ESJP
#22 | 2022

On the Impossibility of Moral Development in John Rawls's *Theory of Justice*

Jeroen de Vries

Consider the following scenario. You are again a 10-year-old child, and there is a family dinner party. All other guests at the party are adults, and have been for quite some time. You are having a blast, engaging in clever conversation, and even learning some new dirty jokes. But then, as the evening draws further, the unimaginable happens: you are told it is time for bed. However, this flies right in the face of your subjective wellbeing: you are having a great time, and you do not feel tired yet, anyway.

So, you do as any child in this situation would: you try to get your way. Being a smart child, and having learned from experience, you opt for rational discourse, rather than tumultuous complaint. It quickly becomes manifest, however, that the adults deem you 'too young' to determine your own bedtime. They are 'older,' have been in your position before, and therefore know what is best for you. Or so they argue.
Besides the initial surge of frustration, you notice something different. The only way to convince the adults that you are, in fact, wise enough to determine, or at least influence, your own bedtime, is through showing them that you are 'mature' enough. In other words: you need to become like them – adopt their *framework*, if you will – before you can raise your voice about such matters. Matters, to be clear, that significantly impact your wellbeing!

Abstractly, this essay considers whether the situation described above constitutes an injustice. It does so by investigating the moral implications of positing normative criteria for rationality. The essay argues that John Rawls's conceptualization of moral persons in his A *Theory of Justice* (Rawls 1999) (*TJ*) is circular, which makes the theory unable to accommodate change[1].

The essay is structured as follows. Section 1 sketches the context in which *TJ* was written. Section 2 investigates Rawls's conceptualization of *moral persons*, and argues that it is circular. Section 3 briefly introduces the notion of *Reflective Equilibria*, and illustrates how the circular conceptualization of moral persons makes *TJ* conducive to the naturalistic fallacy. Section 4 discusses how this can lead to a form of majority rule, and considers two possible consequences. Section 4.1 considers the notion of moral overconfidence, and Section 4.2 relates this to the notion of a moral stalemate. Section 5 offers some nuance, and comments on the limitations and practical relevance of the arguments made.

1. From Natural Law to the Original Position: An Outline.

Historically, the most prominent class of theories of justice, known as natural law theories, have justified inequalities in terms of necessity, based on historical consistency, or from the tacit assumption that the wellbeing of some outweighs that of others (Finnis 2020). *Utilitarianism,* however, which emerged in the 18th century, departed from this view. It perceives all people as identical in their moral worth, as well as (or because of) their identical capacity for experiencing pain and pleasure (E.g., Mill 1998, chap. 4). This was progressive, perhaps even radical, in its time: it was unprecedented for a theory of justice to explicitly place identical intrinsic value on every individual (Driver 2014). Progressive as it was, utilitarianism, as critics were quick to point out, is not without its flaws. One of its largest main flaws is its inability to respect the

[1] N.B., this paper restricts itself to Rawls's work as presented in TJ; it does not consider the possible responses put forward by Rawls in later works, e.g., in (Rawls 1993). Moreover, change (in moral intuitions) will be used interchangeably with moral development; it does not strictly imply improvement.

separateness or inviolability of people (Schneewind 1990, 44–46). No constraints exist regarding who experiences pleasure or pain, and from what source: as long as the outcome in terms of pleasure or pain remains the same, there is no discernable difference between a sadist enjoying the abuse he inflicts upon a victim, and the wholesome joy a good Samaritan experiences after helping the elderly cross the street.

In light of this, Rawls attempted to find the principles of ideal social justice (Rawls 1999, 4–5). In *TJ*, he sought to establish a theory that would respect the inviolability and separateness of persons, placing limits on the sacrifices that could be asked of individuals, whilst maintaining the incentive of increasing one's possessions through productive labor (Wenar 2021). He did so by positing his *Original Position* in which people are to choose the principles of justice for the society in which they will live (Rawls 1999, 15–17). To prevent people from seeking personal advantage over distributive justice, they make this decision from behind a hypothetical *Veil of Ignorance*. This veil makes them forget all that is particular about themselves as individuals, and even their conception of the good (Rawls 1999, 11). They strictly remember the "general facts about human society," namely the "basis of social organization and the laws of human psychology," (Rawls 1999, 119). This should ensure that the agreed-upon principles are acceptable regardless of the position from which they are perceived. Rawls believed the inhabitants of the *Original Position* would agree upon the following principles[2]: those of equal basic liberties, equal opportunities, and the difference principle (Rawls 1999, 65).

Crucially, Rawls sought to prevent majority rule (Rawls 1999, 200–203), where the preferences of a majority justify the violation of a minority's rights. Allowing for this would effectively reduce Rawls's theory to utilitarianism. Therefore, the principles of *equal basic liberty* and *equal opportunities* ensure that individuals' basic rights and liberties are not violated. The *difference principle* guards against the harmful exacerbation of inequalities (Rawls 1999, 53).

Through rational deliberation from shared assumptions, *moral persons* can arrive at *Reflective Equilibria*. This is a state of harmony between considered convictions, moral intuitions, and moral judgments (Rawls 1999, 18). To reach this harmony, convictions are revised to improve coherence between them. It is important to note that no conviction is *a priori* exempt from being revised: in principle, any conviction, at every level of abstraction, can be revised, if it serves to improve the coherence between the convictions, and thus further approximates a *Reflective Equilibrium* (Daniels 2020). The next section illustrates how Rawls's conceptualization of *moral persons* is circular.

2. Moral Persons and Where to Find Them

For Rawls, morality requires *reciprocity*(Rawls 1999, xv, 13; cf. 88). Correspondingly, children are taught morality through reciprocating loving and caring acts they receive from moral adults (Rawls 1999, 433, 436). Furthermore, reciprocity requires *equality* or *equal respect*: there needs to be some recognition of similarity between agents to allow for reciprocity (See Rawls 1999, 513; cf. Pritchard 1977, 60–61).

The importance of equality and reciprocity, then, necessitates some criterion for evaluating equality. Rawls offers two such criteria from his conceptualization of the inhabitants of the *Original Position*. He assumes them to be "*moral persons*, […] creatures having *a conception of their good* and *capable of a sense of justice*" (emphasis added) (Rawls 1999, 17).

This conceptualization is circular. There is no mention of where the first moral persons came from; Rawls postulates the first moral person and writes only that children become moral through reciprocating moral acts (E.g., Pritchard 1977, 61–63; cf. Brennan and Noggle 1998, 212–13). Moral persons are said to

[2] N.B. this is a brief and perhaps superficial representation of the principles; they are more sophisticated, but the representation above suffices for the purposes of this paper.

be capable of a sense of justice, and justice is that which moral persons have a sense of. As such, they constitute each other. When a dominant majority agrees upon a necessary condition for a moral person, or upon what justice entails, this can cause this majority to disregard a minority if this minority does not share their conviction. Since moral persons have a sense of justice, and the minority does not agree with the majority, either can call the other unreasonable. When, for example, two subgroups in a society, agree, among themselves, that their fundamental criteria for morality and rationality are crucial for intelligible communication, this can support the inhabitants of both subgroups in deeming the other inarticulate, unreasonable, and not to be taken into consideration for further discussion.

Alternatively, if the child at the dinner party has some peers sharing their opinion on the necessity of delaying their bedtime, yet the parents nevertheless insist on sending them to bed at this instant, the children have no possibility of demonstrating that they are sufficiently rational and moral, except for conceding to their parents' opinion. Moral persons agree on the shared intuitions that make up moral persons, and reasoning that does not start from those convictions, ought not to be included in deliberations of bedtimes – or any moral issue, for that matter.

Admittedly, there are legitimate reasons for parents to be paternalistic towards their children – they are, after all, their parents. Furthermore, adults and non-adults might still agree on the preferred consequences of a decision – feeling well-rested tomorrow morning versus feeling very tired, or growing strong and healthy versus suffering from over-exhaustion, for example. The point I seek to make is that the insistence on a shared moral conviction as a necessary condition for treating others as subjects of justice can have harmful consequences, and aid polarization rather than pluralism.

I contend that the circular conceptualization of Rawls's *moral persons* makes *TJ* unable to accommodate change in the subjects of justice and their moral intuitions. The next section illustrates how this can be problematic, as it invites both a *naturalistic fallacy* and a form of *majority rule*.

3. Hume's Guillotine and Inhumane Democracy

Reflective Equilibria serve to give insight into justice through the explication and subsequent analyzation of shared moral intuitions (Rawls 1999, 16, 18). Starting from considered judgments – i.e., judgments made under conditions favorable for clear and rational thinking (Rawls 1999, 40–42)[3] - intuitions are scrutinized to improve their consistency with convictions at different levels (Daniels 2020). Through repeated revision, the considered convictions[4] of the members of a society become increasingly more coherent with each other. When this happens, a *Reflective Equilibrium* is attained[5]. Subsequently, judgments can be made in terms of consistency with this equilibrium; if something is inconsistent with it, this is deemed irrational.

Importantly, Rawls points out that moral sentiments stem from natural attitudes (Rawls 1999, 428). The considered convictions are therefore derived from the moral intuitions of moral persons. As such, a theory of justice must correspond with (shared) moral intuitions (Baldwin 2006, 249). I agree with this: God really is dead, and moral transcendentality has become a distant memory. However, I believe Rawls overvalues human intuitions and overlooks that human intuitions can be idiosyncratic, mistaken, or otherwise inadequate. Rawls seems to acknowledge this when he writes how children develop morality through exposure to, and subsequent reciprocity of, moral acts (Pritchard 1977, 68–69).

[3] N.B. considered judgments are not assumed to be infallible, but they are more reliable than those made under e.g., stressful conditions.

[4] N.B. "considered convictions" and "considered judgments" are used interchangeably.

[5] sN.B. A full Reflective Equilibrium is not assumed to be reached. Instead, it is approximated, with improved coherence leading to a "wider" Reflective Equilibrium – See (Daniels 2020).

The following section argues that this overvaluation of human intuitions makes *TJ* conducive to committing the naturalistic fallacy and a broad form of majority rule. Consequently, this makes it conducive to an attitude of moral overconfidence; furthermore, it can lead to moral stalemates when shared intuitions are absent.

3.1 The Naturalistic Fallacy: Turning an "Is" Into an "Ought"

In *TJ*, the attainment of a *Reflective Equilibria* is synonymous with attaining justice[6]. This implies that all that is relevant to justice has been addressed from within this position. Furthermore, it implies that whatever is agreed upon in the *Original Position*, is just.

Throughout history (and geography), however, different moral consensus and norms have existed; entire peoples have condoned and agreed-upon practices that are unambiguously condemned nowadays. There have been societies where slavery was condoned and never seriously questioned, whereas contemporary Western societies wholeheartedly condemn this practice. Similarly, the notion of animal rights is much more prominent nowadays than a mere fifty years ago. There might be no consensus on these matters, but it is demonstrably clear that moral intuitions change over time; consequently, that agreement on intuitions does not equate to perfect and permanent justice.

Hume famously made the *is-ought* distinction (Hume 1985, book 3, Section 1, Part 1), arguing that mere observation cannot deliver a normative judgment[7]. However, using shared intuitions as the sole input for moral deliberation and evaluation risks doing precisely that: perpetuating practices by virtue of their historical existence. It is impossible, then, to move beyond our current intuitions in *Reflective Equilibria*; we can only improve their consistency.

This leads to theoretical lacunae when it comes to explaining development in intuitions over time. If children acquire their moral intuitions through the reciprocating of moral acts, how can it be that contemporary Western societies generally condemn slavery, whereas the Greeks and Romans did not?
Thus, the naturalistic fallacy in Rawls's *TJ* allows for the justification of practices through their historical occurrence, as well as by their being intuitive to *moral persons*.

3.2 Shared Intuitions and their Discontents

Regarding his *Original Position*, Rawls writes that "no complaints will be heard at all until everyone is roughly of one mind as to how complaints are to be judged," (Rawls 1958, 171)[28]. His *moral persons* share a framework of intuitions, according to which they can rationally and transparently engage in moral deliberations. Rawls does not, however, spell out what these basic intuitions are. Furthermore, his theory does not comment on situations where the intuitions are *not* shared, which brings us to the topic of majority rule.

4. Majority Rule

Rawls's *TJ* has not escaped the threat of *majority rule*, because its conceptualization of moral persons excludes whoever does not adhere to the dominant set of intuitions. In a society, there will indubitably be degrees of disagreement about moral matters. Therefore, the intuitions that are to be explicated will not have complete overlap. Because justice is determined by the reasoning of moral persons, who share a set of intuitions, the result is that there is no way for a minority to disagree with the dominant set of intuitions whilst still being regarded as a subject of justice. Recall that moral persons are postulated to have a "*conception of their good*" and

6 N.B. Perfect Reflective Equilibria are unattainable, but the procedure of arriving at them aids deliberations on justice. Cf. (Rawls 1999, 44).

7 Cf. (Hunter, 1962, 149–50).

8 N.B., this citation is from a different work of Rawls, but illustrates my point regarding 'majority rule' clearly.

to be "*capable of a sense of justice*" (Rawls 1999, 15, emphasis added). If the overwhelming majority shares certain intuitions, and the intuitions of a small minority do not match with theirs, this can mean one of two things. Either they are not moral persons – for their intuitions diverge from those deemed constitutive of moral persons – or their intuitions are ill-considered. Essentially, both mean the same: they must "adapt" their intuitions (i.e., conform to the dominant set of intuitions), or they lose their say in matters of morality. Effectively, this means that Rawls's *Reflective Equilibria* are constitutively deductive, and can, in tautological fashion, contribute nothing to the dialogue that was not already implicit in the intuitions of those members of a society that are relevant to justice. Therefore, *Reflective Equilibria* aid in systematizing the status quo, and exclude the possibility of change in moral intuitions, and thus of moral development. Below, I argue that this can lead to moral overconfidence when shared convictions are present, and to moral stalemates when they are lacking.

4.1 Moral Overconfidence

Crucial to *Reflective Equilibria* is the tacit assumption that moral persons possess sufficient information relevant to the topic at hand. Perhaps the most striking illustration of the falsity of this assumption comes from climate science. Virtually all inhabitants of the 21st century have heard mention of climate change. It is striking to consider, then, that the (modern) environmental movement began less than 100 years ago; before that, the general perception was that mankind could not significantly impact the oceans, as it was deemed too vast to be threatened by mere human beings.(E.g., Hays 1981, 219)[39]. This illustrates how an exclusive focus on consensus between *moral persons* can lead to moral overconfidence, where a consensus within a group leads its members to consider themselves morally sufficiently informed, making them less inclined to consider unconsidered perspectives. A society-wide consensus still lacks normative force when it comes to the physical part of reality, and it is important to be attentive to what might be overlooked, both in the moral realm and in the physical realm. Although *TJ* does allow for the incorporation of new variables and considerations in the light of new information, the emphasis on establishing consensus can cause unconsidered issues to remain overlooked.

4.2 Moral Stalemates

In *TJ*, justice has been attained in *Reflective Equilibria*, where the shared intuitions of moral persons are coherent with each other at all levels. However, this implies that there is no possibility of attaining justice when these shared intuitions are lacking. When two parties with different shared intuitions, both deeming themselves moral, each attempt to attain a *Reflective Equilibrium*, this amounts to a moral stalemate, where both parties hold on to their intuitions, and no consensus can be found. Because the subjects of justice base their deliberations on rationality and morality, it implies that any divergent position must lack either, or both, of these qualities.

The assumption of shared intuitions can be conducive to an unwillingness to engage in discussions with those who appear *too* different from us in terms of moral intuitions. Deeming those who diverge from the shared assumptions irrational and immoral for having different opinions places the contemporary intuitions of a given group on a pedestal and precludes the potential insights from vastly different worldviews. Given the turbulence and diversity of the world, it is unrealistic to expect that an overlapping consensus in moral intuitions between different people can always be arrived at. The modern world is markedly plural in the worldviews it offers; there is increased communication, faster travel, and more international trade than there has been in the past. Encountering people with convictions different from one's own is increasingly

9 Cf. Sloterdijk's notion of 'backdrop ontology': (Sloterdijk 2015, 334–35)

common; therefore, the requisite shared intuitions might realistically be unattainable. Consequently, an insistence on shared intuitions can endanger meaningful plurality instead of being conducive to it.

Consider, for example, that a subgroup of children, having experienced a childhood vastly different from that of their parents, come to agree among themselves on different moral convictions than those deemed constitutive of moral persons in their society. If the 'traditional' shared convictions become normative, and thus cannot be questioned by those not subscribing to them – remember that "no complaints will be heard at all until everyone is roughly of one mind as to how complaints are to be judged," – this can cause both subgroups to view the other as barbaric, unsophisticated, and not to be considered in deliberations on matters of morality. Consequently, new and unexpected issues, values, and perspectives might not be addressed, or not be addressed appropriately[410].

5. Pragmatism and Being Reasonable

The points made above reason within the confines of *TJ*; therefore, they are primarily theoretical. The pragmatic consequences, however, might differ. Admittedly, *TJ* is not applied dogmatically; its influence in political philosophy is undeniable, but is not followed to the letter. Therefore, it is unlikely that such strict demarcations regarding *moral persons* will arise in practice. Moreover, *TJ* is a *contract theory*; presumably, the emphasis on public reason[511] invites an attitude of open-mindedness and willingness to listen to what might be neglected. Rawls describes "considered convictions" as well-contemplated and generally undisputable moral positions, such as that discrimination is wrong (Rawls 1999, 17–19). He is silent on where they originate. However, it seems that these *considered convictions* imply the possibility of change in moral intuitions – Rawls endorses the disapproval of slavery, for example – although it is not explicated. Rawls presumably did not deem moral development impossible or undesirable; however, it is not incorporated into *TJ*.

Realistically, therefore, changes in moral intuitions will occur and be incorporated into moral reasoning, and the contractarian spirit stimulates open-mindedness in moral deliberations. Nevertheless, it is important to be aware of transience and the possibility of overlooking matters of importance, in both the subjects of justice and the information considered relevant. Similarly, it is important to consider that wholly different convictions might be rational for reasons currently unconsidered.

Conclusion

I have argued that Rawls's *Theory of Justice* employs a circular conception of *moral persons*, which makes the theory unable to accommodate change or development in the subjects of justice and their moral intuitions. It risks committing the *naturalistic fallacy* and is vulnerable to a form of *majority rule*. Consequently, it can be conducive to an attitude of *moral overconfidence* within groups, and it can lead to *moral stalemates* between groups when shared intuitions are lacking. However, the theory is not applied literally, and the *contractarian* emphasis stimulates an attitude of open-mindedness when engaging in moral deliberations. Therefore, the risk of becoming dogmatic and completely removed from the corporeal realm is small. Nevertheless, the approach might be conducive to overlooking what might be unexpected, as well as disregarding what is uncomplying. This can be particularly problematic considering the increasing interwovenness of people with different convictions, creeds, and cultures. When it comes to deliberations on justice in a diverse and transient world,

10 N.B. although Rawls does discuss intergenerational justice in *TJ* – e.g., Rawls, 251-258 – his approach still seeks a form of consensus, e.g., a savings rate that counts for all generations. The example above emphasizes that such axiomatic forms of consensus, with their consequent normative accounts of rational behaviour, can be harmful to meaningful pluralism of world views within societies.

11 TJ, in contractarian fashion, emphasises the importance of transparent and rational reasoning. For further reference, cf. Horton, 'Rawls, Public Reason and the Limits of Liberal Justification', 9–13.

the inevitable necessity of adapting to change should be embraced, and attentiveness to what falls beyond the current conceptualization of justice should be stimulated.

Regarding the example at the start: there might be legitimate reasons to make an exception to the child's bedtime, or society-wide changes might make a delayed bedtime appropriate. Its unlikeliness need not equate to its impossibility; a valid but unconsidered reason can never be excluded. Precisely *how* seriously the bedtime complaints of children at dinner parties must be taken, however, is beyond the scope of this paper.

References

Baldwin, Thomas. 'Rawls and Moral Psychology'. In *Oxford Studies in Metaethics*, by Russ Shafer-Landau, 247–70. Oxford; New York: Clarendon Press: Oxford University Press, 2006.

Brennan, Samantha, and Robert Noggle. 'Rawls's Neglected Childhood: Reflections on the Original Position, Stability, and the Child's Sense of Justice'. In *The Philosopher's Child: Critical Perspectives in the Western Tradition*, by Susan M. Turner and Gareth B. Matthews, 203–32. Rochester, NY: University of Rochester Press, 1998.

Daniels, Norman. 'Reflective Equilibrium'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2020. Metaphysics Research Lab, Stanford University, 2020. https://plato.stanford.edu/archives/sum2020/entries/reflective-equilibrium/.

Driver, Julia. 'The History of Utilitarianism'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2014. Metaphysics Research Lab, Stanford University, 2014. https://plato.stanford.edu/archives/win2014/entries/utilitarianism-history/.

Finnis, John. 'Natural Law Theories'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2020. Metaphysics Research Lab, Stanford University, 2020. https://plato.stanford.edu/archives/sum2020/entries/natural-law-theories/.

Hays, Samuel P. 'The Environmental Movement'. *Journal of Forest History* 25, no. 4 (1981): 219–21. https://doi.org/10.2307/4004614.

Horton, John. 'Rawls, Public Reason and the Limits of Liberal Justification'. *Contemporary Political Theory* 2, no. 1 (March 2003): 5–23. https://doi.org/10.1057/palgrave.cpt.9300070.

Hume, David. *A Treatise of Human Nature*. Edited by Ernest C. Mossner. Reprinted. Penguin Classics. London: Penguin Books, 1985.
Hunter, Geoffrey. 'Hume on *Is* and *Ought*'. *Philosophy* 37, no. 140 (April 1962): 148–52. https://doi.org/10.1017/S0031819100036809.

Mill, John Stuart. *Utilitarianism*. Edited by Roger Crisp. Oxford Philosophical Texts. Oxford; New York: Oxford University Press, 1998.

Pritchard, Michael S. 'Rawls's Moral Psychology'. *The Southwestern Journal of Philosophy* 8, no. 1 (1977): 59 - 72.

Rawls, John. *A Theory of Justice*. Rev. ed. Cambridge, Mass: Belknap Press of Harvard University Press, 1999.
———. 'Justice as Fairness'. *The Philosophical Review* 67, no. 2 (1958): 164–94. https://doi.org/10.2307/2182612.

Schneewind, J. B. 'The Misfortunes of Virtue'. *Ethics* 101, no. 1 (1990): 42–63.

Sloterdijk, Peter. 'The Anthropocene: A Process-State at the Edge of Geo-History?' In *Art in the Anthropocene: Encounters among Aesthetics, Politics, Environments and Epistemologies*, by Heather Davis and Etienne Turpin, 327–40. translated by Anna-Sophie Springer, 1st ed. Critical Climate Change. London: Open Humanities Press, 2015.

Wenar, Leif. 'John Rawls'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2021. Metaphysics Research Lab, Stanford University, 2021. https://plato.stanford.edu/archives/sum2021/entries/rawls/.