Does Normative Behaviourism Offer an Alternative Methodology in Political Theory?

Eva Erman & Niklas Möller
Forthcoming in *Political Studies Review*

***

Many debates in political philosophy over the last decade have focused intensively on methodological issues, such as the debate on ideal theory vs. non-ideal theory, political moralism vs. political realism, and practice-independence vs. practice-dependence. Recently, Jonathan Floyd has brought up methodological aspects related to theories 'grounded in thoughts' vs. theories 'grounded in behaviour'. It is argued that so-called 'normative behaviourism' offers a better methodology than mainstream so-called 'mentalism'. In Floyd's view, normative behaviourism is a "new way of doing political philosophy" (Floyd 2017: 181).[1] Our claim in this paper is that normative behaviourism does not offer an alternative methodology in political theory in the sense envisioned by Floyd. First, we show that normative behaviourism is as dependent on 'normative thoughts' as mainstream political theory and is therefore also 'mentalist'. Second, we illustrate the structural similarities between normative behaviourism and mainstream political theory from a methodological standpoint by comparing the former to an influential normative theory, namely, utilitarianism.

I. Normative behaviourism as an alternative methodology

As mentioned above, the central distinction for Floyd is between the traditional method of pursuing political philosophy conducted by mainstream theorists, which he calls *mentalism*, and his suggested method, *normative behaviourism*. In mainstream political theory, theorists rely on *patterns of thought*, ideas in the form of intuitions, principles or considered judgments in order to justify their normative theories. "The key idea," Floyd suggests, "is that political philosophy's building blocks, its data, its working material – call it what you will – are thoughts, and in particular what I call normative thoughts. These thoughts are thoughts about what should and should not be the case in the world and about what individuals and groups ought and ought not to do" (100). His own account, on the other hand, instead relies on *patterns of behaviour*. Whereas mainstream political philosophy "is doomed to failure because different reasonable people disagree about the

---

[1] In the rest of the paper, page numbers within parenthesis refer to Floyd 2017.

fundamentals of political normativity" (121), Floyd argues that studying patterns of behaviour fares better:

> "If you study our behaviour, in response to variations in our political environment, then you learn that there are some systems we have little time for, some we have more time for, and perhaps one that suits us best of all. As a result, we learn that there is at least one way of justifying political principles – in this case the principles expressed by that last system – that has nothing to do with mentalism, because it has nothing to do with our normative thoughts." (167)

Floyd uses considerable space arguing that the method of mainstream political philosophy – 'mentalism' – is doomed to fail since it cannot deliver agreement among reasonable people. Patterns of reasoning simply do not converge. Floyd argues that his own method fares better since it focuses on patterns of behaviour where, despite verbal or mental disagreement, we can indeed discern a pattern pointing to a certain form of social-liberal democracy (168-169).

The fundamental methodological move that Floyd takes to distinguish his method is his reliance on *expressed political preferences*, that is "normative preferences that get revealed by certain patterns of behaviour" rather than the traditional method of relying on "normative preferences that get revealed through certain patterns of thought, such as those captured by impartial choice situations like the 'original position' or abstract dilemmas like the trolley problem" (168). Specifically, Floyd argues that there are two behavioural phenomena which differentiates the good political order from the bad: the level of insurrection and the level of crime in a society. To indulge in insurrection and crime, Floyd argues, comes at a high level of personal risk and people tend to "only to partake of them when they find their current lives, which lack goods x, y, and z, unbearable" (168-169). Consequently, Floyd argues, the political systems "that produce less of this behaviour are more justifiable than those that produce more, with the one that produces the least, clearly, more justified than any other" (169).

We remain sceptical about the role of agreement that Floyd assigns to mainstream political philosophy.[2] Neither are we convinced that, on the level of generality on which Floyd's social-liberal democracy is analyzed, there is a more reliable pattern in our behavioural response to variations in our political environment than there is to our 'mental response', i.e. our value commitments. However, in this paper, we will leave these concerns to the side and focus on the fundamental question of whether Floyd is right in his claim that normative behaviourism constitutes an alternative method from mainstream political theory that is grounded on behavioural facts rather than values. Our conclusion is that he is wrong.

---

[2] We do not share the reading that all, or even most, political theorists attribute agreement such a central role in their theories, as Floyd claims. Not even Rawls' concept of 'overlapping consensus' relies on actual consensus for the justification of his theory.

II. Grounding an account on facts instead of values

Floyd is fully aware that on his account, normative conclusions are drawn from factual premises: an empirical premise – an actual pattern of behaviour (minimal crime and insurrection) – justifies a normative conclusion – organising society by way of a certain set of principles (social-liberal democracy). While Floyd acknowledges that moving "from an 'is' of behaviour to an 'ought' of principles" is often considered a 'naturalistic fallacy', he claims that grounding an account on factual premises is actually non-problematic (170).

First, Floyd argues that his account and mainstream political theory are on pair. Mainstream political theory "assumes that we can perfectly easily move from *thoughts* to principles – whether those thoughts be impartial choices, judgements, or intuitions" (170) without there being any fallacy involved. "Why," he rhetorically retorts, "should this fallacy apply to actions if it does not apply to thoughts?" and concludes that "[i]f there is a fallacy here, then it is what I will call the 'mentalistic fallacy', which is that thoughts are somehow non-factual, non-empirical – almost non-existent" (170).

Second, and more centrally, he argues for the positive claim "that facts can ground political principles", regardless of whether those facts are mental or behavioural (171),[3] by objecting to Cohen's argument to the contrary. Cohen argues that *no fact of any kind* can ground a normative principle, unless there is a further normative principle which explains why this is so; in other words, unless there is a further grounding normative principle. At rock bottom, Cohen holds, there will always be a normative principle.

For Floyd, refuting Cohen is important since it is "the strongest case against" normative behaviourism (170-171, n. 3). He argues against Cohen by utilising the following alleged syllogism:[4]

> "If (1) political principles, for mentalism, are only justified for an individual in virtue of the thoughts in that individual's head, and (2) it is a matter of fact whether or not I hold a particular set of thoughts, then (3) it is also a matter of fact whether or not I am bound to such principles. As such, the principles we are bound to depend upon facts about us – which means in this case facts about the thoughts in our head." (171)

As Floyd illustrates: "if I intuitively think that I should rescue a group of twenty people rather than a group of ten, if I can rescue either but not both, then it is a fact that I have that intuition" (171).

The guiding premise in both of Floyd's points seems to be the basic idea that regardless of whether we utilise the 'mentalist' or the 'behaviourist' method, a person's justification status always depends on facts about that individual. This corresponds to the traditional analysis of justified

---

[3] He continues to claim "that behavioural facts do a better job of this than mental facts" (171). For our present purposes, it is the former part of the claim that interests us.

[4] 'Alleged' since syllogisms are logically valid schemas and, as will become clear in the main text, this one is not. In further reference to Floyd's schema, 'alleged' is however suppressed in the main text (for stylistic as well as etiquettical reasons).

belief: a claim X is justified for person P if and only if P has good reasons for believing X. Absent good reasons, P can of course still be totally convinced that X is true, but that conviction is not justified.

While this is correct as far as it goes, it completely misses its mark when it comes to Cohen's main insight. Cohen's argument (Cohen 2003, 2008) is about the necessary *explanatory relation* between different kinds of principles.[5] Cohen distinguishes between two kinds of principles: those that are dependent on facts and those that are fact-independent. On his account, "a principle can reflect or respond to a fact only because it is also a response to a principle that is not a response to a fact" (Cohen 2003: 214). The idea is that if we have a principle P – say, a principle of justice – that is sensitive to facts F, there is another "more ultimate principle that explains why F supports P" (Cohen 2003: 218).

Cohen exemplifies with the principle 'we should keep our promises' (A), which (we assume) depends on the fact that only then can people pursue their projects. Principle (A), Cohen argues, is true only because there is a further principle (B), 'we should help people pursue their projects' (Cohen 2003: 216–17). This further principle (we again assume) is true only because of a further fact, namely that people can achieve happiness only if they pursue their projects. But the latter fact is explanatory salient only if there is yet another principle (C) that states 'we should promote people's happiness'. For utilitarianists, finally, we have now reached a principle which is true regardless of any further facts; it is a fact-free normative principle. If we are not utilitarians, the regress continues at least one more step, since there must be some fact which explains in virtue of what we should promote people's happiness. But eventually, we reach bedrock: some normative principle which holds regardless of any further facts.

We can now see that Cohen's account is not about whether you can justifiably move "from an 'is' of behaviour to an 'ought' of principles", but rather a version of the general point that whenever you make a normative inference from a non-normative premise, its validity depends on the truth of an underlying normative premise. The point is thus not that you make a *mistake* – a fallacy in Floyd's terminology – if you make the inference from "act X helps P pursue her project" to "P should perform act X". The point is that this inference is only valid, if it is valid, in virtue of an underlying normative premise which – directly or 'down the line' – supports it. This underlying value premise explains why it may be valid to make a normative inference from a non-normative premise.[6]

Equipped with this insight, we may return to Floyd's syllogism and grasps its flaw. Floyd's syllogism moves from the premise (1) that a principle is justified for P by a mental object X (a thought) in P's head, and the further premise (2) that it is a fact whether X is instantiated in P, to the conclusion that (3) it is a matter of fact whether P is bound to the principle in question.

---

[5] For an extended analysis of this insight, see Erman and Möller 2016, 2017, 2019.

[6] Note that this insight about the logical structure of valid normative inference is independent on which epistemic theory of justification one endorses.

The problem with Floyd's syllogism is that it does not properly distinguish between *justification* and *boundedness*. Whether or not an agent is bound to a normative principle does not depend on whether she is justified to believe in it, in virtue of mental objects instantiated in her head or otherwise. She is bound to a principle if the principle is right, even if she is not herself aware of it being so.[7] Just like she is legally bound by a law, regardless of believing in or even being aware of it. Hence, even if it is conceivable that a person might be justified in thinking that babies love being tortured, together with more plausible principles about making people (of all sizes) happy, that does not mean that she is bound by the principle 'you should torture babies.' On the contrary, she is bound by the principle not to torture babies. Hence, the syllogism's conclusion (3) does not follow.

Now, Cohen's point expressed in terms of boundedness is that if you are bound by the principle not to torture babies, this is because of either of two alternatives: it depends on some fact, such as that torturing babies hurt them and some further premise that you ought not hurt babies; or it is itself a fact-free, basic principle.

Let us now apply this insight to normative behaviourism. Floyd's normative behaviourist thesis is that we are bound by the principles guiding social-liberal democracy. These principles are fact-dependent in Cohen's terminology, since they depend on a set of facts: that these principles are the ones guiding the kind of society with the least amount of dissatisfaction (measured in crime and insurrection). Now, either there is a further fact which explains why *the set of principles guiding the kind of society with the least amount of dissatisfaction is the right set of principles*, or this principle is a basic, fact-free principle. There are no other options. Hence, contrary to what Floyd seems to think, normative behaviourism, as a theory, is just as dependent on value commitments as any other 'mentalist' account. Just as the political principles of any 'mentalist' political theory depends on whatever fundamental values have been used in order justify them, Floyd's theory is dependent on the value commitment that dissatisfaction is the fundamental disvalue and the principle that the least amount of dissatisfaction (measured in level of crime and insurrection) determines the (politically) best society.

In other words, there is just no getting around the fact that Floyd's account, in order to be normative, must be grounded in a (fact-free) normative premise. Demonstrating the connection between a certain set of descriptive facts (low levels of crime and resurrection) and a certain system of rule only shows that this rule is justified with the help of a value premise, implicit or explicit. Reliance on such value premises is what Floyd takes to be a fundamental mark of the 'mentalist' method. And since Floyd's account is just as much grounded in value premises, it does not mark out a new method in this respect.

---

[7] We will use 'right' in the main text, but the reader may exchange the term for whichever favourite term she has ('true', 'correct', etc) that allows that a person may have a mistaken belief in a normative proposition.

III. Structural similarities with mainstream political theory

We have so far defended our claim that normative behaviourism does not offer an alternative methodology based on (behavioural) facts rather than values, by demonstrating that normative behaviourism is equally anchored to value premises as mainstream political theory and therefore not solely grounded in patterns of behaviour. In this section, we move over to the second step, demonstrating the structural similarities between normative behaviourism and mainstream political theory with regard to methodology. We do so by comparing normative behaviourism to an influential theory, namely, utilitarianism.

According to utilitarianism, a fundamental *value premise* is that well-being, for example, in terms of happiness, desire-fulfilment or preference-fulfilment is the final good. Similarly, according to normative behaviourism, a fundamental value premise is that dissatisfaction among citizens in society is bad. Connected to these value premises are some *fundamental principles*. For utilitarianism, roughly put, those societies are best whose principles generate as much well-being as possible. For normative behaviourism, those societies are best whose principles generate as low level of dissatisfaction among citizens as possible "with both the political system and the ways of life it facilitates" (189). With regard to *measurements* (and thus empirical implications), utilitarianism has been developed in a variety of ways, often measuring the maximation of well-being overall. According to one dominant view, what matters for well-being is the overall level of preference-satisfaction in a person's life as a whole (Crisp 2021). According to normative behaviourism, as we have seen, the amount of dissatisfaction among citizens is measured in terms of the level of crime and the level of insurrection in society. In terms of output, lastly, in both cases it becomes mainly an empirical question which type of society the theories suggest; in fact, it is reasonable to argue that both theories will in practice favour a regime type grounded in egalitarian principles, principles of democratic legitimacy, principles of fairness and justice, and so on.

The structural similarities between the two views are thus quite striking. Both are committed to value premises and fundamental principles that are 'mentalist' in Floyd's terminology. Furthermore, both utilize measurements that are 'non-mentalist', since neither measurement is tied to any ideas of what people have reason to believe or what they are convinced by themselves. Instead, focus is directed at actions, behaviour and well-being. For utilitarianism, in broad strokes, the justification of principles equals the sum of well-being over all individuals. For normative behaviourism, in broad stokes, the justification of principles equals "political commitment X numbers, with numbers meaning the percentage of people who commit a certain type of action under a certain type of political system" (169).[8]

---

[8] While it would take us away from our current methodological trajectory to discuss the point in any detail, it is worth mentioning that this structural similarity with utilitarianism points to a problem that Floyd's account eventually needs to treat: the problem with the *distribution* of dissatisfaction. Just as utilitarianism is criticized for neglecting the distribution of happiness (or preference satisfaction), only counting it total sum, it is clear from the quote in the main text that Floyd's account can be accused of only counting the sum of dissatisfaction, not its distribution. We thank an anonymous referee for bringing up this point.

According to Floyd, a convincing answer to the 'organizing question' (i.e., how should we live?) must not only be politically determinate "in terms of its principles and institutions" (229), but also rationally compelling. Also with regard to the fulfilment of the latter criterion, the two views are structurally similar. Utilitarians typically argue that a welfarist society of the kind that Floyd defends offers a convincing answer to the organizing question because it distributes relatively evenly both fundamental goods and rights, which is empirically proved to generate more happiness, preference-satisfaction, and the like. In a similar manner, Floyd argues that there are four reasons to be convinced by social-liberal democracy as the best answer to the organizing question. In his view, we should be attracted to social-liberal democracy because it is the political system which provides most security, produces least cruelty, generates most human flourishing, and the lowest number of people who find life unbearable (229-231).

In other words, in trying to find support for the preferred society among potential sceptics, Floyd is equally committed to values as mainstream ('mentalist') political theory. As we have seen in the history of political philosophy, what counts as security, cruelty, human flourishing and unbearable are deeply contested and, above all, dependent on what normative framework we apply (conservatism, neo-liberalism, Marxism, Kantianism, virtue ethics, and so on).

IV. Winding Up

In this paper, we have tried to show that Floyd's account does not offer an alternative methodology in political theory that 'by-passes' values and instead is grounded on a special kind of facts, i.e. behavioural facts. Rather, it needs all the same building-blocks as the mainstream 'mentalist' account he criticises. What Floyd's account positively does, however, by focusing on the role of a certain set of behavioural facts, is to bring up new and interesting ways in which empirical data may feed into political theory, such as the importance of taking into consideration the correlation between the dissatisfaction in society and high levels of crime and insurrection. In this respect, Floyd's account may fare better than utilitarianism in taking empirical data seriously.[9] As demonstrated in the previous section, both accounts are in a structurally similar way reliant on empirical data for making claims about which kind of society is preferable. Yet, contemporary utilitarianists have often been accused of failing to give sufficiently detailed empirical evidence to justify their conclusions. Floyd's account may owe us a more detailed account on why levels of crime and insurrection alone should play the fundamental role it does in his theory (i.e. a more solid value argumentation), but he is very explicit in terms of which data he takes to support his conclusions.

References

Cohen, G. A. 2003. "Facts and principles." *Philosophy & Public Affairs* 31(3), 211–245.

---

[9] We thank an anonymous referee for suggesting this point.

Cohen, G. A. 2008. *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press.
Dworkin, Ronald (1986) Law's Empire. Cambridge, MA: Harvard University Press.

Crisp, Roger. 2021. "Well-being". In E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), URL = https://plato.stanford.edu/archives/win2021/entries/well-being/

Erman, Eva and Niklas Möller. 2016. "What Distinguishes the Practice-Dependent Approach to Justice?" *Philosophy & Social Criticism* 42(1): 3-23.

Erman, Eva and Niklas Möller. 2017. "Practice-Dependence and Epistemic Uncertainty.*" Journal of Global Ethics* 13(2): 187-205.

Erman, Eva and Niklas Möller. 2019. "How Practices Do Not Matter." *Critical Review of International Social and Political Philosophy* 22(1): 103-118.

Floyd, Jonathan. 2017. I*s Political Philosophy Impossible? Thoughts and Behaviour in Normative Political Theory*. Cambridge: Cambridge University Press.