun and often funny, not some way of figuring out who is the cleverest (whoever it is, it won't be me). I read a recent interview with a philosopher in which he was offering a checklist of which philosophers were as 'smart as him' and which might be smarter (there was only one candidate): I find that sort of talk inhibiting and damaging. It should be possible to think that we are all in this together, should it not?

What do you think are the challenges the UK university system faces? Do you think they're surmountable?

I am not quite sure where to begin with this! The current funding system is not only fiscally unsustainable, but also deeply damaging, because it promotes a market ideology which is inappropriate to account for the value of the university system as a public good. The ways that this has worked out recently has, it seems to me, also seriously damaged how teaching in universities is done – even although hard-pressed departments try very hard to temper the bad effects of increased numbers of students. There are two other lurking problems (which are connected, of course, to the market ideology). The first is that we still haven't sorted out how to make access to university genuinely open; the problem of inequality is urgent, and I don't think we have worked out yet how to solve it (one thought may be to pluralise the sector more effectively: Martin McQuillan calls this the 'rainbow sector' and I think that is spot on). The second is that university governance has increasingly been taken away from academics and into the control of administrators; and I think this means that the issue of academic principle are not being voiced or heard.

What advice do you have for students who are currently studying Philosophy?

Enjoy it, and make sure not to turn it into a competition – do philosophy together if you can as a conversation, not a contest. Remember that it is a slow and laborious process, much slower than you might expect – don't be surprised if it takes you a week or more to read and understand an important article, don't think that there is somehow a right speed to do things. When you read, let the ideas settle and coalesce – give yourself time for thinking and rethinking. (That means, among other things, not doing philosophy at the last minute). And don't worry about it, don't worry if you don't get the point of something immediately, just allow yourself time to think things through. Just take it easy!

The *BJUP* team is extremely grateful to Professor McCabe for taking the time to answer our questions in such detail.

Interview: John Worrall

London School of Economics

What was your first exposure to philosophy of science; where did things go from there?

So there I was at my North-of-England Grammar School in 1963/4, deciding at the last minute to reverse the plan to go into the 'Scholarship Stream' to apply for Oxbridge. (Think *The History Boys* – it then meant an additional year in the Sixth Form, and so another year in Leigh and I just couldn't face that!) This produced a last minute rush to decide where to apply for entry the following September. I received 15 minutes of careers advice – all I ever got. The outcome was that I should (a) study something involving maths, since that was my favourite subject; (b) aim for a course that might make me rich (my only input into the careers session); and therefore (c) go to the LSE to study mathematical statistics, with a view to being an actuary. (Actuaries apparently having, at any rate then, the highest average salary of all professionals.) So I applied to the LSE for stats, without knowing the first thing about what actuaries do.

I had one optional course in my first year at LSE, and I chose Alan Musgrave's 'Introduction to Logic'. The timetable for this course had an optional lecture series – Karl Popper's 'Problems of Philosophy'. Since Popper's lectures didn't help you do proofs in first-order predicate logic and were in all respects irrelevant to the logic exam, most of my fellow logic students soon dropped them. But I was hooked. First of all they were pretty good theatre: David Miller – his then Research Assistant – would arrive first and write "NO SMOKING" in big capitals on the board. Students – and lecturers – often smoked during lectures in those days, but Popper believed he was allergic to tobacco smoke. Then all the other academics in the Department (all four of them!) would troop in and sit at the front, and finally the great (but tiny!) man arrived. Whatever history may decide, there was no doubt that Popper, in his own quiet way, really believed himself to be a major philosopher. He had, after all, 'solved the problem of induction' – and

this came across strongly in his lectures.

He was also a shameless name-dropper. There was much talk of correspondence with Einstein and conversations with the great logician Alfred Tarski on various park benches in Vienna. This was heady stuff for an 18-year-old working-class lad. I bought, and devoured, *Conjectures and Refutations* and *The Logic of Scientific Discovery*. And, although I had been perfectly happy mastering chi-squared tests and the like, I enquired at the LSE Registry and, yes: it was possible to change to Philosophy instead of Statistics. And so it was that Popper ensured I would never be rich.

Next, another stroke of luck. There was only one option in the Philosophy provision then. This was only the second year it had been going – before that the department had only done 'service' teaching. You could choose either Moral and Political Philosophy or Mathematical Logic. I was the only one in my year to take Mathematical Logic and it turned out that if you made the choice you were assigned Imre Lakatos as your tutor. He got me all sorts of special permissions to go on studying some maths and stats alongside philosophy. He set me a ferocious list of tasks, including working through Stoll's *Set Theory and Logic* and Courant and Robbins' *What is Mathematics?*, telling me not to see him again until I had worked through the list. I think he believed he would not see me again. When he did (shortly before Christmas), he branded me a 'hopeful monster' and from then on took a very keen interest in my studies and, later, my early career. Studying his *Proofs and Refutations* was *the* intellectual event of my undergraduate study.

From there I went into a career entirely at LSE (aside from various visiting positions). In some ways, I regret the lack of variety, but I could not have been in a better place to research philosophy of science.

What do you see as the main purpose of philosophy of science?

Although I never got very much out of reading Locke, I do think he had absolutely the right characterisation of the role of a philosopher – to act as an 'under-labourer to the scientist'. So philosophers of science should be out just to articulate the intuitive principles by which scientists judge theories in light of evidence, and to clarify some of the positions of scientists, because they have not found the time to think through their positions with full logical rigour – or because they have more important things to do! Of course, scientists (especially from the 'softer' end) don't always get it right – witness the enormous exaggeration of the evidential virtues of randomization within the clinical trials community. And then philosophy of science can step in and try to make corrections. But all good philosophy of science is, as we like to put it in our department, 'continuous with science itself".

What spurred your interest in evidence-based medicine?

Well, I wanted to do work that might impact outside of philosophy of science. But the main spur was undoubtedly a host of conversations over dinner with my wife. She is a consultant physician (though her first degree was in Philosophy at LSE). Of course, it would be hard to be against basing *any* body of claims and judgements on evidence. (As Hume said in the *Enquiry*: "A wise man proportions his belief to the evidence [across the board].") The issue was always in the details: What counts as the strongest evidence for the effectiveness of medical intervention and why? What happens when evidence of different types pulls in opposite directions? Is clinical experience a legitimate source of evidence alongside results from clinical trials? It was always clear to us both that, while it undoubtedly had its heart in the right place, evidence-based medicine exaggerated the epistemological merits of randomised studies and hence tended to downplay the real significance of other types of evidence.

I ought also to mention my former colleague, Peter Urbach, who had been thinking about the methodology of clinical trials some years before I began to. I was initially resistant to his Bayesian analysis (as leaving too wide a role for subjective judgement), but I later became much more sympathetic. Bayesian principles are altogether more in line with 'educated, scientific common sense' than those involved in orthodox 'classical statistics'.

What impact did Imre Lakatos have on your work?

It will be clear from my answer to your first question that Imre's impact on me was enormous. Another strand of my work has been to clarify, and I hope improve, on Lakatos's views on what counts as progressive science (through better analysis on what counts as predictive success in science and why such successes provide stronger evidence for a research programme than successfully accommodating empirical results after the fact). But, aside from my having developed some specific Lakatosian themes in my work, I do feel that in general, it was by sitting at Imre's feet as an apprentice (my first job was as his Research Assistant) that I learned *how* to do serious philosophy of science. He was massively generous with his time and efforts to improve my early fumblings.

How has your work on Structural Realism come to impact philosophy of science?

Good question. I did my usual job of publishing the original paper in the most obscure place. (Many of my better papers were written for conferences and published in *proceedings* – nowadays of course, you're not supposed to do that and have to aim for the 'top' journals.) It's always pleasant to get attention, but I have been surprised in this case: I made it completely transparent that what I

was doing in the 'Best of Both Worlds' paper was reviving a position initially fully developed by Poincaré. The papers where I go a bit beyond Poincaré – by defending what some people have called 'Ramsey Sentence Structural Realism' (it's just Structural Realism really; no other version is coherent) – have attracted less attention, but then I did publish them obscurely!

Where do you see philosophy of science in ten years' time?

Popper used to distinguish prophecies and predictions. The latter being based on well-confirmed (sorry: corroborated!) theories. Like Popper, I don't put much store by prophecies: I don't know where the subject will be in ten years' time. But there is one change of direction that I would like to happen. For the past few decades there has been a great deal of excellent work on specific philosophical problems arising from science. For example, on trying to articulate a believable interpretation of quantum mechanics, or working out exactly what is involved in Darwinian evolutionary theory. But the 'grander' problem of the rationality of science and of theory-change in science has largely dropped out of focus since the Popper–Kuhn–Lakatos days. It would be nice to see a revival.

Where do you think traditional areas of philosophy, such as language and mind, stand in light of recent hard-science work in these areas?

Well, I have no trouble with philosophy of language – much of which is continuous with logic. Though I do think that some of the most interesting recent work, for example that of Kim Sternly, is involved with the issue of how various linguistic abilities, both syntactic and semantic, might have evolved. Philosophy of mind is a different issue. I struggle really to find a defensible role for it. I guess I am largely an unreformed positivist. If it isn't a question of clarifying some potentially confusing bits of science but also involves making substantive claims about the 'mind', then those claims ought to be testable – and that would be indistinguishable from neuroscience. To pick one example, I find the debates about the 'modularity' of the mind pretty barren. Whether or not there is an interesting notion of 'modularity' and, if so, whether and to what extent the mind/brain is modular, seem to me clearly questions for science itself – in this case, neuroscience. I just don't see any role for philosophy.

The *BJUP* team is extremely grateful to Professor Worrall for taking the time to answer our questions in such detail.

Issue 9 (2) 10th Anniversary Edition
Autumn Conference 2015

Interviews

Journal of the British Undergraduate Philosophy Society