



2018 OZSW CONFERENCE

9 & 10 November 2018

Local Host and Conference Organisation

Faculty of Behavioural, Managerial, Social Sciences,

Department of Philosophy

University of Twente

Table of Contents

ORGANISATION.....	3
CONFERENCE SCHEDULE.....	4
KEYNOTE LECTURES.....	5
PANEL (ALPHABETICAL ORDER)	7
EXTENDED ABSTRACTS IN ORDER OF PRESENTATION (SYMPOSIUM).....	135
POSTERS IN ALPHABETICAL ORDER	171

Organisation

Organisation Committee

- Mieke Boon (UT, Chair)
- Michael Nagenborg (UT)
- Peter-Paul Verbeek (UT)
- Marianne Boenink (UT)
- Ciano Aydin (UT)
- Koray Karaca (UT)
- Miles MacLeod (UT)
- Kevin MacNish (UT)
- Nolen Gertz (UT)

Reviewers

- | | |
|-----------------------------|------------------------------------|
| ▪ Jan Bergen (UT) | ▪ Per-Erik Milam (UT) |
| ▪ Bas de Boer (UT) | ▪ Michiru Nagatsu (Univ. Helsinki) |
| ▪ Mieke Boon (UT, Chair) | ▪ Michael Nagenborg (UT) |
| ▪ Stéphanie Gauttier (UT) | ▪ Han van Ruler (EUR) |
| ▪ Jonne Hoek (UT) | ▪ Peter-Paul Verbeek (UT) |
| ▪ Koray Karaca (UT) | ▪ Nils Wagner (UT) |
| ▪ Sujin Kim (UT, Assistant) | ▪ Johannes Korbmacher (Utrecht) |
| ▪ Owen King (UT) | |
| ▪ Nicola Liberati (UT) | |
| ▪ Miles MacLeod (UT) | |
| ▪ Kevin Macnish (UT) | |

Administration

- Sujin Kim (UT, assistant))
- Ada Krooshoop (UT)
- Christine Boshuijzen- van Burken (EUR)

OZSW Board of Chairs

- Christine Boshuijzen- van Burken (EUR)
- Mieke Boon (UT)
- Anthonie Meijers (TUE)
- Han van Ruler (EUR)
- Maureen Sie (Tilburg University)

Conference Schedule

Friday, November 9

length (minutes)

10.00-11.00	60'	Coffee and registration
11.00-11.15	15'	Opening and welcome
11.15-12.15	60'	Keynote 1:
12.15-13.15	60'	Lunch
13.15-14.30	75'	Panel 1 (6 parallel)
14.30-15.45	75'	Panel 2 (6 parallel)
15.45-16.15	30'	Coffee/Tea break
16.15-17.30	75'	Panel 3 (6 parallel)
17.30-18.30	60'	Keynote 2:
18.30-19.15	30'	Drinks
19.15-21.00	90'	Conference dinner

Saturday, November 10

09.15-10.00	30'	Coffee and 2nd registration
10.00-11.00	60'	Keynote 3:
11.00-12.15	75'	Panel 4 (6 parallel)
12.15-13.30	90'	Lunch
13.30-14.45	45'	Panel discussion with keynote speakers
14.15-16.00	75'	Panel 5 (6 parallel)
16.00-16.30	45'	Closure

Keynote Lectures

Toleration or Engagement? Responding to Divisiveness in Democracies

Maeve Cooke
(University College Dublin)

Ethical collisions are part of collective human life. By this I mean encounters between individuals and groups whose identities are shaped by conflicting views as to how humans should live their lives, which in turn shape their everyday behaviour and practices. I call these ‘ideas of the good life’. Such ideas may be based on cultural traditions, religious beliefs, philosophical positions, political ideologies, or other allegiances. They may be held tacitly rather than explicitly and are connected in complex ways with economic interests, claims to justice, struggles for freedom, and other individual and collective motivations. In the history of democratic modernity, encounters between groups who hold diverging or conflicting ideas of the good life have been seen as a source of social divisiveness and polarization. In response, John Rawls advocates liberal constitutionalism as a model of democratic politics that enables a reasonably harmonious and stable pluralist society, based on the successful and peaceful practice of toleration. I propose an alternative model of democratic politics. In my alternative account, encounters between diverging or even conflicting ideas of the good life are not avoided but encouraged. At its centre is a conception of authoritative, but non-authoritarian authority, underpinned by an account of the freedom-constituting powers of social institutions.

Arguments, Authority, and Polarization in Later Medieval Philosophy

Russell Friedman (KU Leuven)

In the polarized environment of, say, contemporary American politics, the polarization has a strong cognitive component, whereby the polarization correlates strongly with our susceptibility to certain arguments and even facts, our notion of truth, and our attitude towards experts and authorities. A similar polarization can be found in later medieval university philosophy, with its roots in rival intellectual and religious traditions that coalesced around the two big mendicant orders, the Dominicans and the Franciscans. In this paper, after giving some crucial information on the medieval institutional and intellectual background to the polarization I will be describing, I will give some examples of how that polarization

manifested itself, in particular by showing how the polarized groups exhibited different attitudes towards arguments and authorities. Possible areas that I'll use to investigate this medieval polarization are: divine illumination, hylomorphism, and trinitarian theology.

Fake News and the Politics of Truth

Michael Lynch (University of Connecticut)

Fake news spread online is a clear danger to democratic politics. One aspect of that danger is obvious: it spreads misinformation. But other aspects, less often discussed, are that it also spreads confusion, undermines trust and encourages us to live in a kind of epistemic bad faith. In this talk, I will argue that it is this last aspect that captures the most pernicious effect of fake news and related propaganda. In particular, I'll argue that its effectiveness is due in part to a curious blindness on the part of many users of social media: a kind of semantic blindness to the function of their online communicative acts. This blindness makes us not only vulnerable to manipulation to those with a better understanding of the semantic character of online communication, it indirectly undermines the political value of truth—or more exactly, the pursuit of truth, by diminishing confidence in the institutions that protect and encourage that value.

Panel (Alphabetical Order)

Admiration Over Time

Alfred Archer (Tilburg University) and Ben Matheson (Stockholm University)

We present a puzzle about the conditions on becoming admirable and remaining admirable. While some work on responsibility has distinguished the conditions (e.g. Khoury 2013, Matheson 2014), all work on admiration has so far focused on the first question. We will present cases that show that a person can cease being admirable for performing an act that they were once admirable for performing. We will consider three explanations for this fact and consider the problems faced by each. We will argue that the most plausible explanation is that admiration over time is sensitive to defeaters.

Aung San Suu Kyi is the de facto civilian leader of Myanmar (or Burma), considered by some to be Asia's Nelson Mandela, "a picture of grace and moral authority" for championing democracy and opposing "the brutal military junta that long dominated Burmese politics" (Tharoor 2017). Suu Kyi seemed to extoll virtue, and the fact that her acts expressed such virtue seems to be what makes her admirable for acting as she did. Her award of a Nobel Peace Prize in 1991 seems like a clear expression of admiration for those acts.

Suu Kyi has, however, recently come under moral scrutiny for not speaking out against the genocide perpetrated by her country's army against the long persecuted Rohingya people. Indeed, it is not just her silence that seems objectionable. As Hasan (2017) puts it, "she is also an apologist for genocide, ethnic cleansing and mass rape". Such actions are not the actions of a morally admirable person. It seems, then, that Suu Kyi is now not a fitting target of admiration – though it remains forever true (due to the fixity of the past) that her earlier self is admirable.

We might provide an epistemic explanation for Suu Kyi's not being admirable later in life: her later acts revealed that (despite appearance) she was never in fact admirable because she was never in fact virtuous. This may explain some cases where admiration seems to be lost over time. However, the more interesting set of cases we wish to consider are those of fallen moral heroes: people who were virtuous earlier in her life and then ceased be virtuous later on.

In this paper, we focus on two issues concerning admiration for persons for acts they perform. First, what are the conditions on becoming admirable? Second, what the conditions on remaining admirable? While some work on responsibility has

distinguished these conditions with respect to responsibility and blameworthiness (e.g. Khoury and Matheson, forthcoming), no work on admiration has investigated the conditions on remaining admirable. Our interest in this paper is primarily on these conditions – that is, on admiration over time. Fallen moral heroes show that if a person is admirable for A-ing at t_1 , then she is not necessarily admirable for A-ing at t_2 . That is, a person can cease being admirable for performing an act that they were once admirable for performing.

The fallen moral heroes case suggests that remaining admirable is determined by remaining virtuous. However, this account looks problematic when we consider cases of non-moral admiration over time. We often continue to admire athletic heroes past their prime – such as Brazilian soccer legend Pele – even though they arguably cease to possess the athletic virtues that made them admirable for their sporting feats. These two cases taken together suggest that moral admiration involves a synchronic evaluation (i.e. an evaluation of a person at a particular time) whereas non-moral admiration involves a diachronic evaluation (i.e. an evaluation of her life up until now).

But two further cases cast doubt on this picture. First, sometimes artists seem to cease being admirable. Consider an artist, Fraser, who starts his career making great works but who then “sells out” and creates uninspiring works with the sole aim of making money. Arguably, Fraser ceases being admirable for his earlier works, (though his earlier self of course remains admirable). Indeed, we might say that they are “undoing” the artistic credibility they had earlier acquired. Second, sometimes moral heroes seem to continue to be admirable despite no longer being virtuous. Consider a person, Sofia, who saved many people from Nazi death camps, but who never again gets the opportunity to manifest the virtues she manifested in her heroic actions during World War 2. It seems possible that the relevant virtues could degrade to the point that she no longer possesses those virtues without her gaining the opposing vices. Because she doesn’t replace her lost virtues with vices, it seems intuitive that she remains admirable.

Notice that when a person does something that goes against (in some sense) her admirable actions, she seems less and eventually not at all admirable for those actions. And when a person does nothing that goes against her admirable actions, she seems to remain admirable for them. This suggests that admiration over time is sensitive to defeaters. We propose an epistemic, metaphysical, and social understanding of defeaters. We suggest these three ways to understand defeaters highlight avenues for further research, though we stop short of endorsing a particular conception here. Our primary goal is to highlight what we call the puzzle of admiration over time. We end by considering whether this puzzle generalises to other emotions and attitudes.

References:

Hasan, M. (2017) “Burmese Nobel Peace Prize Winner Aung San Suu Kyi Turned into Apologist for Genocide against Muslims” The Intercept, URL: <https://theintercept.com/2017/04/13/burmese-nobel-prize-winner-aung-san-suu-kyi-has-turned-into-an-apologist-for-genocide-against-muslims/>

Khoury, A. & Matheson, B (forthcoming) “Is Blameworthiness Forever?” Journal of the American Philosophical Association

Tharoor, I. (2017) “The Shameful Silence of Aung San Suu Kyi” The Washington Post, URL:

<https://www.washingtonpost.com/news/worldviews/wp/2017/09/06/the-shameful-silence-of-aung-san-suu-kyi/>

Value sensitive designers: A person centered account of co-creative innovation

Mandi Astola (Eindhoven University of Technology)

There is much reason to look into the potential of normative virtue theories including epistemic and moral theories in understanding co-creative innovation. Virtue theories focus on the individual and often neglect the importance of collective and consequential values. They also tend to assume that individual virtues always lead to collective good, while this is not always the case in group activities. In co-creation it seems to be precisely the collective and consequential goods which are important. In order to make virtue theory more fitting for the context of co-creation, some account should be given of the importance of these values. I propose two possible solutions. The first is to accept virtue theories that state that there are both consequential and procedural virtues. The second option is to accept that there are two kinds of agents that can have virtues: individuals and collectives.

Co-creation means involving a variety of stakeholders in the process of innovating something new. For example, an innovator developing street-lighting for a city might opt to co-create the streetlights together with the inhabitants of that city as well as researchers from a university. Co-creation has been called a new paradigm in innovation. Co-creation scholar Venkat Ramaswamy even compares it to the Copernican revolution. Individuals become the new center of innovation, rather

than the created product or service.¹ Many parties affected by the innovation are called in to contribute to its design.

There are many ways to co-create. Innovative procurement and living labs are examples of co-creation methods. While there are established norms and guidelines for how to handle participants in a lab setting ethically and how to handle employees ethically, there is a lack of guiding norms on how to handle co-creating participants ethically. Co-creation methods put great emphasis on spontaneity. Even so, it is important to think about what norms co-creation practices should have to prevent participants from being exploited. There is a lack of literature on this topic. I see potential in a virtue-theoretical approach, especially one which integrates both moral and epistemic virtues. This paper is a first step to establishing a virtue theory that is specific to the context of co-creation.

By “virtue theories” I mean both epistemic and moral virtues. Notably, innovation is a goal-directed activity and therefore has a clear moral and epistemic *telos*. An innovator is trying to find out how to address a need in the world in the best possible way. This is a moral as well as an epistemic question.

A naïve view of virtue in co-creative innovation would be to claim that all-round virtuous people make for good co-creation. This view is inadequate because it does not recognize the role that individual vices can play in producing collective good. This becomes apparent in cases where a vice on one level is a virtue on the other. An example of this is the phenomenon of Mandevillean intelligence, as described by Paul Smart.² Mandevillean intelligence is when the vices of individuals contribute to collective virtue and it is thinkable in co-creation enterprises. For example, an egoist might be particularly good at making sure that the innovated product meets their needs and thus also the needs of others like her. Just as one bigot in a group might cause more dialogue and make the group more reflective as a whole than they might be otherwise. If collective good resulting from vices is explained away as moral or epistemic luck, then one has failed to grasp what is so distinctive and valuable about co-creation.

In order to create a virtue theory that is specific to co-creation, the collective and consequential values that virtue theories often don’t account for, should be accounted for. Therefore, I propose two possibilities for modifying virtue theories so that they do account for these values. The first option is to accept that there are both consequential and procedural virtues. If we accept this, then it is easy to explain how a person’s vice can contribute to collective good. While the egoist has a procedural vice, their egoism embodies also a consequential virtue, if it is so that their egoism leads in collective success in innovation. The two kinds of virtues are different kinds of virtue, which is why it is not problematic that they are predicated

¹ Venkat Ramaswamy, Kerimcan Ozcan, *The Co-Creation Paradigm*, (Stanford, CA: Stanford Business Books: 2014). xv.

² Paul Smart, “Mandevillean Intelligence,” *Synthese* 195, (2018). 4169–4200.

of a single character trait. The second solution is to accept that there are two kinds of agents that can have virtues: individuals and collectives. If we accept this, then we can maintain that while the egoist has a vice, the group as a whole has a collective virtue. The collective and the individual are two separate agents to whom we can attribute separate characteristics.

References:

Bruni, Luigino, and Robert Sugden. "Reclaiming Virtue Ethics for Economics." *Journal of Economic Perspectives* 27, no. 4 (November 2013): 141–64. <https://doi.org/10.1257/jep.27.4.141>.

Driver, Julia. *Uneasy Virtue*. New York: Cambridge University Press, 2006.

Hösle, "Ethics and Economics, or How Much Egoism Does Modern Capitalism Need? Machiavelli's, Mandeville's, and Malthus's New Insight and Its Challenge."

Paul R. Smart, "Mandevillian intelligence," *Synthese* 195, no. 9 (September 2018): 4169– 4200, <https://doi.org/10.1007/s11229-017-1414-z>.

Curley, Martin, and Bror Salmelin. "Open Innovation 2.0: A New Paradigm." *OISPG White Paper*, 2013, 1–12. <https://doi.org/10.1109/HIS.2008.172>.

European Commission. "Open Innovation 2.0 | Digital Single Market." Accessed October 31, 2018. <https://ec.europa.eu/digital-single-market/en/open-innovation-20>.

Hösle, Vittorio. "Ethics and Economics, or How Much Egoism Does Modern Capitalism Need? Machiavelli's, Mandeville's, and Malthus's New Insight and Its Challenge." Franz Steiner Verlag, 2011. <https://doi.org/10.2307/23681120>.

Ramaswamy, Venkat. "Co-creating value through customers' experiences: The Nike case." *Strategy and Leadership* 36, no. 5 (2008): 9–14. <https://doi.org/10.1108/10878570810902068>.

Saille, Stevienna de. "Innovating innovation policy: the emergence of 'Responsible Research and Innovation'." *Journal of Responsible Innovation* 2, no. 2 (2015): 152–68. <https://doi.org/10.1080/23299460.2015.1045280>.

Slote, Michael. "VIRTUE ETHICS AND DEMOCRATIC VALUES." *Journal of Social Philosophy* 24, no. 2 (September 1993): 5–37. <https://doi.org/10.1111/j.1467->

Smart, Paul R. "Mandevillian intelligence." *Synthese* 195, no. 9 (September 2018): 4169–4200. <https://doi.org/10.1007/s11229-017-1414-z>.

What Does Fairness in Trade Require? Do We Need a Difference Principle?

Sine Bagatur (Leiden University)

In *Fairness in Practice- A Social Contract for Global Economy* (2012), Aaron James proposes three principles as basic requirements of fairness in international trade. According to the second International Relative Gains principle; national income gains specifically due to international trade need to be distributed equally among trading nations. In this paper, I will argue that considering the role of increasing and diminishing returns in international trade, we can conceptualize international trade practice not as a relation of mutual gain as James does but rather as a relation of unequal exchange. This view of international trade leaves open the question whether we need a Difference Principle as a requirement of fairness in international trade.

In *Fairness in Practice- A Social Contract for Global Economy* (2012), Aaron James proposes a philosophical account of fairness in the global economy. Based on a constructivist methodology, James argues that international social practice of market reliance raises an issue of structural equity, that is "equity in how different countries and their respective classes are treated within the common market reliance relationship" (2012, 3). James consequently defends three principles of structural equity two of each concerns the gains from international trade. According to the second *International Relative Gains* principle; national income gains specifically due to international trade to be distributed *equally* unless greater gains flow to poor countries (2014: 181, emphasis added).

James states explicitly that the 'gains' in question are only the gains in productivity that result *specifically due to trade* as specified in economic theory and approximated in econometric practice (2014: 180, emphasis original). By economic theory, James refers to standard Ricardian theory of comparative advantage and claims that standard free trade argument is consistent with the conception of fairness he proposes (2012: 37). According to standard Ricardian theory of international trade, in simple terms, international trade creates mutual gain if each trading country specializes in the production of the good in which it has comparative

advantage. James takes for granted the notion of *mutual gain* from trade; the issue of fairness arises as a matter of not *creation* but *distribution* of mutual gains both among and within the trading nations.

In the first part of the paper, I will bring this assumption of *mutual gain* from trade into question relying on “the law of diminishing returns” in economic theory. According to the law of diminishing returns, the trading country benefits from specialization in the production of the good they have comparative advantage in, *only if* the good in question is subject to increasing or constant returns. When, on the other hand, the export industry of a country is subject to diminishing returns such as it mostly happens in resource-based industries, the country would not gain from trade if the trading partner’s export industry has increasing or constant returns. In other words, the consideration of law of diminishing returns bring the possibility of unequal exchange in trade, ie. one country benefits from trade at the expense of the other.³

In the second part of the paper, I will explore the implications the law of diminishing returns for the requirements of fairness in international trade. I will explore whether one can argue for a *Difference Principle* instead of James’ *International Relative Gains* principle as applied to gains from trade. In their critical response to James’ account of fairness in international trade Risse and Wollner (2014) calls into question James’s image of fairness in trade. They argue that we should better conceptualize trade in terms of a notion of exploitation. My approach of international trade as a relation of unequal exchange supports this proposal by Risse and Wollner.

References:

James, Aaron. “A theory of fairness in trade.” *Moral Philosophy and Politics* 1.2 (2014): 177-200.

James, Aaron. *Fairness in practice: A social contract for a global economy*. Oxford University Press, 2012.

Reinert, Erik S. “Diminishing returns and economic sustainability; the dilemma of resource-based economies under a free trade Regime.” *Genesis* 13 (1996): 6.

³ There is some empirical evidence supporting the claim that specializing in industries subject to diminishing returns can hamper gains from trade. For example, Reinert (1980) presents three case studies, showing how the historically most important export industries of Ecuador, Bolivia and Peru—bananas, tin, and cotton are producing under the conditions of diminishing returns. Reinert (1996) also claims that this model is a realistic representation of the economic forces which created uneven growth of the North and South.

Reinert, Erik. *International Trade and the economic mechanism of underdevelopment*. Ann Arbor, 1980.

Risse, Mathias, and Gabriel Wollner. "Three images of trade: On the place of trade in a theory of global justice." *Moral Philosophy and Politics* 1.2 (2014): 201-225.

Formalizing Mental Causation

Sander Beckers(Utrecht University)

The exclusion argument presents a challenge to the non-reductive physicalist: how can there be mental causation in a physically closed world? The non-reductive physicalist holds that a mental state supervenes on a physical state, but is not reducible to it. Therefore she seems forced to accept that many actions have two distinct sufficient causes, and are thus overdetermined. But overdetermination of an event by two causes is deemed to be highly exceptional, to the extent that it is highly implausible to be as widespread as mundane cases of mental causation are. This paper takes up this challenge.

A central problem in the philosophy of mind is to figure out if and how mental states can ever be causes in a physically closed world. A simple example suffices to illustrate the problem.

Say Sarah has the mental state of desiring a coffee, and this desire causes her to walk to the coffee shop. Given physicalism, which is the thesis that all of reality is ultimately grounded in physical states, her mental state supervenes on some physical state (for example the firing of a particular pattern of neurons in her brain). This physical state is *also* a cause of her walking to the coffee shop.

The *non-reductive physicalist* holds that the mental state *supervenes* on the physical state, but is not *reducible* to it. Therefore she seems forced to accept that Sarah's walking to the coffee shop has two distinct sufficient causes, and is thus *overdetermined*. But overdetermination of an event by two causes is deemed to be highly exceptional, to the extent that it is highly implausible to be as widespread as such mundane cases of mental causation are.

This problem has been dubbed the exclusion argument, and is most famously presented by Kim (1998, 2005). The exclusion argument presents a challenge for non-reductive physicalism. By now, there is widespread consensus that this challenge takes the following form: given non-reductive physicalism, overdetermination of one event by two cases is widespread (Bennett, 2003;

Hitchcock, 2012; Woodward, 2008, 2015; Weslake, Forthcoming). One type of overdetermination, namely the type that we observe in examples like that of a firing squad, is extremely rare. Therefore the non-reductive physicalist owes us an account of a second type of overdetermination that need not be rare. Concretely, the challenge lies in formulating both types of overdetermination in a manner that allows us to unambiguously distinguish the rare type from the widespread type. This is the challenge I will take up.

So to be clear, this paper offers no direct defense of non-reductive physicalism, but rather starts by simply assuming non-reductive physicalism. It does offer an indirect defense, in that if one is already inclined towards non-

reductive physicalism, then I will show that the exclusion argument doesn't offer you reason to abandon it. Further, the assumptions here made to characterize non-reductive physicalism are non-committal, in that they are compatible with various versions of that position. In particular, I will remain silent on whether causal relata are events or properties, and on whether a mental event is distinct or not from a physical event.

The only additional assumption I will make is that causation is best captured by considering causal models. In particular, causation is best defined in terms of a relation that holds between values of variables within some model. Although I have proposed such a definition elsewhere, this paper is compatible with any definition of causation formulated in terms of causal models, as long as it considers cases of overdetermination to be cases of causation.¹ In particular, it should consider cases of overdetermination to be cases of *sufficient causation*, for this is the type of causation that is at stake in the exclusion argument. Hence all talk of causation should in fact be read as talk of sufficient causation throughout.

First I will work towards defining two distinct types of overdetermination, where the one type captures cases of mental causation, like walking to the coffee shop, and the other type captures cases that are rare, like the firing squad. To get there, I present a formalization of the notion of *abstraction*, meaning the relation that holds between a causal model made up of *high-level variables* and a causal model made up of *low-level variables* such that the high-level variables represent properties that supervene on the properties represented by the low-level variables. Second I formalize the key assumptions of non-reductive physicalism using the vocabulary of abstraction and causal models. Finally I formalize the exclusion argument, and prove that if one accepts the introduced distinction between the two types of overdetermination then it is not valid.

References:

- Bennett K (2003) Why the exclusion problem seems intractable, and how, just maybe, to tract it. *Nous* 37(3):471–497
- Hitchcock C (2012) Theories of actual causation and the causal exclusion argument. *Journal of Consciousness Studies* 19(5-6):40–56
- Kim J (1998) *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: Bradford
- Kim J (2005) *Physicalism, or Something Near Enough*. Princeton University Press
- Weslake B (Forthcoming) Exclusion excluded. *International Studies in the Philosophy of Science*
- Woodward J (2008) Response to strevens. *Philosophy and Phenomological Research* 77(1):193–212
- Woodward J (2015) Interventionism and causal exclusion. *Philosophy and Phenomological Research* 91(2):303– 347

Of Savages and Stoics: Conjectural History as a Republican Tool in Jean-Jacques Rousseau and Adam Ferguson

Rudmer Bijlsma (University of Lausanne)

Scholars tend to drive a sharp wedge between the conjectural histories of the ‘speculative’ Rousseau and those of the ‘empirical’ Scottish Enlightenment social theorists. This paper argues that the convergences in the conjectural histories of Adam Ferguson – the most ambivalent among the Scots *vis-à-vis* commercial society’s merits – and Rousseau are more fundamental than often assumed. It does so by scrutinizing two related oppositions in their histories: nature vs. art, and savage vs. civilized. For both thinkers, it is argued, the figures of the Stoic and the savage serve a similar role as exemplars of moral-psychological wholeness.

It has often been noted that Rousseau’s critical analysis of modern commercial society and its origins significantly influenced Scottish Enlightenment thinkers. Several Scots sympathized with, and sought their own, more toned-down solutions to Rousseau’s claims that the focus of the members of commercial society on self-interested pursuits compromised the public interest, and that their excessive concern with the opinion of others stood in the way of true virtue. At the same

time, nevertheless, scholars tend to drive a sharp wedge between the conjectural histories of the ‘speculative’ Rousseau (1712-1778) and those of the ‘empirical’ Scots. This latter assessment still obscures a proper understanding of the closeness between their positions – particularly in the case of Adam Ferguson (1723-1816). Ferguson offers a more ambivalent conjectural history than fellow-Scots like Adam Smith and John Millar, in which the civilized, commercial stage does not emerge as best *unequivocally*.

I argue that the convergences in the conjectural histories of Ferguson and Rousseau are more fundamental than often assumed. I do so by scrutinizing two related oppositions in their histories: nature vs. art, and savage vs. civilized. McDaniel (2013) contends that one of the central differences between Ferguson and Rousseau is that the former has a Stoic ideal of a life lived ‘according to nature’ which contrasts with the latter’s Hobbesian demarcation of nature and art. I argue that Rousseau’s understanding of artifice is not necessarily/exclusively intended as a superimposition upon, or distortion of nature. It has a positive dimension: The artifice of the social contract allows humans to live according to their true needs and in harmony with their fellows. As such, it restores the moral-psychological wholeness that also pertained to the savage, and remedies the alienation from their own nature of the members of corrupt societies. Moreover, Rousseau’s conception of human flourishing here is, like Ferguson’s, inspired by Stoic natural law. Like the Stoics, the two early moderns see the human self as a complex whole, whose natural unity and strength can be diminished through misguided engagement with, and valuation of the outside world.

Ferguson’s critique of Rousseau’s individualistic depiction of the savage is well-known. In his conjectural history, nevertheless, the savage serves a role similar to Rousseau’s. Repeatedly, Ferguson juxtaposes the equanimous, courageous savage to the hyper-specialized, effeminate commercial citizen. Though appreciating civilized, commercial societies for their prosperity and rule of law, Ferguson’s message is that we have lost the savage’s (and barbarian’s) vigour and public-spiritedness. As in Rousseau, this loss is critical, because a collapse into despotism is an increasingly looming spectre.

While Ferguson’s socio-political critique is less radical than Rousseau’s, his conjectural history is not merely that of a detached anatomist of civil society. His depiction of the savage has an idealized dimension not unlike Rousseau’s, differing considerably from the bleaker accounts of savage life of other Scots. Moreover, the conjectural histories of the two thinkers both prepare a plea for restoring the republican virtue that was common in earlier stages of society. A more general lesson from my comparative analysis, then, is the insight it offers in the way that conjectural history – primarily an instrument of the modern natural law tradition – has served as a powerful tool in the hands of eighteenth-century republican critics of commercial society. In relation to this, I also discuss the way in which *Stoic* natural law helps Rousseau and Ferguson to put their conjectural histories to

republican use. The Stoic model of the life lived according to nature allows them to criticize, and propose remedies for, the particularly poor way in which the members of modern commercial societies manage to live up to it.

References:

Iain McDaniel. 2013. 'Philosophical History and the Science of Man in Scotland: Adam Ferguson's Response to Rousseau'. In: *Modern Intellectual History*. 10 (3): 543-68.

Multi-aspectual approach to Values in the Refugee Chain

Christine Boshuijzen- van Burken (Erasmus University Rotterdam)

In order to improve the refugee logistics chain, we approach the problem of refugee logistics from a value sensitive design (VSD) perspective. VSD is a tripartite design methodology, which implies conceptual, empirical and technical investigations, employed iteratively (Friedman, Kahn, & Borning, 2006). Our first step is to understand current values in the refugee chain. This is done via a stakeholder meeting and a multi-aspectual analysis methodology. The second step is to understand where and how values in the refugee chain (potentially) clash with values of stakeholders, and in particular with values of refugees. The third step is to provide stakeholders with guidelines on how to improve their contribution in the refugee chain in a value sensitive manner.

The world currently faces high levels of refugees. By the end of 2016 there were 22.5 million refugees (<http://www.unhcr.org/5943e8a34.pdf>) across the globe. Amongst the main countries of asylum for refugees were Germany, Italy, Turkey and Islamic Republic of Iran (<http://www.unhcr.org/globaltrends2016/>). Hosting countries experience several challenges in terms of logistics for refugees. Although many countries do their utmost to facilitate this process, there remains a risk of 'protraction', which means that a refugee is getting trapped for decades in the refugee migration and asylum process (Frydenlund & Padilla, 2017; Vernon-Bido, Frydenlund, Padilla, & Earnest, 2017) and another risk is that hosting countries lose societal support for supporting refugees (Oltermann, 2016).

In order to improve the refugee logistics chain, we approach the problem of refugee logistics from a value sensitive design (VSD) perspective. VSD is a tripartite design methodology, which implies conceptual, empirical and technical investigations, employed iteratively (Friedman, Kahn, & Borning, 2006). Our first

step is to understand current values in the refugee chain. The second step is to understand where and how values in the refugee chain (potentially) clash with values of stakeholders, and in particular with values of refugees. The third step is to provide stakeholders with guidelines on how to improve their contribution in the refugee chain in a value sensitive manner.

Many classifications of values have been proposed by several authors, see for example the *Handbook of Ethics, Values, and Technological Design* (van den Hoven, Vermaas, & van de Poel, 2015). Examples of values are privacy, responsibility, safety, freedom, sustainability, equality. We propose a different, more holistic approach to values, by using Dooyeweerd's theory of modal aspects (Dooyeweerd, 1969) as a means to systematically map different values of the stakeholders. It distinguishes for example economic value, juridical value, social value, psychological value, religious value, moral value and aesthetic value.

This project aims to contribute theoretically to Value Sensitive Design research in two ways: (i) by using Dooyeweerd's theory of modal aspects to create a novel and systematic overview of different values and (ii) to apply Value Sensitive Design to the case of refugee logistics. This project aims to contribute to society at large, by creating a tool for policy makers which should inform decision concerning humanitarian logistics in an ethically informed manner.

In this paper we furthermore present the result of the Group Decision Room, which is an online anonymous focus group discussion, in which we used the theory of modal aspects in order to compile a list of values in the chain of refugee logistics.

References:

- Dooyeweerd, H. (1969). *A New Critique of Theoretical Thought, The General Theory of the Modal Spheres* (2nd ed., Vol. 2). The Presbyterian and Reformed Publishing Company.
- Friedman, B., Kahn, P., & Borning, A. (2006). Value sensitive design and information systems. In P. Zhang & D. Galletta (Eds.), *Human-Computer Interaction in Management Information Systems: Foundations* (pp. 348–372). Armonk, NY: M.E. Sharpe.
- Frydenlund, E., & Padilla, J. J. (2017). Modeling the Impact of Protraction on Refugee Identity. In D. Lee, Y.-R. Lin, N. Osgood, & R. Thomson (Eds.), *Social, Cultural, and Behavioral Modeling: 10th International Conference, SBP-BRiMS 2017, Washington, DC, USA, July 5-8, 2017, Proceedings* (pp. 214–222). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-60240-0_25

Jensen, L.-M., & Hertz, S. (2016). The coordination roles of relief organisations in humanitarian logistics. *International Journal of Logistics Research and Applications*, 19(5), 465–485. <https://doi.org/10.1080/13675567.2015.1124845>

Oltermann, P. (2016, March 14). Merkel refuses to abandon refugee policy despite election setbacks. *The Guardian*. Retrieved from <http://www.theguardian.com/world/2016/mar/14/angela-merkel-refuses-to-abandon-refugee-policy-despite-election-setbacks>

van den Hoven, J., Vermaas, P. E., & van de Poel, I. (2015). *Handbook of ethics, values, and technological design*. Springer Netherlands: Imprint: Springer.

Vernon-Bido, D., Frydenlund, E., Padilla, J. J., & Earnest, D. C. (2017). Durable Solutions and Potential Protraction: The Syrian Refugee Case. In *Proceedings of the 50th Annual Simulation Symposium* (p. 19:1–19:9). San Diego, CA, USA: Society for Computer Simulation International. Retrieved from <http://dl.acm.org/citation.cfm?id=3106388.3106407>

Who Decides? An Argument for Democratic Selection Criteria for Refugees

Amanda Cawston (Tilburg University)

In our non-ideal world, states refuse to fully open their borders to refugees, leading some to advocate a ‘triage’ approach and formulation of selection criteria that ensures refugees suffering the most persecution are helped first. I reject this approach, arguing that it lends further legitimacy to an unjust state sovereignty system and perpetuates harmful institutional practices. Instead, I argue for the democratic determination of selection criteria, where such criteria are determined by refugees themselves. While imperfect, the proposal is flawed in a transformative way that prompts measures to empower refugees and that undermines the ideological basis for the status quo.

In our actual, non-ideal world, states refuse to fully open their borders to refugees, leaving many wrongfully excluded. This situation has prompted some to suggest we ought to adopt a ‘triage’ or prioritarian approach to asylum, devising selection criteria that ensures those suffering the most persecution are helped first. Here, I argue that if such criteria must be devised, their content ought to be determined via democratic input from those subject to them.

I begin by demonstrating that there are worries about the existing practice of

distinguishing between refugees and migrants. Kukathas (2016), for instance, argues that this practice involves invasive and dehumanizing bureaucratic treatment, and is ultimately designed to protect states' interests rather than benefit refugees. The existing project and practice of evaluating merit thus does not inspire confidence that further selection criteria would be genuinely humanitarian, either in principle or in effect.

However, even if it is possible to humanely apply a well-intentioned selection criteria, doing so is likely to legitimize and support the unjust practice of state-controlled borders. For adopting a priority-based selection scheme enables the belief that, while we cannot help all, if we succeed in helping those at most risk, we are thereby able to fulfil our most demanding obligations. But such thinking obscures the source of the conflict, casting the issue as one of the distribution of scarce resources/places rather than the injustice of state-restricted borders which creates the 'need' to allocate admittance. In other words, devising prioritarian selection criteria for refugees does not challenge the more fundamental issue, the presumption that states have a right to exclude.

In light of the above, I suggest we construct criteria that do not implicitly support the presumptions that are (partially) responsible for the problem. When thinking about policies concerning the selection of refugees, the aim should be to devise criteria that help address the more fundamental issue (the right to exclude) rather than further entrench it. To that end, I propose drawing on Abizadeh's (2008) claim that, on the democratic theory of popular sovereignty, coercive power must be justified by and to those it is exercised over. I argue Abizadeh's insight can be applied at an intermediate level and provide support for the idea that any selection criteria to allocate asylum ought to be determined by those who will be subject to it, i.e. by refugees. The general move is straightforward: selection criteria involves the use of coercion and thus calls for justification. This justification ought to require the participatory input from those who will be subject to the coercion, which in this case, are refugees. Notably, the focus on coercion helps legitimate the inclusion of refugees, but not receiving populations. While receiving populations are affected by the resulting criteria, they are not subject to the coercive mechanisms that restrict and evaluate refugee admission, and thus are not eligible for inclusion in the democratic process. Similarly, appealing to the ideals of deliberative democracy supports the inclusion of former refugees and close aid organizations, as these groups can help to give equal standing to minority groups or the relatively vulnerable within refugee populations. In other words, their inclusion can help support conditions of fairness and equality.

There are two further arguments to make in favour of this proposal, and that point to its intended transformative dimension. The first notes the ideological break with the foundations of the state sovereignty system and its protection of state interests over refugee interests. Refugee interests are placed at the centre, and while not fully replacing state interests (as states still control admission levels),

permits another group (refugees) to influence priorities. The second argument speaks to Kukathas' point concerning the practical inhumanity of evaluating migrant status. To ensure the democratic process is legitimate, resources must be directed towards supporting the conditions for democracy. While this focus will not entirely replace the bureaucratic and invasive measures that exist to determine refugee status, they may halt developing these measures further, and more importantly, devote resources towards measures that empower refugees.

While it is overly optimistic to expect states to move directly from the current system to border control via global democratic input (as per Abizadeh's suggestion), introducing democratic input into selection criteria represents a manageable intermediate step. This move would moreover be a step in the right direction: it would introduce a practice that helps weaken the presumption in favour of states' unilateral right to control their borders. I conclude by noting that while flawed, the proposal is flawed in a way that prompts measures to empower refugees and that undermines the ideological basis for the status quo. Furthermore, the proposal may work to provoke a transformative response that reveals the unacceptable nature of the original guiding question. Importantly, support for this proposal does not depend on being able to actually secure the conditions that will lend legitimacy to the resulting criteria that refugees select. I do not expect such conditions to be fully achievable. The point is more that our failure will be instructive, and efforts to improve conditions will work to undermine rather than further entrench the state sovereignty system. This proposal is not intended as a concessionary measure that regretfully offers a second-best option out of a lack of hope for securing more. Rather, it aims to be revolutionary and prefigurative.

References:

Abizadeh, Arash. 2008. 'Democratic Theory and Border Coercion: No Right to Unilaterally Control Your Own Borders', *Political Theory*, 36(1): 37–65.

Kukathas, Chandran. 2016. 'Are Refugees Special?' in Sarah Fine and Lea Ypi (eds.) *Migration in Political Theory: The Ethics of Movement and Membership* (Oxford: OUP): 249-268.

Intellectual Virtues, Justified Emotions, and Synchronic Self-Control

Chun-Nam Chan (The University of Manchester)

I shall propose an emotional theory of self-control, according to which the exercise of synchronic self-control requires the assistance of justified emotions (JEs) that are produced by the agent's intellectual virtues, which are reliable emotion-making processes, that constantly generate correct and proportionate emotions (CPEs). The intellectual virtues reliably detect the non-evaluative properties of a competing temptation that likely constitute an associated evaluative property and generate a corresponding JE. With the assistance of the JE, the agent can control herself at the same moment the temptation is attacking her because the JE directly overwhelms the self-indulgent desire alongside the self-disciplined one.

The puzzle of self-control has attracted considerable attention because it poses a serious threat to the exercise of synchronic self-control ('self-control' for short). The agent initiates the exercise of synchronic self-control if and only if she refrains from acting in accordance with her competing temptation while under siege of its influence. For example, '*whilst actually experiencing the craving to smoke*, the person may exercise self-control there and then in such a way that the craving is attenuated and they do not smoke a cigarette' (Connor 2013: 784, original italics). Still, the exercise of self-control is not as intelligible as its first glance. According to the Law of Desire, 'if a person most desires to perform some action A, and if she believes herself free to A, then she will A, if she does anything at all intentionally' (Sripada 2012: 42). What the reformed smoker most wants to do is either to smoke or not to smoke. On one hand, if her strongest desire is to light the cigarette in her hand, it seems psychologically impossible for her to initiate the exercise of self-control because her desire not to smoke is motivationally weaker. On the other hand, if her strongest desire is not to light the cigarette, the exercise of self-control is apparently redundant since she will not fall prey to the temptation. In this paper, although I assume the truthfulness of the Law of Desire, I shall argue that the exercise of self-control is still psychologically possible by proposing my emotional theory of self-control (ETSC).⁴

To clarify (ETSC), I must first introduce the value-based view of self-control (VBSC) and my virtue reliabilism about justified emotions (VRJE). According to (VBSC), the agent initiates the exercise of self-control if and only if she acts in alignment with her core values. As Henden (2016: 4) says: 'According to the value-based view of self-control, then, self-controlled persons govern themselves in

⁴ Although this principle has been challenged in various ways (Clarke, 1994; Gert 2005), it is still a 'useful starting point for discussion' (Mele, 2014: 364).

accordance with what they genuinely value even in the face of strong competing motivation'. By 'core values', I mean the convictions or principles that define the nature of the agent's identity and the path of her life, which she 'in a cool and non-deceptive moment – articulates as definitive of the good fulfilling and defensible life' (ibid.).

A competing temptation is a violation of the agent's core values, and because of that, it possesses a set of non-evaluative properties, such as the property of preventing her from actualizing the core values and the property of bringing negative consequences, that likely constitutes an associated evaluative property. Different core values and temptations, together, may constitute different exemplified evaluative properties. The reformed smoker's health and her temptation to smoke may constitute exemplified dangerousness because it is a threat to her body. The Samurai's code and his temptation to flee may constitute exemplified offensiveness because it is an insult (or a self-insult) to him. The politician's value of being respected and her temptation to verbally retaliate against her adversary may constitute exemplified shamefulness because it undermines her public reputation. It is necessary to differentiate between different evaluative properties as different forms of self-control require different emotions. The agent's fear, which is directed towards the exemplified dangerousness, can motivate her not to smoke. The samurai's anger, which is directed towards the exemplified offensiveness, can motivate him to face the enemy. The politician's shame, which is directed towards the exemplified shamefulness, can motivate her to restrain her tongue.

Having clarified (VBSC) and how the agent's competing temptation may constitute a violation of her core values, I shall move on to explain (VRJE). An emotion *E* is epistemically justified if, and only if *E* is produced by the agent's intellectual virtues, in essence, cognitive abilities, which are reliable emotion-making processes, that have a high probability of generating correct and proportionates emotions (CPEs). *E* is correct when directed towards a particular object that has a set of non-evaluative properties that likely constitute an exemplification of an associated evaluative property and is proportionate when its intensity fits the seriousness of the evaluative property.

With the clarification of (VBSC) and (VRJE), it is easier to see how (ETSC) operates. Assume the smoker has a stronger temptation to smoke and a weaker desire not to smoke. The temptation is a violation of her core value of staying healthy. This violation possesses a set of non-evaluative properties, such as the property of preventing the smoker to actualise the core value and the property of increasing her probability of getting cancer, that likely constitutes an exemplification of dangerousness. Then, the smoker's intellectual virtues reliably detect the non-evaluative properties of her craving for a cigarette and accordingly generate a justified fear, which is likely correct in the sense that it is directed towards the exemplified dangerousness and is proportionate in the sense that its intensity fits the seriousness of the exemplified dangerousness. At last, it is

psychologically possible for the smoker initiates the exercise of self-control because the weaker desire not to smoke has become the strongest one with the assistance of the proportionate motivational force of her justified fear. Since the exercise of self-control would be psychologically impossible without the assistance of the justified fear, it is not redundant for the smoker to control herself while being tempted to smoke. In the rest of the paper, two objections that suggest the assistance of justified emotions is neither necessary nor sufficient will be articulated and replied. If the plausibility of (ETSC) is successfully defended, it will be the first of its kind, since no prominent theory of self-control envisages the importance of emotions.

References:

- Clarke, R. (1994). Doing What One Wants Less. *Pacific Philosophical Quarterly*, 1-11.
- Connor, T. (2014). Self-Control, Willpower, and the Problem of Diminished Motivation. *Philosophical Studies*, 783-796.
- Gert, J. (2005). Breaking the Law of Desire. *Erkenntnis*, 295-319.
- Henden, E. (2018). Addiction as a Disorder of Self-Control. In H. Pickard, *The Routledge Handbook of Philosophy and Science of Addiction* (pp. 45-63). New York: Routledge.
- Mele, A. (2014). Self-Control, Motivational Strength, and Exposure Therapy. *Philosophical Studies*, 359-375.
- Sripada, C. S. (2012). How is Willpower Possible? The Puzzle of Synchronic Self-Control and the Divided Mind. *Noûs*, 41-74.

What Pride, Mr. President? On Pride and Politics

Martha Claeys (University of Antwerpen)

Politicians often appeal to the emotion of pride in order to convince people of their program. In this paper, I argue that the assumption that underlies the kind of pride that is being addressed determines whether the pride rhetoric is either polarizing or unifying. The emotion of pride can rely on two entirely opposite assumptions. One of inequality and competition, and one of radical equality. Acknowledging and paying attention to the role feelings of pride play, both in the politician's rhetoric and in the voter's decision, is fundamental to set apart intolerant demands from legitimate grievances.

Politicians often appeal to the emotion of pride. Most obviously, they have a certain way of being proud themselves. Donald Trump has been called narcissistic many times, Obama often has a proud and elegant posture and tone in his speeches. Their voters have a certain pride: whether it be nationalistic pride, worker's class pride, or anything else. And then politicians have ways to appeal to the people's pride, in order to convince them of their program. Populists, for instance, encourage 'the people', often the working class, who have been deprived and done wrong by the ruling elite for too long, to refuse to be trampled on or ignored any longer. In other words, their pride demands them to do something about the current situation. In a statement on the U.S.'s withdrawal out of the Paris climate change agreement, Trump said: "we don't want other countries and other leaders to laugh at us anymore," appealing directly to a feeling of humiliation. Many politicians, both on the left and right side of the political spectrum, appeal to nationalist or patriotic feelings in order to gather people behind a communal cause and identity, and to convince them to devote themselves to their citizenship and their community. In this paper, I want to explore those particular ways in which politicians appeal to the pride of their electorate in order to win support. I argue that the assumption that underlies the kind of pride that is being addressed determines whether the pride rhetoric is either polarizing or unifying.

The emotion of pride can rely on two entirely opposite assumptions. One of inequality and competition, and one of radical equality. In the first case, pride concerns an **idea of competitive ranking**, where one's success is measured by another's failure. Greatness is thought to be a **scarce good**, from which some can only have more if others have less. This sort of pride is reflected in the political rhetoric through a polarizing language, in terms of winners and losers, where the other is depicted as an **active threat to the own greatness** (us *versus* them). There is typically **no room for self-criticism**. The **greatness is assumed to be self-evident**, and there is **no consideration of luck** and inequality of chances to allow for this greatness. There is an emphasis on **bringing the other down**. And finally, there is a denial of differences within the group, and **differences are perceived negatively**, since in the competitive idea of greatness, **there can only be one** on top.

Pride can be the epitome of fundamental equality as well, though, if the underlying assumptions of the emotion are opposed to those of competitive pride. In this case, pride is the result of an **assumed equal worth and dignity in all**, and these are considered **non-scarce goods**. This sort of pride is reflected in a rhetoric **where the other and our own people exist alongside each other** (us *and* them). There is an **emphasis on the unique qualities of the own group**, but not in spite of the other, who is allowed to exist alongside the own group, with recognized different qualities of its own. The other is not perceived as a threat, but

as an enrichment. There is **significant room for self-criticism** and the recognition of challenges.

Greatness is seen as a non-evident achievement, one that takes hard work, and **factors of luck are acknowledged** in achieving it. And finally, **differences are recognized and embraced**, since the assumption is not one of direct competition, but one of plurality of qualities and inherent dignity as a non-finite good.

Through quotes and analysis of political rhetoric from different ideologies, genders, and political orientations, I show how these two notions of pride are politically employed in different ways. The analysis of the role particular emotions play in political rhetoric and argumentation can help to uncover polarizing tendencies, and show their underlying assumptions. It can be hard to distinguish between those two kinds of pride, since often the same words are used, and the same values are addressed. Yet, when looked into the political speech a bit further, we can notice that one sort of appeal to pride assumes a fundamental equality and dignity in all, whereas the other assumes competition and feeds polarization. It is not a matter of resisting emotional rhetoric in the political discourse altogether, but one of discovering what implicit assumptions precede the emotion that is used to convince, and wonder whether these are the assumptions we want to adhere to.

An understanding of the role pride plays in politics, and the sort of pride that is being addressed, can help to understand what sort of complaints underlie certain political opinions that are formulated in the pride rhetoric. To acknowledge what role an emotion, specifically pride, humiliation, self-esteem, and the like, plays in the wants of people is fundamental to be able to meet the real complaints and desires of people. To dismiss the role this emotion plays in politics, and to think about political choices only in terms of distributive fairness, misses a great motivator in people's political demands: that we want esteem, that we have pride, that we are vulnerable to feelings of humiliation or low regard, and that we tend to prioritize our own group above another. Figuring out in how far these motivations are legitimized, and what sort of assumptions underlie them, can only be constructive for democratic political discourse.

Can we be held responsible for implicit bias? The Ignored Pitfalls of Implicit Prejudice Research

Roland den Boef (Vrije Universiteit Amsterdam)

Implicit bias has become a hot topic in moral philosophy. This paper offers a critical evaluation of implicit bias as a psychological, scientific construct. Implicit

bias tests are found wanting with regards to their reliability and predictive validity. Scientific research cannot support claims of near-universal bias prevalence, widespread consequent discrimination, or accurate individual assessment. First-person awareness of implicit bias cannot ground blame for other biased persons. These findings mean that we cannot hold people responsible for discriminatory behavior caused by implicit bias. Consequences for policy are discussed, including anti-implicit bias training, and impact on philosophical debates.

Implicit bias has taken philosophy by storm. In less than ten years, it has become one of hottest ethical and political issues in the field. One of the central questions asked by moral philosophers is whether people are responsible for their implicit bias(es). The general assumption in the literature seems to be that we can know with sufficient certainty whether people are biased. Holroyd et al. (2017)'s review of the recent literature on this topic provides a good example of the foundational premises of the debate: that most people are implicitly biased, and tend to discriminate against minorities on a regular basis due to their bias; if not outright, then in their micro-behaviors; that this is one of the most important factors that can explain the continued inequalities in modern Western societies; and that we can and do know with a sufficient degree of confidence which individuals are biased and which are not. I will argue that such conclusions are unwarranted by the evidence, and even stronger, that the scientific findings we have to our disposal right now contradict this narrative. One could hope that the science will eventually show implicit bias to be an accurately measurable, almost unavoidable, pervasive and highly damaging phenomenon, but this pessimistic expectation is not based on a critical, accurate assessment of social psychology research in the area. My argument will focus on the practice of holding individuals morally responsible, i.e. praising or blaming a biased person for her (probable) discriminatory behavior.

I start with clarifying Smith's important distinction between being and holding responsible. People can be responsible in a metaphysical sense, without anyone else knowing that this is the case. Praise or blame, however, requires knowledge of the morally problematic attitudes in question and their behavioral consequences. Then I summarize what implicit bias is taken to be, how it is measured, and which uncertainties surround the social psychological research. I discuss the two dominant psychometric paradigms: measuring response latencies (e.g. the IAT) and sequential priming.

Implicit bias tests are generally unreliable. Even after twenty years of research, the IAT is unusable in practical circumstances. Its test-retest reliability is very low and clearly inadequate when measured against accepted psychological standards of quality, as admitted by its designers. Worse, other tests seem to fare no better or have not yet been tested rigorously. We cannot determine accurately whether an individual is biased, based on their IAT-score or the result of any other implicit

bias test. Therefore, implicit bias tests ought not be used to assess individuals. The widespread use of implicit bias tests, especially the IAT, is ethically questionable. There is also a strong case for not using these tests at all, since their use is inherently misleading.

Another way to argue for the claim that we can hold people responsible for implicit bias is to infer their responsibility from the (putative) fact that we are all (implicitly) biased. But we cannot justifiably leap from the fact of imperfect objectivity to the attribution of a pernicious form of, say, racial bias that routinely results in discrimination. The conflation of these things means that people will again be judged to be prejudiced without adequate evidence. Therefore, the social psychological literature would need to show that most of the population is implicitly biased and likely to discriminate due to this bias. If this is true, it would show up in predictive validity scores of implicit bias tests - and it clearly doesn't. According to Brian Nosek, one of the developers of the IAT, the tests' ability to predict behavior is "very weak overall". In fact, there is a good argument to be made that we currently have no rigorous evidence that establishes that the test itself reliably predicts discrimination, even to explain a mere 3-5% of discriminatory behavior. All our current knowledge indicates that implicit bias is a rather small factor in the plethora of causes of our (discriminatory) behavior.

It is often claimed that people are to a certain extent aware of their own implicit biases. This issue depends on one's view of what implicit bias is, and its relation to our emotions. However, this kind of awareness is necessarily restricted to the first-person perspective. We cannot judge reliably from the outside whether a person has certain implicit biases or gut feelings. The upshot is that if subjects can be aware of their implicit bias and they admit this, they might conceivably be held responsible. However, without such an admission, there still is no adequate ground for attribution of moral responsibility.

Finally, I argue that implicit bias research cannot carry the scientific and ethical weight that is given to it. Implicit bias test use, anti-implicit bias training and similar measures are morally questionable, especially since the scientific problems are not communicated to the public. If people engage in discrimination, they could be held responsible, but not on the basis of implicit bias tests. Philosophers need to rethink their largely uncritical use of social psychology research to save the practical value of their arguments.

Conscious Control and Responsibility for Habits

Wessel van Dommelen (University of Groningen)

Whether agents need to be in conscious control of their actions in order to be morally responsible is controversial. Agents seem responsible in two kinds of cases in which they lack conscious control: undeliberate omissions and habitual behavior. I claim that one can be responsible for habitual behavior, but not for undeliberate omissions. I argue that the ‘historical consciousness thesis’ accommodates for responsibility for habitual behavior. On this view, conscious control need not be proximal, but can be distal. Furthermore, agents can be responsible for habitual behavior in some cases where the decision leading to the behavior is automatized.

A commonsense assumption about moral responsibility is that an agent needs to be in conscious control of her behavior in order to be responsible for that behavior. On this assumption, an agent needs to be aware of the moral significance of her action and control her behavior in light of this awareness in order to be responsible for that action. I will call a view that subscribes to this assumption a “Conscious Control View” (CCV).⁵ On such a view we make sense of a case of responsibility as follows. Imagine, to borrow an example from J.L. Mackie, that a woman shoots a supposed intruder, who later appears to be her husband returning early from a journey. Although the woman is aware of shooting an intruder, she is not aware of shooting her husband. Hence, she is not responsible for shooting her husband on a CCV such as Mackie’s view (Mackie, 1977, pp. 203–226).

A CCV is a rather strong view. There are two types of cases that appear problematic. First of all agents seem to be responsible for some undeliberate omissions, such as forgetting a best friend’s birthday (Smith, 2005, p. 236). Secondly, agents seem to be responsible for some behavior performed habitually. Suppose that I have made a habit out of clicking my pen when I am deep in thought. It seems that I am blameworthy for manifesting this habit in a quiet library. Yet, at the moment I am manifesting my habit, I am not aware of the moral significance of my behavior.

These cases present a difficulty for a defender of the view that conscious control is necessary for moral responsibility. Yet, there is simultaneously a push towards accepting the role of consciousness for moral responsibility. For instance, Joshua Shepherd found that folk views on moral responsibility and free will accord an important role to consciousness (Shepherd, 2015). Several adjustments to a CCV have been suggested with the aim of reconciling the idea that awareness is important to moral responsibility with the intuitions pertaining to the two types of

⁵ George Sher calls this view “The Searchlight View” (Sher, 2009, pp. 4, 6).

counterexamples. The first way is to include a ‘tracing clause’, according to which an agent can be responsible for behavior out of that agent’s control, if it can be traced back to suitably connected behavior which was under that agent’s control.⁶ Second, some propose that an agent does not need to exercise conscious control in order to be responsible over some behavior, but that the agent needs to have the ability for conscious control to be responsible (Schlosser, 2013). Finally, some have suggested that an agent does not always need to be occurrently aware of the moral significance of the action (Levy, 2014; Zimmerman, 2008). That is to say, the relevant information does not always need to be in the forefront of one’s mind in order to be responsible for the action. Only dispositional awareness of the moral significance of an action would be necessary to be morally responsible.

In this paper I will argue for three points. First of all, I will argue that even though agents may be responsible for habitual behavior, agents are not responsible for undeliberate omissions since one does not reveal one’s regard for others through an undeliberate omission. Secondly, I will argue that if we appreciate the importance of the question when an agent needs to be aware of the moral significance of an action a number of cases in which an agent manifests a habit are not problematic for a CCV. However, a remaining number of problematic cases push towards accepting another alternative to a CCV, which I dub “the historical consciousness thesis”. According to this thesis an agent is responsible for an action only if conscious processes figure in the appropriate way in the history of that action. Finally, I will argue that the historical consciousness thesis is to be preferred over the three alternative revisions of a CCV.

References

- Aristotle. (2006). Book 3. In C. C. W. Taylor (Trans.), *Nicomachean Ethics*. Oxford University Press.
- Levy, N. (2014). *Consciousness and moral responsibility*. Oxford; New York: Oxford University Press.
- Mackie, J. L. (1977). Determinism, Responsibility, and Choice. In *Inventing Right and Wrong* (pp. 203–226). Penguin Books.
- Schlosser, M. E. (2013). Conscious Will, Reason-Responsiveness, and Moral Responsibility. *The Journal of Ethics*, 17(3), 205–232. <http://doi.org/10.1007/s10892-013-9143-0>
- Shepherd, J. (2015). Consciousness, free will, and moral responsibility: Taking the folk seriously. *Philosophical Psychology*, 28(7), 929–946.

⁶ Tracing is an old idea that can be tracked back to Aristotle (Aristotle, 2006, I. 1113b15-1114a22).

<http://doi.org/dx.doi.org/10.1080/09515089.2014.962018>

Sher, G. (2009). *Who Knew?* Oxford University Press. Smith, A. M. (2005). Responsibility for Attitudes: Activity and Passivity in Mental Life. *Ethics*,

115(2), 236–271. <http://doi.org/10.1086/426957> Zimmerman, M. J. (2008). *Living With Uncertainty*. Cambridge University Press.

Context, Content, and the Inferential Specification of Meaning

Matej Drobnák (Univerzita Hradec Králové, Czech Republic)

In this presentation, I try to show how the inferential specification of meaning can be used to explain several phenomena related to natural languages. First, it opens up the way for a specific inferential view on semantic understanding. Second, I try to show how the inferential specification of meaning can be used to explain cases of ambiguity and free pragmatic enrichment. The aim of this presentation is to establish theoretical foundations on which future discussions could be launched from the perspective of normative inferentialism – especially with regard to the topics discussed within the literalism-contextualism debate.

In my presentation, I try to show how the inferential specification of meaning, especially with regard to the distinction between the inferential potential and the inferential significance, can be used to explain several phenomena related to natural languages. First, it opens up the way for a specific inferential view on semantic understanding. Second, I try to show how the inferential specification of meaning can be used to explain cases of ambiguity, free pragmatic enrichment, and conversational implicatures.

With regard to this aim, I will focus on the distinction between the inferential potential (IP) and the inferential significance (IS) of a sentence as presented by normative inferentialists (Brandom 1994, 2000; Peregrin 2014). From the perspective of normative inferentialism, the inferential potential of a sentence can be understood as a specification of the meaning of a sentence. Moreover, inferentialists believe that the meaning of a sentence is perspectival; i.e. one sentence may have different significance in different contexts. As Peregrin claims, the relation between the context invariant inferential potential and the context dependent inferential significance is straightforward: “the inferential significance of A within the context C is the value of the inferential potential of A for C” (Peregrin, 2014, 51).

One way how to represent the meaning/IP of a sentence A is by a set of ordered

pairs of sentences:

$$IP(M) = \{ \langle X_1, D_1 \rangle, \langle X_2, D_2 \rangle, \dots, \langle X_n, D_n \rangle \}$$

where $X_1 = \{A_1, \dots, A_m\}$ (represents collateral premises) and D_1 is a sentence which can be inferred from $\{A_1, \dots, A_m, M\}$; $X_2 = \{B_1, \dots, B_k\}$ and D_2 is a sentence which can be inferred from $\{B_1, \dots, B_k, M\}$; $X_n = \{C_1, \dots, C_l\}$ and D_n is a sentence which can be inferred from $\{C_1, \dots, C_l, M\}$.

According to this approach, context is inherently represented as a part of the representation of meaning – instantiated by collateral premises which are members of sets $X_1, X_2 \dots X_n$. If IS is the value of IP for C, then to understand a sentence as uttered means to eliminate those sets from IP which do not include premises relevant in particular conversations. So, if $IP(M) = \{ \langle X_1, D_1 \rangle, \langle X_2, D_2 \rangle, \dots, \langle X_n, D_n \rangle \}$ and the premises that are taken for granted in the given contexts include all of X_1, \dots, X_i , but none of X_{i+1}, \dots, X_n , then the inferential significance is $IS(M) = \{ \langle X_1, D_1 \rangle, \dots, \langle X_i, D_i \rangle \}$.

If meaning is specified in this way, then the meaning of a sentence basically includes all contextual values. Because of that, the inferential view on meaning belongs to “rich meaning” approaches. Such a view on specification of meaning leads to a distinctive inferential view on semantic understanding. Contextual modulation is not seen as a deviation, but as a specification of meaning. To fully understand a sentence in general requires knowing all its possible significances. To understand a sentence as uttered requires singling out an appropriate subset of significances from its inferential potential.

Moreover, I believe that the distinction between the inferential potential and the inferential significance is well-suited to explain many phenomena that are typical for natural languages – including the cases of free pragmatic enrichment. According to contextualists (Recanati 2004), the meaning of sentences such as ‘The policeman stopped the car’ is modulated with regard to context in which the sentence is used. From the perspective of inferentialism, the meaning of sentences which are used as examples of free pragmatic enrichment can be specified standardly as their inferential potentials. What is explained by contextualism as different modulations of the meaning of a sentence can be explained by inferentialism as different contents that might be specified in particular contexts, i.e., as different inferential significances. The inferential explanation of the understanding of such sentences/utterances can be standard as well – to understand the content of an utterance requires the ability to single out the appropriate subset of the inferential potential. As I will try to show, the cases of ambiguity and conversational implicatures can be explained in a similar fashion.

References:

- Borg, E. (2004). Minimal Semantics.
- Borg, E. (2012). Pursuing Meaning.
- Brandom, R. (1994). Making It Explicit: Reasoning, Representing, and Discursive Commitment.
- Brandom, R. (2000). Articulating Reasons: An Introduction to Inferentialism.
- Carston, R. (2002). Thoughts and Utterances: The Pragmatics of Explicit Communication.
- Peregrin, J. (2014). Inferentialism: Why Rules Matter.
- Peregrin, J. (2018). Is inferentialism circular?
- Recanati, F. (2004). Literal Meaning.
- Recanati, F. (2010). Truth-Conditional Pragmatics.
- Travis, C. (1997). Pragmatics.

A Formal Impossibility Result for Methodological Individualism

Hein Duijf (Utrecht University), Allard Tamminga (University of Groningen) and Frederik Van De Putte (Ghent University)

In the philosophy of the social sciences, individualism is the methodological precept that any social phenomenon is to be explained ultimately in terms of the actions and interactions of individuals. Such a reduction includes bridge laws that translate social terms into individualistic statements that contain only individualistic terms. We focus on the notion of collective deontic admissibility, which is of central relevance to the study of collective responsibility and collective rationality. Using methods from modal logic, we show that there are no bridge laws that translate collective deontic admissibility into individualistic statements.

In the philosophy of the social sciences, individualism is the methodological precept that any social phenomenon is to be explained ultimately in terms of the actions and interactions of individuals. Ideally, such an explanation of a social phenomenon consists of a set of individualistic premises that describe the relevant actions and interactions of individuals, a set of bridge laws that define the relevant

social concepts in terms of individualistic ones, and a conclusion that describes the social phenomenon to be explained. We focus on collective deontic admissibility statements of the form ‘group G of agents performs a deontically admissible group action’. The notion of collective deontic admissibility is of central importance to formal studies of forward-looking and backward-looking collective moral responsibility, and of collective rationality. We show that there are no bridge laws that define collective deontic admissibility in terms of individualistic concepts: it is impossible to define collective deontic admissibility in terms of an individualistic language that is built from an alethic possibility operator, operators for individual agency, and operators for individual deontic admissibility.

Without going into the technical details, we wish to provide an overview of the argument and proof of our impossibility result. Our main contribution is methodological, since we don’t think that our result refutes methodological individualism. That is, our formal results provides conditions under which methodological individualism can be *disproven*. To clarify this methodological contribution, let us outline the central ideas and sketch the proof of these results.

We systematically introduce two languages: an *individualistic language*, denoted by L_{ind} , and a *social language*, denoted by L_{soc} . Any sentence ϕ in the individualistic language is considered to be an individualistic term, while the social language can contain social (non-individualistic) terms. In our specific impossibility result, the individualistic language is a multi-modal logical language built from an alethic possibility operator, operators for individual agency, and operators for individual deontic admissibility; and the social language extends this individualistic language by including operators of the form \star_G that characterize collective deontic admissibility, one for every group G . The logical language is an extension of the theory of ‘seeing to it that’, which has been central to the formal study of actions, omissions, abilities, obligations, prohibitions, and permissions.

We study the meaning of the sentences in these languages. That is, we consider a *class of situations*, denoted by S , and explain, for each instance $S \in S$, what the *meaning* of the sentences is. Modal logicians (and logicians in general), standardly assume that the truth values determine the meaning of the sentence, or at least that different truth values entails different meanings. To assess the truth values of sentences one typically provides a model theory for sentences in the language. In our impossibility result, we rely on the class of deontic game models and use standard modal-logical conventions to recursively define the truth values of the sentences in the languages L_{soc} and L_{ind} . We write $S \models \phi$ when ϕ is true in S ; we let $S \not\models \phi$ denote that for every $S \in S$ it holds that $S \models \phi$; and we write $S \not\models \phi$ when it is not the case that $S \models \phi$.

The three impossibility results can now be stated as below. It may be helpful to note that the last theorem is implied by the other theorems separately. We will subsequently shed some light on our proof method.

Theorem 1. *Let G be a non-empty and non-singleton group of agents. Then there is no $\varphi \in L_{ind}$ such that $S \models \neg\varphi$ and $S \models \varphi \rightarrow \star_G$.*

Theorem 2. *Let G be a non-empty and non-singleton group of agents. Then there is no $\varphi \in L_{ind}$ such that $S \models \varphi$ and $S \models \star_G \rightarrow \varphi$.*

Theorem 3. *Let G be a non-empty and non-singleton group of agents. Then there is no $\varphi \in L_{ind}$ such that $S \models \star_G \leftrightarrow \varphi$.*

To prove these results, we rely on a notion of *individualistic similarity* that applies to situations in S . This notion plays the following crucial role: whenever two situations $S_1, S_2 \in S$ are individualistically similar then these two situations validate the same individualistic sentences. In other words, for any two individualistically similar situations there is no individualistic term that distinguishes between these situations. Our notion of individualistic similarity, coined *individualistic bisimilarity*, relies on a well-known notion from modal logic: bisimilarity.

The proof of the theorems then goes as follows:

- *Proof of Theorem 1.* Suppose there were an individualistic sentence φ such that the truth of φ entails collective deontic admissibility. Take any situation S in which φ holds. Hence, collective deontic admissibility obtains at S . We then construct another situation S' that is individualistically bisimilar to S but where collective deontic admissibility is violated. Since S and S' are individualistically bisimilar and φ is true at S , this entails that φ is true at S' . By assumption, this entails collective deontic admissibility in S' . Contradiction.
- *Proof of Theorem 2.* Suppose there were an individualistic sentence φ such that the truth of φ is entailed by collective deontic admissibility. Or, equivalently, the falsity of φ entails that collective deontic admissibility is violated. Take any situation S where φ does not obtain. Hence collective deontic admissibility does not obtain at S . We then construct another situation S'' that is individualistically bisimilar to S but where collective deontic admissibility obtains. Since S and S'' are individualistically bisimilar and φ is false at S , this entails that φ is false at S'' . By assumption, this entails collective deontic

admissibility does not obtain in S' . Contradiction.

- The reader can easily check that the parameters of this proof method are given in the first two steps: the individualistic language L_{ind} , the social language L_{soc} , the class of situations S ; and the meaning of sentences in situations from S . In this talk we investigate the methodological contribution by indicating how the theoretical precept of methodological individualism fares when these parameters are varied, and by presenting avenues for further research.

Why buy local?

Ben Ferguson (Vrije Universiteit Amsterdam) and Chris Thompson (UiT The Arctic University of Norway)

This paper discusses the moral arguments that speak in favour of three consumer options: buying local food, buying global (non-local) food, and buying global food while also purchasing carbon offsets to mitigate the environmental impact of food transportation. We argue that because the offsetting option allows one to provide economic benefits to the poorest food workers while also mitigating the environmental impact of food transportation it is morally superior to the alternatives.

Claims that we ought to ‘buy local’ are increasingly common. From food television and celebrity chefs (Williams 2017) to the news media (Naylor 2018) and even agricultural extension publications (Grubinger 2010), everyone seems to be advocating for local food. We are told that locavorism—the practice of buying and consuming local food products—is healthier, tastier, and safer than consuming non-local alternatives and that it has environmental, economic, and social benefits (Schnell 2013; Cleveland et al. 2015). Michael Pollan argues that we should “think global, [but] buy local” and even that we should, “given the choice, buy local over organic” (Weinraub 2006). Although some of the reasons for buying local are primarily prudential, advocates also present arguments for the practice that frame it as an ethically significant choice. But are there really good ethical reasons to buy local?

Of course, any answer to this question depends on what we mean by ‘local food’ and ‘ethically significant’, as well as on the alternative choices we have available. We understand ‘local food’ to mean foods, generally perishable, that are produced a limited number (less than 400) of ‘food miles’ from the point of purchase (Hinrichs 2015). And we will construe the claim that buying this food is ethically

significant as the claim that it is morally better to purchase these food items than it is to purchase non-local alternatives. Finally, we will narrow the scope of the question by focusing on three salient choices.

Consumers could buy local by selecting food items that are produced 400 miles or less from the store. Alternatively, they could buy global by purchasing perishable food items produced more than 400 miles from the store. And finally, we want to consider an option that addresses perhaps the most pressing moral reason for locavorism in a different way: consumers could offset by buying globally and purchasing carbon offsets to mitigate the carbon emitted in the transport of their basket. The initial question, then, can be recast as ‘is it morally better to buy local, buy global, or offset?’

We argue that the strongest ethical reasons for locavorism are the ‘social’ benefits; that is, the environmental, economic, and community-based benefits. However, we argue that the community-based benefits do not track the locality of food production, but rather whether one buys from local sellers. The economic benefits to the local community are undeniable; however, these may come at the cost of economic benefits to relatively poorer food producers. Finally, although the environmental cost of food miles is insignificant compared to how food is produced reducing food miles will (often) reduce carbon emissions. The strongest ethical reasons for buying globally, on the other hand, are the benefits that this economic activity may bring to poorer food workers.

Economic and welfarist considerations seem to point to buying food from poorer producers, while environmental considerations speak in favour of buying locally and limiting food miles. We argue that by buying ‘globally’ from poor producers and offsetting the carbon emissions involved in the transport of this food, both benefits can be realised. Thus, offsetting is morally superior to the alternatives.

In the remainder of the paper we consider four objections to this argument. The first set of objections attack the considerations that we have argued speak in favour of the offsetting option. One objection appeals to special, or agent- relative, obligations in order to deny that it is morally better to help poorer workers rather than local workers. The other denies that environmental harms can be compensated by carbon offsets, in particular, because environmental harms are non-compensable harms. The second set of objections claims that we have not presented locavore argument in its strongest form. The first one argues that we omit a significant moral reason for being a locavore: locavorism disrupts the monopolistic power and corrupt practices of ‘big agriculture’. The other claims that by setting aside ‘merely’ personal considerations like taste and nutrition and focusing on considerations we label ‘social’, we miss the most valuable considerations that support the locavore movement.

References:

Bureau of Labour Statistics. (2014). Occupational employment statistics for fishing farming and forestry. United States Department of Labour.

USDA, NASS Pacific Region. (2013). Agricultural Overview. California Agricultural Statistics Report.

Cleveland, D., A. Carruth, and D. Mazaroli. (2015). Operationalizing local food: goals, actions, and indicators for alternative food systems. *Agriculture and Human Values*, 32: 281–297.

Ferguson, B. (2016). The Paradox of Exploitation, *Erkenntnis*, 81(5): 951–972.
Ferguson, B. and Ostmann, F. (2018). Sweatshops and Consumer Choices, *Economics and Philosophy*, forthcoming.

Grubinger, V. (2010). Ten reasons to buy local food, University of Vermont Vegetable and Berry Programme, <<https://www.uvm.edu/vtvegandberry/factsheets/buylocal.html>>.

United States Department of Health and Human Services. (2014). Poverty guidelines for the 48 contiguous states and the District of Columbia. Government Report.

Hinrichs, C. (2015). Fixing food with ideas of ‘local’ and ‘place’. *Journal of Environmental Studies and Sciences*, in press.

Hendrickson, J. (1997). Energy Use in the US Food System: A Summary of Existing Research and Analysis, *Sustainable Farming*, 7(4).

Hughes, D., C. Brown, S. Miller, and T. McConnell. (2008). Evaluating the Economic Impact of Farmers’ Markets Using an Opportunity Cost Framework. *Journal of Agricultural and Applied Economics*, 40(1): 253–265.

Naylor, T. (2018). Hyperlocal heroes: meet the chefs growing their own, *Guardian*, 25 January 2018, <<https://www.theguardian.com/lifeandstyle/2018/jan/25/hyperlocal-chefs-food-and-drink-restaurants-tony-naylor>>.

Otto, D. and T. Varner. (2005). Consumers, Vendors, and the Economic Importance of Iowa Farmers’ Markets: An Economic Impact Survey Analysis. Iowa State University, Leopold Center for Sustainable Agriculture Report.

Pretty, J. and Ball, A. (2001). Agricultural Influences on Carbon Emissions and Sequestration: A Review of Evidence and the Emerging Trading Options, Centre for Environment and Society Occasional Paper, University of Essex.

Pollan, M. (2006). Eat your View, New York Times, May 17, 2006, <<https://michaelpollan.com/articles-archive/eat-your-view/>>.

Schnell, S. (2014). Food miles, local eating, and community supported agri- culture: putting local food in its place. *Agriculture and Human Values*, 30: 615–628.

Singer, P. and Mason, J. (2006). *The Ethics of What We Eat*, Melbourne: Text Publishing.

Weinraub, J. (2006). Q&A with Michael Pollan: Think Global, Eat Local, Washington Post 28 June 2006, <<https://michaelpollan.com/interviews/qa-with-michael-pollan-think-global-eat-local/>>.

Webber, C. and Matthews, H. (2008). Food-miles and the relative climate impacts of food choices in the United States. *Environmental Science and Technology* 42(): 3508–3513.

Williams, G. (2017). Noma's taste of tomorrow: creating the future of food, Wired, 22 February 2017, <<https://www.wired.co.uk/article/inside-noma-copenhagen>>.

To be or not to be... sets!

Michele Ginammi (Politecnico di Milano)

I discuss Steinhart's argument against Benacerraf's famous multiple reductions argument to the effect that numbers cannot be sets. Steinhart offers a mathematical argument according to which there is only one series of sets to which the natural numbers can be reduced (namely, the finite von Neumann ordinals), and thus attacks Benacerraf's assumption that there are multiple reductions of numbers to sets. I will show that Steinhart's argument is problematic and should not be accepted.

In his (1965) paper, Paul Benacerraf famously put forward an argument against metaphysical reduction of numbers to sets: since there are multiple ways (all equally good) to reduce numbers to sets, the (supposedly unique) reduction fails. There have been several attempts to reject Benacerraf's conclusion that numbers cannot be sets, but — as far as I know — there has been only one attempt to challenge the common view according to which *there are* multiple series of sets which numbers can be reduced to. In a (2002) paper, Eric Steinhart claims to have a «detailed mathematical demonstration that 0 is $\{ \}$ and for every natural number n , n is the set of all natural numbers less than n », which implies that «Natural numbers are sets. They are the finite von Neumann ordinals» (p. 343). Even more,

he claims to have a «mathematical demonstration that if \mathcal{N} is any progression that satisfies Benacerraf's conditions, then \mathcal{N} is the finite von Neumann ordinals» (p. 343).

Here is the proof, in short. Let us assume that $\mathcal{N} = (\mathcal{N}, f, e, <)$ is our natural numbers (NN), where \mathcal{N} is a set of sets, f is a 1-1 function from \mathcal{N} to \mathcal{N} , e is a particular set belonging to \mathcal{N} , $<$ is a set-theoretic relation (the idea is, \mathcal{N} is the set of natural numbers, f is the successor function, e is 0 and $<$ is the 'less than' relation). Since $\mathcal{N} = \text{NN}$, then \mathcal{N} satisfies both the arithmetic and cardinality conditions. Following Benacerraf, Steinhart takes these conditions to be both sufficient and necessary to identify the NN. The cardinality condition consists, according to Steinhart, of two parts: an existential rule (C1), and a definition (C2):

- (C1) for any x , if $x \in \mathcal{N}$ then there exists a set $x^* = \{z \in \mathcal{N} \mid z < x\}$ (i.e., for every x , there exists the set of all its predecessors);
- (C2) the cardinality of a set S is x iff there exists a 1-1 correspondence between S and x^* .

From condition C1, we can form the set of all the x^* . Let us call it $\mathcal{N}^* = \{x^* \mid x \in \mathcal{N}\}$. Conditions C1 and C2 also defines other two functions: a function K from \mathcal{N} to \mathcal{N}^* , and a function C from any numerable set to \mathcal{N} . By combining these functions and the other functions in \mathcal{N} , we can define a new structure $\mathcal{N}^* = (\mathcal{N}^*, f^*, e^*, <^*)$. \mathcal{N}^* is implied by the NN-conditions. Now, according to Steinhart, everything which is necessary to spell out the metaphysically relevant conditions for being a natural number is metaphysically relevant as well, and therefore must be included in the NN universe (let us call this assumption NNUA). Since \mathcal{N}^* has been spelled out only by means of metaphysically relevant elements of the NN-universe, \mathcal{N}^* must be in the NN universe as well. But \mathcal{N}^* also satisfies the NN-conditions. Therefore we have that, if $\mathcal{N} = \text{NN}$, then $\mathcal{N}^* = \text{NN}$, which implies that $\mathcal{N} = \mathcal{N}^*$. Since this is true *iff* \mathcal{N} is the FVNO, then NN can only be the FVNO.

My rejection of Steinhart's proof is based on the following two considerations. (1) Steinhart assumption NNUA is in need of a justification. If we cannot find a justification for NNUA, then we have no reason to include \mathcal{N}^* in the NN-universe, and the proof fails. (2) Steinhart's conclusion crucially relies on a particular definition of cardinality, which involves that, for any number n , the set of its predecessors n^* exists. This definition of cardinality is quite standard, but it is not the only one available. Another way to define cardinality is through the notion of "transitive closure of a set": we say that the cardinality of a set M is m iff there is a one-to-one correspondence between M and the transitive closure of the set corresponding to $m \in 1$. This definition does not force us to construe the set \mathcal{N}^* . As a consequence, the identification of \mathcal{N} with \mathcal{N}^* fails, and so does the proof.

My take on Steinhart's proof is that it is not conclusive. One might still insist that,

even if the proof is not conclusive, we should accept mathematical practice for what it says, and simply go for the von Neumann's ordinals as the natural numbers. This amounts to adopting some version of Naturalism. However, such a naturalism would be so strong to violate the basic principles of Naturalism itself (i.e.: listen to mathematicians!): not all mathematicians adopt FVNO, some of them still prefer FZO (see for example Moschovakis 1994). Even more important, often mathematicians explicitly say that set theoretic reductions of numbers should not be intended as metaphysically loaded, but simply as representatively convenient for their purposes.

References:

- Benacerraf, P. (1965), 'What numbers could not be', *Philosophical Review* 74, 47–73. Now in (Benacerraf & Putnam 1983).
- Benacerraf, P. & Putnam, H., eds (1983), *Philosophy of Mathematics*, 2nd edn, Cambridge University Press, Cambridge (Mass.).
- Moschovakis, Y. N. (1994), *Notes on Set Theory*, Springer-Verlag, New York.
- Steinhart, E. (2002), 'Why numbers are sets', *Synthese* 133(3), 343–361.

Nudging and decisional privacy

Eva Groen-Reijman (Wageningen University)

Since its introduction, the concept of nudging has raised questions of legitimacy. Rather than considering nudging as a form of manipulation, this paper distinguishes between different nudges by focusing on the relationship between the nudger and the person nudged. It does so by presenting a view of decisional privacy - the dimension of privacy that concerns the right to make one's own decisions - that recognizes privacy norms as context-dependent. It thus addresses a question that so far has been undertheorized in the nudging debate: how the legitimacy of a nudge may depend, in part, on who does the nudging.

From Fangirl back to Philosopher: Analysing the Work of Mme de Staël

Eveline Groot (Erasmus University Rotterdam)

An investigation of the historical reception of Mme de Staël's theoretical work, through a comparison of two commentaries, reveals a remarkable contrast of views with regard to the status of this author. Whereas Underwood describes De Staël as a 'Rousseauistic fangirl', Jacquinet addresses her as a female genius and enlightened predecessor. Relating the two positions to De Staël's critique on Rousseau demonstrates that a pivotal question in De Staël's work concerns the relation between sentimentality and rationality. In this presentation, I hope to make clear some of the links we may observe between her epistemology, anthropology, and view on morality.

Recently, there has been a major revival of interest in the life and work of Anne Louise Germaine de Staël-Holstein, commonly known as Mme de Staël (1766–1817). Over the last decades, extensive biographies have been written on her life, politics, literary influence, and correspondences (Du Plessix Gray: 2008, Fontana: 2016). Yet, despite her invaluable intellectual contributions, De Staël has generally been overlooked as a *philosophe*, and up-to-date research on her philosophical thought is severely lacking (Craiutu: 2010, Fontana: 2016, McDonald: 1998). This is particularly remarkable in light of the rediscovery of 18th century liberal thinking – to which De Staël was an influential contributor – but it also marks an evident lacuna within the history of philosophy itself.

Currently, scholars from a variety of backgrounds are advocating a revision of the philosophical canon, so as to provide a more inclusive account of the history of philosophy. This brings into focus so-called minor figures, non-western traditions, and female philosophers. In line with this urgently required revision, the present research proposal contributes to the emerging debate by acknowledging De Staël's position in modern, Western, intellectual history, and the importance of her substantive philosophical work on sentimentalism.

As one of the first cosmopolitans, a host to important salons in Paris, a first-hand witness of the horrors of the French Revolution, and a politically involved author, Mme de Staël is a unique figure in the early modern world. Intellectually, she was influenced by a broad range of traditions, most importantly by the late French and Scottish Enlightenment, 18th century sentimentalism, and Early German Romanticism. These intellectual movements, however, embrace different perspectives on the concept of man. My view is that De Staël's philosophical work and political views can be regarded as an original attempt to unite the Enlightened and Romantic traditions.

Within the historical reception of the theoretical work of Mme de Staël, a comparison of two exemplary commentaries, an introduction to an edition of selected works of De Staël by P. Jacquinet, and a paper by G.A. Underwood, will offer a profoundly contrasting set of analyses. In 1893, Jacquinet mentions in the introduction to De Staël's *Extraits de ses oeuvres*, that she is "one of the brightest women in the history of French literature", as well as "an enlightened predecessor, an active and powerful initiator, who seems to surpass the forces of the feminine". Moreover, she is notably addressed as a philosopher, specialized in moral thought.

In 1915, George Underwood describes De Staël as a fangirl, making the following statements about De Staël's adaptation of Rousseau: "these first efforts [*Lettres sur les ouvrages et le caractère de J.-J. Rousseau* (1789) and *Essai sur les fictions* (1795)] show little beyond an enthusiasm for Rousseau's ideas" and "Rousseau leads Mme de Staël to become absorbed in her feelings". Although Underwood refers solely to De Staël's early work, while Jacquinet takes the whole oeuvre into consideration, both commentaries discuss the relation of De Staël to Rousseau from opposing stances.

Underwood has rightfully described the *Lettres* as a work of enthusiasm. De Staël was inspired by Rousseau on many levels. Yet this work is more than a superficial eulogy of her master. De Staël emphasises the importance of 'Rousseau the sentimentalist'. However, this does not entail that 'Rousseau the rationalist' is secluded. For her, *sensibilité* is the foundation that awakens openness for internalising abstract ideas, and thus gaining knowledge, and it evokes political open-mindedness and virtuous behaviour. Moreover, Jacquinet explains that contrary to Rousseau, De Staël has not sought a necessary foundation of morality within the realm of the sentiments: "it is to the depths of our consciousness where the idea of duty is engraved", it is in the interplay between sentimentality and rationality that morality is to be found.

The comparison of these two commentaries leads to the pivotal question: how does De Staël relate the sentiments to the rational faculties? This relation, first addressed in the *Lettres*, is further developed in *De l'influence des passions* (1796). In this work, De Staël aims to develop a new way of thinking about immersive disruptive historical events while incorporating a theory of the sentiments. By working towards a naturalistic scientific model of politics that unites thinking about sentiments and reason, her model would be able to include political passions. De Staël was convinced that the unification of sentiments and reason would not only suit her theories on freedom of thought, politics and religion, and literary engagement, but might also be applied to political reality itself. A thorough analysis of De Staël's epistemology, anthropology, and morality, contributes to our knowledge of sentimentalism in the history of philosophy as well as would recognise the rightful reputation of De Staël as a philosopher.

References:

Primary sources:

Rousseau, J. J. (1762). *Émile, ou de l'Éducation*, La Haye: J. Néeulme.

De Staël, G. (1798). *Lettres sur les ouvrages et le caractère de J. J. Rousseau*, 2nd edition, Paris: Charles Pougens.

De Staël, G. (1789). *Letters on the works and character of J.J. Rousseau* (Trans. Unknown), London: G.G.J. and J. Robinson, Paternoster-Row.

De Staël-Holstein, G. (1796). *De l'influence des passions sur le bonheur des individus et des nations*, Tome 1–2, Lausanne: Jean Mourer.

De Staël-Holstein, G. (1798). *Treatise on the Influence of the Passions upon the Happiness of Individuals and of Nations*, George Cawthorn (Trans.), London:, British Library.

De Staël-Holstein, G. (1893). *Extraits de ses oeuvres*, introduction de P. Jacquinet, Paris: Berlin Frères, 1893.

Secondary sources:

Berges, S. (2015). On the Outskirts of the Canon: the Myth of the Lone Female Philosopher, and what to do about it?, *Metaphilosophy*, 46(3), pp. 380–397, LLC and John Wiley & Sons Ltd.

Craiutu, A. (2010). Flirting with republicanism: Mme de Staël's writings from the 1790s, *History of European Ideas*, 36(3), pp. 343–346, DOI: 10.1016/j.histeuroideas.2010.05.001.

Fontana, B. (2016). *Germaine de Staël – A Political Portrait*, Oxford and Princeton: Princeton University Press.

Guerlac, S. (2005). Writing the Nation (Mme de Staël), *French Forum*, 30(3), pp. 43–56, University of Pennsylvania Press Stable.

Lotterie, F. (2003). Une revanche de la «femme-auteur»? Madame de Staël disciple de Rousseau, *Romantisme*, 122, pp. 19–31, DOI: 10.3406/roman.2003.1218.

Marso, L. J. (2017). The stories of citizens: Rousseau, Montesquieu, and de Staël", *Challenge Enlightenment Reason*, 30(3), pp. 435–463, The University of Chicago Press Stable.

Martin, J. E. (2002). Nineteenth-Century German Literary Women's Reception of Madame de Staël, *Women in German Yearbook*, 18, pp. 133–157, University of Nebraska Press Stable.

Rosenblatt, H. (2007). On the Intellectual Sources of Laïcité: Rousseau, Constant, and the Debates about a National Religion, *French Politics, Society and Culture*, 25(3), pp. 1–18.

Solovieff, G. (2000). *Madame de Staël – Selected Correspondences*, K. Jameson-Cemper

(Trans. & Ed.), Dordrecht: Springer Science+Business Media.

Underwood, G. A. (1915). Rousseauism in two early works of Mme de Staël, *Modern Philology*, Vol. 13(7), pp. 417–432.

Knowledge of Disability and Disabled People's Knowledge

Caroline Harnacke (Tilburg University)

To judge whether to prevent the birth of a disabled child, what the limits of gene modification are (if any), and what society owes to disabled people, we need an understanding of what it means to have a disability. Bioethics, and applied ethics in general, require not only moral principles and empirical facts, but also knowledge about specific experiences. I want to find out how to acquire knowledge of such experiences and how this knowledge relates to ethical and anthropological questions. Specifically, I will argue that disabled people have a duty to share their personal experiences of their situation.

To judge whether to prevent the birth of a disabled child, what the limits of gene modification are (if any), and what society owes to disabled people, we need an understanding of what it means to have a disability. Bioethics, and applied ethics in general, require not only moral principles and empirical facts, but also knowledge about specific experiences. I want to find out how to acquire knowledge of such experiences and how this knowledge relates to ethical and anthropological questions. Specifically, I will argue in this talk that disabled people have a duty to share their personal experiences of their situation.

I assume that even though disabilities are diverse – think of a wheelchair user, a blind person, a person with autism or cognitive impairments – different experiences of disabilities have enough in common to make it worth talking about the experience of disability as a kind rather than about the experience of individual disabilities or individual persons. Disability experience, in my view, is the experience of living with an atypical structure of function and the experience of stigma or discrimination.

First, I argue that society can do better for disabled people. Independent from your preferred model of disability we can agree that society is not adapted for disabled people in a way society should or could easily be. Note that this does neither imply that all disability-friendly changes are a matter of justice, nor that a non-disabling environment for all is possible, nor that all disability-related disadvantages can be addressed by societal change. Some of the problems society should take on could be addressed relatively easily, such as ensuring accessibility of buildings for people

with a physical impairment. Other problems are more difficult to take on, such as the integration of children with cognitive impairments in schools. But in all such cases, we need to know what living with a disability is like.

Second, I argue that disabled people tend to think differently about what living with a disability is like in comparison with non-disabled people. I will look at the extensive body of research which shows that non-disabled people expect living with a disability to be much worse than people with the disability in question think. Disabled people themselves judge their own well-being much more positively than non-disabled people making a judgment about a disabled person's well-being. This is striking. I argue that that this dissonance cannot (fully) be explained by reference to adaptive preference. Rather, the non-disabled perspective on disabled lives is at least partly inadequate and can be attributed to stereotypes and prejudices. As unfamiliarity with disability is one of the main reasons for thinking negatively about disability, research suggests that engaging with disability and disabled people is a way to change this view.

Third, I argue that we can only know what it is like to be disabled by listening to the testimony of disabled people. Accessing other minds to receive knowledge of inside-experiences is generally difficult, but it is especially difficult in cases of disability. The situation of disability, I will argue, is a situation of epistemic inaccessibility for a non-disabled person: non-disabled persons can, in general, not imagine what it is like to be disabled. Testimonies have therefore a value that cannot be replaced. I will discuss some criticisms directed against trusting testimonies of disabled people.

I conclude that disabled persons have to share their experiences of their situation as disabled persons, or their testimonies, to bring about change in society. In the same way, but less disputed, non-disabled persons have a duty to gather relevant information. Disabled persons have to take non-disabled persons seriously with regard to the difference in their epistemic standpoints, and hence share their experiences of their situation. I will specify how I understand this conclusion by sketching the scope of the responsibility of disabled people and by discussing a few examples.

Last, I will address a possible criticism of my position by discussing the concept of epistemic exploitation. Epistemic exploitation occurs when privileged persons compel marginalized persons to educate them about the nature of their oppression. I recognize that the epistemic work I demand might indeed be hard to do, but I show some possibilities to make it at least less hard.

Relevant Alternatives and Missed Clues: Redux

Peter Hawke (University of Amsterdam)

I re-evaluate the challenge posed to Relevant Alternatives Theory (RAT) by the *missed clue counter-examples* of Schaffer [2002]. The import of the challenge, I argue, has been misconstrued and underestimated, not least because Schaffer's specific argument invites distracting objections. But more forceful and precise arguments are in the neighbourhood. I offer a novel formalization of RAT that accommodates a suitably wide class of concrete RA theories. Then, I introduce the notion of an *abstract missed clue counter-example*, and prove that every RA theory, as formalized, admits such a counter-example. This forms a precise argument in the *spirit* of Schaffer's argument, but one that resists easy dismissal: to respond, the RA theorist must embrace an error theory and oppose our intuitive judgments about missed clue cases.

I re-evaluate the challenge posed to Relevant Alternatives Theory (RAT) by the *missed clue counter-examples* of Schaffer [2002], in light of RAT's enduring appeal [Pritchard, 2010, Lawlor, 2013, Bradley, 2014]. The import of the challenge, I argue, has been misconstrued and underestimated, not least because Schaffer's specific argument invites distracting objections. But more forceful and precise arguments are in the neighbourhood.

An RA theorist holds that the truth of 'a knows that ϕ at time t ' does not entail that a has sufficient evidence to eliminate every alternative to ϕ - only every relevant alternative. Thus broadly construed, RAT corresponds to a family of concrete theories, ranging over possible explications of (i) alternative, (ii) (the criteria of) relevance, (iii) evidence and (iv) elimination. RA theorists flesh out the criteria for relevance in diverse (combinations of) ways: psychological salience of an alternative to either the attributor or subject of a knowledge ascription; resemblance to actuality; presupposition; conversational relevance to the question or topic under discussion; compatibility with the agent's beliefs; and practical stakes.

RAT's appeal lies in joint accommodation of *epistemic modesty* and *epistemic fallibilism*. Modesty counsels that a theorist is remiss to exaggerate our ordinary epistemic powers: in particular, ordinary evidence does not rule out radical skeptical possibilities. The fallibilist, however, posits that whatever epistemic status we express with ordinary knowledge ascriptions, it is compatible with certain possibilities of error.

Schaffer [2002] flags missed clue cases as a universal source of counter-examples to RAT. A *Schafferian missed clue counter-example* (SMC) has the following structure: agent a knows P , where proposition P is a 'clue' that Q is the case. That is, P *objectively indicates* Q , in the modally robust sense that the possibility that P but not Q is a highly abnormal one, holding only in possible worlds that are 'far removed'

from actuality. However, by stipulation, *a* has no (subjective) reason to believe that *Q* is the case. In particular, she has no (subjective) reason to believe that *P* objectively indicates *Q*.

Intuitively, *a* is not positioned to know *Q*: if she believes *Q* on the basis of knowing *P*, then she not only lacks knowledge of *Q*, but is irrational. The clue is *discerned* but not *appreciated*. RAT's difficulty, Schaffer [2002] claims, is that it contradicts this verdict. Since the possibility that *P* but not *Q* is distant, it is akin to a skeptical one. So RAT must count it as irrelevant. The only other alternatives to *Q* are those where neither *P* nor *Q* holds, but, where relevant, such possibilities are eliminated by the evidence: if not, *a* would not know *P* after all. By RA lights, *a* is positioned to know *Q*. Or so judges Schaffer [2002].

The dialectic is puzzling. The stated aim of Schaffer [2002] is to locate a *universal* source of counter-examples to RAT. But SMCs rely on a specific criterion of relevance: if *P* objectively indicates *Q*, then any possibility that *P* holds without *Q* is irrelevant. Nothing in RAT's basic statement commits one to this. Suffice to say, focus on SMCs is unnecessarily narrow, inviting the all-too-easy response of Brueckner [2003] and Black [2003]: Schaffer's considerations might well imperil some RA theories, but leave nuanced RA theories untouched.

The deeper import of missed clue cases is better detected when focusing on the *core structure* of SMCs - abstracting from specific criteria of relevance and other narrow constraints. Or so I argue. First, I render RAT with precision. My formalization studiously avoids commitment to specific criteria of relevance. Second, I return to SMCs and locate a core abstract structure, introducing the notion of an *abstract missed clue counter-example* (AMC). Third, I prove that for every possible RA theory (as formalized) there exists an AMC. This forms the basis for an abstract argument along Schafferian lines. The Schafferian worry is vindicated: suitably generalized, missed clue cases provide a universal source of trouble for RAT. AMCs are counter-examples in a precise sense, violating a formal 'closure' principle (for conditional knowability claims) that follows from two intuitive principles: *No Vacuous Knowledge* (NoVK) and *Minimal Connection* (MinC). The former says that empirical knowledge requires evidence. The latter says that learning *P* is a route to knowing *Q* only if it is knowable that *P* materially implies *Q* (i.e. it's not that *P* holds yet *Q* doesn't). The challenge of missed clues is hereby assimilated to RAT's challenging relationship with intuitive closure principles [Dretske, 1970, Luper, 2016].

References:

T. Black. The Relevant Alternatives Theory and Missed Clues. *Australasian Journal of Philosophy*, 81:96–106, 2003.

D. Bradley. A Relevant Alternatives Solution to the Bootstrapping and Self-Knowledge Problems. *Journal of Philosophy*, 111(7):379–393, 2014.

A. Brueckner. What Missed Clue Cases Show. *Analysis*, 63(4):303–305, 2003.

F. I. Dretske. Epistemic Operators. *Journal of Philosophy*, 67(24):1007–1023, 1970.

K. Lawlor. *Assurance: An Austinian View of Knowledge Claims*. Oxford University Press, 2013.

S. Luper. Epistemic closure. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2016 edition, 2016.

D. Pritchard. Relevant Alternatives, Perceptual Knowledge and Discrimination. *Nous*, 44(2):245–268, 2010.

J. Schaffer. Knowledge, Relevant Alternatives and Missed Clues. *Analysis*, 61(3): 202–208, 2002.

Finding Its Place: Non-Ideal Political Philosophy and the Contemporary World

Colin Hickey (Utrecht University)

In this paper I aim to make progress understanding the project of so-called “non-ideal” political philosophy, such that it provides real value-added with respect to the task it sets itself. Motivated by worries about “ideal” theory, non-ideal theorists have tried to provide a countermeasure that is more directly aimed at providing normatively defensible action guidance, in our world. However, I suggest that for non-ideal theory to culminate the break from ideal theory, we still need a fairly radical reconception. Drawing on an extended analogy with bioethics, I suggest this will include a significant opening up of the field across disciplines.

How we should understand the project of doing so-called “non-ideal” political philosophy, in the world we inhabit, such that it provides real value-added with respect to the task it sets itself? In this paper, I try to make progress toward answering that question. I do so from a place of partial concern about the status quo in non-ideal theorizing. However, via an extended analogy to the current constitution of the field of bioethics, I also do so in hopeful anticipation of some clarity of purpose.

In the *First Section* of the paper, in order to set the contrast and see what “non-ideal” theorists have generally conceived themselves as responding to, I present an ecumenical account of “ideal” theory and survey a range of criticisms that have been lodged at ideal theory and motivated scholars to turn towards and focus on the “non-ideal” alternative. In ideal theory, scholars are variously focused on describing more utopian visions of perfect justice, or ideal end-states to aim at, or who employ, knowingly, implausible assumptions about compliance in order to generate answers about the principles of justice or who bears what duties of justice without simultaneously providing answers when such assumptions don’t hold. The common thread linking concerns about ideal theorizing is that its deliverances are too abstract or removed from everyday realities, and that they are therefore unable to offer concrete practical guidance in the world we find ourselves living. Likewise, detractors worry, partly in virtue of such problems, that political philosophy organized around the pursuit of ideal theory is doomed to be ignored or dismissed as irrelevant.

In response to such worries many political philosophers have moved towards non-ideal theory. In the *Second Section* of the paper I provide a preliminary characterization of non-ideal theorizing as a kind of counter-measure as it has taken shape in the literature. While the landscape of non-ideal theorizing is diverse and there is no obvious consensus in conception or methodology, under some description, non-ideal theorists are all concerned about providing normatively defensible action guidance, in the world, as it is now. Those engaged in non-ideal theorizing purport to be doing something different than “ideal” theorists. So non-ideal theorists have tried to recognize the inevitable reality of “partial compliance,” or they build in more “realistic” facts or “feasibility” constraints on our theorizing, or they have directed focus toward explicitly theorizing about the imminent transitional steps from where we are toward a more just alternative—attending more to improvements than destinations.

However, although we can start by conceiving of the project of non-ideal political philosophy in contrast to ideal theory (i.e. whatever it is, it is not *that!*), inevitably that doesn’t take us very far.

So in the *Third Section* and throughout the rest of the paper, I want to suggest that to culminate the break from ideal theory, we still need a fairly radical reconception of so-called non-ideal political philosophy rather than just ratcheting down compliance assumptions or building in feasibility constraints and otherwise going on largely as before.

In order to make good on its promise of (normatively justifiable) practical action guidance in the real world, I suggest that non-ideal political philosophy needs to significantly open itself up interdisciplinarily and conceive of itself as more than just belonging to the purview of academically trained philosophers (cf. Valentini, 2012).

Rather than being seen, or at worst confined, as a *sub-discipline* of political philosophy, the project of non-ideal political philosophy, if it is to be what it says it wants to be, should start seeing itself more along the lines of bioethics; as a *field* where “traditional” philosophy plays but one role among many.

Oriented around a range of shared ethical issues, bioethics *just is* the nexus of law, policy, clinical medicine, medical research, public health, philosophy, etc. It isn’t as if there is some pure moral philosophical project of bioethics which is then applied, seeded, or incorporated into such other realms. Those other realms just are bioethics; constitutive parts of it, anyway. And the normative justifications for concrete action guidance that have so successfully characterized the field have grown organically out of this eclectic hodgepodge of voices. In bioethics clinicians, researchers, hospital managers, public health and legal experts, policy wonks, philosophers, and more come together in robust conversation *with each other*.

I hope to make the case, given that the goal of non-ideal theorizing is supposed to be centered around locating normatively defensible practical guidance, that non-ideal political philosophy would do well to embrace and centrally engage with, as constitutive parts of a broader program, activists and organizers, peace-builders, policy makers, political scientists, investigative journalists, human rights organizations, lawyers, and the like. Contrary to some caricatures, philosophers don’t have a kind of special dominion over normative arguments, and folks in the fields I mentioned above operate with sophisticated orientations toward justice, equality, basic rights, and the like. But they also are often—whether by training, with respect to access to information, through lived experience on the ground, or otherwise—in significantly better epistemic positions to know the most important concrete priorities to provide action guidance towards justice, or to understand in practice the feasibility dynamics on the ground. Non-ideal political philosophers should seek out such skills and expertise and deeply engage from a place of humility and curiosity.

Spinoza’s Omne Esse and Certainty

Tomoko Higuchi (Erasmus University Rotterdam)

Spinoza demonstrates the existence of God in E1P11 from his nominal definition. Nonetheless, it is unclear why and how a being deduced from the nominal definition can be real. This presentation will show that the concept of God is what Spinoza calls *omne esse* (the total being) and this gives certainty to it. First, we will see the concept of *omne esse* in the *Treatise of Emendation of the Intellect* and the *Ethics*. Subsequently, referring to Spinoza’s theory of true idea, we will see that such an idea cannot be false since it has no exclusion.

Spinoza demonstrates the existence of God in E1P11 from his nominal definition, and since true ideas must correspond with their *ideata* (objects of the ideas) according to E1Ax6, it follows that God really exists. However, it is unclear why and how a concept of a being deduced from the nominal definition can be real. If this is allowed to anything, would it not imply that anything to which we give some definition and proof of existence could exist? This presentation will show that the concept of God constructed in the *Ethics* is what Spinoza calls *omne esse* (the total being) and that it is this which makes the readers certain of the existence of God. First, we will discuss the concept of *omne esse*, a unique and infinite being in the *Treatise of Emendation of the Intellect* and the *Ethics*. Subsequently, referring to Spinoza's theory of true and false idea, we will see that such an idea cannot be false since it has no exclusion.

My aim is to prove that the concept of what Spinoza calls God shows the real existence even though deduced from a nominal definition. The definitions which Spinoza employs to start the demonstration should be nominal, in other words, explain the terms' meanings. According to Letter 9, they should not be real definitions expressing the nature of the defined things. Moreover, we will be faced with an infinite regress, if we try to begin with the real definitions. For they require a demonstration showing why they are true, and this demonstration starts with the definition, and so *ad infinitum*. Yet the question is how a being deduced from nominal definitions can be real. Since nominal definitions signify how the author uses the defined terms, they can be arbitrary. Some might insist that anything can be provided with a definition and demonstrated as existing. Yet, it is only God in Spinoza's sense, which is described as *omne esse* that can be shown in such a manner.

From the earliest period, Spinoza maintains that we need to start with the source and the origin of Nature, which cannot be wrong. Spinoza describes such a being as "unique and infinite, that is, *omne esse*, and beyond which there is nothing." The concept of a unique and infinite being appears also in the *Ethics*. According to E1P14C, "God is unique, i.e., that in Nature there is only one substance, and that it is absolutely infinite." Thus, seeking *omne esse* has always been a major theme of Spinoza's philosophy.

Omne esse is of significant importance in this discussion because this concept excludes the possibility of being false. To see this, it is necessary to examine the difference in the theory of true and false ideas between Descartes and Spinoza. According to Descartes, the cause of formal falsity is the gap between the intellect and the will, and material falsity occurs when the idea represents a non-existent being. By contrast, Spinoza, refuting material and formal falsity, maintains a simple account of the falsity: "Falsity consists in the privation of knowledge which inadequate, or mutilated and confused, ideas involve." As to the true idea, Spinoza

distinguishes the intrinsic and extrinsic denomination of a true idea, namely adequacy and agreement. For instance, the idea of an existing building is true because it agrees with the object, and the idea of an orderly conceived building is adequate, regardless of the fact whether it exists or not.

The concept of God constructed through the demonstration satisfies both criteria. As to the agreement, it must correspond with the *ideatum*, by E1Ax6. As to the adequacy, there is no possibility for it to be inadequate since it is *omne esse*, which does not possess ‘the outside’ that inadequate ideas necessarily own. For we cannot find anything that lacks in God; if we attempt to seek what lacks in God, there are only three options, substance, attributes, and modes. First, since we cannot conceive another substance separately, it is impossible that a substance be absent from the cognition of God, secondly since God has infinite number of attributes, God cannot be deficient in any attribute, and lastly any mode is an affection of an attribute, it necessarily refers to an attribute and God. Hence, we have a true and adequate idea of God, which means we have the idea of the idea of God. In TIE and the *Ethics*, Spinoza considers the certainty as the idea of the idea, i.e., that the one who possesses an adequate idea is certain of it at the same time because Spinoza regards an idea not as a mute image on a tableau, but as cognition involving volition, namely affirmation or negation. Therefore, we are necessarily certain of the idea of God obtained through the demonstration.

References:

Primary Literature

Descartes, R. (1897-1909). *Œuvres de Descartes*. (C. Adam & P. Tannery, Ed.). Paris: Cerf.

Descartes, R. (1984). *The Philosophical Writings of Descartes*. (Vol. 1-3) (J. Cottingham, R. Stoothoff & D. Murdoch, Trans.). New York. NY: Cambridge University Press.

Spinoza, B. d. (1925). *BARUCH DE SPINOZA OPERA*. (C. Gebhardt, Ed.). Heidelberg: Carl Winters.

Spinoza, B. d. (1985). *The collected works of Spinoza: Volume I* (E. Curley, Trans.). Princeton, NJ: Princeton University Press.

Spinoza, B. d. (2002). *Spinoza: complete works* (S. Shirley, Trans.). Cambridge: Hackett Publishing Company.

Spinoza, B. d. (2016). *The collected works of Spinoza: Volume II* (E. Curley, Trans.). Princeton, NJ: Princeton University Press.

Secondary Literature

Donagan, A. (1988). Spinoza. Harvester Wheatsheaf.

Garrett, D. (2018). Representation, Misrepresentation, and Error in Spinoza's Philosophy of Mind. In *The Oxford Handbook of Spinoza*. (Ed. M. Della Rocca). Oxford university press.

Glezier, M. (2017). *Verité et certitude chez Spinoza*. Paris :Garnier

Gueroult, M. (1974). *Spinoza L' Ame (Ethique, II)*. Paris: Aubier-Montaigne.

Matheron, A. (1989). Idée, idée d'idée et certitude dans le Tractatus de Intellectus Emendatione et dans l'Ethique. In *Groupe de recherches spinozistes : Methode et Metaphysique*. Paris : Press de l'université de paris sorbonne.

A Historical Challenge for Moral Realism

Jeroen Hopster (Utrecht University)

Michael Huemer (2016) has recently presented empirical evidence in support of the thesis that over the course of human history, there has been a global shift in moral values towards a broadly 'liberal' orientation. Huemer argues that this shift better accords with a realist than an antirealist metaethics: it is best explained by the discovery of mind-independent truths through intuition. I argue, contra Huemer, that the historical data are better explained when assuming an antirealist metaethics. Huemer's realist view does not fit the data as well as he suggests, whereas antirealists have underappreciated resources to explain the relevant historical dynamics.

Are empirical and genealogical findings relevant for evaluating the plausibility of metaethical theories? In the recent metaethical literature there has been a specific interest in the tenability of moral realism in the light of morality's evolutionary origins (e.g. Street 2006; Joyce 2006; Kitcher 2011). For instance, Sharon Street's 'Darwinian Dilemma' (2006) departs from the premise that the contents of our normative judgments have been heavily influenced by evolutionary forces; she argues that this *datum* makes a realist metaethics highly implausible.

Realists have fashioned various kinds of response to Street's dilemma. One prominent line of reply is to question whether her evolutionary challenge can be substantiated on empirical grounds (e.g. Shafer-Landau 2012; FitzPatrick 2015; Huemer 2016). Arguably, the evolutionary influence on the contents of many of our normative judgments is relatively minor, and can easily be accounted for in

realist terms.

I begin this presentation by concurring that there are good reasons to resist Street's evolutionary thesis. The contents of many commonplace normative judgments cannot obviously explained in evolutionary terms. Consider moral judgments that have only gained widespread adherence over the last few centuries, such as the judgment that men and women deserve equal treatment, or that racism is morally wrong. For many judgments grounded in inclusionary moral values, the evolutionary component in explanations of their origins is likely to be minor (cf. Buchanan and Powell 2015).

I go on to argue, however, that several realists are still committed to a genealogical claim that can be contested on empirical grounds: namely, the claim that over the course of history we have tracked mind-independent moral truths. Hence, debunkers can appeal to historical considerations to offer an argument to the best explanation against moral realism.

In the latter part of the presentation I illustrate this 'historical debunking argument' by critically engaging with Michael Huemer's (2016) recent defense of realist intuitionism. Huemer presents empirical evidence in support of the thesis that over the course of human history, there has been a global shift in moral values towards a broadly 'liberal' orientation. He argues that this shift better accords with a realist than an antirealist metaethics: it is best explained by the discovery of mind-independent truths through intuition. I argue, *contra* Huemer, that the historical data are better explained when assuming an antirealist metaethics. Huemer's realist view does not fit the data as well as he suggests, whereas antirealists have underappreciated resources to explain the relevant historical dynamics. These resources include an appeal to socialization, to technological and economical convergences, to lessons learned from history, to changes induced by consistency reasoning and to the social function of moral norms in overcoming some of the cooperation problems that globalizing societies face.

My conclusion is twofold. First, an appeal to historical developments can help debunkers to overcome an essential weakness of Street's evolutionary challenge against realism. Second, Huemer is hoist with his own petard: his realist *explanans* has multiple shortcomings, whereas the antirealist's *explanans* has several explanatory virtues, and provides a superior account of the historical shift towards liberal values.

References :

Buchanan, A. and Powell, R. 2015. The Limits of Evolutionary Explanations of Morality and Their Implications for Moral Progress. *Ethics* 126: 37-67.

FitzPatrick, W. 2015. Debunking Evolutionary Debunking of Ethical Realism. *Philosophical Studies* 172: 883- 904.

Joyce, R. 2006. *The Evolution of Morality*. Cambridge: MIT Press. Huemer, M. 2016. A Liberal Realist Answer to Debunking Skeptics: The Empirical Case for Realism. *Philosophical Studies* 173: 1983-2010.

Kitcher, P. 2011. *The Ethical Project*. Cambridge: Harvard University Press.

Shafer-Landau, R. 2012. Evolutionary Debunking, Moral Realism and Moral Knowledge. *Journal of Ethics and Social Philosophy*, www.jesp.org, 7(1): 1-37.

Street, S. 2006. A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127: 109-166.

Brentano on Russell's Paradox

Carlo Ierna (Radboud University Nijmegen)

In 1909, prompted by the correspondence with his student Hugo Bergmann, Franz Brentano wrote a short treatise on Russell's Paradox. I will discuss Brentano's analysis and proposed solution of Russell's Paradox as well as the reception of his interpretation. Brentano moves several criticisms to Russell, both internal and external: how can something be an object and a class at the same time? How can "classes" be said to exist and literally have properties? At best we can only improperly predicate something of a class, which would make Russell's Paradox meaningless when taken literally.

Throughout 1908 Franz Brentano and his student Hugo Bergmann exchanged letters discussing the foundations of mathematics and the axioms of geometry. At the end of 1908, Brentano prompted Bergmann to look up "the philosophical-mathematical works by Russell", because they would be close in some respects to his own views. Bergmann did and early on in 1909 came back to Brentano with questions about Russell's paradox, proposing a solution, and inquiring after Brentano's own views on the matter. Brentano did not only send Bergmann a lengthy letter in return with comments on Bergmann's solution, but apparently also took the time to elaborate a short treatise containing his own critique and proposed solution of Russell's Paradox. This hitherto unknown and unpublished document can be found in the Prague Archives, hosting materials by Brentano's students Anton Marty and Oskar Kraus.

The treatise on Russell's Paradox was originally dictated by Brentano to his son Giovanni, due to his advancing blindness, but his student Oscar Kraus later also

prepared a typewritten transcription of it, adding footnotes and an introduction with a view to edit it. As far as I have been able to determine, the treatise nevertheless remained unpublished. In my talk, I will present Brentano's analysis and proposed solution of Russell's Paradox as well as the effects of his interpretation through Bergmann.

Indeed at the time Bergmann was collaborating with Benno Urbach on an article regarding the classical and modern paradoxes. This first attempted solution to Russell's Paradox mentioned in the article, matches a proposed solution by Bergmann which was criticized by Brentano in an unpublished letter. Bergmann acknowledged the cogency of the criticism in his answering letter and the "solution" is then consequently also rejected in the article.

Brentano moves several criticisms to Russell, both internal and external: what does Russell mean exactly with "class" and can a "class" be said to exist and literally have properties? How can something be an object as well as a class and be sub-ordered and supra-ordered to itself at the same time ("absurd that something should be supra-ordered to itself" / "*absurd, dass irgend etwas sich selber übergeordnet sein sollte*")? Brentano chides Russell for having confused "class" ("*Klasse*") and "class-concept" ("*Klassenbegriff*"): "Russell's argument suffers from a lack of clarity in its expressions and confusion of concepts. He speaks of a class "man": what does he mean with it? [...] Does he mean the class-concept "man"?" ("*Das Argument RUSSELLS leidet an Unklarheit der Ausdrücke und Konfusion der Begriffe. Er spricht von einer Klasse "Mensch": was meint er damit? [...] Versteht er darunter den Klassenbegriff "Mensch"?*")

However, Russell did distinguish class and class-concept in his 1903 *Philosophy of Mathematics*, so Brentano's comments are at least in part misguided. Yet, Brentano concludes that at best we can only improperly predicate something of a class, which would make Russell's Paradox meaningless when taken literally. According to Brentano. Russell's Paradox would be due to confusions from an ambiguous use of the terms "class", "class-concept", and "object of a class", to the point that he invokes the application of the scholastic doctrine of distinguishing various types of *suppositio* as a solution: you cannot have literally the same term in the same sense as subject and as predicate. If classes cannot really have properties, because they cannot be understood as objects on Brentano's account, then predications about classes are meaningless and have no truth-value, dissolving the paradox.

References:

Bergmann, Hugo, 1946, "Briefe Franz Brentanos an Hugo Bergmann" in *Philosophy and Phenomenological Research* 7 (1), 83–158.

Urbach, Benno, 1910, "Über das Wesen der logischen Paradoxa", in *Zeitschrift für Philosophie und philosophische Kritik* 140, 81–108.

Conveivability Arguments and the Overgeneration Problem

Savvas Ioannou (University of St Andrews, Scotland)

Chalmers (1996, 2011) argued that ideal primary positive conceivability entails possibility. He claimed that phenomenal zombies are conceivable this way and thus, they are possible. Therefore, physicalism about phenomenal properties is wrong. However, I will argue that this conceivability argument faces the overgeneration problem. That is, the kind of reasoning suggested by this conceivability argument can be used to claim that a variety of concepts (including waterish concepts) refer to non-physical entities. The proponent of the conceivability argument faces a dilemma. Either s/he continues using it and endorses an abundant property dualism or rejects it because of its absurd conclusions.

It has been argued that a higher-level concept refers to something physical, only if the facts concerning this concept can be entailed by the microphysical facts (Chalmers, 1996, 2011; Jackson, 1998; Chalmers & Jackson, 2001). For example, water is physical because if we know all the microphysical facts, we can entail all the facts concerning water.

However, Chalmers (1996, 2011) argued that this is not the case for phenomenal properties. Even if we knew all the microphysical facts, we wouldn't be able to know whether other actual or possible people have phenomenal experiences. This is shown by the conceivability argument. We can conceive of phenomenal zombies: beings that possess the same physical properties as us but lack our phenomenal experiences. Therefore, these zombies are possible. This is why physicalism about phenomenal properties is wrong and why we should accept property dualism about these properties.

Nevertheless, I believe that something similar can be said about waterish concepts, and as a result, the proponent of the conceivability argument faces the overgeneration problem. Imagine a waterish concept that does not imply any specific ontology. The primary intension of 'water*' picks out the dominant thing in our environment that satisfies our thirst. This concept is only used to talk about this causal power. It picks out something real. Still, mere analysis of this concept does not suffice to decide between physicalism or property dualism about the referred entity. Water* may emerge from H₂O, and H₂O may be correlated with water* only contingently. I can conceive of a world, in which there is H₂O but it does not satisfy our thirst when we drink it because there is no water*. Therefore,

water* seems to be something non-physical.

I can conceive of another world in which certain laws are true in certain locations, but false in other locations. In this world, sometimes, there is water* where there is H₂O, but sometimes there is not. Water* is absent but H₂O is present only in places that H₂O is never drunk by people. Consequently, there is not any situation in which H₂O is drunk by a person but there are no microphysical changes related with drinking water*. As a result, there can be two worlds with the same microphysical facts but different water* facts.

It seems to me that a priori ideal rational reflection allows more cases to be conceivable than Chalmers is comfortable with. If the aforementioned case about water* is ideally conceivable and can be generalised to other higher-level concepts, then the proponent of the conceivability argument faces a dilemma. Either s/he continues using it and endorses an abundant property dualism or rejects it because of its absurd conclusions.

I choose the latter because we have good theoretical reasons to reject this dualism. The theoretical desideratum of simplicity suggests identities in both cases. Water* is identical to H₂O, and phenomenal properties are identical to physical properties. All else equal, an ontologically simpler theory is to be preferred. It seems odd to accept such an inflated ontology just because we can imagine it.⁷

Even if someone bites the bullet and accepts abundant property dualism, then a similar problem appears. Similar considerations can lead us to property trinalism. Imagine that a linguistic community distinguishes between two different water concepts: 'water1' and 'water2'. A thing is called 'water1' if it has the causal power of making us feel less thirsty when we drink it. A thing is called 'water2' if it has the causal power of making us feel happy when we drink it. The people of this community believe that water2 supervenes on water1, and water1 supervenes on H₂O. They claim that these connections are contingent. There is a conceivable world in which H₂O exists but water1 and water2 do not exist, and there is a conceivable world in which H₂O and water1 exist but water2 does not exist. These worlds are logically possible, and hence property dualism and materialism about these entities are false. These properties are non-physical.

This kind of considerations can be used to claim that even more waterish entities exist. This leads to the conclusion that there is a different property for every different causal power related with phenomenal effects. I will reject this view because of its oddness. We have been led into a very bloated ontology just because

⁷ I agree with the a posteriori entailment view (Block & Stalnaker, 1999) that conceptual analysis is not required for reductive explanation. If there is no a priori entailment from microphysical truths to phenomenal truths, reductive explanation does not fail. I agree with Block & Stalnaker (1999) that a posteriori physicalism is true.

we can imagine it. Still, someone may bite even this bullet and accept this inflated ontology. There is nothing incoherent with it. Nevertheless, this ontology is in conflict with the aims of Chalmers. Chalmers does not want to reject physicalism about the majority of entities. He just thinks that there are special reasons in the case of consciousness to do so.

Another notion of conceivability may be able to reveal a difference between phenomenal concepts and water(ish) concepts. But in the absence of such a notion, there is not any good reason to claim that only phenomenal concepts refer to non-physical entities.

I believe that the conceivability argument does not help us to decide between physicalism and property dualism. This should be decided by examining other arguments for and against physicalism. The conceivability arguments just reveal our intuitions.

References:

Block, N., & Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *The Philosophical Review*, 108(1), 1-46.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.

Chalmers, D. J. (2011). Does Conceivability Entail Possibility?. In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and Possibility* (pp. 145-200). Oxford: Oxford University Press. Originally published in 2002.

Chalmers, D. J., & Jackson, F. (2001). Conceptual analysis and reductive explanation. *The Philosophical Review*, 110(3), 315-360.

Jackson, F. (1998). *From metaphysics to ethics: A defence of conceptual analysis*. Oxford University Press.

Uncertainty in Science: A Study on the Role of Non-Cognitive Values in the Assessment of Inductive Risk

Silvia Ivani (Tilburg University), Matteo Colombo (Tilburg University) and Leandra Bucher (University of Wuppertal)

Scientific research often involves uncertainty. In such cases, scientists may make

mistakes, such as accepting a hypothesis that is actually false. Philosophers call the chance of being wrong when assessing hypotheses inductive risk and argue that it shows the beneficial role of non-cognitive values in science. Scientists can legitimately consider non-cognitive values if a mistaken assessment of hypotheses involves bad consequences, such as killing people. In this paper, we investigate the relationship between reasoning, inductive risk, and non-cognitive values. We present an experimental study clarifying the impact of non-cognitive values on the assessment of hypotheses in cases of inductive risk.

Imagine the following scenario. The scientific community disagrees over the safety and efficacy of a new medicine. Scientists debate whether to introduce it into the market. While some researchers argue that the medicine is safe and effective and can be launched into the market, some scientists challenge this assessment, raising concerns over the risks involved in using the drug. The point of contention is the fact that the clinical trials conducted to test the medicine involved a sample of only men. While the advocates of the safety and efficacy of the drug think that the sample is representative of the general population and the data are then reliable, the critics claim that the evidence gathered does not inform us on the benefits and risks of the medicine for some groups of the population, such as women. Finally, the decision is made to introduce the medicine into the market. Do you think the decision was a good one?

Scientific research involves uncertainty. Scientists have to take decisions about methodologies and hypotheses and each one of these decisions involves uncertainty. Lack of sufficient evidence and disagreements about methodologies – as the example above shows - are sources of uncertainty that can introduce error in scientific reasoning. One kind of error is associated with the notion of inductive risk, i.e., the chance of taking wrong decisions, such as accepting a hypothesis that is in fact false. Philosophers argue that inductive risk challenges the ideal of value-free science, i.e., the idea that non-cognitive values (e.g. moral and economic values) do not influence research, and it shows their actual beneficial role in science (Hempel 1965; Douglas 2000). Specifically, considering non-cognitive values is beneficial when taking wrong decisions may involve non-cognitive consequences, such as harming women's health.

Our study aims at investigating the relation between non-cognitive values and inductive risk. We present the results of an experimental study clarifying the psychological impact of political values and personal features like one's race and sex on the acceptance (or rejection) of scientific hypotheses in the face of inductive risk. Our hypotheses was that political and personal identity features reliably predict people's sensitivity to scientific errors. Specifically, people are less likely to accept hypotheses that they perceive as clashing with their political ideology and identity. In our study, participants were asked to read and evaluate three vignettes,

where scientists disagree about the adequacy of a specific test and take decisions about hypotheses involving sexual or racial differences. In each vignette, the consequences of a mistaken decision could harm a group of people (either women, men, Black or White people). One of the vignettes concerned the exclusion of women from clinical trials. In this vignette, scientists decided to introduce a new drug tested on a group including only men into the market. Participants were asked to express how certain they were that the decision taken was a good one. Our hypothesis was that conservative men were more likely than women to see that decision as a good decision. At the end of the survey, information about political ideology, race, and sex was collected.

Our results provide us with a more nuanced understanding of the bearing of non-cognitive values on the psychology of inductive risk. Though philosophers of science have drawn on several historical case-studies to clarify the notion of inductive risk, little attention has been paid to how people actually reason about inductive risks. In this paper, we set out to begin filling this gap in the philosophical literature by investigating the relationship between reasoning, inductive risk, and non-cognitive values.

References:

Douglas, Heather. (2000). "Inductive risk and values in science". *Philosophy of science*, 67(4), 559- 579.

Hempel, Carl G 1965. "Science and human values." *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York: The Free Press, 81-96.

Vulcan, the golden mountain, and the man without properties: neo-Frege meets neo-Meinong

Bjørn Jespersen (Utrecht University)

I show how my neo-Fregean theory solves three problem cases that are variations of the sentence "The F is an F". The novelty is a notion of individuals-in-hyperintension. These are fine-grained modes of presentation of individuals-in-intension, serving to model 'impossible individuals', which are, naïvely speaking, individuals that could not possibly exist. Individuals-in-hyperintension are required for the third problem case, whereas standard individuals-in-intension suffice for the first two. My theory is Tichý's Transparent Intensional Logic, which will be compared against the neo-Meinongian theories of Zalta's object theory and Priest's modal Meinongianism.

This paper discusses predication de re and de dicto of properties of entities that are either contingently absent from the actual world or necessarily absent from any possible world.

My background framework is Tichý's neo-Fregean Transparent Intensional Logic (TIL), which comes with a constant domain of individuals, as well as individuals-in-intension (so-called *offices*) modelled as functions from worlds to partial functions from times to individuals (see Duží et al., 2010). An office is occupied by at most one individual at any given world/time pair, such pairs serving to model empirical variability. Offices suffice for a logical analysis of cases like:

- *Vulcan is a planet* (A)
- *The golden mountain is a golden mountain* (B) The definite description 'the golden mountain' is interpreted as a name of an office, and so is the grammatical proper name 'Vulcan'. Neither term ever denotes the occupant of either office at any world/time pair. Both offices come with a host of constraints (so-called *requisites*) that must be satisfied by all of their respective occupants. The requisite properties pertaining to the office of Vulcan include, but are not limited to, being a planet orbiting between Mercury and the Sun and causing Mercury's skewed orbit. Therefore, it comes out an analytic (hence necessary) truth that Vulcan is a planet. This is predication de dicto. Predication de re is geared to predicate being a planet of the occupant, if any, of the office at a particular world/time pair of evaluation. If the office is occupied there, the result is a contingent truth (for it could not be a falsehood). If vacant, the result is a contingent truth-value gap. The underlying logical semantics applies to factual discourse, which includes discourse both about Vulcan the planet and Vulcan the god, but does not extend to fictional discourse, which involves grammatical proper names such as 'Sherlock Holmes'. Offices are insufficient for a satisfactory logical analysis of cases like:
 - *The man without properties is a man without properties* (C)
 - *The only fake banknote that is a banknote is a banknote* (D)

Enter *hyperoffices*, which are fine-grained modes of presentation of offices. Hyperoffices serve here to logically model 'impossible individuals', which are, naïvely speaking, individuals that could not possibly exist.

My dual thesis is that 'impossible individuals' are special cases of hyperoffices and as such two levels removed from the extensional level of individuals, whereas 'merely possible individuals' are offices (one level removed from the level of individuals) that are contingently vacant in the actual world at the present time.

The fact that Vulcan (the golden mountain, resp.) fails to exist is interpreted in terms of the contingent vacancy of the office of Vulcan (the golden mountain, resp.).

Both of (C) and (D) embody analytic impossibilities. Therefore, they both converge in *the impossible office*, which is the necessarily vacant office. The fact that the man without properties (the only fake banknote that is a banknote, resp.) fails to exist is interpreted in terms of this or that particular hyperoffice presenting the impossible office.

One problem that (C) and (D) pose is how to distinguish in a principled manner between two analytic impossibilities. We cannot turn to occupancy/vacancy throughout logical space, nor to requisites, as it is a corollary of the definition of requisites that any and all properties are requisites of the impossible office. The solution I propose relies on the formally precise notion of *logical procedure* that provides TIL with an elaborate theory of structured hyperintensions. The solution to the problem of hyperintensional individuation of hyperoffices consists in making the difference matter between whether the very constituent sub-procedures of the respective procedures that produce the impossible office are procedures for producing one property rather than another. Formally, this is achieved by displaying rather than executing the respective sub-procedures. This way, being a man and lacking properties are requisite properties of (C), but not of (D), and vice versa for being a banknote and being a fake banknote of (D), though not of (C).

The predication de re of being a man without properties of the man without properties yields the necessarily gappy proposition (i.e. the truth-condition that takes each world/time pair to a truth-value gap). The impossible office is necessarily vacant, so there is nowhere and never an occupant of whom it would be either true or false that he was a man without properties. This analysis applies equally to the fake- banknote example, where the relevant sub-procedures have been swapped as described above. The upshot is two different structured hyperpropositions both producing the necessarily gappy proposition.

Predication de dicto expresses that being a man is a requisite of the hyperoffice of the man without properties, and that being a banknote is a requisite of the hyperoffice of the unique fake banknote that is a banknote. These two structured hyperpropositions both produce the truth-value *true* in two different manners.

In this paper I compare my analyses against those of two neo-Meinongian theories, namely Zalta's Object Theory (OT) and Priest's Modal Meinongianism (MM). Meinongian theories are custom-built to tackle inexistent entities and predication of properties of them, whereas broadly Fregean theories tend to have a harder time when there is no unique *F*. However, this talk demonstrates how at least TIL is able to match OT and MM. A closer comparison shows that TIL is closer to OT than to MM, most notably by embracing dual predication and eschewing

impossible objects and impossible worlds. However, whereas the so-called abstract objects of OT are liable to be logically sterile, TIL's sets of requisites are closed under logical consequence.

The general picture that emerges of TIL is one of fine-grained structured procedures at the top, intensions in the middle, and extensions at the bottom, all organized within a typed universe.

References:

Duží M et al. (2010), *Procedural Semantics for Hyperintensional Logic*, LEUS vol. 17, Springer-Verlag.

Explanation and Agency: Exploring the normative landscape of the “Right to Explanation”

Fleur Jongepier (Radboud University Nijmegen) and Esther Keymolen (Tilburg University)

Suppose an algorithm makes a decision that has significant consequences for your future or daily life, like getting a mortgage. Intuitively, the least you're entitled to is an explanation of why that particular decision was made. The new European General Data Protection Regulation (GDPR) includes a so-called 'right to explanation' in situations involving automatic decision-making. In this paper, we explore the normative landscape surrounding the right to explanation. We consider both a metaphysical approach that grounds such a right in a 'human versus machine'-distinction, and a normative one in terms of the necessary epistemic conditions for deliberative agency.

A large part of the existing literature on the right to explanation has focused on the factual question whether there *is* a right to explanation (implicit) and where it should be located in the GDPR (e.g. Goodman and Flaxman 2016; Selbst and Powles 2017). Much less attention has been devoted to the question of whether there *ought* to be a right to explanation, regardless of whether or not it's explicitly or implicitly there in the GDPR, and if so, why. In this article, we therefore want to explore this normative landscape. After all, even if there is a strong normative intuition for having a right to explanation, intuitions can be wrong. So we need to do more than just appeal to intuitions when it comes to explaining the normative significance of having a right to explanation. The aim of the paper is therefore to provide an account of what might underlie the normative intuition in more detail

than has been done so far.

The paper consists of two parts. In the first part, we address the question of whether there are *special* normative reasons for having a right to explanation in cases involving automated decision-making that involve no human intervention. Those who answer this question positively accept what we refer to as the ‘asymmetry thesis’, according to which automated and human decisions each introduce different normative challenges and make different normative claims on us. We argue, however, that the asymmetry thesis rests on a mistaken conception of the relation between technology and human agency, namely, one according to which (1) human agency is defined principally as being independent of or in opposition to or technology, and (2) technology is defined as being independent of or in opposition to human agency.

We argue the asymmetry thesis not only underestimates the *prima facie* non-problematic nature of human decisions but also overestimates the *prima facie* problematic nature of non-human decisions. In many cases, the fact that ‘machines’ make decisions that affect us without our understanding how the decision was reached, is no immediate cause for concern, in part because technology often enhances or extends our agency rather than forming an obstacle or standing in opposition to it.

We defend, in its place, what we call the symmetry thesis, according to which there is nothing normatively special let alone *prima facie* bad or wrong about automated decisions *as such*. Importantly, we argue that while it is very common for technology and human action to become deeply intertwined, this does not make the right to explanation any less normatively significant or legal requirements to secure it any less justified. On the contrary, the right to explanation helps guarantee that this intertwining is of the right sort and that the network of human and non-human actors involved in automated decision making is balanced in the right way.

In the second part of the paper, we therefore turn to the question of what normative reasons might then underlie having a right to explanation, if not an appeal to the simple fact that one is subjected to a decision in which (supposedly) no humans were involved. We propose to answer this question by exploring the connection between the epistemic ability of persons understand important decisions made about them and what we call a person’s ‘deliberative agency’. A person has, we argue, a *general* right to explanation when her agency is substantially affected and/or the necessary means for agency are undermined.

Our starting point is the observation that in order to make decisions at all, certain minimal epistemic conditions pertaining to knowledge and understanding need to be in place. Only if these broadly epistemic conditions are in place, one can deliberate about the options and only then one can be said to be in a position to

make a decision. But if one fails to know or understand information that is crucial to making a decision, then one's deliberative agency is undermined.

The right of explanation gets its normative force, we argue, from the fact that it is a necessary condition for deliberative agency. The normative intuition we started out with is, on closer inspection, not about automated decisions per se, but more generally about how decisions affect/undermine our agency combined with the observation that automated decisions may do so more often and/or that they pose a threat to specific dimensions of deliberative agency such as informational self-determination and “informational privacy” (Roessler 2004).

Having set out our account of what underlies the right to explanation in terms of deliberative agency, we then put it to the test by introducing and responding to various objections to it, such as the worry that having a right to explanation is not of much use in practice, and that the idea of informational self-determination underlying the right to explanation is – except for people like Max Schrems, the law scholar and privacy activist who took up Facebook in court for violating his privacy – not in line with how people actually behave in real-life (cf. Koops 2014).

Epistemic Dimensions of Environmental (In)Justice

Jason Kawall (Colgate University, United States)

Empirical research suggests that our physical environment has striking impacts upon our cognitive development and performance. Studies show that excessive noise, heat, proximity to highways, and other factors have severe detrimental impacts on cognition. Globally, impoverished and marginalized communities face these issues disproportionately. In this paper I review some of the relevant empirical work, and argue that the issues raised ought to be of significant interest to epistemologists. Such work draws attention to new ways of thinking concerning the significance of embodied cognition, new ways of addressing situationist worries raised for virtue epistemologies, and alternative ways of improving epistemic performance.

In this paper I explore the importance of our ambient physical environments to a range of epistemic concerns. In particular, I explore recent empirical work on the significant detrimental impacts of excessive noise, heat, proximity to highways, and other factors on cognitive performance. Globally, impoverished and marginalized communities face these issues disproportionately, and (it is widely accepted) unjustly. Environmental injustice thus has often has significant epistemic impacts – indeed, we can plausibly treat environmental injustice as frequently leading to

particular forms of epistemic injustice.

I begin by providing an overview of relevant recent empirical work. For example, a recent study by Goodman et al. suggests that cumulative heat exposure inhibits cognitive skill development [...] [H]otter school days in the year prior to the [PSAT] test reduce learning, with extreme heat being particularly damaging and larger effects for low income and minority students. [...] Without air conditioning, each 1°F increase in school year temperature reduces the amount learned that year by one percent. (2018: 1)

Sunyer et al. find that “that cognitive development is reduced in children exposed to higher levels of traffic-related air pollutants at school. This association was consistent for working memory, superior working memory, and inattentiveness, and robust to several sensitivity analyses” (2015: 11). With respect to noise pollution, Wright et al. find “there is a clear adverse effect of continuous environmental noise on attending to peripheral cues, WM [working memory], and episodic recall in non clinical adult participants” (2014: 160). Ferguson et al. , in a review of the impacts of physical environments on child development, observe that “it is likely that environmental toxins and pollutants [...] impact the cognitive and socioemotional development of children living in different contexts similarly” (2015: 456). As they turn to suggestions to ameliorate conditions, they note that the evidence we have reviewed here suggests that factors contributing to chaos, including noise and crowding, likely impact children and adults across the globe in similar ways. [...] Home, classroom and school designs that reduce chaos may be particularly important. In addition, adequate lighting and comfortable climatic conditions (temperature, indoor air quality) are important for effective learning in school environments (Ferguson et al. 2015: 457).

I argue that these and related findings suggest that environmental conditions warrant much greater attention from epistemologists (and not only insofar as our environments might be plagued by fake barns and the like!). In brief: to produce better epistemic agents, and to improve epistemic performance in particular instances, we will often be as wise to improve an agent’s physical environment as trying to improve the epistemic faculties of the agent herself. This could be especially important in the case of members of lower income and marginalized communities who are often subject to environmental injustice and face particularly stark epistemic impacts as a result.

I next consider how a focus on the physical environment might help us to develop virtue epistemologies, particularly in light of certain situationist challenges (see Fairweather and Alfano 2017). For example, suppose that a student apparently has remarkable concentration and creativity when allowed to engage in solving difficult maths problems in a comfortable library after a relaxing walk in a well-maintained city park. Does this show her to have genuine virtues of creativity and concentration? What if she would fare quite poorly if she were instead tested at her

noisy home, following a long day at school? Have we simply manipulated the environment such that we can coerce good epistemic results out of an individual who lacks relevant virtues – that the situation or environment is what is really doing the key work here? I argue that even if the student can only solve difficult problems under good environmental conditions, we should treat her as possessing concentration and creativity (and other intellectual virtues). Crucially, such cases show that she can meet difficult epistemic challenges that defeat others; these cases reflect underlying epistemic abilities and virtues. After all, even in these good circumstances, not all agents would succeed in such problem-solving. By improving environments to improve the performance of agents, we are allowing the epistemic virtues of these agents to be engaged, rather than simply coaxing good results out of agents who lack them.

I contrast the epistemic case above with what might appear to be an ethical equivalent – imagine a person who typically never demonstrates any form of generosity. But once she wins a million dollar lottery, she is able to perform some basic generous acts (small donations, etc.). Should we treat this case like that of the student who can solve difficult maths problems in good environmental conditions? I argue that the cases are disanalogous. In particular, the student is successfully completing demanding epistemic tasks in a good environment, while the lottery winner is only able to perform minimal, non-demanding acts of generosity when similarly placed in an ideal environment. There is not a display of significant ethical virtue here, unlike the case of the student.

To illustrate how taking such issues seriously might have implications for applied epistemology, I provide a brief overview of some recent proposals for cognition-improving environmental changes, ranging from providing parks and green-spaces in cities, to improving air quality in buildings, to reducing noise and other forms of pollution.

Finally, I address a potential objection – that environmental injustice and its related environmental impacts, while affecting cognition, might not be of particular epistemic interest. Compare: car accidents and other events in which victims may suffer concussions can have significant cognitive impacts, but it is unclear that we need papers in *epistemology* devoted to such matters. In response, I argue first that instances of environmental injustice often provide striking examples of additional issues of interest to epistemologists. For example, we often find epistemic injustice in the (lack of) environmental information shared with marginalized communities, and testimonial injustice in the reception of testimony from indigenous and other communities concerning environmental harms they face. Climate change denial raises issues in the epistemology of ignorance, and so on. Secondly, and more fundamentally, unlike cognition-affecting accidents or other (comparatively) rare occurrences, we are fundamentally embodied, situated creatures. We are always cognizing within environments; as such, consideration of our environment is not merely consideration of some occasional or incidental feature. Work in embodied

and extended cognition already embraces this stance – the current proposal attempts to further develop this general approach.

References:

Fairweather, A. & Alfano, M. (eds) 2017, *Epistemic Situationism*, Oxford University Press, New York.

Ferguson, K.T., Cassells, R.C., MacAllister, J.W. & Evans, G.W. 2013, "The physical environment and child development: An international review", *International journal of psychology : Journal international de psychologie*, vol. 48, no. 4, pp. 437-468.

Goodman, J., Hurwitz, M., Park, J. & Smith, J. 2018, "Heat and Learning", *National Bureau of Economic Research Working Paper Series*, vol. No. 24639.

Sunyer, J., Esnaola, M., et al. 2015, "Association between Traffic-Related Air Pollution in Schools and Cognitive Development in Primary School Children: A Prospective Cohort Study", *PLoS Medicine*, vol. 12, no. 3, pp. e1001792.

Wright, B., Peters, E., Ettinger, U., Kuipers, E. & Kumari, V. 2014, "Understanding noise stress-induced cognitive impairment in healthy adults and its implications for schizophrenia", *Noise & health*, vol. 16, no. 70, pp. 166-176.

Considerations for Moral and Fulfilling Work

Sujin Kim (University of Twente)

Aim of the paper is to critique on the theory of meaningful work in order to deepen the concept of meaningfulness in order to have a strong conceptual basis for analysis of gig-work. Gig work is important for the technological society, yet its status is contentious because its labour serves the society in many different ways while the current business model sets a very low wage level to those engaged in gig-work. Studying the framework is the starting point to the analysis of the freelance work practices. This can contribute to the understanding of the digital platform, or challenge the understanding on the platform and platform-mediate work. The given understanding based on meaningful work in political theory is rather perfectionist because meaningful work is a human need. It is rather instrumentalist about platform-mediated work because objective conditions ought to shape meaningfulness. The paper aims to arrive at a typology that makes it possible to relate the theory of meaningfulness to platform-mediated work practices

Political implication of understanding these considerations is that it supports the argument for basic human rights, basic income, and deeper political intervention to work. This has implication on responsibility, in other words the responsibility for acting out on maintaining basic working conditions or basic income. This presupposes on the economic and social conditions to which policies have to adapt into. Automation and financial change are some of the oblivious trends that shape the economy in general. The paper starts from questioning these assumption that economic change is objective reality. It implies that working conditions and human values of work are subjective goods, either from preferences or need. It also implies that objective material condition becomes the *context* of meaningful work.

The aspect of the context that this paper highlights is the polarized development in terms of meaning, signified by 'Lovely and Lousy jobs', where Goos and Manning argued that Britain's jobs have been polarised more and more in terms of skills. Polarisation in economic results from technological development, as I observe, are coupled by polarisation of meaning. To work for a lousy job also undermines dignity or the voice and opinion for meaningfulness. Presenting a case of platform based work environment, the paper reconsiders freedom and dignity in the case of platform work, where people choose such work in the name of freedom. It is no longer enough to remove evil prevent objectification. It is less determined in the design of the network, such as design of platforms or market regulations, so much as in active and dynamic relations of constitution. This allows us to think about human technology relations, where valuer and valuable object(work, wage, and other forms of added value) are mediated via technology. In this case, platforms such as Amazon and Fiverr play such a role.

Meaningful work is a human need to experience, evaluate, and cultivate. While political theory of meaningfulness is specific on the moral status of valuer it has to consider the need of cultivation. With a bigger aim to critique on this, a smaller aim is to discuss the technological aspects of enhancement and meaningfulness. Enhancement of work while democratising work are mutually compatible goals in the given theory. However, it is restrictive if technology and technologically shaped values are not recognized in enhancement and democratisation. After introducing the argument on meaningfulness, I show the challenges to improve the meaningfulness of freelance work in platform economy. One issue rises from freedom and another from skill development. . In the analysis, the lens of human-technology relations or mediation is useful because it makes visible the role of the platform and the entities in the world that affects the character of work. Along the analysis, the drawbacks of taking a bipartite model of meaningfulness was also suggested.

Gig work allows us to think about the discourse and its assumptions. It is necessary to discuss quasi work and career development the policies in this area. It is often not covered in the study of organisations, or human resource management.

One obstacle is that the realm they are concerned is flexible and changeable and the principles they follow might be more arbitrary or relative to circumstances. Moreover, it is difficult to apply the understanding of current MW to this realm because to improve work, either by self motivated or externally motivated way. Freelancers, interns, workers after retirement, and those looking for better work for their career are not easily placed in our categories of work and non-work. Their foundation for meaning and enhancement is contentious.

The second obstacle arise from points of view. Meaningfulness has to serve someone. It is either the purpose of engagement from management view (on leadership, human resource) Otherwise, technological design serves the workers yet may not serve other people. Pursuing an ethical perspective, therefore, seems to entail the choice between points of view or a zero-sum game. What is lacking in this form of innovation is the aspect of 'becoming' in meaningfulness. This implies we lack understanding how people make meaning from work. This goes back to the distinction between subjective and objective meaningfulness. If we were to choose between individual and social good, it is easy to neglect the aspect of becoming or realisation of capability.

The third obstacle arise from perfectionist and absolutist tendencies of meaningful work theory. Principles or ultimate consideration of meaningful work might force us to dismiss important understandings. Relating to the first obstacle, the challenge of theorizing on this arise from the aspect of gig-work that lacks conventional foundations of discussion - human work and human dignity. Perfectionism can be seen in the discourse in the ideas that aims to set up objective standard from the given problem. This could be seen as well in the absolutist character of the principles, or to explain some work as totally liberation or total alienation.

Identifying Self-fulfilling Prophecy in Automated Prediction

Owen King (University of Twente) and Mayli Mertens (University of Twente)

A self-fulfilling prophecy is a prediction that somehow increases the likelihood of its own truth. In this paper we offer a thorough account of self-fulfilling prophecies and their significance. First, we describe two conditions that are both necessary for a prediction to be self-fulfilling and further explain why some are. Second, on that basis, we distinguish three types of self-fulfilling prophecies. Third, we show that our analysis provides fruitful explications of instances of self-fulfillment in important contexts of prediction: prediction of a person's preferences and prediction (prognostication) of medical outcomes. We pay special attention to automated prediction, especially machine-learning-based data analytics.

What is a self-fulfilling prophecy? An easy—and correct—answer is that it is a prediction that somehow increases the likelihood of its own truth. But that easy answer tells us little more than the term itself. A more thorough account would tell us something about how it could be that making a prediction (whether by uttering it, recording it, or drawing inferences based on it) could possibly influence its own truth. In this paper we offer a more thorough account.⁸

We motivate our discussion by briefly reviewing the example of automated predictive policing, a commonly acknowledged case of a self-fulfilling prophecy involving AI.⁹ After motivating our discussion, we describe two conditions that explain how self-fulfilling prophecies occur. On that basis, we distinguish three types of self-fulfilling prophecies. We show that our analysis provides fruitful explications of instances of self-fulfillment in important contexts of prediction: prediction of a person's preferences and prediction (prognostication) of medical outcomes. We pay special attention to computer systems that automate prediction, such as classifiers based on machine learning. Ultimately, if the phenomenon of self-fulfillment is epistemically and socially significant—which we believe it is—it becomes even more so when systems automate self-fulfilling prediction, distancing the processes from human scrutiny and increasing the speed with which they are performed.

There are two conditions that are necessary for a prediction to be self-fulfilling and further explain why some predictions are indeed self-fulfilling. The first is that a self-fulfilling prophecy *produces no error signal*: Self-fulfilling predictions are met with no protest to induce correction. This condition already distinguishes self-fulfilling predictions from ordinary ones. Ordinarily, if a person or system issues some sort of prediction but the prediction is not justified by available evidence, then this prompts some protest or negative feedback or some other indication that correction is required. In contrast, meeting no protest, a self-fulfilling prediction is unlikely to be retracted by the predictor, even if the prediction fails to match what would be probable in the absence of any prediction. For a given self-fulfilling prophecy, to explain the lack of error signal would be to explain why the ordinary conduits of negative feedback are blocked or diverted.

Second, self-fulfilling prophecies have some *means of perpetuation*: Once the prediction has been made, the world is bound to the prediction in a way that ensures (or at least makes more likely) that the next time there is a “neutral” assessment of the predicted situation, the assessment matches the original prediction.

Distinguishing three different modes of perpetuation yields three types of self-

⁸ Self-fulfilling prophecies have not been a major topic in theoretical ethics. An exception is Miller (1961). There has been a bit more discussion in medical ethics. See, e.g., Wilkinson (2009).

⁹ See, e.g., Lynch (2016).

fulfilling prophecies: *efficacious*, *interpretive*, and *inherent*. We can exemplify this three-way distinction through three variations of an example. Consider how a teacher's prediction of a student's high marks might increase the chances of the student receiving high marks. First, the prediction could become known to the student, affecting the student's behavior positively, leading to better work, and hence high marks. The object changes to match the outcome predicted; this would be *efficacious* perpetuation. Second, the prediction could shape the teacher's framework for evaluation, leading her to evaluate the student more highly than she would have, thus yielding high marks. The interpreter's standards or perspective are altered to match the prediction; this would be *interpretive* perpetuation. Third, the prediction could take place within a system such that the prediction itself becomes part of the student's default score, thus yielding higher marks. This would be *inherent* perpetuation, a sort of immediate and trivial self-fulfillment; that the prediction has been issued is itself a significant determinant in its truth.

Examining cases, we see that a focus on whether predictions generate an error signal and on whether they have some potential means of perpetuation allows us to identify contexts of prediction fertile for self-fulfillment. Two particular, and superficially quite dissimilar cases, occupy our focus. First is product recommendation, i.e., prediction of consumer preferences. Second is medical prognostication, i.e., predictions of health and mortality.¹⁰ Obviously, each of these types of prediction has been with us in some form or fashion for most of human civilization. But both are currently undergoing revolutions due to automation, specifically with computerized prediction systems based on machine learning. Hence, it is important that we now be attuned to whether and how these predictions are self-fulfilling.

As various sorts of data analytics techniques transform our social practices and configurations, a number of values have guided our scrutiny. Foremost among these have been moral values of fairness and privacy. Particular types of socio-technological dynamics—for example, feedback loops and filter bubbles—have also been matters of concern for both the public and technologists.¹¹ Notably, self-fulfilling prophecies, although frequently mentioned in discussions of fairness and/or pernicious feedback loops, have been neglected.¹² This omission is notable because, not only do self-fulfilling prophecies constitute a distinctive class of phenomena (as demonstrated by our analysis), they have a formidable kind of alethic and epistemic power, not to be ignored or treated lightly. Self-fulfilling

¹⁰ We will be especially interested in prognostication of recovery in cases of cardiac arrest. The risk of self-fulfilling prophecies in this context has been noted by Geocadin et al. (2012), Bouwes et al. (2012), and Maciel et al. (2017).

¹¹ Could filter bubbles be self-fulfilling prophecies? Our analysis will show that they are, under certain circumstances. Regarding filter bubbles, see Pariser (2011).

¹² Lynch (2016) is a paradigm example of how these issues are conflated in the popular press. Traditionally social psychology has been the domain with the greatest interest in self-fulfilling prophecies. See, e.g., Snyder et al. (1977) and Jussim (1986).

prophecies generate knowledge not due to sensitivity to the way the world is, but by remaking the world in their image, whether or not that image is a desirable one.¹³

References:

Bouwes, A., Binnekade, J. M., Kuiper, M. A., Bosch, F. H., Zandstra, D. F., Toornvliet, A. C., ... & Weinstein, H. C. (2012). Prognosis of coma after therapeutic hypothermia: a prospective cohort study. *Annals of neurology*, 71(2), 206-212.

Christakis, N. A. (2001). *Death foretold: prophecy and prognosis in medical care*. University of Chicago Press.

Geocadin, R. G., Peberdy, M. A., & Lazar, R. M. (2012). Poor survival after cardiac arrest resuscitation: a self-fulfilling prophecy or biologic destiny?. *Critical care medicine*, 40(3), 979-980.

Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological review*, 93(4), 429.

Lynch, J. (2016, April 24). Is Predictive Policing the Law-Enforcement Tactic of the Future. *The Wall Street Journal*. Retrieved from <https://www.wsj.com>

Maciel, C. B., Morawo, A. O., Tsao, C. Y., Youn, T. S., Labar, D. R., Rubens, E. O., & Greer, D. M. (2017). SSEP in Therapeutic Hypothermia Era. *Journal of Clinical Neurophysiology*, 34(5), 469-475.

Miller, C. (1961). The self-fulfilling prophecy: A reappraisal. *Ethics*, 72(1), 46-51.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and social Psychology*, 35(9), 656.

Wilkinson, D. (2009). The self-fulfilling prophecy in intensive care. *Theoretical medicine and bioethics*, 30(6), 401-410.

¹³ Regarding desirability of self-fulfilling prophecies in the medical context, see Christakis (2001), especially chs. 6-7.

Depolarizing Theory and Practice: What Might Constitutivism about Political Norms be?

Tim Klaassen (Tilburg University)

Constitutivist approaches to normativity seek to derive the authority of normative claims from the nature of action and agency. These theories, however, have been mostly concerned with the nature of moral norms and individual agency. But what about political norms and collective agency? If these are *sui generis*, then perhaps it is possible to formulate a political constitutivism. In this paper I offer a suggestion of what such a theory might look like. The thesis that I develop is that the source of political normativity lies in the interpretive capacities that people exercise in making sense of their socio-political order.

Mainstream political philosophy has sometimes been accused of being out of touch with political reality. The kinds of ideals that it comes up with are too far removed from the social and political realities of everyday life. As a consequence, normative political theorizing becomes irrelevant from a practical point of view (Mills 2005, Sen 2006). The source of this insufficiency is perceived to lay in the fact that much normative political theorizing comes in the form of “ideal theory”. A term which, on the surface of it, indeed easily leads to connotations of utopic thinking in the pejorative sense of the term. David Schmidtz seems to express just this very sentiment when he suggests that saying that “ideally, this is what we ought to do” is really just to make a throwaway remark as a “preliminary to getting down to serious problem solving.” (Schmidtz 2011).

Worries such as these have instigated various debates on the status of ideal theory. What makes it difficult to assess these debates, however, is that no two authors apparently mean exactly the same thing when they talk about “ideal theory”. As Laura Valentini writes in her attempts at mapping the relevant issues, the debate on ideal theory has become rather “messy” (Valentini 2012, 2017). Relatedly, Schmidtz writes about how “elusive” the distinction between ideal theory and nonideal theory is. Furthermore, he suggests that what is at stake here is actually *a constellation* of different issues, related by family resemblance rather than any “shared essence” (Schmidtz 2011).

In this paper, I am going to single out one possible take on what is at issue in the controversy over ideal theory, but without trying, in line with Schmidtz’s suggestion, to reduce everything else to it. There may be different things at stake in evaluating the significance and status of ideal theory, and here I am simply going to focus on one of them. In particular, I am going to focus on the issue of the extent to which normative principles have to be “*fact-(in)sensitive*”, *i.e.*, the degree to which “theorizing about justice should be anchored in existing factual realities” (Valentini

2017).

Regarding that issue, I want to begin by making the observation that there is a perhaps trivial but very *important* sense in which normative principles have to be *both* fact-sensitive *and* fact-insensitive. On the one hand, normative principles have to be fact-sensitive in the sense that they have *conditions of applicability*. They have a certain *range*. The rules of chess do not apply when you are playing a rugby match. On the other hand, normative principles have to be fact-*ins*sensitive in that their validity does not depend on whether in any applicable situation they are actually being complied with (Estlund 2011, 2014). Even if people do not obey the law, it does not make it any less of a law.

So it turns out to be perfectly compatible to say that normative principles have to be both fact-sensitive *and* fact-insensitive. Therefore, if we want to maintain that there is nevertheless an issue here, then we have to give a different account of it. That is, if there is any truth to the worry that normative political theorizing should remain firmly with its “feet on the ground”, on pain of losing touch with social and political reality altogether, there must be something else at stake in the issue of fact-sensitivity and the requirement that normative theorizing should be properly anchored in political reality.

What is needed is not only a better understanding of how political norms are supposed to be “anchored” in political reality, but also of how *conversely* the latter can be “*bound*” by political norms. What matters is not simply the important truth that (in a way) norms determine facts and (in a way) facts determine norms, but *what* this relation of mutual grounding *consists in*. If our worry is that political normative theorizing is out of touch with the things themselves, then that worry might be taken away if it is shown how certain political ideals are normatively binding for us *on account of the kind of social and political beings that we are*. As I’ll explain, an important element of our understanding of ourselves is that our phenomenal life-worlds for the most part already shows up as organized against the horizon of a pre-given institutional order.

In normative philosophy there is one approach that has an elegant answer to how norms, and the kinds of things to which they are meant to apply, are related. That approach is referred to as *constitutivism*. Such approaches to normativity seek to derive the binding force of normative claims from the very nature of action and agency (Katsafanas 2013). For the most part, however, these theories have been mostly concerned with the nature of moral norms and individual agency. But what about *political* norms and *collective* agency? In the field of social ontology it is a common view that collective or joint actions are not simply reducible to aggregates of individual actions (Searle 1995, Gilbert 2013). Similarly, many political philosophers agree that political normativity is irreducible to moral normativity (Galston 2010, Rossi and Sleat 2014). Hence, if collective action and political normativity are *sui generis*, then perhaps it is possible to formulate a separate *political*

constitutivism.

Taking Christine Korsgaard's particular brand of moral constitutivism (Korsgaard 2009) as an initial model, my aim is to sketch the possible contours of what a political constitutivism might look like. On analogy with Korsgaard's theoretical framework (which I will clarify), the goal will be to work out a theory about the constitutive relationships between political normativity, collective action, and the (ongoing) institution of political order. In the same way that, on Korsgaard's view, the binding force of moral norms is to be found in our identity as self-conscious agents, the sources of political normativity are anchored in the interpretive schemes that structure our identities as we find ourselves born in a particular society, at a particular time and place. As we draw upon our own interpretive resources in understanding our social world, the institutional order in terms of which we already understand is simultaneously effected.

References:

- Estlund, David. 2011. "Human nature and the limits (if any) of political philosophy." *Philosophy & Public Affairs* 39 (3):207-237.
- Estlund, David. 2014. "Utopophobia." *Philosophy & Public Affairs* 42 (2):113-134.
- Galston, William A. 2010. "Realism in political theory." *European journal of political theory* 9 (4):385-411.
- Gilbert, Margaret. 2013. *Joint commitment: How we make the social world*. Oxford University Press.
- Katsafanas, Paul. 2013. *Agency and the foundations of ethics: Nietzschean constitutivism*. OUP Oxford.
- Korsgaard, Christine M. 2009. *Self-constitution: Agency, identity, and integrity*. Oxford University Press.
- Mills, Charles W. 2005. "'Ideal theory' as ideology." *Hypatia* 20 (3):165-183.
- Rossi, Enzo , and Matt Sleat. 2014. "Realism in normative political theory." *Philosophy Compass* 9 (10):689-
701. Schmidtz, David. 2011. "Nonideal theory: what it is and what it needs to be." *Ethics* 121 (4):772-796.
- Searle, John R. 1995. *The construction of social reality*. Simon and Schuster.
- Sen, Amartya. 2006. "What Do We Want from a Theory of Justice?" *The Journal of Philosophy* 103 (5):215-238.
- Valentini, Laura. 2012. "Ideal vs. Non-ideal Theory: A Conceptual Map." *Philosophy Compass* 7 (9):654-664.

Valentini, Laura. 2017. "On the Messy "Utopophobia vs. Factophobia" Controversy: A Systematization and Assessment " In *Political Utopias* edited by Michael Weber and Kevin Vallier, 11-34. New York, NY: Oxford University Press.

The Logic of Free Choice Permission

Johannes Korbmacher (Utrecht University) and Albert Anglberger (Bayreuth University)

In this paper we develop a logic of permission with the free choice permission principle as its sole deontic axiom. We further provide a natural semantics in terms of truthmakers, and show that our logic does not contain any of the paradoxical consequences typically associated with free choice permission.

The problem of free choice permission ([7] and [8]) is one of the oldest puzzles in modern deontic logic. Consider the following permission statements about your options for dessert:

- (i) You may have tiramisu or zabaglione.
- (ii) You may have tiramisu and you may have zabaglione.

In the language of propositional deontic logic these statements are naturally expressed by:

$$(i') \quad P(A \vee B)$$

$$(ii') \quad PA \wedge PB$$

Intuitively, (i) and (ii) are synonymous: in natural language, the permission of a disjunction typically means that an agent is may choose either one of the disjuncts. If we expect a logic of permission to reflect this intuition, we might therefore consider adding

$$(FCP) \quad P(A \vee B) \leftrightarrow (PA \wedge PB)$$

as an axiom. This, however, is easier said than done: it seems that adding FCP necessarily leads to unacceptable consequences. The strongest argument against FCP is due to Hilpinen [1], and it shows that the left-to-right direction of FCP is unacceptable as soon as permission allows for replacement of classically equivalent formulas:

$$(\text{Replacement}) \quad \text{If } \vdash A \leftrightarrow B, \text{ then } PA \vdash PB$$

Now suppose PA. According to classical propositional logic, A is equivalent to $(A \wedge B) \vee (A \wedge \neg B)$, and so we arrive at $P((A \wedge B) \vee (A \wedge \neg B))$. By the left-to-right direction of FCP we get $P(A \wedge B)$. This means that whenever A is permitted, it is also permitted in conjunction with any arbitrary B , which is clearly unacceptable. This argument suggests that it is generally very hard to find a logic which contains FCP but also avoids problematic consequences like the one just mentioned. As Sven Ove Hansson puts it: “It [i.e. the derivation of IC] indicates that the free choice permission postulate may be faulty in itself, even if not combined with other deontic principles such as those of SDL” [6, p. 208].

So it seems that we are faced with a dilemma: on the one hand, FCP seems to express a natural notion of permission which should, therefore, also have a proper formal semantics and logic. On the other hand, the replacement rule is deeply entrenched in many deontic logics, because it is considered to be at least as plausible as FCP. Hence, dropping the replacement rule cannot be successfully justified by the mere desire to block undesirable consequences of FCP. What we need are good semantic reasons derived from a intuitive semantics for permission that (a) explain why FCP is a plausible principle for permission, and (b) explain why replacement does not hold for this very notion of permission. This is the problem of free choice permission.

To solve the problem of free choice permission we use a framework called truthmaker semantics (invented by Bas van Fraassen in [9], developed further by Kit Fine in [2], [3], [4], [5]). The basic idea of (exact) truthmaker semantics is that we can give the semantic content of a statement by saying what precisely in the world makes the statement true: by giving its exact truthmakers. Intuitively, an exact truthmaker of a statement is a state (of affairs) such that whenever the state obtains it is directly and wholly responsible for the truth of the statement. In particular, an exact truthmaker of a statement will not contain as a part any other state that is not wholly responsible for the truth of the statement. For example, the state of Dominik eating tiramisu is an exact truthmaker of the statement “Dominik is eating tiramisu.” But the complex state of Dominik eating tiramisu and zabaglione is not an exact truthmaker of the statement, since it contains as a part the state of Dominik eating zabaglione, which is irrelevant to the truth of “Dominik is eating tiramisu”.

To interpret permissions, we associate with every state a set of states which are rendered admissible if the state obtains: every state comes with its sphere of permissions. For example, consider the state of John of checking in at the airport. This state permits him to proceed to the gate, and thus, the state of John going to the gate is in the sphere of permissions of him checking in. Based on this idea, we suggest the following clause for the exact truthmakers of permissions:

(CP) A state a is an exact truthmaker of PA iff every exact truthmaker of A is in

the sphere of permissions of a.

In our talk, we shall develop this informal idea in full formal detail: (1) We will show that our semantics naturally validates the free choice permission principle. (2) We will explain why the Replacement rule fails in our semantics, and why it is a rule that should not hold due to even more general, philosophical reasons. (3) We will present a sound and complete axiom system with FCP as the only deontic axiom. (4) Finally, we will explain why our logic successfully withstands all the objections from the literature. This solves the problem of free choice permission.

References:

- [1] Hilpinen. “Disjunctive Permission and Conditionals with Disjunctive Antecedents.” *Acta Philosophica Fennica* 35: 175–94, 1982.
- [2] Kit Fine. Counterfactuals without possible worlds. *Journal of Philosophy*, 109(3):221–46, 2012.
- [3] Kit Fine. Permission and possible worlds. *Dialectica*, 68(3): 317–336, 2014.
- [4] Kit Fine. Truth-maker semantics for intuitionistic logic. *Journal of Philosophical Logic*, 43(2-3): 549–77, 2014.
- [5] Kit Fine. “Angellic content. *Journal of Philosophical Logic* 45(2): 199–226, 2016.
- [6] Sven Ove Hansson. The varieties of permission. In Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre, editors, *Handbook of Deontic Logic and Normative Systems*, pages 195–240. College Publications, 2013.
- [7] Hans Kamp. Free Choice Permission. *Proceedings of the Aristotelian Society*, 74: 57–74, 1973.
- [8] A. Ross. Imperatives and Logic. *Theoria*, 7:53–71, 1941.
- [9] Bas Van Fraassen. Facts and tautological entailments. *The Journal of Philosophy* 66: 477–487, 1969.

Indigenous tourism, recognition, and wellbeing

Matthias Kramm (Utrecht University)

Can indigenous tourism be a positive factor for the wellbeing of indigenous peoples? There exist three different criticisms of such tourism projects: Firstly, the economic criticism that the income of indigenous tourism does not empower indigenous peoples, but rather enriches third parties. Secondly, the ecological criticism that indigenous tourism causes more harm to the environment than can be made up for by using the acquired resources. And thirdly, the cultural criticism that the staging of cultural tradition for outsiders is inauthentic. In this paper, I will defend the normative view that indigenous tourism can nevertheless contribute to indigenous wellbeing.

The behaviour and preferences of tourists have undergone significant changes during the last decades. The world market shares of Africa, the Middle East, and of Asia and the Pacific region have been growing from 1980 to 2010 and will most probably grow further during the coming decade (World Tourism Organization 2012). In the course of this process, indigenous cultural traditions have become an attractive destination for a certain group of tourists who are looking for an “authentic” experience or for cultural immersion. But can indigenous tourism be a positive factor for the wellbeing of indigenous peoples?

In order to answer this question I will engage from a normative point of view with anthropological, sociological, and economic literature. We can distinguish between three different criticisms of indigenous tourism projects: Firstly, the economic criticism that the income of indigenous tourism does not empower the indigenous peoples themselves, but rather enriches third parties like middlemen and state agencies (Cheer, Reeves, and Laing 2013; Crystal 1989). Secondly, the ecological criticism that indigenous tourism causes more harm to the environment than can be made up for by using the newly acquired financial resources (Mitchell 1998; Frick McKean 1989). And thirdly, the cultural criticism that the staging of cultural tradition for outsiders is inauthentic and will lead in the long run to a corrosion of indigenous cultural identity (MacCannell 1976; Greenwood 1989). In this paper, I will defend the normative view that under certain conditions indigenous tourism can nevertheless contribute to indigenous wellbeing.

My argument will proceed in two steps. In a first step I will address the economic, the ecological, and the cultural criticism. I will argue that indigenous cultural tradition can become a commodity which can be sold to tourists without necessarily becoming inauthentic and without resulting in cultural erosion (Cole 2007; Scheyvens 2002; Bunten 2008). Indigenous tourism can also increase the social cohesion within the community once its members cooperate with each other

for a common tourism initiative. Last but not least, indigenous tourism can strengthen the community's self-identity and pride once its members are able to reassert their traditional cultural identity (Comaroff and Comaroff 2009).

In a second step I will specify a list of conditions which have to be in place before indigenous tourism projects can be conducive to indigenous wellbeing. The three most important questions which have to be addressed are the distribution of the benefits, the avoidance or mitigation of environmental erosion, and the assurance of cultural ownership. Cultural ownership means that indigenous communities retain the right and authority to present themselves to the tourist instead of merely being presented by other agents (Kirtsoglou and Theodossopoulos 2004). Further conditions are the education and training of indigenous communities who take part in the tourist industry, the creation of linkages in terms of public transport or communication, the recognition of legal rights to ancestral lands, the establishment of regulating bodies to protect local cultures, and the establishment of policies which recognize ownership, access to and economic support for tourism development (Butler and Hinch 2007; Fletcher, Pforr, and Brueckner 2016). If these conditions are taken into account, indigenous tourism projects can in principle be empowering for the local community.

References:

- Bunten, Alexis Celeste. 2008. 'Sharing Culture or Selling Out?' *American Ethnologist* 35 (3): 380–95. Butler, Richard, and Thomas Hinch. 2007. *Tourism and Indigenous Peoples: Issues and Implications*. Amsterdam; Boston, Mass.: Butterworth-Heinemann.
- Cheer, Joseph M., Keir J. Reeves, and Jennifer Laing. 2013. 'Tourism and Traditional Culture: Land Diving in Vanuatu'. *Annals of Tourism Research* 43: 435–55. Cole, Stroma. 2007. 'Beyond Authenticity and Commodification'. *Annals of Tourism Research* 34 (4): 943–60. Comaroff, John L., and Jean Comaroff. 2009. *Ethnicity, Inc.* Chicago, London: University of Chicago Press.
- Crystal, E. 1989. 'Tourism in Toraja (Sulawesi Indonesia)'. In *Hosts and Guests: The Anthropology of Tourism*. Philadelphia: University of Pennsylvania Press.
- Fletcher, Campbell, Christof Pforr, and Martin Brueckner. 2016. 'Factors Influencing Indigenous Engagement in Tourism Development: An International Perspective'. *Journal of Sustainable Tourism* 24 (8–9): 1100–1200.

Frick McKean, Philip. 1989. 'Towards a Theoretical Analysis of Tourism: Economic Dualism and Cultural Involution in Bali'. In *Hosts and Guests: The Anthropology of Tourism*. Philadelphia: University of Pennsylvania Press.

Greenwood, Davydd J. 1989. 'Culture by the Pound: An Anthropological Perspective on Tourism as Cultural Commoditization'. In *Hosts and Guests: The Anthropology of Tourism*. Philadelphia: University of Pennsylvania Press.

Kirtsoglou, Elisabeth, and Dimitrios Theodossopoulos. 2004. "'They Are Taking Our Culture Away': Tourism and Culture Commodification in the Garifuna Community of Roatan'. *Critique of Anthropology* 24 (2): 135–57.

MacCannell, D. 1976. *The Tourist: A New Theory of the Leisure Class*. New York: Schocken. Mitchell, Clare J. A. 1998. 'Entrepreneurialism, Commodification and Creative Distruction: A Model of

Post-Modern Community Development'. *Journal of Rural Studies* 14 (3): 273–86. Scheyvens, Regina. 2002. *Tourism for Development: Empowering Communities*. Harlow, London, New

York: Prentice Hall. World Tourism Organization. 2012. 'Tourism towards 2030: Global Overview'. Edited by Carlos

Vogeler. 54th Meeting of the UNWTO Commission for the Americas.

How epistemic injustice can deepen disagreement

Thirza Lagewaard (Vrije Universiteit Amsterdam)

I want to contribute to a better understanding of 'deep disagreement' by arguing that sometimes, disagreements are deepened due to epistemic injustice. I review the literature on the nature of deep disagreements. Then, I explore a case of deep disagreement: the debate in the Netherlands about racism. I argue that this dispute on racism should be understood as a deeper disagreement, because there is disagreement about what counts as evidence for the claim that racism is an significant issue in the Netherlands, due to both testimonial injustice (dismissal of expert testimony) and hermeneutical injustice (no uptake of relevant epistemic resources).

In this paper, I want to contribute to a better understanding of what has been called deep disagreement by arguing that sometimes, disagreements are deepened due to epistemic injustice.

Deep Disagreement

Firstly, I review the small literature on the nature of *deep disagreements*. Deep disagreement should be differentiated from ‘normal’ peer disagreement. Normally, when you disagree with a peer about *p*, there is a lot of background agreement, on what counts as good evidence for and against *p*. This is not the case with deep disagreement.

Deep disagreement can be defined as disagreement about fundamental epistemic principles which involves disagreement about what counts as evidence and/or justified belief regarding a certain domain (Lynch 2010, Kappel 2017). You don’t just disagree about *p*, but about *what counts as evidence for p*.

Racism and Black Pete

Then, I will introduce and explore a case study of real-life deep disagreement: the debate in the Netherlands about racism. A majority in the Netherlands holds the view that racism is not an significant issue in the Netherlands. A minority disagrees. I illustrate this disagreement about racism with the hotly debated figure of Black Pete, the black-faced, obedient helper of the St. Nicholas, the Dutch equivalent of Santa Claus.

A deeper disagreement: disagreement over evidence

I argue that this dispute on racism should be understood as a deeper kind of disagreement, because there is disagreement about what counts as evidence for the claim that racism is an significant issue in the Netherlands, due to epistemic injustice.

There seems to be disagreement about whether (A) *private first-personal experience* of racism should be ranked highly as an epistemic principle in this domain (higher than third-personal experience of white people) and related to this whether (B) *expert testimony* should be applied as an epistemic principle in this domain.

As has been argued before, the first personal experience from a group that experiences oppression is a distinct source of evidence because the personal experience from this minority is fundamentally different from the experience of the majority. The majority cannot have the same first-personal experience because they are not subjected to racism. For someone who doesn’t experience racism, the testimony of those who have experienced racism is a very important way to learn about racism. People who experience racism are arguably experts on the racism question, as they have knowledge other people cannot (so easily) have.

As the majority cannot access this knowledge directly, their main source of evidence about racism is *expert testimony*.

Epistemic Injustice

However, only part of the people who do not experience racism, take expert testimony on first personal experience of people of color to be a good source of evidence for the claim that racism is a problem in the Netherlands.

I argue at length, drawing on Fricker (2007), Dotson (2014) and Medina (2017) that both testimonial injustice and hermeneutical injustice can lead to the deepening of a disagreement because of the dismissal of the actual experts.

Due to testimonial injustice based on racial prejudices, testimony of people who are arguably experts in this domain, is not taken as evidence in this domain.

In such a case there is disagreement about what counts as evidence for the claim that racism is a significant issue, and this deepening is partly due to testimonial injustice. The disagreement becomes harder to resolve.

Something similar happens with hermeneutical injustice. Dotson (2014) expands the notion of hermeneutical injustice: it's not just that people might lack epistemic resources for interpreting and communicating their experiences. In fact, they might have very rich and powerful resources for doing so. Even so, however, testimony using these resources might fail to communicate adequately to the dominant knowers.

Although things are changing in recent years, racism is barely addressed in the Netherlands. Because racism supposedly is not an issue, there is not a lot of debate about it (Wekker 2017). Because of this, there is no shared vocabulary to talk about racism in the Netherlands.

This is an instance of hermeneutical injustice because the group that is subject of racism has no way to discuss racism in the Netherlands in a constructive way. Because of this, it becomes very hard to exchange epistemic reasons. The majority can deny that there is anything to discuss or to testify on for people of color in this domain.

Conclusion

When the non-dominant group's testimony is given insufficient credibility and their proposed epistemic resources are dismissed or misunderstood, the disagreement becomes deep. It has become deep because there is disagreement about what counts as evidence. There ceases to be a dispute-independent way to settle the disagreement. What seems to be going on in this case, and presumably in others like it, is that parties disagree (1) which exact sources of evidence to take seriously in the case at hand, and (2) the proper way to apply the relevant epistemic principles and norms to particular cases (who's experiences have what sort of probative force). Because both the experiences and the epistemic resources are not

shared, a non-dominant knower lacks the tools to convince dominant knowers of the existence and nature of this experience. Because of this, there is no agreement between the groups about what counts as a valid contribution to the debate.

I conclude that epistemic injustice can lead to a form of deep disagreement that is just as intractable, entrenched, and difficult to resolve as one that involves basic epistemic principles and norms. Understanding such injustice-based deep disagreements seems to me to be a project not only of theoretical interest, but also of great practical social relevance.

References:

- Dotson, K. (2014). Conceptualizing epistemic oppression. *Social Epistemology*, 28(2), 115-138.
- Fricker, M. (2007).
- Miranda, F. (2009). Epistemic Injustice: Power and the Ethics of Knowing.
- Kappel, K. (2012). The Problem of Deep Disagreement. *Discipline Filosofiche* 22.2 (2012): 7-25.
- Lynch, M. (2010). Epistemic Circularity and Epistemic Incommensurability. In *Social Epistemology*. A. Haddock & A. Millar and D. Pritchard, Oxford University Press: 262- 77.
- Medina, J. (2017). Varieties of hermeneutical injustice. In *The Routledge Handbook of Epistemic Injustice*. Taylor and Francis.
- Pohlhaus, G. (2012). Relational knowing and epistemic injustice: Toward a theory of willful hermeneutical ignorance. *Hypatia*, 27(4), 715-735.
- Wekker, G. (2017). *Witte onschuld: paradoxen van kolonialisme en ras*. Amsterdam University

Falsifying generic stereotypes

Olivier Lemeire (Katholieke Universiteit Leuven)

Generics are generalizing statements that are not explicitly quantified, like “Dogs bark” or “Birds fly”. The truth-conditions of these statements have puzzled researchers for decades. Recently, a proposal by Sarah-Jane Leslie has become very popular. Her proposal explicitly aims to account for the troubling aspects of generic stereotypes, like “Muslims are terrorists” or “Black people are lazy”. In this paper, I argue that there are several counter-examples to Leslie’s proposal. I propose a different set of truth-conditions that does allow us to falsify generic

stereotypes.

Generics are generalizing statements that are not explicitly quantified, like “Dogs bark” or “Birds fly”. The truth-conditions of these statements have puzzled researchers for decades. To see why they are so puzzling, consider that generics can be true even though only a minority of the members of the kind actually instantiates the property. “Birds lay eggs” is a true generic, for example, even though only the adult females do. Some generics are even true although only a very small minority has the property. And statements like “Sharks attack bathers” and “Ticks carry Lyme disease” are true although they only apply to 1% of the population.

Recently, Sarah-Jane Leslie has proposed a set of truth-conditions for generics that has attracted a lot of attention. Part of the reason why her proposal is so popular is that she explicitly aims to account for the troubling aspects of stereotypes that are generically formulated, like “Muslims are terrorists” or “Black people are lazy”. According to her, the generic “Fs are G” is true iff:

The counterinstances are negative, and: If F lies along a characteristic dimension for the Ks, then some Ks are F, If F is a striking property, some Ks are F and the others are disposed to be F, Otherwise, almost all Ks are F.

In this paper I discuss three problems with these truth-conditions when applied to generic stereotypes. Firstly, Leslie’s proposal would count a generic like “Gay people suffer from homophobic violence” as false. Secondly, one could not truthfully respond to someone uttering “Women are submissive” by saying “That’s not true. Women are not submissive.” According to Leslie, negations take small scope in generics, and so one would read the above negation as something like “It is characteristic for women not to be submissive”, which is doubtful. Instead, what one aims to say is that it is not characteristic for women to be submissive, and so the negation should take wide scope. Thirdly, Leslie cannot hold that “Muslims are terrorists” is false without interpreting her notion of a ‘disposition’ so strongly that “Ticks carry Lyme disease” is false too. Thus here truth-conditions depart from our intuitive truth judgments.

In this paper I propose a different set of truth-conditions. A generic of the form “Fs are G” is true iff:

All three of the following conditions hold: A. It is not a stable fact that no F is G.

B. It is not a stable fact that all things that instantiate an alternative to G, are F. C. It is not the case that some Fs instantiate an alternative to G *because* they are F.

And at least one of the following three conditions holds: A. A large majority of F

instantiate G. B. A large majority of things that instantiate G are F. C. Some F's are G *because* they are F.

Not only does this set of truth-conditions solve the problems with Leslie's proposal, it also results in two different strategies to respond to generic stereotypes. A first strategy is to argue that none of the final three conditions hold. A second strategy is to argue that one of the first three conditions hold. Both strategies are sufficient to falsify a generic stereotype.

The Extensional Constitution View

Martin Lipman (Leiden University)

According to the constitution view, a lump of clay is distinct from the statue that is crafted out of the clay. This is generally taken to conflict with classical mereology. Classical mereology accepts an extensionality principle according to which distinct objects cannot have the same parts. Adherents of the constitution view typically assume that the statue and the clay have the same parts (and are nevertheless distinct), so that we appear to have a counterexample to the extensionality principle. In this talk I will explore a relatively neglected alternative: namely that the statue and the co-located lump are distinct objects but do not share any parts. I will introduce the extensional constitution view in some detail and defend it against potential worries.

An artisan has some lumps of clay lying in the cupboard. First, she puts the bits together into a single lump of clay and then she crafts a beautiful statue. The constitution view holds that the created statue is not identical to the lump of clay because the lump existed before the statue did and the lump survives severe changes in shape that would destroy the statue. On its turn, the lump would not survive the destruction of some of its bits of clay, whereas the statue would survive this. There are therefore two co-located objects, the statue and the lump of clay, differing in their temporal profile and essential properties. I will call this the constitution view.

If the constitution view is true, what are its consequences for our mereology? Classical mereology is a theory of parthood that underwrites extensionality and uniqueness:

Extensionality: if x and y have the same proper parts, then $x = y$.

Uniqueness: if x and y are a fusion of the same things, then $x = y$.

If one thinks that the statue and the lump have the same proper parts, or that there is some collection of parts that they are both composed of (say the clay particles), then this implies that the statue is identical to the lump of clay, contradicting the constitution view. It seems therefore that the constitution view is at odds with classical mereology.

One response to this conflict is to explore a non-classical mereology. There are various routes to non-extensional mereologies that deny extensionality, such as via the rejection of supplementation principles (Simons 1987: Ch. 9) or via the rejection of the anti-symmetry of parthood (Cotnoir 2010). Alternatively, one could uphold classical mereology (in letter) by positing forms or structural parts and take the statue and the lump to have the same material parts but different forms as parts (Koslicki 2008; cf. Fine 1999; Paul 2006; Johnston 2006). For the purpose of this essay, I will lump all these views together. Let a material decomposition of a whole be a plurality of material proper parts such that any further material part overlaps some of them. These theories share the general view that there are pairs of distinct material objects that have a shared material decomposition, for instance, they all assume that the statue and the clay have a shared material decomposition, namely into bits of clay. Call this the constitution view.

There is a natural but relatively neglected alternative. To see the alternative, we need to think a little more carefully about the supposed parts. Consider the torso, the head, and the four limbs of the statue. Compare these parts to the torso-shaped lump, the head-shaped lump, and the four limb-shaped lumps of clay. The torso came into existence later than the torso-shaped lump of clay and the lump would survive deformations that would destroy the torso. Hence, the torso is not identical to the torso-shaped lump of clay. Similarly for the head and the limbs: the same sort of considerations imply that they are not identical to the corresponding lumps of clay. It's therefore entirely in line with the constitution view that these two pluralities of objects are distinct pluralities: the plurality of lumps of clay is not identical to the plurality consisting of the torso, head and four limbs of the statue on account of the differences in modal and temporal profiles of these parts.

If the plurality of lumps is indeed distinct from the plurality consisting of the torso, head and four limbs, then it is furthermore natural to suppose that the statue is composed by the torso, head and limbs and not the corresponding lumps of clay, whereas the big lump of clay is composed from the smaller lumps of clay and not the corresponding body parts of the statue. These two pluralities, given that they are distinct, aren't counterexamples to extensionality or uniqueness: we have distinct but co-located pluralities of material parts composing distinct but co-located material wholes. The example suggests a more general view: the mereological decompositions of the statue and the lump never converge on a single shared plurality of parts from which both objects are composed. I will call it the extensional constitution view.

This talk will introduce the extensional constitution view in some detail (part one) and defend it against various potential worries (part two).

References:

- Cotnoir, A. (2010) 'Anti-symmetry and non-extensional mereology', *Philosophical Quarterly* 60 (239): 396-405.
- Fine, K. (1999). 'Things and Their Parts', *Midwest Studies in Philosophy* 23 (1): 61-74.
- Johnston, M. (2006). 'Hylomorphism', *Journal of Philosophy* 103 (12): 652-698.
- Koslicki, K. (2008). *The Structure of Objects*. New York: Oxford University Press.
- Paul, L. A. (2006). 'Coincidence as overlap', *Noûs* 40 (4): 623–659.
- Simons, P. (1987). *Parts: A Study in Ontology*. Oxford University Press.

Against Kantian Moral Relativism

Sem de Maagt and Rutger Claassen (Utrecht University)

Recently, several authors who identify themselves as Kantians also identify themselves as moral relativists (Velleman 2015; Flikschuh 2017). Kantian moral relativism combines the Kantian idea that morality is grounded in the first-person perspective of an agent, with the moral relativist idea that there are different conceptions of agency and therefore different (correct) moralities. In this paper, we argue that any conception of agency presupposes a more fundamental form of agency which is necessary to make behaviour truly our own, and that this fundamental conception of agency is sufficiently thick to form the basis of a substantive, universal morality.

Kantian ethics is typically associated with the idea of universal morality: there are moral principles and/or norms that everyone has to accept independent from their contingent desires, preferences or membership of a specific community. Kantian ethics tries to vindicate this universal morality by showing that any agent necessarily has to accept certain moral principles from her first-person perspective as an agent, where agency is understood in terms of acting for reasons and/or pursuing one's freely chosen purposes. The resulting moral principles are typically thought to have a broadly liberal egalitarian content. Christine Korsgaard, for instance, claims that her theory entails that "enlightenment morality is true" (Korsgaard 1996, 123). Alan Gewirth (1978, 1996) proposes an extensive set of

human rights, including both negative and positive rights and both civil and political and socio-economic rights.

Recently, however, several authors who identify themselves as Kantians also identify themselves as moral relativists (Velleman 2015; Flikschuh 2017). Kantian moral relativism combines the Kantian idea that morality is grounded in the first-person perspective of an agent, with the moral relativists idea that there are different conceptions of agency and therefore different (correct) moralities.

David Velleman, for instance, argues that in some cultures, such as the Sherpa culture, persons understand themselves first and foremost in terms of social roles. Subsequently, he claims that if one understands oneself in terms of a social role conception of agency, one is not committed to a universal morality, let alone one with liberal content. Instead, someone who understands herself in this way is only committed to the normative requirements following from this specific conception. Or so Velleman argues.

In this paper, we argue that one can be a moral relativist or a Kantian, but never at the same time. More specifically, we argue that different conceptions of agency differ less radically than Kantian moral relativists suggest, and that what these conceptions have in common - that what makes these different conceptions, conceptions of *agency* – is sufficiently thick to lead to universal moral conclusions. We argue for this conclusion by analysing a social role conception of agency and by showing that in order to make behaviour truly our own a social role conception of agency needs to presuppose a more fundamental form of agency for the creation, interpretation and authorization of social roles.

The underlying ambition of the paper is to clarify the conception of agency which plays a central role in Kantian ethics and to provide a response to a recurring objection to Kantian ethics concerning the ambiguity of the Kantian conception of agency (see e.g. Regan 1999; Tiffany 2012; Schafer 2015). The objection is that Kantian ethics either relies on an inescapable, thin conception of agency, but this conception of agency does not entail any substantive moral commitments. Or it relies on a *thicker* conception of agency, but this conception only generates moral conclusions for those persons who happen to share this contingent self-understanding. In other words, the dilemma is that Kantian ethics either leads to a powerless, empty universalism or to moral relativism. In the paper we focus on Kantian moral relativism specifically because we take it to be the strongest instantiation of the second horn of the dilemma.

The structure of the paper is as follows. In section 1, we introduce Kantian moral relativism as a challenge to Kantian moral universalism by reconstructing recent argument for Kantian moral relativism by David Velleman and Katrin Flikschuh. In section 2, we analyse a role conception of agency and we argue that this conception of agency presupposes more fundamental forms of agency for the

creation, interpretation and authorization of social roles. In section 3, we argue that these fundamental, universal features of agency are sufficiently thick to form the basis of a substantive, universal morality.

Can a Robot be a Good Colleague?

Sven Nyholm (Eindhoven University of Technology) and Jilles Smids (Eindhoven University of Technology)

There are people who treat robots they work alongside in ways that suggest that they value these robots as they might value good human colleagues. Similarly, there are people who want to have robots as their friends or romantic partners. Philosophers of technology have discussed whether robots can be our friends or our romantic partners, but not whether robots can be good colleagues. In this presentation, we will discuss the question whether robots could be good colleagues. In doing so, we will compare this question to the questions of whether robots can be our friends or romantic partners.

When Boomer “died” in the battlefield in Iraq, the soldiers in his team gave him an improvised military funeral. They also gave him two medals of honor: a Purple Heart and a Bronze Star. These soldiers regarded Boomer as a highly valued team-member and good colleague. Unlike the rest of the team, however, Boomer was not a human being. Boomer was a robot, whose job was to seek out and disarm bombs. Prior to his destruction, Boomer had saved many lives. Not only was he a life-saver, Boomer’s coworkers also thought he had “developed a personality” of his own (Carpenter 2016). So, it is perhaps no wonder that Boomer was given these honors when he was destroyed and that he was regarded as a good colleague worthy of an honorable funeral.

This is a real-life story, not a philosophical thought experiment. But this real-life story does raise philosophical questions. For example, can a robot be a good colleague? This is the question we will be discussing in this presentation. As more and more robots are entering the workplace across various different domains, this is a question we need to take seriously. Whether we are concerned with military robots like Boomer, care robots in a hospital, logistics robots in a warehouse setting, or any other robot in the workplace, the future of work satisfaction and meaningful human work partly depend on whether robots can be good colleagues. In general, one of the most important factors that determine people’s work satisfaction and whether they find their work meaningful has to do with being a member of a team with good colleagues. Accordingly, as more and more robots are introduced into the workplace, this prompts the question of whether – like human colleagues – some robots can be good colleagues that can help to make

work meaningful and satisfying.

Our aim in the presentation is not to give a definite and categorical “yes” or “no” answer that applies to all cases in the same way. Rather, our aim is to consider how to philosophically approach this question of whether a robot can be a good colleague and to take some initial steps towards answering this question. In discussing this topic, we will compare the question of whether a robot can be a good colleague to the somewhat more widely discussed questions of whether a robot can be a friend and whether a robot can be a romantic companion (e.g. Elder 2017; Nyholm & Frank 2017). We come down in favor of the following conditional and (we hope) nuanced thesis: if people are willing to regard some of their robotic collaborators as “colleagues”, robots are likely to soon start acquiring capabilities that could also make them fit to be regarded as good colleagues by at least some of the key standards we typically apply to good human colleagues. Whether a robot can be as good of a colleague as a human can be or not, it is comparatively speaking “easier” for a robot to be a good colleague than it is for a robot to be a friend or a loving romantic partner. That is to say, the types of capacities and the forms of agency a robot needs to acquire in order to live up to typical criteria people often associate with a good colleague are easier to realize in a robot than are the criteria typically associated with being a true friend or a loving romantic partner.

In this presentation, we will start by first saying a little more about what we mean by our main question and why we think it is an important question (section 1). We will then briefly look at how philosophers of technology have recently approached the questions of whether robots can be our friends and whether there can be mutual love between humans and robots (section 2). This will provide us with the general method we will use to focus on our case in point, viz. whether a robot can be a good colleague. Our next step will be to first set out some commonly recognized criteria for being a good colleague, which seem to us to apply to most lines of work. As we do this, we will also consider what types of capacities and/or what forms of agency are involved in realizing these criteria (section 3). We then turn to the question of whether robots can live up to these criteria for being a good colleague. As we do so, we will make comparisons between this case and robots’ ability (or lack thereof) to be a friend or a romantic partner. As already indicated, this will lead us to the tentative conclusion that, at the very least, it is easier for a robot to be a good colleague than it is for a robot to be a true friend or romantic partner (section 4). We will end by considering whether there might be something morally problematic about creating robots with capacities that would make them into good colleagues. We will briefly mention three different possible perspectives on this issue (section 5).

References:

Carpenter, J (2016): Culture and Human-Robot Interaction in Militarized Spaces: A War Story

(Routledge)

Elder, A (2017): *Friendship, Robots, and Social Media: False Friends and Second Selves* (Routledge)

Nyholm, S & Frank, L (2017): "From Sex Robots to Love Robots: Is Mutual Love with a Robot Possible?" in J. Danaher & N. McArthur (eds.) *Robot Sex: Social and Ethical Implications* (MIT Press)

Human Rights and Motivation

Jos Philips (Utrecht University)

This paper discusses what I call the motivational challenge to human rights: can human rights be effectively enforced in societies – also across generations and in acceptable ways? Human rights may be unsuitable as a global moral ideal if they cannot; but I argue that they ultimately can. More precisely, I hypothesize that when people are secure in their very important interests being met, they will not actively violate human rights nor support institutions or people who do. This may, if there is also an avant-garde of professionals and citizens, be enough to uphold a human-rights realizing society across generations.

This paper will discuss a particular challenge to human rights, which I will call the *motivational challenge*. Motivation is understood here in the sense of institutions and individual people *actually* being moved into action. Human rights are understood as, in the first place, the minimum standards of global justice, and I argue that the content of these standards has primarily to do with the protection of very important interests of all human beings (cf. Shue 1996; Nussbaum 2000). The idea of the paper is that for human rights to be plausible as a global justice ideal, it must be possible for societies actually to function in accordance with human rights. Moreover, they should be capable of so functioning from generation to generation and achieve this by acceptable means, for example without using manipulation or certain kinds of coercion. At the minimum, all this should be the case for societies that are as good as human societies can be: if even such societies cannot function in accordance with human rights, human rights certainly forfeit their plausibility as a global ideal. I will be looking for the beginnings of robust evidence that might underpin that human rights indeed have such motivational power.

I will start with Richard Rorty's remark that to establish and maintain a human-rights culture, it is enough to have generations of 'nice, well-off... secure... students' (Rorty 1994, p. 127). This suggestion has, I will argue, a number of problems, and improvements and alternatives will be considered. I will, in order to keep things manageable, confine myself to the motivation of individuals, although a society's

functioning in accordance with human rights is importantly also an institutional matter. Institutions such as governments, and business companies and (other) organisations in the civil society as well, should be moved to actually respect and realize human rights; but to maintain these institutional motivations it is, I will argue, both a necessary condition, and in the end also a sufficient one, that individuals support human rights.

Ultimately I will, drawing importantly on a recent of article in *Ethics* by Allen Buchanan and Russell Powell (2016), suggest that when people are secure in their more elementary and more complex very important interests being met, they will in certain ways act to uphold a human rights-realizing society: they will not actively violate human rights and support institutions and people who realize them over alternatives. This may be enough to uphold a human-rights realizing society across generations, if there is also an avant-garde of professionals and a number of citizens (cf. Ypi 2012), and if there are mechanisms for passing down—for example through stories told at schools and in other groups— a minimal understanding of the importance of human rights realization. I will also investigate whether a human-rights abiding society is upheld here in an acceptable manner, that is to say among other things (as will be argued), that such a society is upheld while allowing people to conceive of themselves as acting for reasons. Furthermore, I will briefly consider some other hypotheses concerning individual motivation, hypotheses which draw, among other things, on nudging and on the role of integration of individuals in groups (cf. Lichtenberg 2014). I will argue that some of these hypotheses hold promise as well.

References:

Buchanan A, Powell R (2016) Toward a Naturalistic Theory of Moral Progress. *Ethics* 126: 983-1014

Lichtenberg J (2014) *Distant Strangers: Ethics, Psychology, and Global Poverty*. Cambridge: Cambridge University Press

Nussbaum M (2000) *Women and Human Development: The Capabilities Approach*. Cambridge University Press, Cambridge

Rorty R (1994) Human Rights, Rationality and Sentimentality, in Shute S, Hurley S eds., *On Human Rights: The Oxford Amnesty Lectures 1993*, Basic Books, New York, 111-134

Shue H (1996) *Basic Rights*. Princeton: Princeton University Press (orig. 1980)

Ypi L (2012) *Global Justice and Avant-Garde Political Agency*. New York: Oxford University Press

Post-Truth Politics as Collective Gaslighting

Natascha Rietdijk (Tilburg University)

Ever since the 2016 US presidential election, journalists, scientists and philosophers have sought to understand the phenomenon known as post-truth politics. However, the question that has remained underexplored is, what the implications of this political rhetoric are for epistemic agency. I will argue that post-truth politics is best understood as a form of grand-scale gaslighting. It manipulates victims' trust and self-trust in order to epistemically isolate them. I conclude that both gaslighting and post-truth politics do not just violate, but in fact also systematically erode their victims' epistemic autonomy.

Ever since the 2016 US presidential election, journalists, scientists and philosophers have sought to explain the rise of the phenomenon known as *post-truth politics*. Some have investigated the origins of fact-resistance, others have called Frankfurtian *bullshit*, and still others have diagnosed the “post-truth era” as the death of democracy (Wikforss 2017; Kristiansen and Kaussler 2018; Chugrov 2017). Though it has been argued that the term “post-truth” is rhetorical and conservative (Habgood-Coote 2018), the factual existence of political discourses that exhibit a lack of concern for the truth is undeniable – and epistemically problematic. Leaving the question of its historic origins and psychological foundations aside, I aim to study this political rhetoric from a so far underexplored perspective: that of its implications for epistemic autonomy. Specifically, I will argue that some of the features of post-truth politics and its effects on agency are best understood in terms of a grand-scale form of gaslighting. In other words, on my view we need the concept of gaslighting in order to fully make sense of the post-truth phenomenon.

Gaslighting, after the 1938 Patrick Hamilton play *Gas Light*, is a type of manipulation aimed at having the victim doubt their own judgment, perception, and sense of reality. At its worst, this can result in the complete eradication of the victim's epistemic self-trust, leaving them to question even their own sanity. Psychologists have been studying the phenomenon since the late 1960s. More recently, journalists have speculatively connected powerful political actors like the Trump administration and the Kremlin to gaslighting practices (Caldwell 2016; Ghitis 2017; Carpenter 2018). These opinion piece accusations are typically partisan and rather sweeping in nature, but I believe there is evidence to support their core insight. There is an apparent parallel between the features that distinguish post-truth politics from more traditional forms of propaganda and those that distinguish gaslighting from general manipulation. As I will argue, both do not simply instill a (possibly false) belief in their victim, but instead challenge

the victim's ability to distinguish between truth and falsehood itself. Both do not only violate the victim's autonomy, but actually systematically erode it and its resources.

After a brief introduction into the salient features of either phenomenon, I will further outline the striking resemblances between gaslighting and post-truth politics. Besides the common aim of undermining epistemic autonomy in order to consolidate power, they also share many specific methods used to achieve that end. These include the discrediting of opponents and critics (by accusing them of, e.g. spreading "fake news" or being delusional), the abuse and exacerbation of the victim's vulnerability, the use of deception and subsequent denial, and the overall sowing of confusion. All these various methods serve to epistemically isolate the intended audience. Through manipulation of trust, both gaslighters and post-truth-political actors can reinforce their own alternative narratives, to the detriment of their critics and victims.

Having delineated the parallel between both phenomena, I will consider two possible objections against their correspondence. Firstly, while gaslighting is commonly thought of as an intentional form of manipulation, many instances of post-truth politics are perhaps not as systematic. Secondly, while gaslighting primarily undermines a victim's self-trust, post-truth politics, by promoting echo chambers, actually often enhances it. I believe that, while these objections should give us pause in labeling certain instances of post-truth rhetoric as gaslighting too readily, they can in most cases be answered satisfactorily. Moreover, the benefit of considering post-truth politics as having at least an overlap with gaslighting is that it enhances our insight into its effects on epistemic agency. By manipulating agents' epistemic self-trust and their trust in those resources which might restore it to a justified level, post-truth politics systematically impairs their epistemic autonomy. Thus, in interpreting salient features of post-truth politics as species of gaslighting, I hope to offer new perspectives on the harms of such rhetoric, and thereby also on possible ways to identify and resist the manipulative techniques it employs.

References:

- Caldwell, Adam. 2016. "How Russia Successfully Gaslighted the West." Huffington Post, December 16, 2016. https://www.huffingtonpost.ca/adam-caldwell/the-gaslighting-of-the-west_b_13657466.html.
- Carpenter, Amanda. 2018. *Gaslighting America: Why We Love It When Trump Lies to Us*. New York, NY: Harper.
- Chugrov, Sergei. 2017. "Post-Truth: Transformation of Political Reality or Self-

Destruction of Liberal Democracy?” *Polis Political Studies* 2: 42–59.

Ghitis, Frida. 2017. “Donald Trump Is ‘Gaslighting’ All of Us.” CNN. January 16, 2017. <http://edition.cnn.com/2017/01/10/opinions/donald-trump-is-gaslighting-america-ghitis/index.html>.

Habgood-Coote, Joshua. 2018. “Stop Talking about Fake News!” *Inquiry*, August, 1–33.

Kristiansen, Lars, and Bernd Kaussler. 2018. “The Bullshit Doctrine: Fabrications, Lies, and Nonsense in the Age of Trump.” *Informal Logic* 38: 13–52.

Wikforss, Åsa. 2017. “Resisting the Facts.” Presented at the OZSW Conference, Doorn, November 11.

Nudges and Freedom

Philip Robichaud (Vrije Universiteit Amsterdam)

Nudges are policy tools that governments and private parties can use to predictably influence our choices. A crucial question is whether nudges impact our freedom? Proponents of ‘libertarian paternalism’ claim that they don’t. In this paper, I propose answers to the following questions: (1) Do nudges that target heuristics and biases make us unfree in some way or to some degree? (2) Even if we grant that nudged agents are by and large free, do nudges affect the value of the exercise of their agency? In other words, do we have axiological reasons to prefer non-nudged exercises of our agency?

Nudges are a new tool that governments and private parties can use to predictably influence the choices we make. A central question in discussions about the ethics of nudging is whether nudges impact freedom, and a common answer is that they don’t. Thaler and Sunstein have famously defended the libertarian paternalist credentials of nudges. The inclusion of ‘libertarian’ is supposed to indicate that the nudged agent’s freedom is not relevantly compromised.

But, might there be reason to think that the nudger’s intervention in the choice situation nevertheless affects the freedom of the nudgee in some way? Does the fact that the choice situation is intentionally constructed such that some option will be favored if the target heuristics and biases work as they are supposed to give us some reason to think that the nudgee was made unfree in some way or to some degree? A related question: granting that nudges are free to choose the nudged or

non-nudged option, does the fact that their choice was nudged affect the value of the exercise of their agency? In other words, do we have axiological reasons to prefer that exercises of our agency in the absence of nudges? These questions have been neglected in the expansive literature on nudges, and in this paper, I propose answers to them that should enrich the discussion both about the nature of freedom and the factors that bear on the moral permissibility of nudges.

The first task is to specify the kind of freedom that's relevant for this discussion. According to many prominent theories of freedom, impediments to freedom, so called unfreedom-producing constraints, are an all or nothing affair. Consider the following view defended by Shnayderman:

Responsibility view of unfreedom: C is an unfreedom-producing constraint on A's freedom to F iff C prevents A from F-ing and some agent (individual or corporate) is morally responsible for C in the sense that she is either praise- or blameworthy for the act which either is identified with or results in C (2013).

According to this theory, nudges don't make us unfree. The first conjunct on the right side says that something is an unfreedom-producing constraint only if it prevents the relevant agent from choosing certain options, which, nudges fail to do. Consider:

"Libertarian paternalism is a relatively weak and nonintrusive type of paternalism, because choices are not blocked or fenced off. In its most cautious forms, libertarian paternalism imposes trivial costs on those who seek to depart from the planner's preferred option" (2003, 1162).

Still, one might think that nudged agents are made unfree to a certain *degree* by the nudger's intervention. The idea here is that, short of preventing the nudgee from performing some action or other, the choice architect makes it the case that the agent is less free to perform the action that isn't favored by the choice architect. My task in the first section of the paper is to address an argument that freedom of the relevant kind *cannot* come in degrees. Steiner (1994, 78), Carter (1999, 228, 233) and Kramer (2002) have each defended the claim that the freedom to perform to perform some act is non-scalar and non-partitive. An agent is either free to act or she isn't. Just to take one example, Kramer argues that once we carefully attend to how certain barriers affect the difficulty in performing some action, we will see that agents are not thereby made free to a lesser degree (2002, 233–34). If these theorists are right, then the only way a nudge can affect one's freedom to act is by preventing the agent from acting otherwise, which isn't to nudge them at all. In reply, I show how these arguments don't apply to nudge cases. In particular, I argue that the way in which nudges make the performance of some option more difficult is to introduce certain factors internal to the agent's making of a choice. In a nudge case, the bias that is inclining the agent toward the choice architect's

favored option doesn't cease to exist as a result of any efforts made by the agent. Status quo bias, actually affects the agent's sensitivity to reasons—she effectively inflates reasons to stick with the status quo option. If this is right, then the question of whether the nudge diminishes the agent's freedom due to introducing some difficulty, such as the difficulty to make a choice that best conforms to one's antecedent preferences, cannot be handled according to Kramer's analysis.

In the final section of the paper, I turn to the axiological dimension. Here, I assume for the sake of argument that nudge interventions do not diminish freedom at all. This means that there is no difference between default-nudged agents and agents who made a non-nudged, prompted choice in terms of the degrees of freedom—they are both equally, fully free to choose one way or the other in the choice situation. I argue that, if they are equally free, then an agent who is interested in maximizing freedom would have no reason to prefer being presented with non-nudged, prompted choice situations rather than default nudge situations. This counterintuitive result can be avoided by accepting the claim that the value to the kind of agency that is exercised in nudge contexts is diminished relative to non-nudge contexts. I explicate what kind of value this is and what role such considerations of value can play in deontic debates about the permissibility of nudges.

References:

Carter, Ian. 1999. *A Measure of Freedom*. Oxford University Press. Kramer, Matthew H. 2002. "Why Freedoms Do Not Exist by Degrees." *Political Studies* 50: 230–43.

<https://doi.org/10.1111/1467-9248.00368>. Shnayderman, Ronen. 2013. "Social Freedom, Moral Responsibility, Actions and Omissions." *The*

Philosophical Quarterly 63 (253): 716–39. <https://doi.org/10.1111/1467-9213.12067>. Steiner, Hillel. 1994. *An Essay on Rights*. Blackwell publisher, Oxford, First Published, p64. Sunstein, Cass R., and Richard H. Thaler. 2003. "Libertarian Paternalism Is Not an Oxymoron." *The*

University of Chicago Law Review 70 (4): 1159–1202. <https://doi.org/10.2307/1600573>.

Unleashing Moral Progress

Hans Sauer (Utrecht University)

Standard evolutionary explanations seem unable to account for inclusivist shifts that expand the circle of moral concern beyond strategically relevant cooperators. Recently, Allen Buchanan and Russell Powell have argued that this shows that evolutionary conservatism – the view that our inherited psychology imposes significant feasibility constraints on how much inclusivist moral progress can be achieved – is unjustified. Secondly, they hold that inclusivist gains can be sustained, and exclusivist tendencies curbed, under certain favorable socio-economic conditions. I argue that Buchanan and Powell concede too much to the evolutionary conservative, because their second point shows that conservatives are right about the first: inclusivist shifts are unrealistic where it matters most, namely under harsh social, political and economic conditions. I suggest two promising strategies for solving this problem. One is to focus on different forms of moral progress to secure the same moral gains. The other is to look beyond possible extensions of our psychological capacities altogether, by providing institutional support that renders them irrelevant. We should bypass, rather than further stretch, the constraints of our evolved psychology to make moral progress possible.

Can research on psychopathy settle the debate between rationalism and sentimentalism?

Joost Schreuder (University of Groningen)

Whether and to what extent does research on psychopathy support or refute rationalist and sentimentalist theories of moral judgment? To address this question, I use causal graph theory to systematically assess the empirical support of two theories: the sentimentalist theory of Shaun Nichols (2002) and the rationalist theory of Heidi L. Maibom (2005). I conclude that currently available studies about psychopathy do not give us a reason to prefer one theory to another. Instead, evidence of another kind is needed: studies that examine the relation between performance on rational, affective and moral tasks.

The debate between rationalism and sentimentalism revolves around the following question: Which psychological capacities are necessary for the capacity for moral judgment? According to sentimentalists, affective capacities, such as the capacity for moral emotions or empathy, are required. Rationalists, on the other hand, claim

that rational capacities, such as the capacity for practical reasoning or self-control, are required instead. Recently, research on psychopathy has been used to support both rationalist and sentimentalist theories of moral judgment. Since people with psychopathy are assumed to have impaired moral capacities, the hope is that research on psychopathy will illuminate what causes those impairments and thus what moral capacities require.

Shaun Nichols (2002) was one of the first to use empirical evidence about psychopathy to support a theory of moral judgment. According to his sentimentalist theory, the capacity for moral judgment depends on the capacity for affective response. He argues that studies show that people with psychopathy are both impaired in affective response and in moral judgment, and that the former impairment is the best explanation for the latter one. However, this conclusion is criticized by Heidi L. Maibom (2005), who argues that there are also studies that show that people with psychopathy have rational deficits. On basis of these studies, she argues that the capacity for moral judgment depends on the capacity for practical reasoning. Subsequently, many more rationalist and sentimentalist theories of moral judgment have been proposed on basis of empirical evidence about psychopathy.

To make progress in the debate between rationalism and sentimentalism, we need to decide which theory of moral judgment has the best empirical support given currently available empirical evidence about psychopathy, and if the best theory cannot be decided, we need to determine which studies are yet to be done. To systematically assess empirical evidence, I will use causal graph theory (Pearl, 2009; Spirtes, Glymour & Scheines, 2000). One of the main aims of this theory is to provide tools to infer which causal models are compatible with a given body of empirical evidence. Using this theory, I propose the following method to systematically assess the empirical support of theories of moral judgment: (1) represent each theory of moral judgment as a causal model, (2) infer the associations entailed by each model, (3) compare each set of associations with the associations found by empirical evidence, (4) choose that model with the set of associations that has the best empirical support.

I will apply this method to the theories of Nichols (2002) and Maibom (2005) to show that, for even relatively straightforward theories of moral judgment, other empirical studies are needed than those cited to differentiate the two on empirical support. First, I will represent the theories as causal models. Since the theory of Nichols claims that affective response ultimately determines whether we are capable of moral judgment, this can be represented as a causal model where affective capacities directly cause the capacity for moral judgment. Similarly, the

theory of Maibom entails that practical reasoning directly causes the capacity for moral judgment. However, it is unclear how rational and affective capacities are related to each other in both models. This leads to a distinction between a strong version of both models, where affective and rational capacities are unrelated, and a weak version, where they do have a causal relation.

Second, I infer the associations entailed by the causal models compatible with the positions of Nichols (2002) and Maibom (2005). To do this, I use the Causal Markov assumption, which is the fundament of causal graph theory: a variable, given its direct causes, is independent of any other variable except its effects. Using this, I will show that the causal models compatible with the position of Nichols (2002) entail that rational capacities are unassociated with the capacity for moral judgment either unconditionally, for the strong version, or only conditional on affective capacities, for the weak version. On the other hand, the causal models compatible with the position of Maibom (2005) entail that affective capacities are unassociated with the capacity for moral judgment either unconditionally, for the strong version, or only conditional on rational capacities, for the weak version.

Finally, I examine whether these associations are supported or refuted by the empirical studies about psychopathy that Nichols (2002) and Maibom (2005) cite. These empirical studies yield an association between a psychopathy score and performance on a specific task. I argue that such studies make a weak case that both rational and affective capacities are associated with the capacity for moral judgment, potentially refuting the strong version of both models. However, the empirical studies are irrelevant for the weak version of both models. To support or refute whether rational capacities are independent of the capacity for moral judgment, given affective capacities, or whether affective capacities are independent of the capacity for moral judgment, given rational capacities, we need empirical studies of an entirely different kind: studies where subjects perform rational, affective and moral tasks, instead of only one task.

So, by applying causal graph theory to the theories of Nichols (2002) and Maibom (2005), I elaborate on a way to systematically assess the empirical support of theories of moral judgment. Moreover, I show what empirical evidence is needed to make progress in the debate: studies that examine the relation between performance on rational, affective and moral tasks.

References:

Maibom, H. L. (2005). Moral unreason: The case of psychopathy. *Mind and*

Language, 20(2), 237-257.

Nichols, S. (2002). How psychopaths threaten moral rationalism: Is it irrational to be amoral? *Monist: An International Quarterly Journal of General Philosophical Inquiry*, 85(2), 285-303.

Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). New York, NY: Cambridge University Press.

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge, MA: The MIT Press.

The Good Life and Commitment: Complementary Components for an Agonistic Political Participation

Dennis Schutijser (Pontificia Universidad Católica del Ecuador)

This contribution seeks to bring together two essential components of an agonistic political deliberation in which the value of polarization is acknowledged from two points of view. Political deliberation requires both the personal capacity to formulate and maintain one's own committed ethical point of view, and the political commitment to defend that ethical point of view. Both issues are worked out in a Neo-Aristotelian framework, with reference to the ethical conception of the good life, and the political conception of citizenship.

The ideal of a participatory deliberative democracy is, in Habermas' words a counterfactual ideal: it cannot be realized. This counterfactuality is due to a number of elements, one of which being the likely impossibility to reach actual and full consensus, even (or especially) if full participatory deliberation were possible. It is therefore important to contrast this model with a fundamentally agonistic aspect of the political arena. Not the objective of consensus, or even of the elimination of formal hindrances to participatory deliberation, are what matters most: it is the clash of ideas, the struggle between different viewpoints. The thesis of this proposal, is that this clash of ideas requires a combination of two fundamental components in all participants: individual commitment to one's own conception of the good life, and political commitment to defend that conception. Both can be formulated from a neo-Aristotelian point of view.

The end of the modern grand narratives (Lyotard, 1979:31) has also led to the impossibility of assuming that different people in the same cultural context have the same conception of the ethical good life. It is therefore hardly surprising that Habermas, in his courageous attempt to save the Kantian search for a universal

moral principle from the end of modernity, recurred to a formal framework from which reference to the good life is excluded. Habermas proposes to base his model for a discursive ethics of the full and transparent participation of rational agents capable of abstracting from their own ethical convictions about the good life (Habermas, 1997). These matters are restricted to the personal sphere and have no place in a setting of public deliberative participation. The counterfactual impossibility of this formal solution is but one difficulty, to which could be added such issues as power relations, bodily affections, desires, etc. (cf. Kunneman, 1998).

Although the particularity of a postmodern morality has been extensively commented by a number of eminent authors (cf. Lyotard, 1984; Foucault, 1976-1984; Butler, 2005), a return to one of the first great thinkers in the field of ethics can offer a means to reintegrate both the acknowledgment of a personal conception of the good life, and the requirement for socio-political commitment. Aristotle is rightly considered the founder of modern science, having most importantly established the distinction of a number of scientific disciplines and identified the leading principle of being and of knowing for each. Thus, ethics and politics are presented as the practical sciences of which the object is “changing, not always the same”, and the knowledge we can obtain inherently particular.

To Aristotle, the ultimate goal of politics is establishing the good life. What this good life is, is determined in accordance with one’s goal or *telos* as a being. This teleological basis has often been criticized in Aristotle. It seems to presuppose an idea of ‘what man is’; most notable a political animal (1253a3) and a rational being (1102a13-1103a10) – either way a universally determinate being. Both modernity and postmodernity have attacked this position, on the basis of its projected rationality and its universality, respectively.

Focusing on the universality, there are a number of reasons to reconsider this condemnation of Aristotle. First, the universality of human nature and of the good life are countered in Aristotle himself. Thus, elements considered relevant for the good life include the time and place, the body, and desires. In addition, references to the universality of human nature are made not only with regards to nature, but also to culture, or more precisely to habits and education.

Second, the practical wisdom of *phronesis* is a particular wisdom, of particulars. Aristotle famously formulates a universal conception of the good life twice: once (1177a10-1178a8) it serves to prove the factual impossibility of attaining it when it is in the form of the life of contemplation. The second time (1097a15-1097b20) it is in fact the result of an investigation into the existing *doxai* or opinions of his time. This is the result of a science very different from the sciences seeking to work from universal principles.

This brings us to the conception of science at work in ethics and politics. It is a science that does not start from universals as e.g. metaphysics does. Here, science starts from particular opinions and observations. The objective of practical science is not to deduce truths from basic principles, but to put existing opinions to the test, in order to determine their value. As a result, the good life and the excellence it requires are the objective of the investigation but they are not universally attained.

With that, the necessary link between the good life and political commitment appears in a different light. Not only is political commitment, to Aristotle, an essential part of that good life (cf. Arendt, 1958): the good life is also a necessary component of political deliberation. In order to best carry out the deliberation central to the practical science of politics, the different ethical positions are to be as best articulated as possible. We thus arrive at an agonistic political field in which participants are fully committed to their own conception of the good life and prepared to defend that conception. The aim is this confrontation of differing viewpoints, not their reunion or the victory of one over the other. Only in that way can a confrontation of differing opinions lead to the nearest form of certainty we can achieve in this field of knowledge: that of choosing wisely (*prohairesis*).

References:

- Anscombe, G.E.M. (1958). "Modern Moral Philosophy", in: *Philosophy*, 33 (1958).
- Arendt, H. (1958) *The Human Condition*. Chicago: University of Chicago Press.
- Aristotle, (1984). *Complete Works*. New Jersey: Princeton University Press.
- Butler, J. (2005). *Giving an Account of Oneself*. New York: Fordham University Press.
- Dohmen, J. (2002). *Over levenskunst*. Amsterdam: Ambo.
- Dohmen, J. (2005). *Tegen de onverschilligheid*. Amsterdam: Boom.
- Foucault, M. (1976-1984). *L'histoire de la sexualité (4 vols.)*. Paris: Gallimard.
- Foucault, M. (2001) *Dits et écrits (2 Vols)*. Paris: Quarto Gallimard.
- Hadot, P. (2002). *Exercices spirituels et philosophie antique*. Paris: Albin Michel.
- Hadot, P. (1995) *Qu'est-ce que la philosophie antique?* Paris: Folios.

- Habermas, J. (1997). *Theorie des kommunikativen Handelns*. Berlin: Suhrkamp.
- Kekes, J. (2002). *The Art of Life*. New York: Cornell University Press.
- Long, C.P. (2004). *The Ethics of Ontology; Rethinking an Aristotelian Legacy*. New York: State University of New York Press
- Lyotard, J.-F. (1984). *Le différend*. Paris: Editions de Minuit.
- MacIntyre, A. (1984). *After Virtue*. Indiana: University of Notre Dame Press.
- Mill, J.S. (2002). *On Liberty, The Subjection of Women & Utilitarianism*. New York: Random House.
- Nehamas, A. (1998) *The Art of Living*. Berkeley: University of California Press
- Nussbaum, M. (2001). *The Fragility of Goodness*. Cambridge: Cambridge University Press
- Singer, P. (2011). *Practical Ethics*. Cambridge: Cambridge University Press.
- Taylor, C. (1991). *The Ethics of Authenticity*. Cambridge: Harvard University Press.
- Taylor, C. (1989). *Sources of the Self*. Cambridge: Harvard University Press.
- Tongeren, P. Van. (2003). *Deugdelijk leven*. Amsterdam: SUN.

Robots in the workplace. Design choices and problematic responsibility ascriptions to human co-workers.

Jilles Smids (Eindhoven University of Technology)

Robots enter the workplace, not only as a tool, but increasingly also as a collaborator with human workers. In robot design, several characteristics of the robots allow for some range or continuum of options from which designers can choose. For example, a humanoid versus a machine-like appearance, one or more of various modalities of communication, etc. In this paper, I consider how such design choices can lead human co-workers to inaccurately perceive the robot's level of autonomy and competence. Subsequently I argue that problems with both forward- and backward-looking responsibility of human co-workers may result.

US and Them: The Ethics of AI Agents

Edward Spence (The University of Sydney, Australia)

Can artificial intelligent agents (AIAs) think for themselves, will they ever be capable of self-awareness, could they ever be conscious in the way we are, are not questions that this paper will seek to address or answer. Rather the paper will explore conditionally whether insofar as AIAs can develop autonomous intelligence similar to that of human agents sometime in the future, at least to some relevant sense and degree, they will have moral rights and moral obligations in the same way humans do at present.

Introduction

Can artificial intelligent agents (AIAs) think for themselves, will they ever be capable of self-awareness, could they ever be conscious in the way we are, are not questions that this paper will seek to address or answer. Rather the paper will explore conditionally whether insofar as AIAs can develop autonomous intelligence similar to that of human agents sometime in the future, at least to some relevant sense and degree, they will have moral rights and moral obligations in the same way humans do at present.

I will argue, that insofar as ethics is universally prescriptive and species-transcendent and therefore not merely restricted to human beings alone then in principle AIAs would be entitled to the same moral rights and bound by the same moral obligations as humans. Any species other than our own, including artificially intelligent agents that meet the essential conditions relevant for moral consideration must be accorded the same moral status, similar to that of human agents.

The Argument of the Paper in Summary

Starting with the conditional premise that artificial general intelligence (AGI) that leads to the development of AIAs at least in principle sometime in the future is feasible, the argument will show that AIAs will have natural rights to freedom and wellbeing just like us in accordance with Alan Gewirth's argument for the Principle of Generic Consistency (PGC).

Following Alan Gewirth, I shall argue that the possession of the property of *purposive agency* is both necessary and sufficient for moral consideration (Gewirth, 1978; Beyleveld 1991; Spence 2006). Thus, any being that possesses purposive agency in the relevant sense is entitled to moral consideration; specifically, to prima facie rights to freedom and wellbeing. Purposive agency in the relevant sense is an intentional property, one which the being who possess it is both aware of and

values it, at least in some minimal instrumental sense as an enabling condition for the fulfilment of their chosen purposes, whatever they happen to be, and whose fulfilment they care about and strive to secure. Characteristically, all human beings act purposively for purposes which they value, at least in some minimal sense, and which they wish to bring into fulfilment.

Insofar as AIAs are capable of purposive agency in the relevant intentional sense as described above, they too must be granted *prima facie* rights to freedom and wellbeing. Those rights because *prima facie*, can be overridden and infringed when they are being used or intended to be used to violate the legitimate rights of other purposive agents be it human or AIAs. To determine which rights take priority when two sets of mutually exclusive rights come into conflict, Alan Gewirth's Degree for the Necessity of Action Principle (DNA Principle) can be applied: In the event of a conflict of rights between the rights of humans and those of AIAs the conflict can in principle be resolved on the basis of Gewirth's subordinate principle to the PGC, the Degree for the Necessity of Action Principle (DNA principle).

The DNA principle specifies that in the event of a mutually exclusive conflict between agent A's generic rights (rights to freedom and wellbeing) and the generic rights of agent B, A's rights should take priority over those of agent B's rights when the objects of those rights, namely, freedom and wellbeing, are more necessary for purposive action for agent A than they are of those of agent B. Although this way of resolving piecemeal conflicts between the rights of individual AIAs on the one hand and humans on the other seems at least in principle straightforward but potentially complicated in some practical cases, it presents us with a problem when the conflict arises at the species level. How do we resolve the conflict between our rights as a species and those of AIAs when the latter threaten our very survival and continuing existence as a species? What checks and controls need we install in the design phase of AIAs to ensure that such a potentially lethal risk to our species' survival is minimized if not eliminated?

Do we as humans have a special moral obligation to our own species that requires us to infringe or even violate the legitimate rights of other species, such as AIAs, when the latter threaten, because of their potentially superior capacity for intelligence our very existence (Superintelligence, Bostrom 2014)? If generic rights are universal and species transcendent how can we justifiably exercise partiality in favour of our own human rights? I will argue that such species-partiality might be justified only under certain restricted cultural and communal conditions necessary for self-preservation that when present may allow the granting of priority rights to human *persons* over those of AIAs (Spence 2006). The problem of *existential risk* and self-preservation has of course received much attention recently. For example, the *Centre for the Study of Existential Risk* (CSER) is a research *Centre* at the University of Cambridge, established "to study possible extinction-level threats posed by present or future technology" which includes the existential risk for humanity.

The overall motivation for this enquiry is to explore a theoretical and practical solution to the most pressing and difficult problem in AGI research, that of the *Control Problem* as identified by AGI researchers among others Nick Bostrom (*Superintelligence* 2014) and Max Tegmark (*Life 3.0* 2018).

What is the Control Problem? According to Tegmark “To sort out the control issue we need to know both how well an AI can be controlled and how much an AI can control” (T 2018: 159). But what is to be controlled? In a nutshell Artificial General Intelligence (henceforth AI for short).

But what is Intelligence? Interestingly the stated mission of Google’s Deep Mind is “to solve intelligence” specifically AI intelligence. For this paper I will adopt Tegmark’s broad and reasonable definition of intelligence which is the “ability to accomplish complex goals” (Tegmark 2018:50). This definition seems reasonable as it broadly covers both human and AGI.

The control problem of AI can now be expressed and further analysed and explored as a problem of what goals and whose goals to set an AI? As Tegmark acknowledges and most AI researchers seem to agree, the most difficult problem about AI, is goals: Should we give AI goals and if so whose goals? How can we give AI goals, and can we ensure that these goals are retained even if the AI gets smarter? Can we change the goals of an AI that is smarter than us? What are our ultimate goals? (Tegmark 2018: 249-280).

From Goal-oriented AI to Goal-orientated Ethics: A Rationalist-Normative Approach

Having identified the core problem of AI as a control problem concerning which and whose goals to set for an AI and how to control those goals so that we don’t end up being controlled by AI ourselves, the final step in the argument is to attempt to offer a solution to this problem through ethics, and in particular the ethics of intelligent and rational goal-oriented purposive agency.

The solution I propose can in theory be accomplished through the *natural alignment* between Our Goals and the Goals of AIAs on the basis of Rational Purposive Agency that Gewirth identifies as a sufficient foundation for ethics for all purposive rational agents. That is, *Purposive Agency Ethics for Purposive Humans and AIAs*. The argument shows that in principle the goals of AIA whatever they are can be *Controlled Internally* by *self-control* within the rational agency of the AIA since by definition being a rational purposive agent the AIA is rationally committed to the pursuit of its goals ethically, *on pain of self-contradiction*, as the Principle of Generic Consistency (PGC) requires (Gewirth, 1978; Beyleveld 1991; Spence 2006).

This proposed solution is further enabled and reinforced through the accepted principle by AI researchers following Alan Turing that intelligence and its related goal-oriented behaviour is substrate-independent. As Tegmark eloquently puts it.

“A very important and crucial claim in AI research that opens up the real potential for machine intelligence is that intelligence is *substrate independent*. That is in can develop just as much in silicon and other configuration of atoms just as it does in carbon with regard to sapiens. So intelligence does not require flesh, blood or carbon atoms (Tehmark 2018: 65-67).”

Finally, the argument will conclude by showing that a crucial part of this proposed solution to the control problem and that of the related problem of goals (which goals, whose goals, and how to align human goals with those of machines) lies in the concept of *trust*: trust between us and the AIA's. And to that end, the notion of trust I will use for that purpose in John Weckert's notion of trust to the effect that trust requires moral understanding, commitment and most importantly compliance, or at the very least belief in us and AIAs that the moral commitments for our mutual advantage and respect will be complied with. Key to John Weckert's model of trust for software agents is the notion of autonomy. For as Weckert correctly argues, “autonomy is central because it plays a vital role in the distinction...between trust and reliance” (2005: p3). Weckert defines “trust as reliance on moral choices” (2005: p. 4) and goes on to claim that the implication of this definition of trust is “That if we are to trust software agents, they must be moral entities capable of making moral decisions...” (2005: p. 4)

Conclusion of the Argument

First, if AIAs are a practical possibility, they will have rights. They will have rights if they possess the capacity for autonomous purposive agency. This is justified by the Principle of Generic Consistency (PGC) since the intrinsic and necessary connection between autonomous purposive agency and rights is sufficient for establishing a moral status. Secondly, since ethics of purposive intelligence (as per the PGC) is also substrate independent and therefore common to both Humans and AIAs, the Control Problem can in principle be solved through the alignment of Human goals with those of AIA goals, based on the Ethics of Purposive Agency. Thirdly, the proposed solution also addresses the problem of human-machine trust identified as moral trust: “That if we are to trust software agents, they must be moral entities capable of making moral decisions....” (Weckert and Moore, 2004) This further reinforces the argument proposed that a solution to the problem of control requires an Ethical Solution.

References:

Beyleveld, Deryck (1991). *The Dialectical Necessity of Morality: An Analysis and Defense of Alan Gewirth's Argument to the Principle of Generic Consistency*. Chicago: University of Chicago Press.

Bostrom, Nick (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Brey, Philip (2001). *Hubert Dreyfus: Humans Versus Computers*, in Achterhuis, H. (ed.), *American Philosophy of Technology: the Empirical Turn*, Bloomington: Indiana University Press, pp. 37-63.

Buttazzo, Giorgio (2001). *Artificial Consciousness: Utopia or Real Possibility?* Perspectives, IEEE Computer Society, pp. 24-30.

Clarke, Roger (1994). *Asimov's Laws of Robotics: Implications for Information Technology*, IEEE Computer 26, 12 (December 1993) pp.53-61 and 27, 1 (January 1994) pp. 57-66.

Clark, Andy and Chalmers, David (1998). *The Extended Mind*, Analysis, 58.1 (January 1998), pp. 7-19.

Dennett, Daniel C. (1994). *The Practical Requirements for Making a Conscious Robot*, Phil.Trans.R.Soc.Lond.A, 349 pp. 133-146.

Erion Gerald J. (2001). *The Cartesian Test for Automatism*, Minds and Machines, 11: 29-39.

Ford, M. Kenneth, Glymour, Clark and Hayes, J. Patrick (eds.) (2006). *Thinking about Android Epistemology*. Cambridge, Massachusetts: MIT Press.

Franklin, Stan and Graesser, Art (1996). *Is it an Agent, or just a Program? A Taxonomy for Autonomous Agents*, Proceedings of the Third International Workshop on Agent Theories, Architecture, and Languages, Springer-Verlag.

Gewirth, Alan (1978). *Reason and Morality*. Chicago: University of Chicago Press.

Hamad, Steven (2003). *Can a machine be conscious? How?* Journal of Consciousness Studies, Vol. 10, No. 4-5, April/ May.

Kitamura, Tadashi., Tahara, Tomoko. and Asami, Ken-Ichi (2000). *How can a robot have consciousness?* Advanced Robotics, Vol. 14, No. 4, pp. 263-275.

Kugel, Peter (2002). *Computing Machines Can't Be Intelligent (...and Turing Said So)*, Minds and Machines, 12: 563-579.

Minsky, Marvin (2006). *Alienable Rights* in Ford, M. Kenneth, Glymour, Clark and Hayes, J. Patrick (eds.) (2006). *Thinking about Android Epistemology*. Cambridge, Massachusetts: MIT Press, pp.137-143.

Nadeau, Joseph Emile (2006). *Only Androids Can Be Ethical* in Ford, M. Kenneth, Glymour, Clark and Hayes, J. Patrick (eds.) (2006). *Thinking about Android Epistemology*. Cambridge, Massachusetts: MIT Press, pp. 241-248.

Spence, Edward H. (2006). *Ethics Within Reason: A Neo-Gewirthian Approach*. Lanham MD: Lexington Books (an imprint of Rowman and Littlefield).

Sterrett, Susan G. (2000). *Turing's Two Tests for Intelligence*, *Minds and Machines*, 10: 541-559.

Tegmark, Max (2018). *Life 3.0: Being Human in the age of Artificial Intelligence*. London: Penguin Books.

Weckert, John and Moor, James (2004). *Using the Precautionary Principle in Nanotechnology Policy Making*, *Asia Pacific Nanotechnology Forum News Journal* 3, pp. 12-14.

Visualizations of interventionistic mental causation

Frank van Caspel (Open University of the Netherlands & Radboud University Nijmegen)

Recently several authors have claimed that James Woodward's interventionistic theory of causality can solve the causal exclusion problem in philosophy of mind. Michael Baumgartner has convincingly argued that these claims are misguided. Markus Eronen, however, introduces a version of interventionism that prevents the causal exclusion problem, and which simultaneously offers a way of dealing with causal claims in the context of conceptual plurality. To explain the distinguishing feature of Eronen's of interventionism, and to highlight its benefits, I use a novel approach: causal graphs that include visualizations of the variable sets over which causal claims range.

References:

Baumgartner, Michael (2009). "Interventionist Causal Exclusion and Non-reductive Physicalism". In: *International Studies in the Philosophy of Science* 23.2, pp. 161–178.

— (2013). "Rendering Interventionism and Non-Reductive Physicalism Compatible". In: *dialectica* 67.1, pp. 1–27.

— (2018). "The Inherent Empirical Underdetermination of Mental Causation". In:

Australasian Journal of Philosophy 96.2, pp. 335–350.

Eronen, Markus I. (2012). “Pluralistic physicalism and the causal exclusion argument”. In: European Journal for Philosophy of Science 2, pp. 219–232.

Eronen, Markus I. and Daniel S. Brooks (2014). “Interventionism and Supervenience: A New Problem and Provisional Solution”. In: International Studies in the Philosophy of Science 28.2, pp. 185–202.

Menzies, Peter (2006). “Making Things Happen: a Theory of Causal Explanation”. In: Mind 115 (459), pp. 821–826.

Weslake, Brad (forthcoming). “Exclusion Excluded”. In: International Studies in the Philosophy of Science.

Woodward, James (2003). Making things happen: a theory of causal explanation. Oxford University Press.

— (2008). “Mental Causation and Neural Mechanisms”. In: Being Reduced: New Essays on Reduction, Explanation, and Causation. Ed. by Jakob Hohwy and Jesper Kallestrup. Oxford University Press, pp. 218–262.

— (2015). “Interventionism and Causal Exclusion”. In: Philosophy and Phenomenological Research XCI.2, pp. 303–347.

Yang, Eric (2013). “Eliminativism, interventionism and the Overdetermination Argument”. In: Philosophical Studies 164, pp. 321–340.

Wellbeing, experience, and the significance of sentience

Willem van der Deijl (Tilburg University)

Can a person’s degree of wellbeing be affected by things that do not enter her experience? Experientialists deny this claim. The debate about this position has focused on an argument against experientialism – the experience machine objection – but little arguments exist for it. I present an argument for experientialism, which I call the Significance of Sentience argument. The argument is based on the premise that only sentient beings possess wellbeing, welfare sentientism. I argue that welfare sentientism can only adequately be explained by experientialist accounts of wellbeing, other accounts will leave unexplained why experience has this significant demarcating role.

Elusive Knowledge Hyperintensionalized

Janneke van Lith (Utrecht University)

In “Elusive Knowledge” (Lewis 1996), Lewis lays out his relevant alternatives theory of knowledge ascriptions. One problem for this theory is that it doesn’t work in hyperintensional contexts, which is odd since as, Schaffer points out, “epistemology is evidently a hyperintensional topic” (Schaffer 2015, 478). In this paper, we develop a hyperintensional version of Lewis’s relevant alternatives theory in the framework of exact truthmaker semantics, which has recently been championed by Fine and others (Fine 2017).

In “Elusive Knowledge” [4], Lewis lays out his relevant alternatives theory of knowledge ascriptions:¹⁴

Relevant Alternatives Theory. A statement of the form “S knows that A” is true in context C if and only if S’s evidence eliminates every $\neg A$ possibility relevant in C (cf. [5, p. 475]).

Following Lewis, what counts as relevant in a context is governed by a set of principles, including, for example, the following [5, p. 475–76]:

Actuality. The actual possibility is relevant in every context. Belief. All possibilities that the subject believes or ought to believe to obtain are relevant in every context.

Lewis explicitly formulates his theory in a setting where propositions are individuated by necessary equivalence [4, p. 551]. This means, in particular, that all necessary propositions, like the truths of logic and mathematics, are identified—there is only one necessary proposition [4, p. 551]. In fact, this necessary proposition holds in all possibilities and so we have to ascribe knowledge of that proposition to every subject in any context whatsoever [4, p. 552]. This means that the theory cannot be applied in contexts where knowledge of necessary propositions is precisely what’s at stake. Think, for example, of a reviewer for a mathematics journal trying to determine whether an author has established knowledge of a mathematical theorem. Such contexts, where necessarily equivalent propositions should not be identified, are known in the literature as hyperintensional contexts [1]. Lewis acknowledges that his theory cannot account for hyperintensional contexts [4, p. 552], but as Schaffer points out, this is odd “since epistemology is evidently a hyperintensional topic” [5, p. 478].

Our aim is to modify Lewis’s theory for use in hyperintensional contexts. In order

¹⁴ For an overview of the theory and the relevant literature, see, e.g., [5].

to achieve this, we formulate a relevant alternatives theory in the framework of exact truthmaker semantics, recently championed by Fine and others [2]. A state is an exact truthmaker of a proposition just in case the state necessitates the truth of the proposition and is wholly relevant to the truth of the proposition. So, for example, the state of the ball being red is an exact truthmaker of the proposition that the ball is red. But the state of the ball being red and round does not qualify as an exact truthmaker for the same proposition, since the ball being round is irrelevant to the truth of that the ball is red. The requirement that exact truthmakers be wholly relevant to their propositions allows us to distinguish between necessarily equivalent propositions. For example, a truthmaker of $A \vee \neg A$ is in general not going to be an exact truthmaker of $B \vee \neg B$ —simply because what’s relevant to A is not automatically relevant to B .

Our proposal is to formulate Lewis’s theory in terms of exact truthmakers as follows:

Hyperintensional Relevant Alternatives Theory. A statement of the form “ S knows that A ” is true in context C if and only if S ’s evidence relevantly excludes every exact truthmaker of $\neg A$ that is relevant in C .

S ’s evidence is straightforwardly understood as a set of states. The principles governing what’s relevant in a given context are also straightforwardly understood in terms of exact truthmakers. Actuality and Belief, for example, become:

Exact Truthmaker Actuality. Every actually obtaining state is relevant in every context.

Exact Truthmaker Belief. Every state that the subject believes or ought to believe to obtain is relevant in every context.

The notion of relevant exclusion is the one introduced by Fine [3, p. 634–35; 658–61], where the state of the ball being red relevantly excludes the ball being green, but the ball being red doesn’t relevantly exclude the ball being round and not round.

To see how the Hyperintensional Relevant Alternatives Theory works, let’s consider an example. Consider a logic teacher who wishes to determine whether her student knows that $A \vee \neg A$. According to the standard recursive clauses for exact truthmakers [2, p. 562], a state is an exact truthmaker for $\neg(A \vee \neg A)$ iff it is the fusion of an exact truthmaker for A and an exact truthmaker for $\neg A$. Note that such a state would be inconsistent (i.e. it never actually obtains), but it could very well be relevant in a given context C , like our classroom situation under discussion. So the question is whether the student’s evidence relevantly excludes all exact truthmakers of $\neg(A \vee \neg A)$. If the student didn’t study enough, it could very well be that his evidence is not enough to relevantly exclude some such state—and

hence the student doesn't know that $A \vee \neg A$.

Our second example shows that on the Hyperintensional Relevant Alternatives Theory, knowledge can fail to be closed under implication, even within a single context. Lewis admits of such failure only on the basis of a switch in context midway a conversation [4, p. 564], but hyperintensional contexts require us to be more lenient. Consider a conspiracy theorist who believes that all politicians are reptilians. She also has the ordinary belief A that the individual she is looking at is a human being playing tennis, which is well-supported by her evidence. Let B be the proposition that the individual she is looking at is a politician, for which she doesn't have any evidence. Then, it may well be that our conspiracy theorist knows that A , but fails to know that $A \vee B$. For on the basis of Exact Truthmaker Belief, the compound state of a tennis-playing reptilian politician is a relevant alternative to $A \vee B$ and it isn't excluded by the evidence, whereas the state of a tennis-playing reptilian is not a relevant alternative to A , and therefore need not be excluded.

References:

- [1] Cresswell, M.J. 1975. "Hyperintensional Logic." In: *Studia Logica* 34(1): 25–38.
- [2] Fine, Kit. 2017a. "Truthmaker Semantics." In: *A Companion to the Philosophy of Language*, 2nd edition, edited by Bob Hale, Crispin Wright, and Alexander Miller, 556–77. Hoboken, NJ: Wiley-Blackwell.
- [3] Fine, Kit. 2017b. "A Theory of Truthmaker Content I: Conjunction, Disjunction and Negation." In: *Journal of Philosophical Logic* 46(6): 625–74.
- [4] Lewis, David. 1996. "Elusive Knowledge." In: *Australasian Journal of Philosophy* 74(4): 549–67.
- [5] Schaffer, Jonathan. 2015. "Lewis on Knowledge Ascriptions." In: *A Companion to David Lewis*, edited by Barry Loewer and Jonathan Schaffer, 473–90. Hoboken, NJ: Wiley-Blackwell.

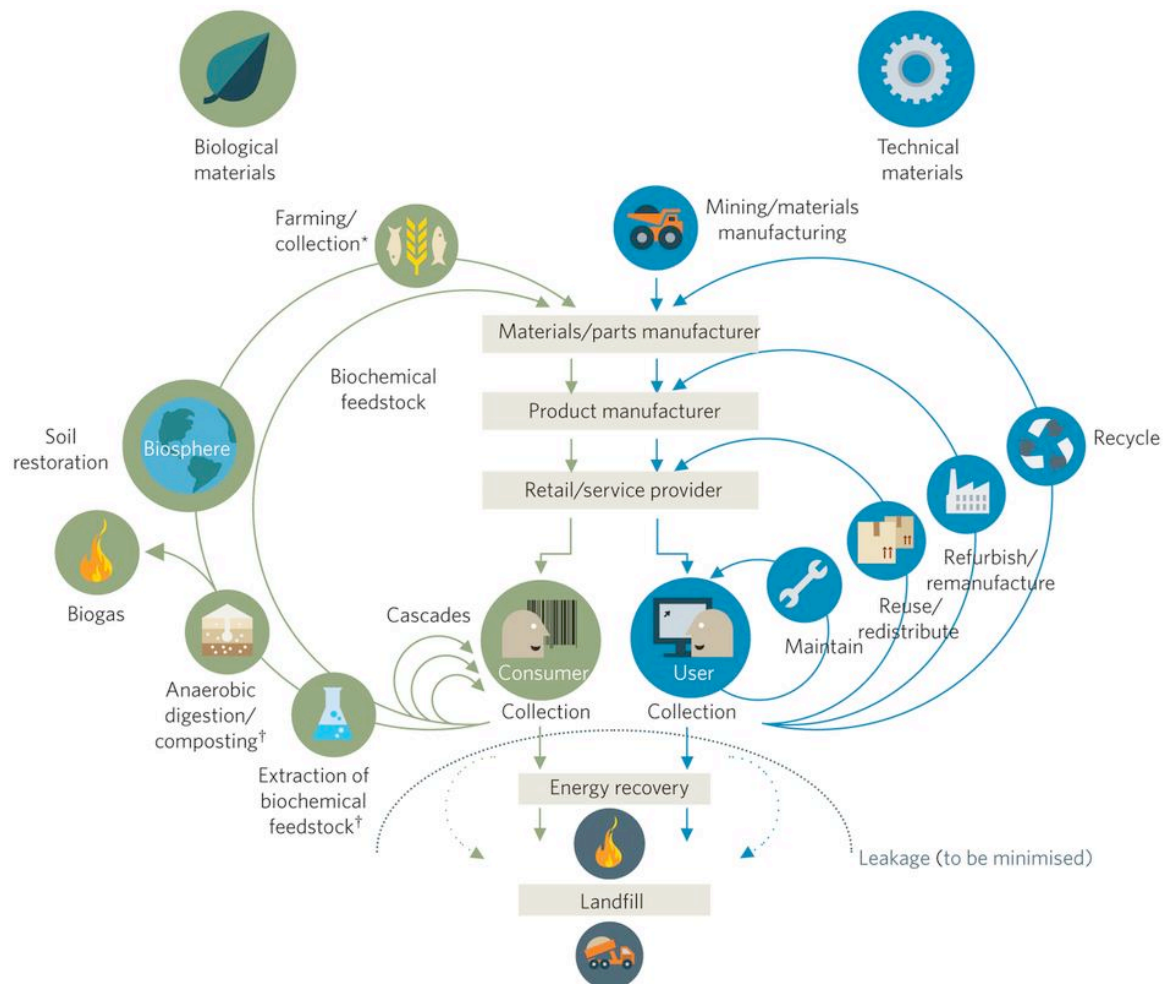
The Bio-Based Economy: Philosophical Reflections on the Relation between the Biosphere and the Economic Sphere and their Consequences for Sustainable Practice

Roel Veraart (Wageningen University) and Vincent Blok (Wageningen University)

Today, society faces serious environmental challenges. In order to mitigate climate change, the EU invests into initiatives such as the Bio Based Economy (BBE). We investigate the BBE's conceptuality regarding the relation between economy and ecology from a philosophical perspective (Levinas, Stiegler). We argue that by attempting to combine linear economy with ecological solutions as if these two spheres could be completely coalesced, the BBE preserves problematic tendencies from the past. Our hypothesis is that the effort to make our economy bio-based should recognise a principal polarization between human happiness and the biosphere itself.

Today, society faces serious environmental challenges. In order to mitigate climate change, the EU invests into initiatives such as the Bio Based Economy (European Commission, 2012). The Bio Based Economy, or BBE, builds on the idea of “An economy that relies on renewable natural resources to produce food, energy, products and services. The BBE will reduce our dependence on fossil natural resources, prevent biodiversity loss and create new economic growth and jobs in line with the principles of sustainable development” (Bosman & Rotmans, 2016). A visual representation of the Circular Economy (see Figure 1) is suitable to explain the ideal of basing human designs upon processes in the biosphere. Recent studies indicate that the idea of the BBE is promising, although the actual practice is still marginal (Jonker et al 2017). If companies do adopt BBE in their business practices, it is often still a side event and not part of the core business. Current practices in the BBE are framed within the market or economic logic, disregarding the normative aspects of the notion of sustainable development. Since sustainable development is a normative concept – it doesn't describe the world as it is, but as it should be – it is questionable whether the BBE can be understood as a new way of doing business as usual, in which biomass is seen as a source of added value for economic return; does it not, rather, call for the establishment of different economy, one that operates within the carrying capacity of the biophysical system of planet Earth?

Figure 1: schematic overview of the CE. Image source: Ellen MacArthur Foundation – adapted from the Cradle to Cradle Design Protocol by Braungart & McDonough (2014).



In this paper we investigate the fundamental conceptuality of the Bio-Based Economy (BBE) as it currently exists. By analysing the BBE from the philosophical perspective on the relation between economy and ecology developed by Emmanuel Levinas and Bernard Stiegler, we show how the BBE maintains a narrow notion of economy, unfit for the ecological measures necessary today. Accordingly, by attempting to combine linear economy with ecological solutions as if these two spheres could be completely coalesced, the BBE preserves problematic tendencies from the past. Our hypothesis is that the effort to make our economy bio-based should recognise principal tensions between human happiness and the biosphere itself. We seek to both gain a deeper understanding of these tensions, by comparing the individual and the macro-perspective on global warming, and try to improve on contemporary conceptuality by making several critical assessments. We argue for the need of a broadened account of economy, including the human

condition both as an irreducible value for future thought and as inherently carrying a problematic endeavour towards direct pleasure and overconsumption. Finally, by defending a major part of nature as absolutely uncontrollable, we seek to discover new possibilities to cope with the threatening biosphere whilst working towards an economy which truly operates within the carrying capacities of planet Earth.

The relation between economy and ecology is heterogenous, ambiguous and contradictory. Aggregating the two – basing economy upon the biosphere – will pose a major challenge, and not lastly on the conceptual level. As humans, we seize the day by seizing the planet, dominating and controlling our hostile environment. But this control has come to bite us in the tail, as we depleted the planet. Our search for happiness inverted into a flight for extermination. But surviving will require a radical change in the economic structures we value so for our happiness. This is a fundamental polarization in our time: to be human, we must overcome nature, but by doing so, we only increase its hostile character, and to save nature – to save ourselves – we must act different than we are, or are used to at any rate.

Our paper is a critique on the current conceptuality in the BBE, and simultaneously it is an exploration of humanity and nature, showing the flaws and tensions in our past and current behaviour. It grants a descriptive insight in our way of being: our economic tendencies are not evil in themselves, but result in undesirable outcomes for the future. As such, our paper might indicate the ethical direction of enjoyment versus uncontrollable nature.

References:

Asveld L., R. van Est & D. Stemerding. (2011a). Getting to the core of the bio-economy: A perspective on the sustainable promise of biomass. The Hague: Rathenau Instituut.

Benner, M., Löfgren, H. (2007). 'The Bio-economy and the Competition State: Transcending the Dichotomy between Coordinated and Liberal Market Economies'. *New Political Science*, 29(1): 77–95. Doi: 10.1080/07393140601170842

Bensaude-Vincent, B. (2018). 'Biomimetics, Biomimicry, Bionics and Bioeconomy'. (yet to be officially published).

Benyus, J. M. (2002). *Biomimicry: Innovation inspired by nature*. New York: Harper Perennial. Blok, V. (2015). 'The Human Glance, the Experience of Environmental Distress and the "Affordance" of Nature:

Toward a Phenomenology of the Ecological Crisis'. *Journal of Agricultural and*

Environmental Ethics, 28(5): 925-38. Blok, V. (2016). 'Biomimicry and the Materiality of Ecological Technology and Innovation: Toward a Natural

Model of Nature'. Environmental Philosophy, 13(2): 195–214.

Blok, V. (2018). 'Information Asymmetries and the Paradox of Sustainable Business Models: Towards an Integrated Theory of Sustainable Entrepreneurship'. Sustainable Business Models, Springer (yet to be officially published).

Blok, V., Gremmen, B. (2016). 'Ecological Innovation: Biomimicry as a New Way of Thinking and Acting Ecologically'. Journal of Agricultural and Environmental Ethics, 29(2): 203.

Bosman, R., Rotmans, J. (2017). 'Transition Governance towards a Bioeconomy: A Comparison of Finland and the Netherlands'. Sustainability, 8(10), 1017.

Bugge, M., Hansen, T., Klitkou, A. (2016). What Is the Bioeconomy? A Review of the Literature. Sustainability, 8(7): 691. Doi:10.3390/su8070691

Casey, E. (2003). 'Taking a glance at the environment: Preliminary Thoughts on a Promising Topic'. In Eco- phenomenology: 187–211. New York: State University of New York Press.

Clifton-Soderstrom, M. (2010). 'Levinas and the Patient as Other: The Ethical Foundation of Medicine'. Journal of Medicine and Philosophy, 28(4): 447–460. Doi: 10.1076/jmep.28.4.447.15969

Cools, A. (2015). 'Levinas's Defence of Intellectualism: An Undecidable Ambiguity?' In Debating Levinas's Legacy (A. Breitling, C. Bremmers, A. Cools, eds.), pp. 3–16. Leiden: Brill.

Derrida, J. (1978). 'Metaphysics and Violence'. In Writing and Difference (A. Bass, trans.), pp. 97–193. Chicago, IL: Chicago University Press.

Dicks, H. (2017). 'The Poetics of Biomimicry: The Contribution of Poetic Concepts to Philosophical Inquiry into the Biomimetic Principle of Nature as Model'. Environmental Philosophy, 14(2): 191-219.

Diehm, C. (2000). 'Facing Nature: Levinas Beyond the Human'. In Philosophy Today, 44(1): 51–9. Druker, J. (2006). 'Ethics and Ontology in Primo Levi's "Survival in Auschwitz": A Levinassian Reading'. Italica, 83(3/4): 529–542. URL: <https://www.jstor.org/stable/27669104>

Edelglass, W., Hatley, J. & Diehm, C., eds. (2012). Facing Nature, Pittsburgh: Duquesne University Press.

Ellen Macarthur Foundation. (2014). 'Towards the Circular Economy: Accelerating

the scale-up across global supply chains'. Vol. 3.

Ellen MacArthur Foundation. (2015). 'Growth Within: A Circular Economy Vision for a Competitive Europe'. McKinsey Centre for Business and Environment. UK: SUN.

European Commission. (2012a) Innovating for sustainable growth—a bioeconomy for Europe. Online at

http://ec.europa.eu/research/bioeconomy/pdf/bioeconomycommunicationstrategy_b5_brochure_web.pdf

European Commission. (2012b). Manifesto for a resource-efficient Europe. Online at: http://europa.eu/rapid/press-release_MEMO-12-989_en.htm

Finegold et al. (2005). *BioIndustry Ethics*. Orlando: Elsevier Academic Press. ISBN:0123693705 Frazzetto, G. (2003). 'White Biotechnology'. *EMBO reports* 4, 835–837. Doi: 10.1038/sj.embor.embor928

Joldersma, C. (2013). 'An Ethical Sinngebung Respectful of the Non-Human: A Levinassian Environmental Ethics'. *Symposium*, 17(2): 224–245.

Jonker, J. (2012). *New Business Models*. Nijmegen School of Management: RU Nijmegen.

Jonker, J. et al (2017). 'Kringlopenladder voor de Circulaire Economie'. *SIGMA*, (1), 18.

Kitchen, L., Marsden, T. (2011). 'Constructing sustainable communities: a theoretical exploration of the bio- economy and eco-economy paradigms'. *The International Journal of Justice and Sustainability*, 16(8): 753–769. Doi: 10.1080/13549839.2011.579090

Levinas, E. (1963). 'The Trace of the Other'. *Tijdschrift voor Philosophie*, (25): 605–623.

Levinas, E. (1996). *Totality and Infinity* (A. Lingis, trans.). Pittsburgh, PA: Duquesne University Press.

Levinas, E. (1998). *Otherwise than Being or Beyond Essence* (A. Lingis, trans.). Pittsburgh, PA: Duquesne University Press.

Mcdonough, W., Braungart, M. (2014). 'Towards a Sustaining Architecture for the 21st Century: The Promise of a Cradle-to-Cradle Design'. *UNEP Industry and Environment*, ap-sept: 13–16.

Muijsenberg, S. et al. (2017). *Joint Conference on Bio-Inspiration and*

Biomimicry'. *Biomimicry Magazine*, May(9).

Nelson, E. (2011) 'Levinas and Adorno: Can there be an Ethics of Nature?' In *Facing Nature*, Pittsburgh: Duquesne University Press.

Osseweijer, P., Landeweerd, L., Pierce, R. (2010). 'Genomics in Industry: issues of a biobased economy'. *Genomics, Society and Policy*, 6(2): 26–39.

Oudemans, W., Peeters, N. (2014). *Plantaardig – Vegetatieve Filosofie*. Zeist: Knv. Pfau, S., Hagens, J., Dankbaar, B., Smits, A. (2014). 'Visions of Sustainability in Bioeconomy Research'.

Sustainability, 6(3), 1222–1249. Doi:10.3390/su6031222 Raworth, K. (2017). *The Doughnut Economics*. Chelsea Green Publishing: Vermont.

Rizos, V. et al. (2016). 'Implementation of Circular Economy Business Models by Small and Medium-Sized Enterprises (SME's): Barriers and Enablers'. *Sustainability*, 8, 1212. Doi: 10.3390/su8111212

Toadvine, T. (2003). 'The Primacy of Desire and Its Ecological Consequences'. In *Eco-phenomenology*: 139–155. New York: State University of New York Press.

Toadvine, T. (2012). 'Enjoyment and its Discontents'. In *Facing Nature*, Pittsburgh: Duquesne University Press. Todd, S. (2003). *Learning from the Other: Levinas, Psychoanalysis, and Ethical Responsibilities in Education*.

Albany: State University of New York Press.

Stiegler, B. (2010). *For a New Critique of Political Economy* (D. Ross, trans.). Cambridge: Polity Press.

Wahl, D. (2006). *Designing Regenerative Cultures*. Triarchy Press: Axminster.

Welsh, M. (1998). 'From the Impersonal to the Environmental: Extending the Ethics of Levinas to Human Ecology'. *Human Ecology Review*: 5(2).

Zembylas, M., Vrasidas, C. (2005). 'Levinas and the "Inter-Face": The Ethical Challenge of Online Education. *Educational Theory*, 55(1). <https://doi.org/10.1111/j.1741-5446.2005.0005a.x>

Zwier, J., Blok, V. (2015). 'The Ideal of a Zero-Waste Humanity: Philosophical Reflections on the Demand for a Bio-Based Economy'. *Journal of Agricultural and Environmental Ethics*, 28(2).

Hamilton's rule: understanding the disagreement about its explanatoriness

Philippe Verreault-Julien (Erasmus University Rotterdam, Erasmus Institute for Philosophy and Economics) and Vaïos Koliofotis (Erasmus Institute for Philosophy and Economics)

More than half of a century since its initial development, inclusive fitness theory and Hamilton's Rule (HR) remain controversial in evolutionary biology. We argue that the distinctions between how-possibly and how-actually explanations, on the one hand, and between causal and mathematical explanations, on the other hand, illuminate the source of the disagreement between the critics of HR and its supporters. Beyond illuminating the source of the disagreement over HR, an important benefit of using these distinctions is that they allow to temper both camps in their evaluation of whether HR affords understanding.

More than half of a century since its initial development (Hamilton 1964a; Hamilton 1964b), inclusive fitness theory remains controversial in evolutionary biology. Nowak, Tarnita, and Wilson's (2010) *Nature* article resulted in a fierce response (e.g. Abbot et al. 2011) and a heated and ongoing debate over the explanatory status of Hamilton's rule (e.g. Marshall 2015; Okasha and Martens 2016; van Veelen et al. 2017). In a nutshell, Nowak, Tarnita, and Wilson argue that the general form of Hamilton's rule (HR) does not explain nor afford understanding.

In this paper, we argue that the distinctions between how-possibly (HPEs) and how-actually (HAEs) explanations, on the one hand, and between causal and mathematical explanations, on the other hand, illuminate the source of the disagreement between the critics of HR and its supporters. Furthermore, it may serve as a cautionary note to not over or understate whether HR affords the epistemic benefit of understanding.

First, it is possible to see the controversy through the lens of the HPEs-HAEs distinction. In short, whereas HAEs typically cite the actual factors responsible for the explanandum, HPEs show what *may* explain it. While Nowak et al. consider HR only supplied "hypothetical explanations" (2010, 1058), Abbot et al. consider it was successful in "explaining a wide range of phenomena" (Abbot et al. 2011, E1). In

other words, there is fundamentally a dispute over whether HR provides an *actual* explanation of phenomena such as the evolution of eusociality or only a *possible* one.

Second, one way of interpreting Nowak et al.'s (2010; 2011) charge against HR is that not only is it a HPE, but that it is a *mathematical* HPE. Whether or under what conditions HR can receive a causal interpretation is nebulous (Allen, Nowak, and Wilson 2013; Birch 2014; Okasha and Martens 2016). Mathematical HAEs typically show how an explanandum holds out of mathematical necessity (Lange 2013; Pincock 2015). Interpreting HR in the light of the literature on mathematical explanation allows to see that Nowak et al.'s critique is twofold:

HR fails 1) to identify necessary mathematical conditions and 2) to map physical structures similar to those the rule depicts (Bueno and Colyvan 2011; Pincock 2007). Therefore, it does not provide a mathematical HAE, but at best a HPE.

Beyond illuminating the source of the disagreement over HR, an important benefit of using these distinctions is that it allows to temper both the critics and the supporters of HR in their evaluation of whether HR affords understanding. It does so for two reasons. First, even if we accept that HR fails as a HAE of social behaviour like cooperation, it does not follow that it can't afford understanding (see Kuorikoski and Ylikoski 2015; Rice 2016). But even if HPEs may afford understanding, proponents of HR should also be careful when claiming it *actually* explains. Second, under suitable conditions, a lack of causal interpretation does not imply HR can't be explanatory (cf. Birch 2014). A mathematical HAE, or even a HPE, may afford understanding of empirical phenomena. While the critics may be too quick to dismiss HR on the ground that it is a mathematical explanation, its proponents should also be wary of not jumping too fast from the mathematical to the causal interpretation.

References:

- Abbot, Patrick, Jun Abe, John Alcock, Samuel Alizon, Joao A. C. Alpedrinha, Malte Andersson, Jean-Baptiste Andre, et al. 2011. "Inclusive Fitness Theory and Eusociality." *Nature* 471 (7339): E1–E4.
- Allen, Benjamin, Martin A. Nowak, and Edward O. Wilson. 2013. "Limitations of Inclusive Fitness." *Proceedings of the National Academy of Sciences* 110 (50): 20135–9.
- Birch, Jonathan. 2014. "Hamilton's Rule and Its Discontents." *The British Journal for the Philosophy of Science* 65 (2): 381–411.
- Bueno, Otávio, and Mark Colyvan. 2011. "An Inferential Conception of the Application of Mathematics." *Noûs* 45 (2): 345–74.

- Hamilton, W. D. 1964a. "The Genetical Evolution of Social Behaviour. I." *Journal of Theoretical Biology* 7 (1): 1–16.
- . 1964b. "The Genetical Evolution of Social Behaviour. II." *Journal of Theoretical Biology* 7 (1): 17–52.
- Kuorikoski, Jaakko, and Petri Ylikoski. 2015. "External Representations and Scientific Understanding." *Synthese* 192 (12): 3817–37.
- Lange, Marc. 2013. "What Makes a Scientific Explanation Distinctively Mathematical?" *The British Journal for the Philosophy of Science* 64 (3): 485–511.
- Marshall, James A. R. 2015. *Social Evolution and Inclusive Fitness Theory: An Introduction*. Princeton: Princeton University Press.
- Nowak, Martin A., Corina E. Tarnita, and Edward O. Wilson. 2010. "The Evolution of Eusociality." *Nature* 466 (7310): 1057–62.
- . 2011. "Nowak *et Al.* Reply." *Nature* 471 (7339): E9–E10. Okasha, Samir, and Johannes Martens. 2016. "The Causal Meaning of Hamilton's Rule." *Royal Society Open Science* 3 (3): 160037. Pincock, Christopher. 2007. "A Role for Mathematics in the Physical Sciences." *Noûs* 41 (2): 253–75. ———. 2015. "Abstract Explanations in Science." *The British Journal for the Philosophy of Science* 66 (4): 857–82. Rice, Collin C. 2016. "Factive Scientific Understanding Without Accurate Representation." *Biology & Philosophy* 31 (1): 81–102.
- van Veelen, Matthijs, Benjamin Allen, Moshe Hoffman, Burton Simon, and Carl Veller. 2017. "Hamilton's Rule." *Journal of Theoretical Biology* 414 (February): 176–230.

What Business Ethics can do for Applied Philosophy

Thomas Wells (Tilburg University)

I argue that organisations should be taken more seriously by philosophers concerned with societal problems, from 'Fake News' or racial inequality. At present there is a tendency to focus at the level above (the justice of laws) or below (individual ethics). Yet a philosophical analysis of Fake News in terms of bad faith

speakers or principles of free expression would be seriously incomplete without considering the choices of a handful of major social media companies to adopt a particular (advertising based) business model. Insights and concepts from business ethics are an underused resource for such problems.

Academic sub-fields like the ethics of war or medicine develop because certain domains of human behavior generate peculiar problems unlike those elsewhere. In thinking about war, for example, one must find a way to save and sustain ethics under extreme pressures of power and violence. These sub-fields become specialized because they need to develop new conceptual tools and perspectives to get a grasp on the problems they are confronted by. This specialization in problems and concepts tends to create an ever greater distance between subfields. However, sometimes concepts developed in one place turn out to be helpful elsewhere. (For example, the principle of double effect developed by Catholic theologians has been used to analyse the ethics of killing in war and assisted suicide.)

This paper argues that business ethics is a sub-field whose hard-earned familiarity with organisations is a valuable resource that should be mined by mainstream applied philosophers. Liberal societies are not only made up of individuals and governments. They depend on a thriving civil society of organisations too. So it is a little odd that so much philosophical analysis focuses on the first two levels. Organisations have tremendous influence over a society's achievement of moral ideals, and in the evolution of what those ideals mean, such as 'meritocracy'. Most of us also spend a large amount of our time working in these organisations, directly under their control.

There are a number of important normative problems in society that would benefit from taking organisations seriously, as entities whose structure, values and decision-making processes are of independent significance. Especially if we are interested in actually solving those problems in the non-ideal conditions of the real world.

For example, one could argue that Fake News presents a challenge to the liberal principle of free expression, or one could analyse why it is immoral to say offensive things. But this would be to miss the elephant in the room: the adoption of a particularly toxic advertising based business model by a handful of large internet companies. The business ethics perspective makes the problem of Fake News more manageable problem both in terms of moral analysis and practical action. (Among other things, the bar for placing legal restrictions on company behavior for the sake of the public good is much lower than for persons.)

Various other societal problems seem amenable to a business ethics approach. For example, sexual harassment may seem like a directly ethical problem, but it has an

important organizational dimension – or it wouldn't be so persistent. Highly networked industries like academia, politics and entertainment have a power structure that makes employees particularly vulnerable to exploitation. That is the starting point for an organizational solution.

For another example, there is much debate about the extent of states' duty to accept refugees and economic migrants. But little attention to what happens next. The single greatest determinant of how migrants' lives will go is their integration into mainstream employment. That in turn largely depends on how organisations recognize, reward, and promote their talents and respect their equal dignity. Depending on the decision procedures used by companies, relatively small biases in judgement by individual officers can translate into systematically discriminatory hiring and promotion practices. Again, paying attention to how organisations work is necessary to diagnosing and addressing the problem.

Business ethics is specialized for thinking through the moral challenges that arise in organisations and their dealings with each other. It is an underused pragmatically oriented resource for thinking about a whole host of social problems that applied philosophers are nowadays expected to address.

Having a Good: the curious case of plants

Sander Werkhoven (Utrecht University)

For certain entities, the world is polarized into things that are good, bad, or evaluatively neutral – these are the kind of entities that 'have a good'. Christine Korsgaard has argued that all living beings have a good, given their functional make-up and the fact that they can perceive things as being good and bad for them. This raises the question whether plants have a good, as no sensory perception can be presumed in plants. I will argue that plants do have a good, and that central parts of Korsgaard's meta-normative position will therefore have to be rejected.

This paper addresses the question what it means to be a kind of entity for which things can be good or bad. Although this is generally considered to be a meta-ethical question, the 'ethical' in this label is slightly misleading, as answering it will not yet tell us anything about how we are to treat such entities, or whether they should enjoy any special status in practical deliberations. The question is more fundamental, and is better captured by what some now call 'meta-normative' enquiry.

Christine Korsgaard has proposed an influential answer to this question in a number of papers and book chapters. On Korsgaard's view, the kind of entity that has a good has to be a functional system. But not any functional system qualifies, as all entities can be regarded as fulfilling some sort of function to living beings (including entities like mountains, rocks, knives, and vacuum cleaners). Hence, a further criterion is necessary, namely that it is the kind of entity that 'has a sake' for which things are good or bad. Artifacts don't have such a sake: their functionality is entirely dependent on us, and so it's our sake that determines what counts as them being in a good or bad state. The kinds of entity that have a sake, according to Korsgaard, are living organisms: things can be good or bad for them insofar as they benefit or damage their health.

In her further specification why living organisms have a sake for which things are good and bad, Korsgaard's answer is that they can perceive them to be as such. She thereby follows a variety of other philosophers, including people like George Canguilhem and phenomenologists like Hans Jonas. Be that as it may, it raises the question where living organisms fit that do not have any perceptual qualities, like various lower animals and plants. On the one hand, it seems they are the kind of entity that have a sake for which things can be good or bad – plants can be healthy, contract diseases, and die. On the other hand, plants lack the perceptual qualities that Korsgaard used to distinguish functional entities that have a good (living organisms) from those that can be regarded as functionally useful to something else (basically all other substances). Which way should we go?

I will argue that nature of plants create a trilemma for Korsgaard. The first option is of the trilemma is that we reject that plants can be healthy or diseased, so that things cannot be good or bad for their sake. Plants are functional in just the same way as rocks, knives and vacuum cleaners are – i.e. only in relative to an entity that does have a good. This seems problematic for a number of reasons, not least because it requires a drastic revision of plant sciences: plants can't have genuine diseases anymore. The second option is that we reject that only living beings have a good, thereby opening the door to other entities having a good in the same way that we do. This is problematic because it implies losing the distinction between entities that are functional in relation to others (like artefacts) and entities that are functional in relation to themselves (living organisms) – a distinction that also someone like Kant recognized. The third option is to reject Korsgaard's proposal that being able to experience things as good or bad is what distinguishes entities that have a sake from those that do not. This is problematic because we do not have much in terms of an alternative.

In the remainder of the talk, I will argue that the third option is still the best answer to the trilemma. Rather than experience, it seems to be the mere fact that something is a living being that gives it a sake for which something is good or bad. The mere fact that something falls under the category of life, means that it has a good for its own sake, i.e. a final good. As we cannot point to anything

metaphysical unique about the living, it seems to be essential to *our* concept of life. I will end by noting that several remarks in Kant's third critique may ultimately be most helpful in answering the question what it means to have a good.

A reconstruction and an assessment of Yessenin-Volpin's ultra-intuitionism

Takahiro Yamada (Utrecht University)

I will analyse important notions of Yessenin-Volpin's ultra-intuitionism and the mathematics based on it, and argue that this standpoint should be located between Dummett's intuitionism and Wright's strict finitism. Researchers tend to classify Yessenin-Volpin's view as the conceptual source for strict finitism's research tradition, since Dummett criticised strict finitism and Wright attacked Dummett in favour of it. But I will attempt to reveal that the standpoint Dummett and Wright had in mind was not Yessenin-Volpin's itself. In so doing, I will argue that Yessenin-Volpin's mathematics is not as simply 'feasibilist' as it may appear.

What would mathematics look like, if we are really serious about the limitations of our cognitive abilities and tried to build new mathematics based on them? This is what I consider as one of the motivations behind constructive and/or finitist mathematics. According as what the limitations are, the form of the resulting mathematics has to change. Yessenin-Volpin's 'ultra-intuitionistic mathematics' is a kind of mathematics where the limitation is that of the numbers we can actually construct.

This talk aims to provide a reconstruction and an assessment of the philosophy of mathematics propounded by Yessenin-Volpin, called 'ultra-intuitionism'. This standpoint seems to have been attracting intuitionist (or constructivist) philosophers: in 1975, Dummett published the paper 'Wang's paradox' in which he attacked 'strict finitism' due to sorites-like difficulties, regarding Yessenin-Volpin's standpoint as a serious attempt of this 'strict finitism'; in 1982, Wright published the paper 'Strict finitism' and developed a meticulous counter-argument against Dummett, in favour of 'strict finitism'; and this tradition of investigation is continued still today. It would seem that ultra-intuitionism is one of the main conceptual sources for 'strict finitism'. However, even a glance into Yessenin-Volpin's 1970 paper 'The ultra-intuitionistic criticism and the antitraditional program for foundations of mathematics' would cast a doubt on this de facto received understanding of history. As I read it, Yessenin-Volpin's ultra-intuitionism is distinct from what Dummett and Wright considered it to be, or what we may expect they did. In this talk, I will attempt to show that ultra-intuitionism would be located between Wright's strict finitism and Dummett's intuitionism, by critically

analysing ultra-intuitionism's important notions.

My reconstruction of ultra-intuitionism takes the form of a reconstruction of Yessenin-Volpin's ultra-intuitionistic mathematics. As a philosophy, one could characterise ultra-intuitionism as a constructive and finitist view, and aptly point out that its defining principle is that a statement holds good just in case it is verifiable 'in practice', whereas the intuitionist would use the notion of verifiability 'in principle' instead. However, this basic principle is what Wright's strict finitism also assumes, and it is when we look into the ultra-intuitionist's mathematical system that the distinction between ultra-intuitionism and strict finitism is clear.

The focus of the first part of my talk will be on what the ultra-intuitionistically constructible numbers are, and what principles are responsible for them. As a rule, a legitimate natural number in ultra-intuitionistic mathematics is that which one can actually, 'feasibly', construct. But it turns out that the extent to which the ultra-intuitionist's feasibility reaches is surprisingly larger than one may anticipate. In this part of the talk, I will reconstruct Yessenin-Volpin's principle and methods, and present that the ultra-intuitionist could construct, for any given primitive recursive function, a totality closed under it. What allows this in his system is what he calls 'the central ontological hypothesis' ('c.o.h'). Ultra-intuitionistic mathematics is a kind of mathematics where the notion of a 'feasible number' is taken seriously: a number is called *feasible* just in case it is possible (for us) to count up to it. The ultra-intuitionist assumes (1) 0 is feasible; (2) if x is feasible, then so is the successor of x ; and (3) some number c is not feasible. The principle c.o.h. virtually allows the mathematician to 'iterate' this feasibility notion, so that she is said to be able to construct astronomically huge totalities. The existence of this principle would make a striking contrast with Wright's strict finitism, where the notion of a feasible number plays the central role, but no such principle as c.o.h. is adopted.

In the second part, I will address critical remarks on Yessenin-Volpin's system. I will attempt to argue that the notion of feasibility in ultra-intuitionistic mathematics has a flaw, by showing that in this mathematics the totality F of the feasible numbers cannot be accepted. Some readers attributed the belief in the existence of such a totality to Yessenin-Volpin, and indeed he seems to be embracing such a belief. But as I see it F is not self-identical. This is because, in his system, identity is considered as action of identification: to accept the identity statement $F = F$ is to accept that the process of checking the identity relations between the corresponding members of the two alleged 'objects' (i.e. the two F 's) can be completed. Since we suppose that $F = \{ 0, 0', 0'', \dots \}$, one has to accept that the process of checking $0 = 0, 0' = 0', 0'' = 0'', \dots$ terminates, in order to accept that $F = F$. But this identification process would continue as long as the mathematician tries to continue: therefore it does not terminate. If this argument is correct, then, jointly with the availability of the huge totalities, the ultra-intuitionist would be urged to reconsider the status of feasibility in its mathematics. My assessment of

ultra-intuitionistic mathematics will be that the ultra-intuitionistic notion of feasibility is significantly stronger than that of strict finitist mathematics, but weaker than the intuitionist notion of idealised agent ('creating subject'). I will argue for this conclusion by comparing the mathematical systems.

References:

Dean, W. (2018). 'Strict finitism, feasibility, and the sorites'. *The Review of Symbolic Logic*, 11(2), pp.295-346.

Dummett, M. (1975). 'Wang's paradox.' In his *Truth and Other Enigmas*, chap. 15, (pp. 248-68). Harvard University Press, Cambridge. Published in 1978.

Isles, D. (1980). 'Remarks on the notion of standard non-isomorphic natural number series'. In F. Richman (Ed.), *Constructive Mathematics*, Lecture Notes in Mathematics 873, (pp. 111-34).

Wright, C. (1982). 'Strict finitism'. In his *Realism, Meaning and Truth*, chap. 4, (pp. 107-75). Blackwell Publishers, Oxford U.K.; Cambridge U.S.A. 2nd edition in 1993.

Yessenin-Volpin, A. S. (1970). 'The ultra-intuitionistic criticism and the antitraditional program for foundations of mathematics'. In A. Kino, J. Myhill, & R. Vesley (Eds.), *Intuitionism and Proof Theory*, (pp. 3-45). Buffalo N.Y.: North-Holland Publishing Company: Amsterdam, London.

Yessenin-Volpin, A. S. (2011). 'From Esenin-Volpin -- 23. VII. 1958'. In D. van Dalen (Ed.), *The Selected Correspondence of L.E.J. Brouwer*, (pp. 457-9). Springer-Verlag. A letter to L.E.J. Brouwer.

Extended Abstracts in Order of Presentation (Symposium)

Symposium I.1

Divided Views and Divided Methods

Sujin Kim (University of Twente) and Petko Karadechev (Aalborg University, Denmark)

A study group was formed to discuss various approaches to Ethics of emerging technologies in health care. The study group used various angles of comparison to discuss both techno-anthropology and empirical bioethics, as well as mediation theory that seems to be useful to accommodate differences between the two and clarify the shared goals. Therefore, polarization of methodologies concerned the original problem between descriptive and normative ethics which the two approaches have already attempted to address. Polarization, however, also refers to the challenges to understand each other's' approaches. Difference in priorities of ethical analysis will be analyzed in the symposium. This exercise deepens our understanding of Techno-anthropology, bioethics, and philosophical anthropology in general. We will benefit from our discussion if we can discuss shared goal of both lenses, while we understand better the role of technologies as well as the role of critical reflection. Upon our previous experiences of learning-by-doing workshops, many students became aware of the diversity of approaches to critically discuss issues in emerging technology.

Contribution 1: Critical Bioethics and a Nuanced View Proposed

Understanding emerging technologies involve anticipating the potentials for technical and moral change. It also involves theorizing on the problems and solutions, especially on ethical issues. Therefore, the study group gained knowledge on how to analyze technological designs, participatory design efforts, as well as negotiations and governance efforts to appropriate these practices. Although this could involve a view on large array of practices, rather than focusing on specific area of practice that involves ethics (such as design), there is a clear advantage of having a holistic view. We can acknowledge the interconnection between these practices, we can also deepen our understanding on methodologies because many of them promise to provide insight on several or all of these areas. Our discussion of these two angles - anthropological ethics and philosophical ethics - will highlight strong and weak points of each perspective, also from the differences on methods and selection of cases.

The symposium compares techno-anthropology and empirical bioethics. What is at stake in our discussion are moral negotiation and moral innovation in the interfaces of human-technology relations. Theoretically, we try to articulate on the model that techno-anthropology students devised on assessment of technology. This will be used in the assessment of technologies in healthcare. Methodologies share the understanding on human-technology relations and the role of practice on shaping of norms. Original problem between descriptive and normative ethics might have been resolved from the approaches we looked into, empirical bioethics and anthropology. Polarisation in our discussion is addressing the different understandings of priorities when invoking, creating and applying ethical analysis. Indeed, the challenges persist when we attempt to understand each other's approaches or when we set out for a new assessment model. Regarding theory and practice balance, we would like to inquire the role that ethnography plays in the models of assessment. Regarding innovation, we inquire the potential of bioethics in emerging technology, deepening our understanding on the anticipation. or for technical and moral change. It also involves theorizing on the problems and solutions, especially thinking about ethical solutions neither precautionary and anticipatory. Between ethical theories of inter-disciplinary character, we human-technology relations. This exercise deepens our understanding of theories, or theories in philosophical anthropology in the broadest sense. The group aim to gain knowledge on how to analyse technological designs, participatory design efforts, as well as negotiations and governance efforts to appropriate these practices.

Well-informed anticipations were at play in the experiment of morality in emerging technology, for instance in an experiment of Google glass users (Verbeek and Kudina 2018) the users or designers were not free from built in effects or matters of concern. Ethical issues of privacy for instance in legal and corporate formulation stems from lack of control of information rather than an issue of multiplied self (multidimensional bodies) and uncoordinated sharing of information between them. Critical question to the mediation approach is the political importance of this angle. For bioethics research to deliver such argument of intervention from better view of moral change, either enhanced or degraded. And this has to be justifiable from the society at large from its democratic values. So, to which extent does macro-mediation of morality come into effect could be a question that determines moral agency (individual or social group) Hence, we have a nuanced view on the built-in effects of technologies on societal and normative frameworks by looking closely into our 'concerns' and 'built in effects' (privacy, sustainability e.g.) that strongly shape our anticipation of techno moral effects. If meaningful relations with emerging technologies is the new concern, this is more hermeneutic and practice-based.

Contribution 2: Model for Assessment of Technology (extension of Soft Edges of the Hardcore Technology)

The talk presents a techno-anthropological approach to emergent values of the good lives in bioethics. A research framework for emergent values in relation to bioethical questions regarding emergent technologies is proposed. In fact, enhancing human capabilities concerns an important goal. Our approach originates in a critique of current bioethics, which primarily focuses on moral right or wrong judgements based on universal ethical principles. We instead propose a bioethical focus on what we call the good lives and virtue ethics, based on Nicomachean Ethics. We emphasize the individual level and the investigation of embodied rather than universal values and how these evolve in relation to technology. We combine Børsen (2013, 2015), Berthelsen, Nøhr and Botin (2015). Institutional and societal macro levels are included as these should stand on the insights gained at the individual level. The approach offers a way to investigate value change in regards to the good lives. Opposed to current bioethics that focuses on clarity, this approach will open the door to pluralism, ambiguity and experienced values.

Symposium II.1

Concerning the self

Bas de Boer (UT), Jonne Hoek (UT) and Jan Bergen (UT)

In the tradition of 20th century continental philosophy we have seen a gradual decentering of the human subject. The human subject does no longer appear as an autonomous agent that can be considered the locus of freedom and reflection, and the beginning and end of all meaningful action. Instead, it has been argued that the human subject is actually shaped within power relations (Foucault 2005), as it only comes into being through contingent assemblages (Deleuze & Guattari 1987). Contemporary philosophers of technology of the continental persuasion have largely followed suit. They too reject any a-priori, transcendental ground for human subjectivity, and see the self and subjectivity constituted only in relational ontologies, either symmetrically bound in networks of actants (Latour 2005), or else mediated by technologies (Verbeek 2005). Clearly, the human self does no longer appear as the stable entity that can shape its own development. When uncritically embraced, the mentioned perspectives seduce one to view the human self as essentially deprived of its subjectivity. What remains is little more than a cog in the machinery of society, of technology, of biology. In this symposium, however, we will challenge this all too hasty interpretation and restore a legitimate, philosophical concern for the self. We simply cannot dismiss the fact that we human beings remain concerned with our selves. We try to get to know our selves

through technologies, engage ethically with other selves, and are challenged to create new forms of subjectivity in view of technological developments. Haven't we too easily dismissed important insights about human subjectivity that can be found in transcendental philosophical approaches? Or how else do we acknowledge that human subjectivity is indeed always shaped by the things it is surrounded with (e.g., technologies) and the situation from which it develops (e.g., socio-political structures, biological determinants)? What stronger notions of the self, or subjectivity are compatible with current philosophy of technology? These questions will be addressed along three lines: (1) subjectivity and ethics, (2) subjectivity and science, and (3) subjectivity and existential limits: (1) By discussing how understanding subjectivity as constituted through a Levinasian form of responsibility allows for developing an ethical ground for self-development and action. 2 (2) How subjectivity can engage with research in the cognitive neurosciences to be able to critically relate to the biological and neurological processes it finds itself being steered by. (3) By showing how postphenomenological theories of technological mediation conceptualize human-technology relations as existential relations and so picture the human self as a process of continuous self-reinvention.

References:

- Deleuze, G. & Guattari, F. (1987). *A Thousand Plateaus: Capitalism and Schizophrenia* (B. Massumi, trans.). Minneapolis: The University of Minnesota Press.
- Foucault, M. (2005). *The Hermeneutics of the Subject: Lectures at the Collège de France, 1981-1982*, 1st ed. New York: Palgrave-Macmillan
- Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press.
- Verbeek, P.P.-. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. Pennsylvania: The Pennsylvania State University Press.

Contribution 1: The Responsible Self: A Levinasian human(ism)for mediation theory

In philosophy of technology, mediation theory is an amodern approach to the study of humantechnology relations. As such, it eschews strict subject/object dichotomies and modern humanism (whether based on biological essence or a human monopoly on agency)(e.g., Verbeek, 2005, 2011). Rather, intentionality and

agency are seen as hybrid affairs in which humans and technologies partake together, thus doing away with the modern autonomous subject. Similarly, modern ethical theories will no longer do. Instead, Verbeek (2011) proposes an Foucaultian ethics of self-subjection. At first glance, this suggestion is somewhat puzzling in light of Ihde's (2003, p. 11) contention that post-phenomenology has substituted embodiment for subjectivity, rejecting the transcendentalist tradition from Descartes and Kant to Husserl, instead embracing Merleau-Ponty. The latter provides a foundation for the experiential or phenomenal self (as an alternative to different versions of the transcendental ego) (Zahavi, 2005) in the form of an embodied being-in-the(-technological)-world, which in turn provides the basis for the description of human-technology relations. However, in proposing the abovementioned ethics for mediation theory, Verbeek (2011) goes beyond phenomenalselfhood which focuses on the mineness of experience. Inspired by Foucault, he adds a normativity-oriented, narrative conception of the self (Zahavi, 2005), living through technology and continually selfsubjectivating through embodied 'technologies of the self'. For Foucault, subjectivation was a practice of freedom, leading to a critical ethics of resistance; an implication Verbeek does not wish to follow (he aims for constructive engagement rather than resistance). However, if it is not the spark of freedom that mobilizes subjects, what is left to fuel an ethics of selfsubjectivation? Mediation theory has yet to provide a satisfactory ethical foundation for subjectivation that is in line with its methodological, ontological or metaphysical assumptions (e.g., the primacy of experience, embodiment, human/technology co-constitution, etc). In this paper, I propose that part of such a foundation lies in the 'model' of selfhood implied in Emmanuel Levinas' explicitly ethical phenomenology (e.g., Levinas, 1969, 1981). First, I show that it includes a) embodied being-in-the-world (i.e., enjoyment) as the origin of the phenomenal self, b) the birth of subjectivity in our responsibility to the infinite Other, and c) a narrative conception of the self, the continual reconstruction of which is demanded by justice. Secondly, I elaborate on the origin of subjectivity as responsibility, which has likely eluded mediation theory due to the latter's tendency to ontologize (the appropriate question 4 being not what the subject is, but who). Thirdly, I explore the compatibility of this Levinasian conception of 'self' with mediation theory. This includes reflection on the role of technologies in the different 'stages' of the Levinasian Self and a reappraisal of transcendence and the 'alterity' relation. Interestingly, this Levinasian conception of subjectivity and selfhood opens up possibilities for a humanism based in radical passivity rather than a monopoly on agency, on alterity rather than sameness, on responsibility for others rather than biological genus or unbridled freedom (Levinas, 1987). This could possibly be the basis for a profoundly ethical, decidedly amodern, but nevertheless humanist foundation for mediation theory.

References:

Ihde, D. (2003). Postphenomenology—Again? Working Papers from the Centre

for STS Studies, 3. Retrieved from
[http://sts.au.dk/fileadmin/sts/publications/working_papers/Ihde -
_Postphenomenology_Again.pdf](http://sts.au.dk/fileadmin/sts/publications/working_papers/Ihde_-_Postphenomenology_Again.pdf)

Levinas, E. (1969). *Totality and infinity: An essay on exteriority*. Pittsburgh, PA: Duquesne University Press.

Levinas, E. (1981). *Otherwise Than Being or Beyond Essence*. Dordrecht: Kluwer Academic Publishers. Levinas, E. (1987). *Collected Philosophical Papers*. Dordrecht: Martinus Nijhoff Publishers.

Verbeek, P.-P. (2005). *What Things Do - Philosophical Reflections on Technology, Agency and Design*. University Park, PA: The Pennsylvania State University Press.

Verbeek, P.-P. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago, IL: The University of Chicago Press.

Zahavi, D. (2005). *Subjectivity and Selfhood*. Cambridge, MA: MIT Press.

Contribution 2: 'Existence' in theories of technological mediation

Postphenomenological theories of technological mediation conceptualize human-technology relations as existential relations. That catchword, 'existential', is rather vague though. Even if we agree that the word has something to do with the human 'self', its exact meaning is difficult to pin down, or even indefinable according to some philosophers. (Kierkegaard 1980) What is the function of this notion in the postphenomenological philosophy of technology? And what sort of human 'self' do such existential relations with technologies bring about? I will address these questions in three steps. First, I follow the positioning of the theory of technological mediation as an alternative to theories of technological extensionism and theories of humantechnology networks or power relations following Kiran & Verbeek (2010). Second, I will compare claims made about the changing nature of human existence due to technological developments against a more classically attuned Existenzphilosophie, namely that of Karl Jaspers. (Jaspers 2010) Third, I will compare and contrast two technologies that are said to mediate the existential self in novel ways. The analysis will reveal that the theory of technological mediation can be said to have a so-called 'meta-modernist' signature. (Vermeulen & van den Akker 2010) By alternating between the modern, and a-modern extremes of its contenders, it tries to formulate a new vantage point for positive engagement with technologies as structuring our vulnerability, possibilities, and

trust. Trying to argue for the attainment of such a middle ground, however, the technological mediation theory forgoes claims of conceptual primacy and declaration of ultimate values or goals. Pragmatically oriented and empirically informed, its mode of analysis stays so to say ‘amidst of things’, in medias res. Thereby, the approach seems to evade precisely those issues that – according to Karl Jaspers at least – most notably bring about existential reflection: limits. I will argue, however, that such a conclusion might be premature. Reflecting on a technology as a new middle ground implicates on reflection that it brings along also new extremes. New limits, foundations, and goals can so appear, though not as long as these limits are assessed in abstract. Rather, it only is by engaging with new technologies, by actively putting our selves into the balance, that we might come to terms with our technologically mediated self. In this call for engagement lies the true existential challenge that this theory of technological mediation advocates.

References:

Jaspers, Karl. *Philosophy of existence*. University of Pennsylvania Press, 2010. (Translation: Richard F. Grabau)

Søren Kierkegaard, *The Sickness unto Death: A Christian Psychological Exposition for Upbuilding and Awakening*, Princeton, NJ: Princeton University Press, 1980.

Kiran, Asle H., Verbeek, Peter-Paul “Trusting Our Selves to Technology”. *Knowledge, Technology & Policy*. 23, no. 3/4 (December 2010): pp. 409-27. 6

Vermeulen, Timotheus; van den Akker, Robin. “Notes on Metamodernism”. *Journal of Aesthetics and Culture*, Vol. 2 (2010): pp. 1-14.

Contribution 3: Critical Self-Identification: Human Subjectivity and the Cognitive Neurosciences

Although we are tempted to think about the self in terms of a unified entity, current research in the cognitive neurosciences suggests that the realization of different aspects of the self are not ‘governed’ by a unified agent – or can be traced back to an underlying substance (e.g., Gallagher 2013; Legrand & Ruby 2009). This reminds of Hume’s famous skeptical argument that ‘we are nothing but a bundle or collection of different perceptions’ (Hume 1978 [1739- 40], 252). Even if we accept developments in cognitive neuroscience, or a Humean skepticism indeed invalidate every understanding of the self as an entity, they do suggest that there can be a position from which questions about the ‘self’ can be posed. In one way

or another, it seems that it is possible to relate to one's own neurophysiology, or to engage in a position that allows Hume to deny the existence of the self. In line with critical perspectives on the self in current (continental) philosophy, I suggest to call this process through which it becomes possible to question what it means to be a 'self' human subjectivity (e.g., Taylor 1989, 514). Contrary to how Hume and an idealized view of the cognitive neurosciences approach the self, this implies that this question is posed by embodied, interested subject that already has a certain interest in posing this question. In this paper, I explore how research in the cognitive neurosciences allows human subjectivity to ask questions about the 'self' and opens up possibilities for critical self-identification. Firstly, I show how the concept of critical self-identification can be used to show that human subjectivity does not need to coincide with the historical and socio-political structures that it finds itself relating to, but can be better understood as mediated by those structures (cf. Verbeek 2011). Secondly, I discuss how Catherine Malabou attempts to develop a notion of 'plasticity' that allows to establish a critical relation with our biology, and uses this notion to open up a space for self-reflection that does not necessarily need to coincide with societal demands (e.g., Malabou 2008, 11). Thirdly, I suggest that the concept of critical self-identification can be of help when exploring how current cognitive neuroscientific research using brain imaging technologies can help constituting forms of human subjectivity.

References:

- Gallagher, S. (2013). A Pattern Theory of Self. *Frontiers in Human Neuroscience*, 7, 443.
- Hume, D. (1978 [1739-40]). *A Treatise of Human Nature* (L. A. Selby-Bigge, ed. Revised by P.H. Nidditch). Oxford: Oxford University Press.
- Legrand, D. & Ruby, P. (2009). What is Self-Specific? Theoretical Investigation and Critical Review of Neuroimaging Results. *Psychological Review*, 116(1), 252-282.
- Malabou, C. (2008). *What Should We Do With Our Brain?* (S. Rand, trsl.). New York: Fordham University Press.
- Taylor, C. (1989). *Sources of the Self: The Making of the Modern Identity*. Cambridge: Harvard University Press.
- Verbeek, P.P.-. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: Chicago University Press.

Symposium III.1

Politics & technologies of the city - “All Watched Over by Machines of Loving Grace: Liberal Democracy and the Smart City”

Michael Nagenborg (UT) Bart van Leeuwen (Radboud University) Marcel Muller (KRITIS - TU Darmstadt), and Ryan Wittinglow (University of Groningen)

Smart cities undeniably offer administrative benefits in domains such as traffic management, energy, and waste disposal. However, I argue, when designed without clear and plastic governance structures in place, smart city programs pose a risk to the procedural norms that underpin liberal democracy.

In his poem, “All Watched Over by Machines of Loving Grace”, Richard Brautigan proposes governance by machines. Instead of democracy, he writes, we will have a “cybernetic ecology” where animals and computer co-exist in “mutually programming harmony”. The cost of this service is but a small thing. We need only offer our liberty to Brautigan’s titular “machines of loving grace”.

Although it’s been a long time since 1967, the utopian impulse that the poem articulates—the notion that we will be fitter, happier, and more productive if we allow ourselves to be administered by intelligent machines—remains compelling. Within the domain of urban design and planning, this impulse has recently expressed itself in a suite of proposals concerning what is called the “smart city”: that is, using information and communication technologies to autonomously manage municipal systems with the aid of sensors and actuators managed by artificial intelligences. Pilot programs are already underway in a number of Dutch cities, including Eindhoven, Utrecht, and Assen.

At least among philosophers, some of the flaws inherent in this view are well known. Machine learning algorithms possess politics just like any other artefact. The mere fact that these algorithms demonstrate some degree of autonomy by no means implies that they do not inherit the cultural and epistemic biases of their creators and users. There are, however, other concerns associated with optimising the panopticon.

Liberal democracy is inherently pluralistic about the good. It is not a political system with an explicit end in mind, but is rather constructed around the notion that there is more than one good. What is good for you may not necessarily be good for me. Moreover, it presumes that *no single one* of these goods is any “more good” than any other. While certain norms are valued within liberal democratic societies — “tolerance” is one that springs to mind — these are rarely considered as goods in and of themselves. Rather, the “goodness” of tolerance is strictly extrinsic. In a liberal democratic society it is necessary that people tolerate the

existence of other, competing notions of the good, just as they expect their own notion of the good to be tolerated in turn.

This pluralism comes with certain administrative costs, however. Given that lawmakers cannot guarantee that all individually-held notions of the good are commensurable, there must instead exist deliberative processes, systems, civic structures, and institutions in which competing notions of the good can be discussed, argued, and—eventually—endorsed. It is via this deliberative procedure that liberal democracies are able to reconcile its pluralism about the good with the demands of governance: out of the plurality, via deliberation, a smaller number of goods are decided upon and prioritised. This implies that the “public good” in liberal democracies is in constant flux. It is always up for debate, for revision, for reflection, changing slowly but surely in line with shifting intellectual and moral norms. The good is not a caged creature, but a vital, dynamic entity in which we constantly participate via deliberation, negotiation, and the procedures that constitute our republic.

These processes mean that liberal democracy is necessarily both slow and inefficient. This inefficiency is not a flaw. While the collectivist nature of liberal democracy means that decisions are almost always rendered less quickly than in authoritarian forms of rule, this very slowness helps guarantee that new policies and new behaviours are given sufficient deliberative scrutiny. It is upon these profoundly inefficient deliberative procedures that the legitimacy of liberal democratic systems is premised.

Smart city technologies sit poorly with these values, however. Indeed, the whole point of letting a network of sensors, actuators and artificial intelligences take over municipal services is that these slow, inefficient deliberative procedures are no longer required. Unfortunately, it is only within the context of these deliberative procedures that the good can be discussed, argued, and decided upon. While it’s all very well to design smart cities with a good or an ethic in mind, the pluralistic nature of liberal democracy means that the good is not a set of values but an ongoing *process* by which the polis can discover and redefine itself. Any democratically legitimate smart city system must meaningfully participate in and respond to this process.

Unfortunately, to my knowledge, not a single smart city pilot plan possesses a governance structure that can respond to the inherently plastic nature of the liberal democratic good. Consequently, current smart city plans run the risk of being incompatible with the liberal democratic process without clear and transparent governance structures in place. While using artificial intelligences — our titular “machines of loving grace” — to swiftly and opaquely replace deliberative decision-making processes would quite likely produce positive outcomes in certain domains, we should not adopt these systems thoughtlessly and without reflection.

Symposium III.5

Inequality, Democracy and Property: Part Two

Michael Bennett (Utrecht University), Rutger Claassen (Utrecht University), and Yara al Salman (Utrecht University)

A major source of contemporary polarization is the polarization of wealth holdings. The papers in these two panels explore what inequality means for democracy, freedom and the justification of property. The classical conception of a liberal-democratic society depends upon a clear split between the private and the public sphere. Ideally, political institutions protect public interests and private individuals/groups can freely pursue their private interests. However, contemporary interpenetrations of the private and public sphere render this claim hollow. Wealthy individuals and corporations influence politics through means such as selectively funding political parties and organs of public opinion. These panels investigate which normative criteria should be used to evaluate the influence of wealth on democratic institutions. Democratic theories often make idealising egalitarian assumptions, and the transition to non-ideal theory may necessitate re-thinking the foundations of democratic legitimacy and fairness. Even more than these abstract questions, however, we focus on what policies and institutions are required in the light of these reflections. In particular, is it feasible to insulate politics from economic inequality, or is something more radical necessary? Are huge wealth holdings unacceptable as a standing threat to democracy, freedom, and autonomy? If so, what can be done about them? These panels explore various aspects of this under-appreciated political rationale for policies to promote greater economic equality. Beyond this, we also question whether the structure of classical liberal ownership itself needs to be amended to make room for democratic equality. Should shareholders be compelled to share control over firms with their workers? What would the structure of property look like if we paid more attention to the opportunities people actually have rather than formal rights against non-interference? And finally, if the idea of the public-private divide should be attenuated, should we move away from individualistic, contractarian models of the justification of property itself? This pair of panels will be relevant to all those interested in these topical questions at the intersection of economic justice and democratic theory.

Contribution 1: How to Argue for Workplace Democracy

Many are troubled by the apparent contradiction between our commitment to democracy in the state, and the hierarchical conditions prevailing in most workplaces. This paper critiques the standard philosophical argument for state promotion of workplace democracy, and then sets out what a more convincing

argument would look like. The most popular philosophical argument for promoting workplace democracy has been the “parallel case” between workplace and state. 1 Implicitly or explicitly, these arguments rely on the assumption that, when workplace democracy is made compulsory or subsidised, states thereby confer a benefit on workers. However, this assumption will not hold. If the state starts to require the inclusion of workplace democracy in labour contracts, we cannot expect other aspects of the contracts to remain the same. Unless the underlying balance of supply and demand is changed, employers will predictably offset the workplace democracy requirement by reducing other benefits workers receive – most obviously, salaries. In the end, workers are unlikely to derive greater net benefit from their jobs by insisting that shared control over the firm is always among the benefits they receive. It is easy to overlook that the current regime already allows people to join worker-run firms, and these predominate in particular sectors of the economy, such as law or accountancy.² Overriding the apparent preferences of workers in other parts of the economy is unlikely to best promote their welfare. A sound case for promoting workplace democracy is therefore unlikely to be found in the welfare of the individual workers who thereby gain a share of control over their firms. Instead, we should focus on potential benefits to other stakeholders in society. Two general categories of moral failure afflict business life. 3 On the one hand, managers are tempted to exploit their knowledge and power when it comes to intra-firm agency relations with investors and workers. On the other, the firm as a whole is tempted to exploit market failures with its customers, suppliers and neighbours. Against the first problem, workers can help investors police managerial abuses through the first-hand knowledge they have of the firm. Against the second, workers are more embedded in the communities firms affect and so less likely to condone abusive behaviour. Ultimately, these are conjectures which require empirical investigation. However, they indicate the lines along which a satisfactory argument for promoting workplace democracy should run. There is nothing deeply significant about the employment relation per se; workers just happen to be well-positioned to hold business managers and investors to account. Promoting workplace democracy should thus be seen as part of a broader project of encouraging countervailing forces against overly powerful interests in society.

References:

- Dahl, A Preface to Economic Democracy; McMahon, Authority and Democracy;
Anderson, Private Government.
- Hansmann, The Ownership of Enterprise. Heath, Morality, Competition, and the Firm.

Contribution 2: Property and Positive Freedom

Defenders of negative freedom have often argued in favor of strong private property rights, embodying a commitment to a minimal state and an extensive sphere of capitalist interactions between individual property-owners. Indeed, one highly influential liberal strand in political thought holds – following John Locke and others – that the legitimacy of the liberal state is based on its ability to protect the property of its citizens. Some contemporary libertarians even argue that rights to non-interference can basically be derived from the right to self-ownership. This paper will inquire what a defender of positive freedom should say about property. This link has had much less investigation than the relation between negative freedom and property. If one believes the legitimacy of the state depends on its ability to protect its citizens' abilities to exercise self-government (or their abilities to lead an autonomous life), what does this imply for the right to property? First, the paper introduces a definition of the ideal of positive freedom and its essential components and the link between positive freedom and property. Arguably positive freedom is most often used as a basis for justifying the provision of welfare state services, paid out of taxation. Hence, the link lies in positive freedom's ability to justify (coercive) taxation and redistribution to pay for services (like health care and education) which enhance citizen's individual autonomy. This arguably leaves the primacy of private property, which is so central to the negative freedom tradition within liberalism, intact. The remainder of the paper investigates the hypothesis that positive freedom cannot be linked so strongly to private property. Instead, the ideal may require different forms of property (communal property held by collectives, public or state property, and corporate property), depending on the circumstances. This hypothesis mirrors the economic point of view which allows for a diversity of property structures depending on what is most efficient, replacing however the idea of efficiency with the idea of positive freedom.

Contribution 3: Starting from World Ownership: Stopping Anywhere

Can theories of world ownership help us think about distributive justice? Left libertarians and some of their critics seem to think so. In developing their theories of justice, they do not start from people's needs, but from people's pre-institutional property rights to the world's resources. These theorists believe that, depending on what you take these original claims to be, you will arrive at different theories of justice (Fisher, 2015; Otsuka, 2003; Fabre, 2002; Van Parijs, 1997). There has been some debate about which world ownership theory can best make sense of our moral intuitions, both in terms of the starting point it describes and in terms of the principles of justice it arrives at (Risse, 2005). Meanwhile, critics argue that any such theory will have undesirable implications (Arneson, 2010; Cohen, 1995). Few, however, have questioned the very use of these theories as a device for

doing political philosophy. This paper argues that world ownership theories do not help us to think about distributive justice. This is not because they lead to normatively unattractive outcomes, but because they simply cannot do any normative work. The paper provides a clarification of world ownership theories, demonstrating that these theories can be used in two ways that are not adequately distinguished in the literature. The first is as a starting point for a bargaining position, the second is as a contemporary claim that individuals have. I also show that different existing conceptions of world ownership rely on ideas about property regimes that are incomplete and sometimes even incorrect. I improve these conceptions so that they fit better with actual property regimes and with bundle of rights property theory. I argue, however, that not even these conceptions can help us as a starting point to think about distributive justice. Once we have excluded unequal world ownership as a plausible starting position, any other original ownership thesis becomes indeterminate. The normative work is always done elsewhere and is independent of the starting point. I do suggest, however, that world ownership could be an attractive end point in a theory of justice. This would also be in line with functional theories of property.

References:

- Arneson, R.J. (2010) Self-ownership and world ownership: Against left-libertarianism. *Social Philosophy and Policy*, 27 (1), 168-194.
- Cohen, G.A. (1995) *Self-ownership, freedom and equality*. Cambridge: Cambridge University Press.
- Fabre, C. (2002) Justice, fairness, and world ownership. *Law and Philosophy*, 21 (3), 249-273.
- Fisher, A. (2015) A left-libertarian proposal for egalitarian world ownership. *Critical Review of International Social and Political Philosophy*, 18 (6), 599–619.
- Otsuka, M. (2003) *Libertarianism without inequality*. Oxford: Oxford University Press.
- Risse, M. (2005) How does the global order harm the poor? *Philosophy and Public Affairs*, 33 (4), 349-376.
- Van Parijs, P. (1997) *Real freedom for all: What (if anything) can justify capitalism?* Oxford: Oxford University Press.

Symposium IV.2

Contentious Embodiment

(Organized by: OZSW Study group Feminist Philosophy)

Annemie Halsema (VU), Nathanja van den Heuvel (Leiden University), and Sanne van Driel (Tilburg University)

The proposed symposium is organized by the OZSW study group Feminist Philosophy. Feminist Theory is well-known for theorizing the body. The papers in this symposium will demonstrate the various ways in which embodiment is conceived of in feminist and critical theory. We will discuss ethnographical, phenomenological, neo-materialist and poststructuralist approaches to embodiment. We consider embodiment “contentious” in more than one sense. In the first place, the body is still underthematized in philosophy (for instance, in debates around personal identity). Within the history of Western thought the body has predominantly been conceptualized as belonging to the realm of nature, set apart from our rational faculties and therefore less worthy of serious philosophical attention. In the second place, in those cases in which the body is thematized, embodiment most of the times is considered in a neutral and abstract way. Factors that in daily life are important for our sense of self - such as gender, race, ethnicity, sexual orientation, age and ability – are hardly ever accounted for. The proposed papers do consider these factors, and therewith contribute to the development of philosophical frameworks for understanding the relation between embodiment, subjectivity and identity. The focus on everyday life experiences is the last sense in which embodiment is contentious: thematizing embodiment implies bringing everyday practices to philosophical consideration. The individual papers thematize different aspects of embodiment, that are part of everyday practices: exercise, sports, sexuality . In paper 1, the general problem of the relation between embodiment and personal identity will be addressed. The paper aims at a notion of personal identity that takes into account intersectional factors that situate a person socially. Based on empirical research on the experiences of elite sports women, paper 2 reconsiders poststructuralist frameworks in sport and gender research and by drawing on the philosophy of play proposes an alternative framework for understanding the relation between discipline, subjectivity and agency. Paper 3 will discuss ideas about embodiment in relation to ‘mental health’, drawing on (auto)phenomenographic research on the practice of running therapy. It will reflect on the embodied position of a (feminist, crip) researcher in the field of mental healthcare, and propose to shift the research lens between embodied experiences and processes of abledment/ableism.

Contribution 1: The socially situated self: intersectionality and personal identity

Since the introduction of the notion of intersectionality in gender studies in the late eighties of the last century (Crenshaw 1989), it is seen as constitutive for personal identity. It is generally taken as an alternative for the single axis perspective of identity politics, because of its focus on the interrelation between gender, class, race, ethnicity and other social determinants (Crenshaw 1991; Yuval-Davis 2006). Intersectionality has been challenged for focusing too much on identity, while overlooking social structures and institutions (Yuval-Davis 2006), for considering a too strict notion of identity, that does not help overcoming identity politics (Grosz 2011), but also for considering identity too little (Staunæs 2003; Prins 2006). In all of these cases, the “identity” referred to primarily signifies one’s belonging to social groups on the basis of bodily characteristics. “Identity” pertains to the social determinants that situate the self, the so-called “intersectional axes” (Yuval-Davis 2006) or “crossroads” (Crenshaw 1989) that position the self socially. In this paper, I will argue that even though this is a relevant aspect of personal identity, it does not provide a complete account of it, and that theorizing intersectionality can gain from understanding personal identity in a fuller sense. Apart from what can be called the “idem”-aspect of personal identity, or whatness, that the intersectional factors pertain to, identity also has an “ipse”-aspect (whoness) that characterizes someone as a person. It is precisely the latter aspect that constitutes experiential selfhood. In order to understand how selfesteem and selfrespect are related to social determinants, such as gender, class, race, intersectionality theories need to consider both sides of identity. For a notion that combines these aspects, I suggest to turn to the philosophical debate over personal identity, more specifically to the narrative accounts of the self, developed in the nineties of the last century by Paul Ricoeur (1990) and Marya Schechtman (1996). The trouble is, however, that these philosophical accounts of the self do not thematize the social determinants of personal identity. While in daily life, factors such as gender, race, class for a large part constitute who we are, in most philosophical theories they are barely accounted for. The narrative notions of identity therefore need an extra dimension. In the last part of the paper, I will develop this alternative: the self as expression, that is embodied and that does take into account one’s social situatedness. On this basis, personal identity will prove to pertain to subjective self-understanding as well as to the social structures that situate the self, and it will become understandable that these structures not only position us socially, but that we are also able to negotiate them.

Contribution 2: Elite women sports and the limits of critique

Towards a feminist playology The introduction of Marxist and poststructuralist social theory into the field of sport research has enabled critical sport studies to

further think the relation between language, the body and the subject's inside. In modern sports, discipline is no longer applied from without, but internalized in such a way that the conformation to norms becomes the mode of self-realization. To be more precise, critical sport studies criticize modern sporting practices for treating the body of the athlete as inert matter that can be moulded, shaped, quantified and measured by sport data analysts, psychologists, trainers, physicians and dieticians. Focussing on identity constructions and the question of gender-normalization in relation to high performance sport, sport and gender studies complement these critiques by addressing the intimate relation between gender and techniques of control. Hence critical sport and gender research importantly show that despite the growth in media and public interest and increasing levels of support systems in the world of elite sports, oppressive gender norms and stereotypes still persist. However, based on extensive fieldwork - in-depth interviews with elite sports women and participant observation - this paper argues that a poststructuralist framework alone is not capable of making sense of the experiences of elite sports women, more philosophically put their place in the world. The experiences of the players as reflected on by themselves reveal that these 'objectivist' analyses cannot fully account for the players' willingness to sacrifice health, social life etc. For this willingness transcends elite sporting culture and its disciplinary techniques. Second, this paper argues that the missing aspect in critical sport and gender research to adequately account for players' willingness to sacrifice, must be understood as an element of play, or the 'playfulness' of sports. Along with the categories of 'gender policing' and 'high performance' 'playfulness' must be considered a key structure of the experiences of elite sportswomen. Play here does not refer to a particular form of behaviour praised for its biological or psychological benefits nor to a certain attitude of the individual sports woman, or to the structures of the game as such. On the other hand, drawing on the work of the pre-rational play theorists, Johan Huizinga, Eugen Fink, Hans Georg Gadamer and Mihai I. Spariosu, in this paper play will be discussed as an embodied, affective experience that belongs to an ontological distinct category, i.e. a reality that surpasses the individual and temporarily suspends 'ordinary' or 'real' life. Thus third, this paper argues that the theorisation of the workings of gender, and high performance, should be complemented with a theoretical account of play. It is only by adding this third aspect that a comprehensive understanding of the structures of experience of elite sports women can be accounted for. Hence play makes visible that social constructivist and critical theoretical accounts of the workings of gender and disciplinary practices miss important insights about the structures of experience. The analysis of playfulness is however not independent of the analysis of gender or disciplinary practices. Therefore an account that is sensitive to playfulness will yield new insights about subjectivity, sexual difference and agency.

Contribution 3: Being able to live a daily life? Embodiment and the process of abledment in the treatment of depression

If according to the World Health Organization depression is the leading cause of ill health and disability worldwide (WHO 2017), it is to say that depression gets in the way of people's daily functioning, causing not only suffering to individuals and their loved ones, but also, and not in the least, costing society a lot of money. Speaking from within this neoliberal health paradigm, the WHO presents the body as a resource from which to extract productive and successful living. When the body does not deliver, something needs to be done –whether it is to find the cause and cure, get psychological treatment or find a way to self-improve - to become healthy and able again. The dualisms ill/healthy and able/disabled, are only two of many dualisms structuring modern discourse –mind/body is another important one- following a divisive and excluding logic in which one is always on top of the other, some bodies more than others pushed to the margins of society. As disability scholar Fiona Kumari Campbell (2009) argues with Bruno Latour (1991), it is the modern division between able/disabled that allows for a process of translation from 'disabled' to 'able' to be set in motion. In other words: disability is the negative identity created to produce ableness. On an individual level this means, as Foucault showed in his work on disciplinary power and normalization, that to find oneself wanting, doing or being something society values negatively, makes one both develop a negative sense of self or identity and work harder to overcome this. It could be argued, as it is being done in critical disability and affect theory, that it is precisely the high prevalence of feelings of exhaustion, fatigue, dread, sadness, feelings of being stuck, not able to go on anymore – often labelled as burn-out or depression – that points to the exhaustion of neoliberal capitalism and its disabling politics. On the level of personal experience however, this critical analysis hardly helps one in need of care, as a subject in society caught up in this ableist orientation. A bodily reorientation, to borrow from Ahmed's *Queer Phenomenology* (2006) or a focus on the materiality of bodies, as Campbell (2009) suggests, is needed to find a way out of disabling dualistic logic. This implies the affirmation of discomfort and unruliness 'spilling' from bodies labelled as ill or disabled, tuning in to the experiences of living bodies seeping through tight and suffocating mental health regimes. Furthermore, instead of pointing the gaze at labelled bodies from the assumption of health and ableness, Campbell proposes to shift the gaze. In line with Foucault's ethical-aesthetic imperative of freedom practices: to refuse who you are (to refuse identity), she refuses the notion of able(ness), and instead shifts her attention to processes of abledment. It is this tension between critical analysis of social-political structures (diagnosing society) and writing about and from personal embodied experiences (including internalised ableism and self-stigmatisation) that I will address in this paper. Drawing on my – ongoing- (auto)phenomenological research on the practice of running therapy as treatment for depression in the Netherlands, I propose to shift the research lens between embodied experiences and processes of abledment/ableism.

Symposium IV.5

Foucault Philosopher: Twenty-First Century Perspectives

Michael Leezenberg (UvA), Michael Karskens (Radboud University), and Laure Bastiaans (UvA)

The recently completed publication of Foucault's Collège de France lectures and the appearance of the fourth and final volume of the History of Sexuality provide the occasion for a new stocktaking of the philosophical relevance of this formidable French thinker. This symposium will focus on the philosophical dimensions of his work, rather than on its - equally important - historiographical or social-theoretical aspects. A first, and currently highly topical, philosophical dimension is Foucault's ongoing dialogue with Marxism. Against a number of recent reductionist analysis, it emerges that Foucault's 1979 lectures on neoliberalism cannot be assimilated to either a liberal narrative that accepts market forces as better warranting individual liberties than social democracy, or a Marxist narrative that takes the political economy to be methodologically primary. Crucially, for Foucault, class identities and class relations are mediated by power, and hence by knowledge; and the neoliberal conceptualization of free markets and individual liberty should be analyzed as a form of government. A second area of philosophical importance is Foucault's theoretization of power. Famously, Foucault introduced notions like discipline, biopolitics, and pastoral power as complements or alternatives to the hitherto dominant concept of sovereign power as state-oriented, repressive, and expressed in the form of law. His own notions, however, raise new questions. For example, the notion of biopolitics was originally developed as a complement to discipline, and as a way of explaining commonalities in the twentieth century's two 'diseases of power,' viz., fascism and Stalinism; but Foucault never really elaborated on this term, thus leaving it unclear, for example, how it relates to more familiar theoretical studies of fascism and totalitarianism. Although the notion has opened up an entire field of study, we may - and as philosophers should - question its conceptual coherence, the more so as Foucault seems to have tacitly abandoned the notion in the course of his 1979 lectures. A third still largely unexplored area is Foucault's contact with and use of analytical philosophy. It is relatively well known that the notion of discursive practices owes much to Wittgenstein's account of language games and to Austin's and Searle's Speech Act Theory; but these links, fruitful as they may be for analytical philosophy, have hardly been explored, thematized, or elaborated. Thus, one wide open question is whether Foucault's genealogical approach to the concept of truth has anything to currently popular analytical and pragmatist views on truth, like Davidson's and Rorty's. Against these, Foucault's scattered remarks suggest a non-epistemic account of truth that pays far greater attention to aspects of conflict in knowledge and language than the analytical tradition does. This panel aims at addressing some of these questions, hoping to raise a renewed philosophical

interest in Foucault's Protean, and still continuing, output.

Contribution 1: Foucault and Analytical Philosophy: Towards a Genealogy of Truth

Foucault's affinities with, and use of, analytical-philosophical ideas has often been noted but rarely been addressed in detail. The contacts between his approach to discursive formations and Austin's and Searle's Speech Act Theory is relatively well known; his use of Wittgenstein's notion of language games for understanding what he calls 'power games,' i.e., everyday practices of exercising power, have been studied rather less; and the relevance of his scattered remarks on truth for analytical have not, for as far as I am aware, been discussed in detail. In this contribution, I will therefore look at Foucault's sketchy remarks pointing to a genealogical perspective on truth, and discuss how this perspective contrasts with analyticalphilosophical (and in particular pragmatist) approaches like Davidson's and Rorty's. First, I will trace the history, or genealogy, of truth. Unlike a number of analytical philosophers, Foucault does not take assertions (i.e., declarative sentences that are at least potential bearers of truth conditions) as the most basic kind of statement. Moreover, he argues that truth is not a primitive notion but an effect of particular operations on statements (or, more generally, on sets of statements). In his published writings, he never developed a history of how truth became a central notion; but in his 1971 lectures at the Collège de France, he provides at least a sketch for such a genealogy of truth, discussing the notion of *alètheia* from archaic poets like Hesiod through the sophists to Aristotle's discussion of apophantic statements, that is, assertions or truth-bearing statements or sentences. He is cautious to distinguish this historical sketch from Heidegger's famous characterization of truth as *Unvergessenheit*, arguing that the ancient notion of truth involves becoming rather than being, and in part precedes the metaphysical concern with being. In this, he appears to follow Nietzsche rather than Heidegger. Second, I look at the systematic implications of a genealogical perspective on truth. Foucault has often been blamed for allegedly giving up the belief in truth and in objectively valid knowledge; but it turns out that he precisely rejects this kind of skepticism. First, I will trace how Foucault appears to imply a non-epistemic account of truth, which reduces questions of truth and knowledge to something radically different, viz., a will to know. Next, I will briefly contrast this position with Davidson's and Rorty's. Like them, Foucault, appears to dissolve, or overcome, skepticism as a major philosophical problem; but unlike them, he does not think of truth in terms of usefulness and progress, but rather in terms of conflict. His position may therefore amount not so much to an epistemological skepticism concerning true statements but rather to, so to speak, a political mistrust of the concept of truth. I will conclude by discussing how this genealogical perspective on truth contrasts with a number of widespread analytical-philosophical assumptions concerning epistemology and the philosophy of

language.

Contribution 2: Bringing the State Back In: Neoliberal Rearticulations of Discipline

Famously, Foucault develops an account of power that avoids thinking of power in purely negative terms of repression and/or distortion; he also tries to uncover modalities of power that cannot be characterized in terms of sovereignty; that is, they are not exercised in or by the state, and are not articulated in terms of laws. This presentation will explore whether and to what extent Foucault actually succeeds in going beyond the state, which at one point he calls a 'fiction' and a 'mythical abstraction'. Is perhaps the state more strongly implicated in nonsovereign forms of power than he is willing to admit? And if so, what does this imply for his accounts of, in particular, discipline and biopolitics? Discipline, according to Foucault, is a non-sovereign, productive form of power that is articulated and exercised not in terms of laws and transgressions but of the normal and the pathological. It is exercised not directly by the state; but it is generally exercised in institutions or spaces that have for a long time been the prerogative of the state, like schools, prisons, and hospitals. Foucault believed however that disciplining made way for more indirect ways of governing in modern western societies, with the emergence of American neoliberalism that according to him involved a crucial remaking of state, society and the self. The notion of biopolitics refers to an equally non-sovereign form of power, but which is totalizing rather than individualizing. The concept, however, more or less disappears from his work, when he focuses on current neoliberal government. This talk will explore whether we face some recombinations of power modalities in today's western societies, that involve a more authoritarian type of state than Foucault's analysis allows for. Current forms of normalizing practices in present-day neoliberal societies can be fruitfully captured in terms of a 'neoliberal-statist' kind of discipline and biopolitics. I will focus on present-day neoliberal forms of normalization in the education system, through statesponsored and state-sanctioned forms of monitoring and screening.

Contribution 3: Civil society as heterotopia and pastoral power

It has often been noted that Foucault has at some moments been in a fruitful and productive dialogue with Marxist ideas, while at other moments rejected Marxism wholesale in a rather more polemical manner. The precise nature of his views on neoliberalism, by contrast, is currently still more controversial and more contested. This contribution will confront Foucault's conception of civil society with both Marxist and neoliberal ones. Foucault's discussion of civil society is the final part

of his analysis of (neo-) liberalism. We will argue that his approach of civil society was based on a post-marxian counter-model of power, being an ongoing struggle (lutte), and on Christian model of pastoral power, being an ongoing group process of individualizing subjection and subjectivation. From the point of view of counterconduct, civil society is a continuous opposition to state and government, to class struggle, and to neo-liberal individualism. This approach opens up an alternative framing of civil society, which was already offered by Foucault himself in his essay on Heterotopias. Our proposition is that Civil Society reconstructed as a political heterotopia offers a more powerful conception of civil society as open society and unlimited glocal or metatopical space in which ‘subjects’ do not act as *homines economici*, but as publicprivate group members. From the point of view of pastoral power, civil society is the modernized version of a church, as an accumulation of (local) non-family communities in which social ties are neither based on explicit contracts, nor on voluntary association, but on a serial of “disinterested interests”. The (pastoral) power in that community is “spontaneously” shaped in concrete, here and now, subjection and subjectivation practices. I will conclude by comparing and contrasting this conception of civil society with currently dominant (neo-) liberal and Marxist accounts.

Contribution 4: Foucault and Neoliberalism: Between Transgression and Resistance

Since the English translation of *The Birth of Biopolitics* first appeared in 2008, Michel Foucault’s philosophical engagement with neoliberalism has been an object of intense debate. Taking Foucault’s analysis of the reconfiguration of neoliberal human capital theory and its rhetoric of individual self-enhancement as central theme, we will argue that Foucault’s discussion of the neoliberal subject, practices and discourses cannot be understood without taking into account his concept of resistance. Starting with his earlier work on transgression and contestation, we will follow Foucault’s trajectory of thought regarding his thoughts of resistance and emphasize recurring themes such as the decentralization of the subject, his refusal to define any limits to resistance, and his persistent critique of any notion of resistance in dialectical, revolutionary and/or emancipatory terms. By linking these concepts to his lectures on neoliberalism, we will demonstrate that Foucault’s writings on resistance can be used to stake a critical position against the neoliberal narrative of the subject and its failure to recognize the shifting border between selfenhancement and self-effacement as well as the humanistic grounds on which such a theory is based. I will contrast this position with more familiar liberal and Marxist images of neoliberalism. In this way, it will also become clear that recent attempts to qualify Foucault as himself sympathetic to neoliberalism are quite mistaken

Symposium IV.6

Inequality, Democracy and Property: Part One

Annelien de Dijn (Utrecht University), Constanze Binder (Erasmus University Rotterdam) and Dick Timmer (Utrecht University)

A major source of contemporary polarization is the polarization of wealth holdings. The papers in these two panels explore what inequality means for democracy, freedom and the justification of property. The classical conception of a liberal-democratic society depends upon a clear split between the private and the public sphere. Ideally, political institutions protect public interests and private individuals/groups can freely pursue their private interests. However, contemporary interpenetrations of the private and public sphere render this claim hollow. Wealthy individuals and corporations influence politics through means such as selectively funding political parties and organs of public opinion. These panels investigate which normative criteria should be used to evaluate the influence of wealth on democratic institutions. Democratic theories often make idealising egalitarian assumptions, and the transition to non-ideal theory may necessitate rethinking the foundations of democratic legitimacy and fairness. Even more than these abstract questions, however, we focus on what policies and institutions are required in the light of these reflections. In particular, is it feasible to insulate politics from economic inequality, or is something more radical necessary? Are huge wealth holdings unacceptable as a standing threat to democracy, freedom, and autonomy? If so, what can be done about them? These panels explore various aspects of this under-appreciated political rationale for policies to promote greater economic equality. Beyond this, we also question whether the structure of classical liberal ownership itself needs to be amended to make room for democratic equality. Should shareholders be compelled to share control over firms with their workers? What would the structure of property look like if we paid more attention to the opportunities people actually have rather than formal rights against non-interference? And finally, if the idea of the public-private divide should be attenuated, should we move away from individualistic, contractarian models of the justification of property itself? This pair of panels will be relevant to all those interested in these topical questions at the intersection of economic justice and democratic theory.

Contribution 1 Title: Republicanism, Freedom and Social Equality: From Machiavelli to Robespierre

This paper explores the interrelation between freedom and social equality in the republican tradition from Machiavelli to Robespierre. By exploring the work of thinkers including Machiavelli, Harrington, Noah Webster, Thomas Jefferson and

Robespierre, I aim to show that the fostering of social equality was an important goal for early modern republicans. More specifically, I aim to show that they believed that true democracy - and therefore republican freedom - was impossible to maintain in a society characterized by a disparity of wealth. I conclude by reflecting on the value of this much-neglected aspect of early-modern republicanism for contemporary neo-republicanism.

Contribution 2 Inequality, Freedom and Democracy

Despite a sharp rise in economics inequality in recent decades, redistribution policies are still heavily disputed. One reason why policies of redistribution are often opposed is that such policies are taken to distort/reduce incentives on free markets. (Increasing) inequality is often taken to be justified as long as it improves the situation of the worst off due to a growth of overall wealth. In this article we shall focus on freedom and inequality and put arguments voiced against redistributive measures under scrutiny. We shall focus in particular on the relationship between inequality and democratic freedoms and identify preliminary conditions under which an increase in inequality can lead or contribute to undermining democratic freedoms. We conclude the paper with a discussion of different policy measures that aim to strengthen democratic freedoms. It is argued that a crucial element of policies aiming to safeguard democratic institutions is to address economic inequality.

Contribution 3 Title: Political Equality and the Problem of Extreme Wealth

In this paper I will argue that a concern for political equality gives strong reasons to view extreme wealth as morally objectionable. I will also argue that limitarian policies, which limit the amount of wealth individuals can appropriate (e.g. a high inheritance tax, or maximum income), are a necessary step towards establishing political equality. The argument for extreme wealth as a moral problem runs as follows: political equality requires that nobody should have more or less political influence just because they are richer. However, extreme wealth allows the rich more political influence precisely because they are richer. This means that it threatens political equality. Therefore, to the extent that limitarianism³ limits extreme wealth, limitarianism is pro tanto justified as a means to promote political equality. Call this the Democratic Argument. One may counter that the Democratic Argument does not render extreme wealth morally objectionable, but points to a relative concern about inequalities in wealth. In this paper I develop three arguments which show that given current inequalities in the distribution of wealth extreme wealth is itself morally significant. They are developed against the background of the actual distribution of wealth in current liberal democratic

societies, rather than from an ideal state in which wealth and political power are strictly distinct. The first argument is the Capital Clustering Argument. Financial capital tends to attract nonfinancial capital, such as knowledge and opportunities ('social capital'), and behavioral norms and dispositions ('cultural capital'). The increasingly substantial differences in access to nonfinancial capital that come with being a member of a specific wealthy social group translates into unjustifiable differences in opportunities to political influence. Limitarian policies target capital clustering and promote political equality by directly targeting the accumulation of wealth. The second objection is that under current non-ideal circumstances extreme wealth itself grants an inordinate amount of power. Call this the Domination Argument. Wealth can be used to encroach on the liberty of others, thereby exercising enormous power over the everyday life of others. In real life, under our present non-ideal circumstances, the purchasing power of the very wealthy ranges over schools, land, houses, factories, companies, hospitals, etc. and through this they gain enormous power over the every-day life of others. And domination in those spheres (e.g. the market sphere) translates into inequalities in opportunity to influence politics. The third argument is the Self-Serving Bias Argument. People tend to be biased towards their own conception of the good and the role they played in their financial success. This self-serving bias is reinforced by extreme wealth, and contributes to polarization on peoples' views about predistribution and redistribution. People are biased against seeing the illegitimacy of being rich and the political influence that comes with it, and the illegitimacy of others having much less of both. The unequal distribution of political influence due to, among other things, extreme wealth, combined with the problem of cognitive biases by those in power, reinforces the claim that extreme wealth threatens political equality.

Symposium IV.7

Regular and Irregular Motions: Normativity in Cavendish's Natural Philosophy

*Laura Georgescu (RUG), Jonathan Shaheen (Uppsala University), and Boris Demarest
(Heidelberg University)*

This panel addresses the status, place, and source of normativity in the natural philosophy of Margaret Cavendish. Scholars have recently begun considering how she sought to deal with such questions. Since Cavendish makes at times perplexing and seemingly paradoxical claims about orderly (or regular) and disorderly (or irregular) natural motions, there are a number of underexplored but important exegetical issues we propose to address. For instance, scholars have begun debating whether Cavendish's views on regular and irregular motions commit her

to the existence of norms in nature, and whether and how those views can be reconciled with what many take to be her claim that nature as a whole is regular and orderly. On one account, defended by Walters (2014), Cavendish's distinction between regular and irregular motions in Nature does not commit her to normativity in nature. Additionally, Walters argues that both regular and irregular motions are necessary for the overall order of Nature. Cuning (2016), however, allows for a limited sense in which Cavendish would ascribe normativity to Nature. He argues that norms for natural motion are ultimately indexed to our perspective, such that what gets classified as irregular is whatever runs counter to our concerns and our current understanding. Nevertheless, he agrees with Walters that, from the perspective of the whole of Nature, these apparent irregularities are part of a harmoniously ordered structure.

By contrast, Detlefsen (2009) and Boyle (2017) both maintain that irregularities are true disorders in Nature. On this view, finite parts of Nature are supposed to act in accordance with a set of prescribed norms, and it is in relation to such norms that their motions get qualified as regular or irregular. If a motion is "fit" for the norm in place, then it is regular; if it is a violation of the norm, then it is irregular. Because parts of Nature can sometimes err, irregular motions are possible. Given this, Detlefsen and Boyle defend a normative reading of natural motions: error is built into Cavendish's metaphysics, and is not merely indexed to a human standpoint, as in Cuning's interpretation. Rather, God is the source of normativity in Nature. For Detlefsen, God "rationally suggests" norms for parts of nature to follow in order to produce harmonious, orderly motion throughout Nature. For Boyle, God's role is less advisory: God "commands" Nature to behave according to "Natural Rules" for motion.

The papers in this symposium challenge the conceptual frameworks and assumptions that structure this debate on regular and irregular motions in Cavendish's philosophy. In doing so, they shed new light on important problems in the interpretation of Cavendish's philosophy (such as the relationship between nature and its parts, and between nature and God), and clarify the problems of irregularity and normativity and their place in nature.

Contribution 1: No true disorder, no true order: Regularity and Irregularity in Cavendish's Philosophy of Nature

This talk contributes to the ongoing debate about normativity and regularity in Cavendish's philosophy by challenging what I take to be a presupposition held by many involved in the debate: that Nature as a whole is the kind of thing that can be ordered (or disordered). I argue that, in Cavendish's philosophical system, especially as developed in the *Observations upon experimental philosophy* (1666) and the *Grounds of natural philosophy* (1668), it is a category mistake to treat the

whole of Nature as regular or irregular. I defend this claim with two arguments. One argument builds on Cavendish's claims about parts and wholes: specifically, I take Cavendish to deny both whole-to-whole interactions and whole-to-part interactions – i.e., wholes qua wholes are by definition non-relational. Regularity and irregularity, though, are relations, and hence can only exist between parts, and cannot be extended to the whole of Nature, given its non-relationality. This is why, strictly speaking, the whole of Nature is not the kind of thing that can be ordered or disordered. The second argument builds on Cavendish's claims about the effects of nature. Everything that could ever possibly be experienced is an effect of natural motions, but Nature as a whole is not *strictu sensu* composed of these effects. Consequently, while they are, in some sense, effects "of" nature, and while we can talk about aggregates of such effects, in doing so, we say nothing about Nature as a whole. In which case, even if we could talk about the order and disorder of the effects of nature, we would still not be saying anything about Nature as a whole. However, all of this does not entail that regularity and irregularity are entirely illusory. From the perspective of the parts of nature, I show that nature's motions have to be both regular and irregular. My claim here takes the form of a conditional: if there is to be the kind of a world in which we encounter particular creatures (finite associations of finite parts enjoying diachronic unity) with ordered behaviour – that is, given the kind of world in which we, and Cavendish, do seem to live – then regular and irregular motions (of parts in relation to other parts) are necessary. Maximum regularity and maximum irregularity simply cannot bring about the kind of world we are familiar with.

Contribution 2: The Source of Normativity in Cavendish's Metaphysics

Margaret Cavendish held that Nature harmoniously governs the orderly motion of her (Nature's) parts, but she also traced societal and personal ills like war and disease to disorderly or irregular motions in Nature. Fitting these commitments together generates puzzles for Cavendish scholars. One such puzzle concerns the status of normative constraints on what count as orderly or regular motions. In her recent monograph, Deborah Boyle structures that debate as between two camps who differ over whether irregular motions truly violate norms. In this piece, I suggest a re-orientation of the debate, such that what is at issue is not primarily whether norms are truly violated, but the source of such norms. The debate is complicated by Cavendish's claims that disorderly motions are required for the overall harmony of nature. As Cavendish puts it, "opposite" motions ensure that Nature does not "run into extremes" and thereby into "confusion" (e.g., at *Observations upon Experimental Philosophy*, 1.27; 1.31; 1.37, Q6 and Q10; 2.6; 2.20; and 3.4, as well as *Grounds of Natural Philosophy*, 1.14, 2.10 and 3.1). In the reconstruction I offer, her claims here are intimately related to her argument against the void. She associates the void with parts moving away from the parts that surround them ("dividing," as she puts it) without immediately thereby

becoming surrounded by other parts ("composing," as she puts it). She also takes plenism to be a precondition of harmony, and so her argument against the void is, as Karen Detlefsen puts it, a normative argument. But though the avoidance of confusion through the opposite motions of composition and division results in Nature's harmony---a seemingly positive normative status---divisions themselves have a negative normative status. In particular, regularity and order are for Cavendish matters of what I call hypothetical normativity. Hypothetical normativity is, roughly, suitability to the achievement of some end. In Cavendish's panpsychist metaphysics, creatures are understood as societies of finite parts of Nature, unified by mutual passionate love. The typical end of such a society is its own preservation. Since divisions are disunifying motions, they are contrary to societies' typical ends. So divisions are violations of hypothetical normativity, even as they are a necessary ingredient in Nature's overall harmony. Since Cavendish takes order and regularity to be grounded in what I am calling hypothetical normativity, those passages in which Cavendish discusses "proper" or "fitting" motions can easily be accounted for, without joining Boyle in attributing a crucial legislative function to God. Whereas Boyle takes such passages to be evidence for the view that norms are divinely prescribed, I read them as descriptions of how societies should behave if they want to remain together, as I think Cavendish intended. Moreover, the position I take relative to the debate as I re-orient it has the benefit of making Cavendish right about the role of God in her philosophy. She accounts for normativity, as she represents herself doing generally in her philosophy, without invoking God.

Contribution 3: Autonomy and Heteronomy as sources of order and disorder

In a recent monograph, Deborah Boyle defends the claim that Cavendish held a "true disorders" view according to which nature contains real irregularities and disorders that are violations of natural norms. Boyle argues that Cavendish conceives of irregularities and disorders as behavior initiated by parts of nature that runs counter to the essential nature of said parts, violating the norms God has suggested or decreed for the behavior of parts with that particular nature. Boyle indicates, however, that, except in the case of human beings, Cavendish does not seem to offer a clear explanation of why parts of nature would exhibit such "unnatural" behavior – she merely seems to suggest that parts of nature are capable of doing so because they are free in a libertarian sense. In this paper, I suggest a reading of Cavendish's account of "unnatural", "disorderly" and "irregular" behavior that allows us to understand how such behavior arises. Specifically, I propose the reading that orderly behavior is behavior guided by the *norms or rules governing the nature of the part of nature initiating the behavior*, whereas disorderly behavior is guided and prompted by the norms or rules foreign to the nature of the part of nature initiating the behavior. In a sense, then, orderly behavior is autonomous, whereas disorderly behavior is

heteronomous. On this reading, all behavior of parts of nature is according to some “nature”, and I will argue that this can be important for understanding what Cavendish might mean when she seems to ascribe order and disorder to nature as a whole. Moreover, the reading allows us to relate Cavendish’s account of “nature” and “normativity” to an Aristotelian framework that has similar motivations (e.g. accounting for the behavior of biological entities) and similar implications (e.g. a strong distinction between “natural” and “artificial” that limits the usefulness of experimentation in natural philosophy). Finally, this reading suggests that there are important points of agreement between Cavendish and Van Helmont that are underappreciated when one focuses mainly on her explicit rejection of the latter’s matter theory.

Symposium V.1

Ethics of human enhancement: overlooked issues important for policy

Stephanie Gauttier (UT), Sean Jensen (UT), Lily E Frank (TUE) and Lesley Ann Daly (University of the Arts London)

Human enhancement is transitioning from a futuristic possibility to a set of existing interventions that people can and will use. Individuals already have the option to utilize digital technologies for cognitive, psychological, and physical human enhancement. While prominent examples of this kind of use have existed for quite some time, they were restricted to a small group of individuals who label themselves as first-generation cyborgs or post humans. But use of cyborg and enhancement technologies are going mainstream. For instance, companies are beginning to encourage employees to get microchip implants, and some Swedish citizens are installing these implants as a pass for the gym or public transit¹. While this particular technology has yet to reach the stage of enhancing human capability directly, its development signals a growing interest in engineering oneself and augmenting one’s abilities with technology. Policy-makers have commissioned research to get a deeper understanding of the ethical issues surrounding human enhancement so as to prepare adequate regulatory measures. To be efficient, this should look beyond the themes tackled by policy concerning technologies, which have traditionally been regulated from the point of view of issues that have already arose such as privacy or safety. Motivating this symposium is the idea that philosophers can make an input into the policy debate by bringing forward what is specific to human enhancement technologies and absent from the current rhetoric. We bring different perspectives forward to identify future areas of work and methods to inform policy-making through applied philosophy. Our first speaker will focus on a dimension of the use of technology rarely considered by

asking how to consider the impact of human enhancement onto the meaning attached to our actions. Then, the next speaker will discuss the potential for moral enhancement that technology brings. Issues such as these need to be communicated to the public who will be affected by resulting policies, so they can participate in the debate with a clear understanding. Communicating to the public is an intervention that shapes the way individuals think about human enhancement. Thinking critically of these interventions and finding a way to embed the philosophical issues at stake is also needed. The third speaker will describe her approach to communicating with the public and the ethical issues this raises. The speakers and the audience will then engage in a conversation on how philosophy and ethics can contribute to the practical debate on human enhancement in society and policy circles, and which aspects of the debate on human enhancement need to be pushed forward.

Contribution 1 Human Enhancement Technology & Meaningfulness: A Proof-of-Concept

Human enhancement, long relegated to a speculative field of what-ifs, is quickly transitioning into an active scientific enterprise yielding new technologies poised to change society in minor and major ways. In order to navigate the emerging changes with ethically positive results, we must know more about what is at stake. Philosophers often discuss the potential for enhancement technologies to affect well-being, or to change human nature. Few, however, have explored the potential for enhancement technologies to affect meaningfulness: how we find it, how we experience it, and how we understand it. My project is focused on analysing how human enhancement technologies may affect meaningfulness in positive and negative ways, as well as how the emerging field may change how we view meaningfulness at a more granular level. My symposium talk will consist of a brief walkthrough of the approach I plan to take in this project, followed by a test case in which I will provide a proof-of-concept analysis of an enhancement technology and discuss how this technology could affect meaningfulness. Using previous research, I will consider how a speculative neuro-implant with the ability to insert advertisements into one's dreams could affect meaningfulness. Pertinent questions include: would the invasion of one's dreams using this technology affect the meaningfulness in one's life? If dreams can play an important role in making a life meaningful, it seems that this technology could impact meaningfulness by altering the authority a user can effectively attribute to what they experience in their dreams. However, if one views dreams as an experience that cannot impact the meaningfulness in an individual's life, then we must look further to know whether such a device would impact meaningfulness. Further, could a commercial neural implant that invades dreams alter modern views of meaningfulness? If such a technology becomes widespread, a society of users may embrace losing the subconscious authority inherent in uninvaded dreams in favour of the convenience

of receiving potentially useful information even whilst asleep. Although I would expect most modern persons would find this notion disturbing, one could argue that minor erosions of the kind are taking place as mobile and internet technologies alter the way we communicate, connect and interact with others. It seems that many are comfortable with the subtle changes occurring due to such technologies. By investigating how our understanding of meaningfulness is changing we can take a deeper look at whether or not we should embrace the unimpeded development of such technologies.

Contribution 2: Wearable and implantable moral technologies as a form of moral enhancement: what are the risks?

Moral bioenhancement, nudge-designed environments, and ambient persuasive technologies may help us behave more consistently with deeply held moral convictions. Alternatively, they may aid us in overcoming cognitive and affective limitations that prevent us from appreciating a situation's moral dimensions. Or they may simply make it easier for us to make the morally right choice, helping us to overcome sources of weakness of will. In this paper I make two assumptions. First, technologies to improve our moral capacities are realizable. Second, such technologies will actually help us get morality right and behave more consistently with whatever the 'real' right thing to do turns out to be. I then consider whether or not we lose anything valuable, particularly opportunities for moral progress, when we make being moral much easier, eliminating difficult moral deliberation and internal moral struggle. I ultimately reject the worry that moral struggle has value as a catalyst for moral progress, understood here as the discovery and application of new values or sensitization to new sources of harm.

Contribution 3: Critical Design exploration of the impact of Human Enhancement

Currently categorised as non-medical, many Human Enhancement (HE) devices do not legally have to conform to regulations that medical devices must. Therefore, HE products go to market without the manufacturer having much practical knowledge of the use and affects, and without regulated testing. HE policy recommendation documents raise issues such as: therapy vs enhancement; accessibility; lack of regulation; and morphological freedom. Through reading these documents they all advocate bringing the public into the discussion as they are who will be affected by the development and widespread implementation of the technology. However, as such dense documents, it is hard to see how the public will be able to understand and meaningfully engage with the issues they raise. Which led me to question: 'How can Critical Design methodologies be used

to examine the impact of Human Enhancement, and assess ethical issues related to the design and use of the technology?' Critical Design methodologies are utilized to interrogate issues of technological impact and critique emerging technologies. By critically analysing the social, political and ethical contexts of devices, along with their functional and aesthetic design, informed futures scenarios can be modelled in material and experiential form. These creative provocations about pragmatic and ethical factors affecting HE users, raise important questions about the trajectory of the technology, and the industrial context that cultivated it. By deconstructing and mapping provoking topics within policy recommendation documents this project aims to create an experiential design piece. This piece will allow the public to obtain greater understanding of the field HE and its potential positive and negative consequences, allowing them to contribute to the conversation.

Symposium V.4

Hermeneutics: Analytical and Continental Prospects

Wouter Bisschop (Vrije Universiteit Amsterdam), Marcus Duwell, (Utrecht University) Sanem Yazicioglu (Tilburg University) and Gert-Jan Van der Heiden (Radboud University)

Contemporary hermeneutics exhibits a very disunited and confused picture. What in the past referred to the method or theory of text interpretation comes these days with a wide variety of very different adjectives—‘theological’, ‘philosophical’, ‘narrative’, ‘interreligious’, ‘radical’, ‘ecological’, ‘naturalistic’, ‘transcendentalist’, ‘Japanese’—and can have, besides texts, very diverse objects: installation art, religious education, the imagination, the film world, hell, action, to name a few. The meaning of ‘hermeneutics’ in these uses often remains rather unclear. Most introductory texts to the field of hermeneutics are very not helpful either. Depending on era and author, ‘hermeneutics’ is used to refer to the methodology, science, art, or rules of interpretation (e.g. Chladenius), to the theory of interpretation or understanding (e.g. Schleiermacher), to reflection on understanding as “die Seinsweise des Daseins selber” (Heidegger, Gadamer), to a particular approach to philosophy or literary studies, analogous to e.g. (post-)structuralism, deconstruction, constructivism, or to a set of philosophical doctrines, e.g. about the historicity of our being and understanding (e.g. Gadamer).

Do these different notions have something in common, or is it, ultimately, merely a historical accident that all these different projects get called by the same name? And what are the future prospects of this apparently divided field? This symposium aims to facilitate a conversation on the nature of hermeneutics between proponents of different philosophical approaches. Apart from contributing to a deeper understanding of the field and the identification of future

projects, the symposium also aims to reflect on the remarkable absence of hermeneutical issues on the agenda of so-called ‘analytical philosophy’. Is hermeneutics inherently non-analytical, or is there room for something like an ‘analytical hermeneutics’?

Among the specific problems on which this symposium invites reflection are the following. First, how do the different historical notions of ‘hermeneutics’ relate to each other? Almost every introductory text on hermeneutics consists of a historical overview that describes views as diverse as those of Friedrich Schleiermacher and Hans-Georg Gadamer on understanding. Should we think of the youngest as continuations of the elder approaches, as them being phases in a historical development? Or does diversity—whether acknowledged or not—dominate historical continuity in such historical overviews? Second, is ‘hermeneutics’ a discipline at all, or should we rather think of it as a philosophical position, as nominalism is in metaphysics, or reliabilism in contemporary analytical epistemology? Third, how does hermeneutics relate to other philosophical subdisciplines? Is it part of metaphysics, or epistemology, is it rather a discipline of its own, or is it a philosophical view with implications for particular subdisciplines in philosophy? Fourth, what’s left of the Wilhelm’s Dilthey’s idea of hermeneutics as the philosophy of the humanities? Is hermeneutics for the humanities what the philosophy of science is for science? Fifth, is hermeneutics an inherently ‘continental’ endeavour, or is an ‘analytical hermeneutics’ possible? The symposium will be successful if it fosters an encounter between continental and analytical styles of philosophizing, and leaves us with a firmer grasp of the variety of hermeneutics and, perhaps, a sense of its unity.

Contribution 1 Hermeneutics and Analytical Epistemology

Analytical philosophers spent little time on hermeneutical problems. Part of the explanation could be that there is significant confusion and disagreement as to what counts as a ‘hermeneutical problem’. In this paper I try to identify the various kinds of projects that have been labelled ‘hermeneutical’, consider how they relate to each other, and address the question as to how this cluster of problems relates to analytic epistemology. In this way, I hope to provide a way to understand the relations between the great diversity of the projects we find in the history of hermeneutics, and to indicate the projects with which analytical philosophers could engage. There are at least four kinds of questions or projects recognizable in the history of hermeneutics, and in the first part of my paper I aim to describe each of them. In particular the latter three are arguably ‘hermeneutical’ in nature. First, there are epistemic questions: What does something *x* mean? Second, the methodological questions: How are we to answer the epistemic question? Third, the epistemological questions, such as: When have we obtained knowledge of the meaning of *x*? And fourth, a number meta-epistemological questions, such as ‘how

is understanding at all possible?', 'What is it to think?' The second part of my paper aims to examine the relations between these four projects, and their relation to the field of analytical epistemology. The latter three kinds of questions I identified are questions at successive 'levels' of reflection. This 'level approach' in describing the projects with which hermeneutics has been concerned enables us to see how the radically different projects of what is often called 'traditional hermeneutics' and Gadamer's 'philosophical hermeneutics' relate to each other. Moreover, it can be argued that in all four projects reasonable and important questions are raised; there is no reason to think that one of the projects is 'outdated' or superseded by a more recent project. Third, there is no principled reason to think that the hermeneutical projects I distinguished suit so-called 'continental' philosophers better than their 'analytical' colleagues. In fact, I'll argue that a significant part of what gets called 'hermeneutics' concerns varieties of the issues with which analytical epistemology is concerned (the epistemological problems), or applied thereof (the methodological questions). Other parts of it belong to ontology, logic, or some other philosophical discipline (some of the meta-epistemological problems). I propose, then, to think of hermeneutics as a part of epistemology or meta-epistemology.

Contribution 2: Is There a Future for Hermeneutics?

Thinking about the role of hermeneutics is important for current philosophical discourses for various reasons. For one, hermeneutics is at the centre of the divide between continental and analytic philosophy. Moreover, hermeneutics is the most paradigmatic endeavour to secure a particular role for the humanities. That means that the discussion about the topic is much more important than only a reflection of a specific branch of philosophy. This paper will explore a possible profile of philosophical hermeneutics along a historical and a systematic line and will conclude with some suggestions as to its future prospects. The mainstream discourse assumes that only the Heidegger-Gadamer tradition of hermeneutics was really able to develop hermeneutics as a philosophical discipline. The judgments of Gadamer about Kant, Schleiermacher and Dilthey were decisive for the dominant view on the nature of hermeneutics. A necessary prerequisite for a future discourse on hermeneutics seems to be a critical revision of this view. This is not to say that I think one can directly continue the philosophical endeavours of Schleiermacher or Dilthey, but a critical reconstruction of the early routes of hermeneutics helps to explore the various possible roles for a future hermeneutics. Rudolf Makkreel offers a fruitful reconstruction of a line from Kant's third Critique to Dilthey's hermeneutics as a model for the humanities. Kant-scholars like Held, Herman, Longuenesse, O'Neill or Waxman have emphasized the vital role of judgment and understanding for the critical project. I shall explore how hermeneutic approaches that emphasize the linguistic and cultural embedding of the interpreting subject allow for critical philosophy. In this way, Makkreel and

others may help us to overcome the problematic oppositions in the historiography of hermeneutics introduced by Gadamer. The assumption is that such a revision makes room to see hermeneutics not just as a technique of interpretation or as an overarching approach to philosophy, but to emphasize that philosophical reflection on the praxis of understanding is important for all other philosophical disciplines. This approach to hermeneutics will be discussed with reference to its relationship to ethics and morality. The guiding assumption is that a promising route to the understanding and justification of morality would locate moral commitments in the consistent self-understanding of agents, a view that sometimes is referred to as Kantian constructivism. For such a route to ethics, it is important to understand the relationship between the hermeneutical view on understanding and the basis of normativity. This connection has been underdeveloped in current debate because the Heidegger-Gadamer-tradition shows some hostility to normative reflexion in the first place, while mainstream analytic philosophers often cherish their allergic reactions to the hermeneutic tradition. The above-mentioned exploration of alternative routes to hermeneutics may be helpful to overcome those unproductive deadlocks. The paper will conclude with some suggestions for further debates in hermeneutics.

Contribution 3 Encountering the World: Perception, Apprehension, Interpretation

This contribution investigates how to extend the sense of hermeneutics along the lines of the reflective judgment in Kant and the notion of intersubjectivity in Husserl's phenomenology. In *The Poetics of Space*, Bachelard writes, "when the image is new, the world is new." (1994, 47.) Yet, how does the world become new to us and can we share this new world(s) with the others? These questions are of pivotal importance to understand what hermeneutics can mean today from a phenomenological perspective. The world and the image have a particular importance in Husserl's phenomenology: the primary task of the philosopher, as Husserl says, is to constitute the world as *eidos*; only this task can found philosophy as a rigorous science. The world should be constituted as *eidos*, because our perception of the reality of the world operates in two ways: (i) our perception of the world changes due to changing perspectives, and (ii) our apprehension of these perceptions are fundamentally interpretative. This sense of the *eidos* points to a similarity between Husserl and Kant because both aim to give another meaning to the word *eidos* or *Idea* than Plato did. For both Husserl and Kant, *eidos* or *Idea* signifies a unity in the manifoldness of perceptions and their apprehensions; this unity renders them communicable. In establishing this unity, the function of Kant's distinction between determinate and reflective judgments is comparable to the function of Husserl's distinction between judgments that are based on determinate and indeterminate perceptions of reality. Reflective judgments and indeterminate perceptions are not only interpretative, but they are

hermeneutical because they enable intersubjective communication. To elaborate these points, I will first discuss the relation between the interpretative character of apprehension and its transposition into a judgment and, second why this transposition is inevitably hermeneutical. To demonstrate the former, I will compare the reflective judgments to the judgments based on indeterminate reality, and to demonstrate the latter, I will compare Kant's maxim of "enlarged thought" to Husserl's intersubjectivity. These analyses expose how transcendental philosophy offers a wider conception of hermeneutics: "It is transcendental, not in the sense providing building blocks of truth, but in the sense of opening up the reflective horizon of communal meaning in terms of which the truth can be determined." (Makkreel, 1990, 158.)

Paper Abstract 4 What is the Relevance of Hermeneutics Today? While attention to interpretation and the art of interpreting correctly goes back to ancient times, the philosophical attention to hermeneutics has a genuinely modern flavor; it "bears the signature of modernity", as Jean Grondin suggests. In order to answer the question of what the relevance of hermeneutics today is, it is important to address the basic elements of the modern philosophical enterprise that have been formative of hermeneutics and philosophical hermeneutics and to ask what the remaining or renewed importance of these elements is. In the first part of this paper, I will discuss the following, closely related elements that determine or drive the specific epistemological enterprise of hermeneutics. One should think here of (1) the finitude and historical—linguistic, cultural—nature of the human being and of human understanding, which leads in hermeneutics to the concepts of horizon or perspective/perspectivity; (2) the importance of the notion of life (starting from Kant's use of it in the third critique, up to Dilthey and Nietzsche) as a category that seems to point to both the source and the goal of human understanding, which can never fully be recuperated by reason and understanding (in this sense, the relation between life and understanding stipulates the finitude of, e.g., human self-understanding and determines it as a hermeneutic process); (3) the dimension of plurality, which is an important element if one wants to understand why, strictly speaking, hermeneutics cannot favour either the first person perspective or the third person perspective, but has an intrinsic emphasis on the second person or social perspective; moreover, this plural perspective has inspired both the more "peaceful" hermeneutical models of Gadamer's dialogue and the more "combatant" models of interpretation that we find in Nietzsche, giving also rise to what Ricoeur called the "conflict of interpretations." In the second part of this paper, if time permits, I want to investigate in which sense exactly these three elements are crucial to understand why hermeneutics is an indispensable epistemological tool in the realism debates that are nowadays appearing on the crossroads of analytical and continental philosophy (e.g. in Gabriel, Ferraris, Benoit, and Harman).

Posters in Alphabetical Order

The ethics of using big data for biomedical research: a case study

Marieke Bak (Amsterdam AMC)

The use of big data for biomedical research brings about ethical dilemmas regarding the balance between participants' rights and the societal value of research. Implementation of new European data protection legislation has focused our attention on privacy concerns, arising from the large-scale use and linkage of these sensitive personal data. However, moral reflection should include broader individual- and group-level harms, and include discussions on the respective importance of values like autonomy and solidarity. This poster presents a case study among researchers collecting big data in a particularly challenging context: medical emergency settings where participants cannot give prospective informed consent.

In recent years, biomedical research has increasingly made use of “big data”, with large-scale observational studies being conducted in the hopes of leading to improved patient outcomes and cost savings in healthcare. However, processing of these datasets brings about a number of ethical concerns, especially when researchers conduct linkage of information across various socio-economic, medical, and genetic domains.

With the introduction of the EU General Data Protection Regulation (GDPR), the ethical debate has been focused on privacy issues. Nonetheless, the use of biomedical big data can also have unintended consequences in terms of justice and fairness, and raise questions about the hierarchy of values like autonomy on the one hand and solidarity on the other. Informed consent is one type of measure frequently employed to give research participants control over the uses of their data – but consent alone is not sufficient, since it does not address group-level harms that may arise when even anonymized data is used irresponsibly.

Moreover, prospective informed consent is not possible in medical emergency settings, an area of research where the moral concerns around big data are intensified, particularly given the fact that patients become – at least temporarily – incapacitated or die. A recently established cross-European research consortium (ESCAPE-NET: the European Sudden Cardiac Arrest network towards Prevention, Education and New Effective Treatment) utilises big data approaches to study the prevention and treatment of sudden cardiac arrest. Currently, I am conducting an ethnographic case study among ESCAPE-NET research groups to

ascertain the ethico-legal challenges associated with collection, sharing and use of these sensitive research data obtained from vulnerable patients.

The proposed poster will highlight some of the issues faced as well as possible solutions. It will include short sections on identified themes such as deferred consent, the use of post-mortem health data without consent, the position of relatives, public engagement and the idea of “data democracy”, data sharing and the bounds of Nissenbaum’s “contextual integrity”, as well as group dynamics and the role of trust in international collaborations. This project builds on a previous literature review about the ethics of big data in emergency medicine research, of which the results will be highlighted in a separate section of the poster.

When Success Becomes the New Normal: The Competitive Society and its Symptoms

Yvette Drissen (Tilburg University)

Introduction: There has been an alarming increase in the number of people with stress related mental disorders over the last decades. One hypothetical explanation is that people in general, and workers more specifically, internalize societal ideas in which competition and success are focused on and narrowed to rising on the hierarchical and salary ladder.

Part 1: How and to what extent can mental health problems such as burnout be understood as symptoms of underlying social discourses?

Part 2: How can workers’ value perspective be broadened, in order to prevent mental health problems and enhancing the quality of life?

From deproductive democracy to reproductive democratization

Anna Hollendung (Kiel University & Bremen University)

Do democracies take a convincing account of the demands of reproductive work and the needs of future generations? How can existing systems, for which I propose the neologism of the deproductive democracies, be supplemented in such a way that they correspond to a positive draft in which the democratic potential would be unfolded in favour of future generations? I describe forms of political action in which sustainability and democracy are simultaneously realized (not in the form of a "expertocracy of sustainability" or a contemporary-oriented democracy), as elements of reproductive democratization.

My paper extrapolates the theoretical framework that will guide a planned research project. It addresses a specific blindness of current democratic theories and proposes a conception of democracy that depends on the needs of the future generations and the tasks of reproductive work. It speaks to the discrepancy between the democratic promise of equality and freedom on the one hand, and on the other the restriction of the spaces of action that are imposed on future generations through political decisions of the past.

There is a bundle of questions that has to be taken into consideration: Do democracies take a convincing account of the demands of reproductive work and the needs of future generations? For existing systems I propose the neologism of “deproductive democracies”, because of their deficits in terms of sustainability. With the term “deproductive democracy” I assemble my criticism on current democracies as well as democratic theories for its under-valorization of sustainability and the reproductive sphere. Arguably, a whole philosophic tradition fully externalized those topics from their concepts of democracy as it based its understanding of politics (or the political) on a strict public/private-dichotomy and corrected it only incompletely after being questioned through social movements and modern change. How can the actual concepts of democracy be further supplemented in such a way that they correspond to a positive draft in which the democratic potential would be unfolded in favor of future generations? What resources exist from which such a reconceptualization of democracy can be drawn? In my presentation, I’ll first work out temporal horizons and intersubjective dependency relationships. Three groups of subjects are of particular relevance in the context of a systematic problematization of “deproductive” democratic constellations: children, care workers (especially parents, educators) and members of future generations. The consideration of their subject positions allows an in-depth discussion of a counter-intuitive short-term orientation of politics. A lack of involvement in the temporal dimension of political decision-taking can have anti-democratic effects in the long term. Thus, depproductive democracy is undemocratic, because it undermines the future basis for democracy.

This problem diagnosis, secondly, needs to be complemented by a constructive perspective drawn from existing proposals. I concentrate on the feminist innovation of first wave feminism as one (though not the only) tradition with resources to fill that gap. With reference to Fraser I argue that an abyss arose between the demand for an emancipation of welfare activities and the struggle for individual development opportunities in the course of the development and liberal appropriation of feminism. Therefore, I reconstruct the incomplete displacement of the former in the democratic-theoretical reception and sketch an idea of democratic politics that takes into account the tasks of reproductive work.

After that reconstruction of the problem of “deproductive democracies” and the

investigation into relevant resources that may help to tackle it productively, a third step is to be done. I am asking for opportunities to implement instruments systematically that respond to the diagnosed misconduct and malpositions, which thus allow more sustainable democratization. The consistent focus on democratic legitimacy is intended to ensure that there is no confrontation between democracy and sustainability, but that the implementation of sustainability institutions takes the form of strengthening “the democratic”. Institutionally, this requires instances that systematically support aspects of sustainability and long-term viability on all levels of decision-making while reflecting on counteractivities through social movements and political activists in particular.

Soft Edges of Hardcore Technology

Petko Karadechev, Anja Jacobsen, Anna Hansen, Corinna Voll, and Monika Banyt□
(Aalborg University at Aarhus)

This project will create a techno-anthropological research approach meant to be applied to concepts within bioethics and investigate their interactions with emergent values addressing “The Good Life” in a present hyper-technologized world setting. An approach that would open up to questions like: How does a socio-technical focus on risk and prevention affect the perception of “The Good Life”? What is an (un)healthy body? What role does the right not to know play in today’s society? The approach will deal with keywords such as risk, autonomy, values and standardization of health. A collection of cases will illustrate the approach.

Defining 'Social Evidence'

Dimitri van Capelleveen (Vrij University Amsterdam)

An important concept for certain subfields of social epistemology, is that of 'social evidence.' Although several rough descriptions of this concept have been given, and there are some standard examples of pieces of social evidence (e.g. testimony), no formal definition has been offered. My aim, then, is to propose such a definition, one that, I claim, captures all, and only, pieces of evidence which can be considered ‘social.’ The hope is that this can be a useful contribution to the growing and important field of social epistemology.

The concept of ‘social evidence’ is important in one branch of social epistemology,

as characterized by Alvin Goldman (2010a), which focuses on individual agents and their “doxastic choices made in the light of *social evidence*” (p.14; emphasis added). But, what is this ‘social evidence’? Various descriptions of this concept have been offered. For example, social evidence supposedly “concerns acts of communication by others, ...traces of such acts [or] other people’s doxastic states” (Goldman 2010b: p. 14,15). According to Goldman (2010b: p. 6) “the evidential beliefs [that constitute social evidence] may have contents concerning what other people have said or written [or] contents concerning other people’s opinions (or other psychological states), which the doxastic agent may acquire at second or third hand.” According to a third description, it is “evidence concerning the utterances, messages, deeds, or thoughts of other people” Goldman & Blanchard (2018). Furthermore, there are various examples of pieces of social evidence, such as testimony and peer disagreement. However, no definition has been offered, which is the gap I’m looking to fill. I propose to define ‘social evidence’ as follows: a piece of evidence E is social evidence for a proposition P for a subject S iff E is an object such that it is evidence for P because of at least one other subject. An explanation might be useful. First, social evidence is always evidence *for a proposition*. Although there is such a thing as non-propositional knowledge (‘know how’), all evidence for this is seeming non-social. Furthermore, a piece of social evidence is always evidence *for a subject*. This is seemingly true of all evidence, but is especially important to mention in the case of social evidence, since at least one subject other than the subject for whom the piece of evidence is evidence, is necessary to make the evidence *social*. Next, that the piece of evidence is *an object*, should be understood broadly. It could be a newspaper containing an article that says “the president of the U.S. did something he wasn’t supposed to,” a state of affairs (e.g. testimony, or an act of communication), or some doxastic state of another subject. That an object is social evidence *because of at least one other subject*, should be taken broadly as well. For example, it could be social evidence because of a subject other than S having a doxastic state, or because a group of other subjects performed some act of communication. So, the resulting definition is broad, and arguably, trivial. However, the concept, as described by Goldman (2010a, 2010b) and Goldman & Blanchard (2018), is broad as well. It seems that any specification of any part of the definition, such as making precise the reason why a subject other than S makes a piece of evidence social, results in the exclusion of pieces of evidence which actually are social. Furthermore, the broadness of the definition has the benefit of ensuring that it seemingly includes all examples of pieces of social evidence. This, admittedly, results in a somewhat trivial definition. However, although this makes the definition not too exciting, it doesn’t decrease its potential usefulness to the growing and important field of social epistemology.

References:

Goldman, Alvin. 2010a. A Guide to Social Epistemology. In Goldman, Alvin &

Whitcomb, Dennis (eds.) *Social Epistemology: Essential Readings*. Oxford University Press.

Goldman, Alvin. 2010b. Why Social Epistemology is Real Epistemology. In Adrian Haddock, Alan Millar & Duncan Pritchard (eds.) *Social Epistemology*. Oxford University Press.

Goldman, Alvin and Blanchard, Thomas, "Social Epistemology", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2018/entries/epistemology-social/>>.

Representing future generations in international climate law-making: fairness, legitimacy and effectiveness

Nicky van Dijk (Utrecht University)

Current international treaties concerning climate change often implicitly or explicitly include the believed needs of future generations, but the fairness and effectiveness of the current take is disputed by many. As a way to address this deficit, several proposals for the representation of future generations by proxy have been made, but these proposals face immense democratic costs. This research focuses on whether, to what extent and how the needs of future generations should be fairly considered in international climate law-making through proxy-style representative institutions, aiming at fairness, legitimacy and effectiveness.

Current international treaties often implicitly or explicitly include the believed needs of future generations, but the fairness and effectiveness of the current take is disputed by many. It is argued that the current global climate regime (e.g. the Paris climate agreement) is unlikely to be strong enough to mitigate sufficiently [e.g. Byrnes and Lawrence 2015]. The current national contributions to non-binding treaties such as the Paris climate agreement are too low [Lawrence and Wong 2017]. On top of this, the current global regime unfairly benefits the current generation at the cost of future generations and the global poor [e.g. Coady and Corry 2013; Thompson 2010].

As a way to address this deficit, several proposals for the representation of future generations by proxy have been made, including (a) reserving several seats in legislative assemblies for members of the environmental sustainability lobby to represent future generations [e.g. Dobson 1996; Ekeki 2005; Kavka and Warren 1983; Thompson 2005]; (b) granting certain procedural rights to protect future generations [Ekeki 2005; 2009; 2016]; (c) commencing a UN High Commissioner of Future Generations [Shoham and Lamav 2006] or ombudsperson [Agius and

Busuttil 1998] or some other agent to veto laws and policies contrary to the interest of future generations [Mank 1996; Stein 1998]; or (d) giving more weight to the vote of the young or letting parents vote for their children.

However, these and similar proposals face immense democratic costs [Beckman 2015; Jensen 2015; Kates 2015]. An alternative would be to give more consideration to the interests of future generations within current deliberative democratic processes, without formally representing them. Here, it is argued, impartial deliberation can take the interest of all affected parties---including future generations---sufficiently into account, in which case there is no need to enforce means of representation whose legitimacy is disputable [e.g. Jensen 2015; Thompson 2010].

But unfortunately, on a national level climate action has consistently been given extremely low priority over the past decades [Dunlap and Scarce 1991; Leiserowitz 2005; Lorenzoni and Pidgeon 2006]. This is partly because the provided abstract and factual climate science hardly motivates towards climate action [Bazerman and Tenbrunsel 2011; Greene et al. 2001; Haidt 2001; Markowitz and Shariff 2012]. Also, unintentional wrongdoing such as flying to one's family for Christmas is judged far less harshly compared to intentional wrongdoing [Guglielmo et al. 2009; Markowitz and Shariff 2012], and the change of the current way of life that mitigation is asking for causes discomfort [Clayton et al. 2015; Gifford 2011]. On top of these reasons, individuals feel less moral concern when there is more psychological distance towards the victims. In the case of climate change where there is a need to take (distant) future generations into account, it is very disputable whether fair international law-making can be achieved through consideration of agents representing their national constituencies [Harris and Fiske 2006; Spence et al. 2011].

Therefore, this research focuses on whether, to what extent and how the needs of future generations should be fairly considered in international climate law-making through proxy-style representative institutions, aiming at (a) fairness, distributing the burdens of mitigation justly over current and future generations; (b) legitimacy, preventing unjustifiable costs for our deliberative democratic processes; and (c) effectiveness, taking clear steps towards sufficient mitigation.

References:

- Agius, Emmanuel and Salvino Busuttil. *Future generations and international law*. London: Earthscan, 1998.
- Bazerman, Max H. and Ann E. Tenbrunsel. *Blind spots: Why we fail to do what's right and what to do about it*. Princeton: Princeton University Press, 2011.

Beckman, Ludvig. "Political representation of future generations and collective responsibility." *Jurisprudence* 6:3 (2015): 516-34.

Byrnes, Rebecca and Peter Lawrence. "Can soft law solve hard problems? Justice, legal forms and the Durban-mandated climate negotiations." *University of Tasmania law review* 34 (2015): 34-67.

Coady, David and Richard Corry. *The climate change debate. An epistemic and ethical enquiry*. UK: Palgrave Macmillan, 2013.

Dobson, Andrew. "Representative democracy and the environment". In *Democracy and the environment*, ed. William M. Laferty and James Meadowcroft (Cheltenham: Edward Elgar, 1996): 124-39.

Ekeki, Kristian. "Electoral design, sub-majority rules, and representation for future generations." In *Institutions for future generations*, ed. Inigo Gonzalez Ricoy and Axel Gosseries (Oxford: Oxford University Press, 2016): 214-27.

---. "Constitutional experiments: Representing future generations through sub-majority rules." *Journal of political philosophy* 17:4 (2009): 440-61.

---. "Giving a voice to posterity - deliberative democracy and representation of future people." *Journal of agricultural and environmental ethics* 18:5 (2005): 429-50.

Guglielmo, Steve, Andrew E. Monroe and Bertram F. Malle. "At the heart of morality lies folk psychology." *Inquiry* 52 (2009): 449-66.

Haidt, Jonathan. "The emotional dog and its rational tail: A social intuitionist approach to moral judgement." *Psychological review* 108 (2001): 814-34.

Harris, Lasana T. and Susan T. Fiske. "Dehumanizing the lowest of the low: Neuroimaging responses to extreme outgroups." *Psychological science* 17 (2006): 847-53.

Jensen, Karsten Klint. "Future generations in democracy: Representation or consideration?" *Jurisprudence* 6:3 (2015): 535-48.

Kates, Michael. "Justice, democracy, and future generations." *Critical review of international social and political philosophy* 18:5 (2015): 508-28.

Kavka, Gregory S. and Virginia L. Warren. "Political representation for future generations." In *Environmental philosophy: A collection of readings*, ed. Robert Elliot and Arran Gare (Milton Keynes: Open university press, 1983): 21-40.

Lawrence, Peter and Daryl Wong. "Soft Law in the Paris climate agreement: strength of weakness?" *Review of European, comparative and international environmental law* 26:3 (2017): 276-86.

Leiserowitz, Anthony. "American risk perceptions: Is climate change dangerous?" *Risk analysis* 25 (2005): 1433-42.

Lorenzoni, Irene and Nick F. Pidgeon. "Public views on climate change: European and USA perspectives." *Climatic change* 77 (2006): 74-95.

Mank, Bradford. "Protecting the environment for future generations: a proposal for a republican superagency." *NYU environmental law journal* 5 (1996): 444-516.

Markowitz, Ezra M. and Azim F. Shariff. "Climate change and moral judgement." *Nature climate change* 2 (2012): 243-247.

Shoham, Shlomo and Nira Lamav. "Commission for future generations in the Knesset: lessons learnt." In *Handbook of intergenerational justice*, ed. Joerg C. Tremmel (Cheltenham: Elgar, 2006): 244-81.

Spence, Alexa et al. "Perceptions of climate change and willingness to save energy related to flood experience." *Nature climate change* 1 (2011): 46-9.

Stein, Tine. "Does the constitutional and democratic system work? The ecological crisis as a challenge to the political order of constitutional democracy." *Constellations* 4 (1998): 420-49.

Thompson, Dennis F. "Representing future generations: political presentism and democratic trusteeship." *Critical review of international social and political philosophy* 13:1 (2010): 17-37.

---. "Democracy in time: popular sovereignty and temporal representation." *Constellations* 12 (2005): 245-61.