Alexandros Manolatos*

Rawls, Humanity and the Concept of Expression

https://doi.org/10.1515/mopp-2023-0091
Received October 23, 2023; accepted April 9, 2024; published online May 27, 2024

Abstract: In this article I present two possible interpretations of Rawls's assertion in *A Theory of Justice* that human beings have a desire to express their nature as free and rational. My reading hinges on different accounts of the Kantian conception of the person and of the Aristotelian principle and its companion effect. According to the first interpretation, this desire is a kind of natural predisposition inherent in all persons irrespective of the society in which they live. It has a universal and ahistorical aspect. The second interpretation sees our free and rational nature as an ideal that we strive to fulfill. This ideal appeals only to citizens of modern liberal democracies and entails a more qualified universalism. I argue that there is strong textual support for both interpretations but the second one is more consistent with the methodological framework of justice as fairness.

Keywords: Rawls; Aristotelian principle; Kantian interpretation; stability; basic structure

1 Introduction

The idea that we have a desire to express our nature as free and equal rational beings is one of the most enigmatic claims (Nielsen 1977) of Rawls's book *A Theory of Justice* (henceforth, *TJ*).[1] It is first stated in the famous chapter on the "Kantian Interpretation" as a necessary complement to Kant's ethics. In the third part of *TJ* it is further elaborated as an essential part of our good. In this often-overlooked section of the book Rawls asserts that this expression of our nature is a fundamental element of our rational plans of life. And from this assertion he goes on to construct his main argument for the congruence of the right and the good.

1 Nielsen (1977) writes that "one is even troubled about what such obscure talk about realizing one's nature comes to."

***Corresponding author: Alexandros Manolatos**, Department of Political Science and Public Administration, National and Kapodistrian University of Athens, Themistocleus 6, 10678 Athens, Greece, E-mail: manalex@uoa.gr. https://orcid.org/0000-0002-6328-2100

But Rawls doesn't analyze thoroughly this desire to express our nature in any part of his magnum opus and he seems to abandon this idea in his later writings, at least in the way he formulated it in *TJ*.[2] The careful reader of *TJ* must glean its intended meaning from various references in different parts of the book. This is puzzling in view of its apparent importance in the overall argument for the stability of the two principles of justice and its connection with the Kantian interpretation of justice as fairness.[3] It is also puzzling considering the metaphysical weight that it carries as a claim about human nature. One could even argue that Rawls uses this idea as a philosophical trump card[4] that puts an end to any discussion about our moral psychology and our deepest motives and the corresponding principles of justice.[5]

In this article I intend to shed light on this idea and highlight the surrounding interpretive knots and ambiguities.[6] One ambiguity has to do with the nature of this desire to express ourselves as free and equal rational beings. Is it an innate desire which regulates a person's plan of life in any historical society irrespective of its time and place? Is it a desire acquired only by citizens of a liberal democratic society? This ambiguity is related with how we understand the concept of human nature in *TJ*. One way to conceive it is as a conception of naturalness which describes the essential characteristics of human psychology. Another possibility is to treat it as a moral ideal which regulates our conduct. There are some passages in *TJ* which seem to support the first meaning and others which support the second. Rawls also never explains thoroughly the notion of expression and how it is connected with his concept of human nature. It is not clear if the term refers to a process of external manifestation of something latent in our nature or the regulation of our conduct in order to fulfill an ideal.[7]

---

2 According to Freeman (2007, 161) Rawls ceased referring to persons' "nature as free and equal rational beings" after the publication of *TJ* "perhaps to avoid a misconstrual of the Kantian elements of his view." As I will note later, in *Political Liberalism* (henceforth *PL*) Rawls makes an analogous but more modest and restricted claim, without any reference to human nature.

3 "Surprisingly, there is no one place in the text at which Rawls explains why he thinks that members of the WOS would have that desire" (Weithman 2010, 103).

4 Bernard Yack (1993, 243 n. 30) argues that Rawls's assertions about our nature "act in his theory more as a placeholder for an absent justification than as that justification itself."

5 One could even argue that the invocation of a desire as a shared human end turns justice as fairness into a teleological theory like utilitarianism. Guyer (2013) contends that Rawls's reliance on the desire to express our nature as free and equal beings means that his theory has a "teleological foundation," something that could be said that it sets it apart from Kant's strictly nonempirical, deontological theory.

6 I will call this idea the "concept of expression."

7 Rawls (1971 [1999], 253/222) says that "to express one's nature as a being of a particular kind is to act on the principles that would be chosen if this nature were the decisive determining element," but this

I will attempt to untangle these ambiguities by presenting two plausible interpretations of Rawls's claim that we have a desire to express our nature as free and equal rational beings. My thesis is that one of these interpretations is more coherent and leads to fewer contradictions with other elements of Rawls's overall theory. Nevertheless, one can find strong textual support for both of them as I will try to prove in the rest of this article. But before I examine these two interpretations I will first outline Rawls's references to this desire in various passages of *TJ* and highlight its role in the congruence argument.

2 Congruence and the Kantian Interpretation

The first mention of the desire to express our nature in *TJ* is found in the chapter "The Kantian Interpretation of Justice as Fairness." Rawls correlates this desire with the desire to follow the two principles of justice. His argument is that we express our nature as beings of a particular kind when we act on principles that would be chosen if this nature was the determining element of our choice.

> By acting from these principles persons express their nature as free and equal rational beings subject to the general conditions of human life. For to express one's nature as a being of a particular kind is to act on the principles that would be chosen if this nature was the decisive determining element … But when we knowingly act on the principles of justice in the ordinary course of events, we deliberately assume the limitations of the original position. One reason for doing this, for persons who can do so and want to, is to give expression to one's nature. (Rawls 1971 [1999], 255/224)

Rawls insists that the *concept of expression* complements Kant's concept of the categorical imperative by overcoming an apparent difficulty noted by Sidgwick, namely that "on Kant's view the lives of the saint and the scoundrel are equally the outcome of a free choice (on the part of the noumenal self) and equally the subject of causal laws (as a phenomenal self)" (Rawls 1971 [1999], 255/224).[8]

---

assertion still leaves open how the notion of expression is connected with the concept of human nature.

8 The problem for Rawls here is the challenge posed by the confusion of neutral and good freedom which Sidgwick famously ascribed to Kant. According to Sidgwick "moral or neutral" freedom is the freedom we exercise when we choose between good and evil. "Good or rational" freedom is the freedom we exercise when we choose to act morally and we are not overtaken by our passions. Following the view of neutral freedom, the liberty to choose doesn't necessarily lead to compliance with justice. One possible solution is to appeal to a "good or rational" view of freedom.

> Thus if a person realizes his true self by expressing it in his actions, and if he desires above all else to realize this self, then he will choose to act from principles that manifest his nature as a free and equal rational being. The missing part of the argument concerns the concept of expression. Kant did not show that acting from the moral law expresses our nature in identifiable ways that acting from contrary principles does not. (Rawls 1971 [1999], 255/224)[9]

Rawls notes also that the concept of expression is reflected in the conditions in the original position and the motivation of the contracting parties.

> The parties qua noumenal selves have complete freedom to choose whatever principles they wish; but they also have a desire to express their nature as rational and equal members of the intelligible realm with precisely this liberty to choose, that is, as beings who can look at the world in this way and express this perspective in their life as members of society. (Rawls 1971 [1999], 255/225)

In the end of the chapter Rawls makes a strong case for the congruence of our desire to express our nature as free and equal rational beings[10] and our desire to act justly. Invoking Kant, he argues that when we act unjustly we have feelings of shame because we have acted in a manner that fails to express our nature as free and equal rational beings. Our self-respect is damaged because we have acted "as beings of a lower order."

> Properly understood, then, the desire to act justly derives in part from the desire to express most fully what we are or can be, namely free and equal rational beings with a liberty to choose. It is for this reason, I believe, that Kant speaks of the failure to act on the moral law as giving rise to shame and not to feelings of guilt. And this is appropriate, since for him acting unjustly is acting in a manner that fails to express our nature as a free and equal rational being. Such actions therefore strike at our self-respect, our sense of our own worth, and the experience of this loss is shame (§ 67). We have acted as though we belonged to a lower order, as though we were a creature whose first principles are decided by natural contingencies. (Rawls 1971 [1999], 256/225)

Almost 200 pages later Rawls takes up again the issue of expressing our nature to claim that this expression is a fundamental element of our good.[11] He doesn't provide

---

9 In this passage Rawls makes the further statement that the expression of our nature or "our true self" is something that we "desire above all else." It is in a way an overriding or highest order desire.

10 Throughout *TJ* Rawls makes slight changes in the definition of the desire to express our nature as free and equal rational beings. For example, in one instance he talks about "moral persons" instead of "rational beings." On another occasion, he writes that we have a desire to "realize" our nature.

11 Since the desire to express our nature as free and equal rational beings is an element of our good and since the desire to act justly is the main way for a person to express this nature then Rawls can conclude that the sense of justice is part of our good. "Thus first of all, the Kantian interpretation enables us to say that everyone's acting to uphold just institutions is for the good of each. Human beings have a desire to express their nature as free and equal moral persons, and this they do most adequately by acting from the principles that they would acknowledge in the original position. When

any explanation for this proposition apart from the cryptic remark that it is something we can infer from the Aristotelian principle.

> First of all, the Kantian interpretation of the original position means that the desire to do what is right and just is the main way for persons to express their nature as free and equal rational beings. And from the Aristotelian principle it follows that this expression of their nature is a fundamental element of their good. (Rawls 1971 [1999], 445/390)

This proposition is used by Rawls as a precondition for his argument for the congruence of the good and the right. In the chapter "The Morality of Principles" he argues that we shouldn't view the sense of justice as "a form of arbitrary obedience" because the Kantian interpretation shows that by acting on the two principles of justice we express our nature as free and equal rational beings and this is part of our good.

> Finally, the Kantian interpretation of these principles shows that by acting upon them men express their nature as free and equal rational beings (§ 40). Since doing this belongs to their good, the sense of justice aims at their well-being even more directly. It supports those arrangements that enable everyone to express his common nature. (Rawls 1971 [1999], 476/417)

Subsequently, in the preface of the last section of the book on "The Good of Justice" Rawls adds an institutional element to the concept of expression. He seems to claim that our nature as free and equal rational beings can be expressed or realized only in the context of the appropriate social environment.

> Finally, by an examination of the contrast between justice as fairness and hedonistic utilitarianism, I attempt to show how just institutions provide for the unity of the self and enable human beings to express their nature as free and equal moral persons. (Rawls 1971 [1999], 513/450)

The final mention of the desire to express our nature comes in the very last pages of *TJ* where Rawls presents his final argument for the congruence between the right and the good. There, Rawls insists that one special feature of this desire strengthens the conclusion that the citizens of a well-ordered society regulated by the two principles of justice would choose to give regulative priority to their sense of justice. This feature is that this desire can't be realized if it is treated "as but one desire to be weighed against others." It can be fulfilled "only to the extent that it is likewise regulative with respect to other desires." So, in order to express their nature citizens have no alternative but to plan to preserve their sense of justice as governing their other aims.

---

all strive to comply with these principles and each succeeds, then individually and collectively their nature as moral persons is most fully realized, and with it their individual and collective good" (Rawls 1971 [1999], 528/462).

> This sentiment cannot be fulfilled if it is compromised and balanced against other ends as but one desire among the rest. It is a desire to conduct oneself in a certain way above all else, a striving that contains within itself its own priority. Other aims can be achieved by a plan that allows a place for each, since their satisfaction is possible independent of their place in the ordering. But this is not the case with the sense of right and justice … For this sentiment reveals what the person is, and to compromise it is not to achieve for the self free reign but to give way to the contingencies and accidents of the world. (Rawls 1971 [1999], 574/503)

The above quotes from *TJ* indicate that the concept of expression is intended to play an important role in the congruence argument. This is why Rawls claims that this concept is an answer to Sidgwick's remark that in Kant's view freedom and respect for the moral law do not necessitate each other. Sidgwick's objection raises the challenge of moral motivation. Freedom of choice doesn't guarantee that persons will act according to the principles of justice. The concept of expression aims to overcome this challenge by equating the desire to act justly with the desire to express our nature as free moral persons and by ascribing to this desire regulative priority over other ends.[12] But in the above passages Rawls doesn't state explicitly how we come to acquire the desire to express our free and moral nature as he does with the desire to act justly. He also doesn't say a lot about the conditions in which we express our nature as such beings. There is no equivalent to the laws of moral psychology for the desire to express our nature as free moral persons.

In what follows I will present two competing views of how we should understand this desire and its development, *the essentialist view* and the *ideal-based view*. These views are based on different accounts of the Aristotelian principle and its companion effect and of the conception of the person as free, rational and equal. This conception can be interpreted as a metaphysical or anthropological conception or as normative ethical and political conception. It can be interpreted as metaphysical or anthropological if we see it as an account that draws on answers to general questions about human nature, practical agency and moral psychology. The focus here is on depicting the invariable and essential features of human behavior and psychology. It can be interpreted as a normative ethical conception if we see it as an account that specifies a moral ideal which guides our conduct in a wide range of issues. And it can be interpreted as a normative political conception if we see it as an account that specifies a moral ideal which guides our conduct as citizens (Weithman 2010, 33). In *PL* Rawls made clear that his revised theory was based on a political conception of the person as a free and equal citizen endowed with two moral powers, a capacity to form an idea of the good and a capacity for a sense of justice. He presented it as an intuitive idea which is embedded in the political institutions of a liberal democracy.

---

12 As Rawls wrote in a lecture on equality in 1968, moral personality is the capacity to "take up a moral point of view and express it in one's conduct" (Galisanka 2019, 163).

But in *TJ* the status of this conception is not clear. In "Justice as Fairness: Political not Metaphysical" Rawls admits that his conception of justice may seem to depend on "claims about the essential nature and identity of persons" (Rawls 1999, 388) and that "the idea of the original position and the description of the parties may tempt us to think that a metaphysical doctrine of the person is presupposed" (Rawls 1999, 239). Although he claims in this article that the conception of the person in *TJ* is not based on metaphysical premises,[13] there is ample room for such an interpretation,[14] as there is for seeing this conception as an ethical ideal.

3 The Essentialist View

Although Rawls mentions at a number of points in *TJ* that we have a desire to express our nature as free and equal rational beings, he never explains precisely who *we* are and how we come to develop this desire. He doesn't also explicate the term *nature*. This omission leaves room for varying interpretations.

One possible interpretation could be that Rawls uses the concept of nature as analogous to the state of nature in social contract theory. From this point of view, the term refers to a universal human essence invariant in all cultures and places.[15] Humans – as we observe them up to this point in history or as we can logically infer[16] – are by nature free and equal rational beings. They are naturally free in the sense that they "can revise and alter their final ends and do not think of

13 "The description of the parties may seem to presuppose some metaphysical conception of the person, for example, that the essential nature of persons is independent of and prior to their contingent attributes, including their final ends and attachments, and indeed, their character as a whole. But this is an illusion caused by not seeing the original position as a device of representation. The veil of ignorance, to mention one prominent feature of that position, has no metaphysical implications concerning the nature of the self; it does not imply that the self is ontologically prior to the facts about persons that the parties are excluded from knowing" (Rawls 1999, 238).

14 This doesn't necessarily mean that Rawls relies on a distinction between phenomena and noumena like Kant or on a certain doctrine about free will but rather that he relies eventually on some features of human psychology and behavior which are presented as invariable and essential.

15 In some instances he presents this nature as what is common and essential in all humans, like when he writes that the "nature of the self as a free and equal moral person is the same for all" or that "a person realizes his true self by expressing it in his actions" or that we feel shame when we have acted unjustly because we have acted "as beings of a lower order."

16 In *TJ* Rawls doesn't make any explicit reference to Kant's position that persons are free, equal, and rational beings by their nature as moral agents. However, one could presume that Rawls implicitly invokes this position. Sandel, for example, attributes to Rawls such a view of the person. Guyer (2018), on the other hand, states that although Kant asserts that the fact that every rational being is an end in itself is "simply a fact that cannot be rationally denied," Rawls "might well have balked at using such a foundation for morality as a foundation for the principles of justice, preferring to leave it as an

themselves as inevitably bound to the pursuit of any fundamental interests that they may have at any given time" (Rawls 1999, 131–132). They are naturally equal in the sense that they have an equal moral status as persons and are "self-originating sources of claims" (Rawls 1999, 331). And they are naturally rational in the sense that they take effective means to their ends, choose courses of action that satisfy more rather than fewer of their purposes and regard life as a whole without giving preference to any particular period. The fact that at one point Rawls talks of "free and equal moral persons" instead of "free and equal rational beings" could be an indication that the meaning of the term rational is broader and includes also our reasonable nature.[17]

The analogy with the state of nature is consistent with an understanding of the desire to express our nature as a deep and widespread sentiment. The Kantian language that Rawls uses certainly implies such a reading. In the chapter "The Idea of Social Union" for example he writes that "*human beings* have a desire to express their nature as free and equal moral persons." Here Rawls seems to accept as a fact that humans generally desire to express their nature as free and equal rational beings.[18]

This reading is supported by Rawls's claim that "from the Aristotelian principle it follows" that the expression of our nature as free and equal rational beings is a fundamental element of our good.[19] Although Rawls stops short of expounding the reasoning behind this syllogism, one could assume that it is grounded on the pleasure we take from the exercise of more refined and complex activities. The Aristotelian principle says that we desire to engage in activities which give us scope to develop our higher and mature capacities. The capacity of a free and equal rational being to formulate and revise a life plan and participate in a moral community of equals admits of such complex development. It involves the development of its two moral powers – the capacity for a conception of the good and the capacity for a sense of

unadorned empirical fact that people (as in *A Theory of Justice*) or citizens in a democracy (as in his later writings) just want to express their nature as free and equal rational agents."

17 Persons as reasonable when they "are ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so" (Rawls 2005, 49).

18 As I remark in footnote 21 this fact can be empirical or logical. Guyer (2013) argues that Rawls describes this desire as an empirical fact. Robert Taylor (2011, 234) understands it as a logical fact.

19 It is not clear if Rawls means that it is a fundamental element of our good in the sense that it has regulative priority in our plans of life. If this is the case then this syllogism is connected with the final argument for the congruence of the good where Rawls says that the sense of justice can be fulfilled "only to the extent that it is likewise regulative with respect to other desires." The Aristotelian principle doesn't explain only why this desire is part of our good but also why it has regulative priority compared to other desires.

justice – and hence it is part of its good.[20] Another possible interpretation is that this syllogism is grounded on the pleasure we derive from the exercise of our natural faculties. In a footnote in *TJ* Rawls writes that "the exercise of our natural powers is a leading human good" (Rawls 1971 [1999], 426/374 n. 20). The important factor here is the *naturalness* of our two moral powers not their *complexity* or *development*.[21] The expression of our nature as free and equal rational beings is part of our good because it involves the exercise of natural faculties, the capacity for a conception of the good and the capacity for a sense of justice.

What still remains unclear, however, is under what circumstances a person develops a desire to express their nature as a free, equal and rational being. Is this, in some sense, an inborn and ahistorical desire?[22] Is it a desire all persons have even if they live in a premodern and illiberal society where freedom and equality are not widespread values? Is it so to speak a kind of natural predisposition? The analogy with the state of nature could imply such an interpretation and at moments Rawls seems to present this desire in strict universalistic terms. To speak of "human beings" points to a view of human condition that goes beyond specific times and places. This reading is also consistent with the general Kantian language in *TJ* and Rawls's talk of natural duties, natural capacities and natural feelings, attitudes and sentiments.[23]

Yet, it is difficult to imagine how such a desire could be a natural feeling or attitude like love, altruism, benevolence and friendship. This is not a desire for a certain object like food, drink or sleep or a desire to engage in pleasurable activities. It resembles more an "artificial" motive[24] like the sense of justice.[25] It is a desire that presupposes some kind of reflection on lived experiences and moral concepts like freedom and equality and as such is dependent on contingent social and historical processes. And if in a very Kantian spirit we assumed that all human beings have

20 One important change in the revised edition of *TJ* is the redefinition of the parties in the original position as having a higher interest to develop their two moral powers in order to secure the priority of basic liberty.

21 Weithman (2010, 101) makes this point about the two conjunct readings of the Aristotelian principle.

22 Dierksmeier (2021) calls it an inborn proclivity and an anthropological postulate.

23 This reliance on desires is mentioned frequently by those who question his Kantianism. Guyer (2013, 559) attempts to show that although it doesn't seem that Kant's ethics could be founded on anything like a desire to express our nature as free and equal persons, Kant "at least entertained the idea of something like a desire for freedom as the foundation for morality in his earliest reflections on moral philosophy."

24 Artificial in the way Hume understands motives that depend "on a quite complex cognitive appraisal and interpretation by the agents of their practical circumstances."

25 The sense of justice is derived from what Rawls calls natural sentiments like love and friendship but is more complex and reflective.

implicitly as moral agents a conception of themselves as free, equal and rational,[26] it would still push the point too far to derive this desire analytically from the concept of morality or to impose it as a fact of reason. Even Kant implies that although the moral law is universally valid the same doesn't hold for our predisposition to moral personality. The efficacy of this predisposition is strengthened through public moral education and the general development of our rational capabilities (Stern 1986).

A more plausible hypothesis is that for Rawls the desire to express our nature as free, equal and rational beings is not totally independent of certain historical and social prerequisites. It is part of a developmental account of human nature which conceives the essential features of our nature as characteristics which can flourish under proper circumstances and not as fixed properties. Or as Rawls says for Rousseau: "The principles of human nature are like a function: given social and historical conditions, they assign the kinds of character that will develop and be acquired in society" (Rawls 2007).

This hypothesis is consistent with the idea of social and historical evolution implicit in Rawls's theory,[27] and the parallel that might be drawn with the three-stage theory of moral development. Of course Rawls doesn't formulate, like Habermas, a systematic theory of moral development at the phylogenetic level. But, as Müller suggests, Rawls seems at times to espouse "a progress story from self-interest to moral principle" (Müller 2006). In the context of such a story, we could become susceptible to the desire to express our nature as free, equal and rational beings as history reveals this nature to us. And we could affirm this desire as we become reconciled with this nature. From this perspective, the acquisition of this desire is triggered by historical circumstance but reflects elements of our nature which wouldn't have been expressed in another social environment. In this sense, Rawls holds that citizens of liberal democratic societies have a desire to express their nature as free, equal and rational beings because they have been reconciled with this nature. The citizens of illiberal or traditional societies who may not have this desire

26 Taylor (2011, 234) argues that this conception of the person must be interpreted as a "necessary presupposition or postulate of practical reason." According to him, in *TJ* this conception of the person is justified as a self-evident principle, either by showing that it is "something we must presuppose if we are to conceive of ourselves as agents," or by showing that it is presupposed in the "fact of reason." In *PL* Rawls seems to make an attempt to derive the concept of the free, equal and rational person from the idea of practical reason, but in a letter to his publisher he recognizes that this was a mistake. I will go into more detail regarding this issue at the end of this article.

27 Rawls doesn't present a philosophy of history like Hegel. He doesn't espouse the metaphysical thesis that history is purposive, rational and fully intelligible. But in some instances he implies that there is moral progress in history. He affirms as progress the abolition of slavery, religious toleration, the rule of law and the guarantee of human rights. He also affirms that there is progress in our moral theories which is reflected in our considered judgments and "the historical consensus about what so far seem to be the more reasonable and practicable moral conceptions" (Rawls 1971 [1999], 581/509).

still share this nature as an abstract potentiality which can be expressed under specific historical circumstances. Freedom, equality and rationality are intrinsic features of human nature, but their expression depends on history.

4 Expression and Development

From what I have said up to this point *expression* could refer either to the development of natural capacities or to the determinate manifestation of abstract potentialities of our nature. In the first case, the decisive element is the development of "abilities that we find latent in our nature." We express our nature as free, equal and rational beings when we exercise the abilities which are "the decisive determining element" of this nature. Of course, we shouldn't understand this as a direct and unmediated activity like the expression of our sentiments and feelings nor like the product of something like a moral libido. The exercise of our two moral powers requires deliberation on the ends we pursue and the moral principles we follow. And this is not an activity that can be fulfilled at one determinant moment. Rawls remarks that "a person realizes his true self by expressing it in his *actions*" and that "a moral person is a subject with ends he has chosen, and his fundamental preference is for conditions that enable him to *frame a mode of life* that expresses his nature as a free and equal rational being as fully as circumstances permit" (Rawls 1971 [1999], 561/491). His reference to "the plural actions" and to "a mode of life" implies that this realization is an ongoing activity in which we are engaged in our whole lives.

In the second case, *expression* could refer to the determinate manifestation of certain properties of our nature which could have been expressed in alternative forms or suppressed under different social circumstances. This understanding of expression has an institutional element which emphasizes the historicity of human nature. Freedom, equality and rationality are abstract features of our nature that can be realized only in specific social and historical circumstances. A society regulated by the two principles of justice enables human beings to express these features by providing the appropriate institutional framework. In a society regulated by the principle of utility our nature would be expressed in a different way.[28] In this sense, we express our nature as free, equal and rational beings and not as hedonistic or benevolent creatures when we live in a specific social order and endorse its principles.

28 Rawls (1971 [1999], 513/450) remarks that by an examination of the contrast between justice as fairness and hedonistic utilitarianism he attempts "to show how just institutions provide for the unity of the self and enable human beings to express their nature as free and equal moral persons."

5 The Ideal-Based View

I have presented one possible interpretation of Rawls's claim that we have a desire to express our nature as free, equal and rational beings. This interpretation is based on an understanding of freedom, equality and rationality as natural, *essential* features of human beings. There is, however, another possible interpretation which is centered on a different view of our status as free, equal and rational persons.

In this interpretation freedom, equality and rationality are not seen as invariable elements of our nature but as aspects of an ideal. According to the ideal-based view we realize such an ideal of the person when we express our nature as free, equal and rational beings. In contrast with the essentialist view, the notion of expression doesn't lie in the development of natural faculties nor the determinate manifestation of abstract potentialities. It lies instead in the struggle to become a certain kind of person that we find desirable and inspiring. We express our nature as free, equal and rational beings when we try to live according to a plan of life which befits our status as such persons. We form, execute and revise a plan of life by having as a guide this conception of freedom, equality and rationality.[29] This process has an existential dimension. It is a way of being in the world which shapes and transforms who we are and requires commitment.[30] This is why Rawls (1971 [1999], 574/503) says that the desire to express our nature is "a desire to conduct oneself in a certain way above all else, a striving that contains itself its own priority." It is a striving because how far we succeed to live up to this ideal "depends upon how consistently we act from our sense of justice as finally regulative" (Rawls 1971 [1999], 575/503).

The appeal of this ideal of the person is restricted to liberal democracies. The possessors of the desire "to express our nature as free, equal and rational beings" are the citizens of a well-ordered society regulated by the two principles of justice. As I noted above, Rawls never states explicitly who *we* are and how we come to acquire this desire. But we can assume that he may be referring to the citizens of a well-ordered society since he mentions this desire mainly in the passages where he

29 This view is very close with Korsgraad's concept of practical identity. According to Korsgaard (1996, 101) the concept of practical identity "is better understood as a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking. You are a human being, a woman or a man, an adherent of a certain religion, a member of an ethnic group, a member of a certain profession, someone's lover or friend, and so on. And all of these identities give rise to reasons and obligations. Your reasons *express your identity, your nature*; your obligations spring from what that identity forbids" (my emphasis).

30 On the transformative dimension of the ideal of the person, see Weithman (2010). See also Lefebvre (2022) and Button (2010). Lefebvre makes the interesting claim that the original position is not just a thought experiment for choosing principles of justice but could be seen as a spiritual exercise that helps us realize this ideal.

examines the stability of justice as fairness and the congruence between the right and the good. In these passages he makes clear that he is concerned "only with the special case of a well-ordered society as characterized by the theory [of justice as fairness]."[31] At first, this seems to come at odds with the first-person plural point of view. The personal pronoun *we* could refer to the readers of *TJ*, the citizens of liberal democratic societies or human beings in general but not to the citizens of an ideal society. One possible explanation is that Rawls implicitly admits that this desire is one that we can identify with as readers of *TJ* and citizens of liberal democracies. After all, the well-ordered society of justice as fairness is modelled as a liberal democratic society. It is very likely that when Rawls uses the personal pronoun *we,* he tacitly admits that citizens of liberal democratic societies generally desire to live as free, equal and rational beings or could come to acquire such a desire under favorable conditions.[32]

But why would citizens of liberal democracies acquire such a desire? As I mentioned above, Rawls's argument about the reasons we find attractive to express our nature as free, equal and rational beings is based on the Aristotelian principle. In the essentialist view the important factor for the connection of the Aristotelian principle with the goodness of living as free, equal and rational beings is the complexity or naturalness of our two moral powers. In the ideal-based view the key to understand this connection can be found in the companion effect of the Aristotelian principle. According to the companion effect, as we witness the "exercise of well-trained abilities by others, these displays are enjoyed by us and arouse a desire that we should be able to do the same things ourselves" (Rawls 1971 [1999], 428/376). The companion effect of the Aristotelian principle triggers our admiration for expressions of human excellence and fuels a desire to be like those persons who exhibit this quality. In a sense, it prompts us to endorse ideals of personal conduct and to follow moral exemplars. The expression of our nature as free, equal and rational beings is such an expression of moral excellence. This is why we have feelings of shame when we act unjustly (Rawls 1971 [1999], 445/391). So if this companion effect is true, it follows that when we witness our fellow citizens to live as free, equal and rational beings we "have a desire to be like those persons who can exercise the abilities that we find latent in our nature" (Rawls 1971 [1999], 428/376). This desire

31 In the third part of *TJ* the whole analysis of the sense of justice and of the relevant moral psychology is framed in the context of a well-ordered society and so is that of our desire to express our nature as free, equal and rational beings.

32 Of course this desire is not so widespread as in the ideal case of well-ordered society.

occupies a central place in our rational plans of life and hence it is a fundamental element of our good.[33] It is "good from the standpoint of ourselves as well as from that of others" (Rawls 1971 [1999], 445/390).

6 Textual Evidence

As I will try to show in the next paragraphs, there are many passages in *TJ* which support both views, the ideal-based and the essentialist. The essentialist view finds textual support in passages of *TJ* where Rawls's Kantianism, the idea that principles of justice apply to all persons as moral persons,[34] is combined with a strong naturalistic and universalistic language. In one passage in the chapter "The Basis of Equality" Rawls (1971 [1999], 505/441) writes for example that the capacities for a conception of the good and for a sense of justice are "natural attributes" which "can be ascertained by natural reason pursuing common sense methods of inquiry." They are natural in the sense that if a person lacks them this is regarded as "a defect or deprivation" (Rawls 1971 [1999], 506/443) or as a lack of a "part of our humanity" (Rawls 1971 [1999], 489/428).[35] The naturalness of the sense of justice is corroborated in other passages of *TJ* by an appeal to evolutionary theory. In these passages Rawls (1971 [1999], 495/433) says that "the capacity for a sense of justice and the moral feelings is an adaptation of mankind to its place in nature" and that "beings with a different psychology either have never existed or must soon have disappeared in the course of evolution." The naturalness of these attributes imputes a universal status to the moral person. Rawls (1971 [1999], 506/443) states that "there is no race or

33 Rawls notes that someone who is liable to moral shame prizes as excellences of his person those virtues that his plan of life requires and that to possess these excellences and to express them in his actions are among his regulative aims.

34 As Galisanka shows, Rawls started in the late 1950s and early 1960s to redescribe the idea of the original position in Kantian terms. At first he gave emphasis in respecting persons as ends and later as moral persons. "In 1958, as Rawls once again started drawing on Kant, the "original position" became a point of view from which one could analyze what it means to respect moral persons. Justice was a way of respecting persons, and this respect was informed by Kant's formula of humanity: respect persons as ends, never as means only. Only in the mid-1960s did Rawls interpret this respect in terms of autonomy. Respect now required treating persons solely as moral beings, independent of the particular contingent facts about themselves and the societies in which they lived. The decisions they actually made, if these were not consistent with their moral nature, were to be rejected" (Galisanka 2019, 9).

35 Rawls also states that "this tendency is a deep psychological fact. Without it our nature would be very different and fruitful social cooperation fragile if not impossible. … If we answered love with hate, or came to dislike those who acted fairly toward us, or were averse to activities that furthered our good, a community would soon dissolve" (Rawls 1971 [1999], 495/433).

recognized group of human beings that lacks" the capacity for moral personality and "that the nature of the self as a free and equal moral person is the same for all" (Rawls 1971 [1999], 563/493). This universal point of view is aligned with the point of view of the original position and its Kantian interpretation. The parties in the original position choose principles of justice in a situation where they don't know "the particular kind of society" (Rawls 1971 [1999], 252/222) in which they live and the "decisive determining element" of their choice is their nature as free and equal rational persons (Rawls 1971 [1999], 253/222). In the final paragraph of *TJ* Rawls remarks that when we adopt this view we see our place in society *sub specie aeternitatis*, regarding "the human situation not only from all social but also from all temporal points of view" (Rawls 1971 [1999], 587/514).

Further support for the essentialist view can also be found in Rawls's lecture notes from 1968 in which he presents the Kantian conception of the person with strong naturalistic and universalistic underpinnings. In these lectures on Kantian philosophy he wrote that we recognize the moral law as "characterizing the conduct that most adequately expresses our nature as free and equal rational beings" and that it "springs from our intelligence and so from our proper self" (Galisanka 2019, 170). He also wrote that every creature seeks to express its nature and this expression is related with the exercise of its higher capacities: "All creatures of whatever kind desire to express their nature, or to realize their nature: to exercise their higher realized capacities" (Galisanka 2019, 170).

But one can find equally robust textual support for the ideal-based view, not only in *TJ* but in subsequent articles. In *TJ*, in the chapter "The Kantian Interpretation of Justice as Fairness," Rawls mentions that "the desire to act justly derives in part from the desire to express most fully what we are or *can be*" (Rawls 1971 [1999], 256/225). The reference to what we "can be" points more to a conception of freedom, equality and rationality as an ideal we want to fulfill than to a conception of how we are by nature. In other passages Rawls makes a direct appeal to an ideal of the person. In the chapter "The principle of perfection" he says that "a certain *ideal* is embedded in the principles of justice" (Rawls 1971 [1999], 326/287) and that the principles "manage to define an ideal of the person without invoking a prior standard of human excellence" (Rawls 1971 [1999], 327/287). In the chapter "Justice in Political Economy" he writes that "the principles of justice define a partial ideal of the person which social and economic arrangements must respect" (Rawls 1971 [1999], 261/231) and that justice as fairness and perfectionism "establish independently an ideal conception of the person and of the basic structure" (Rawls 1971 [1999], 262/232). Later in *TJ* he claims that the strength of the sense of justice is affected by the attractiveness of the ideals of a moral conception (Rawls 1971 [1999], 501/438).

In the paper "A Kantian Conception of Equality," published shortly after *TJ*, Rawls states that "to accept the principles that represent a conception is at the same

time to accept an ideal of the person, and in acting from these principles, we realize such an ideal." In two papers he wrote after *TJ*, "Some Reasons for the Maximin Criterion" and "Reply to Alexander and Musgrave," he refers to the conception of the person as a self-conception of the citizens of a well-ordered society (Rawls 1999, p. 230, 245). In his article "The Independence of Moral Theory" he writes that "a moral conception incorporates a conception of the person" and those "who are raised in a particular conception become in due course a certain kind of person, and they express this conception in their actions and in their relations with one another" (Rawls 1999, 293). And thus "a basic form of moral motivation is the desire to be and to be recognized by others as being a certain kind of person" (Rawls 1999, 293). In "Kantian Constructivism in Moral Theory," he says that in a well-ordered society regulated by the two principles of justice "the derivation of citizens' rights, liberties and opportunities invokes a certain conception of their person" and citizens are "educated in this conception" (Rawls 1999, 339–340). The public political culture of the well-ordered society "provides the social milieu" within which "its ideal of the person can elicit an effective desire to be that kind of person" (Rawls 1999, 339–340). In "Justice as Fairness: Political not Metaphysical" Rawls responds to Dworkin's claim that the original position is based on a right to equal respect by saying that justice as fairness is an ideal-based view (Rawls 1999, 401). In an unpublished letter he wrote to Dworkin in 1973 he says that "that it is incompatible with the coherence view of justification to say that the fundamental assumption of justice as fairness is that moral persons have a basic right to equal respect" (Rawls 1973).

7 The Political Turn and Conception-Dependent Desires

It appears that we can find strong arguments for both views in many passages of *TJ* and in subsequent articles before the so-called political turn. Can we obtain more conclusive evidence in favor of one or the other view by looking into *PL*?

The concept of expression didn't survive the transition to *PL*, at least in the way it was presented in *TJ*. In *PL* Rawls abandoned the congruence argument and with it the idea that we have a desire to "express our nature" as beings of a certain kind. In the text of *PL* there is nowhere any direct mention that we want to express our nature as free, rational and equal beings. However, Rawls makes the more limited claim that under favorable circumstances citizens of a liberal democratic society would acquire a desire to be regarded as free and equal. In particular, in *PL*

Rawls states that justice as fairness uses a moral psychology which includes conception-dependent desires.[36] According to this "reasonable moral psychology" citizens have a capacity for moral motivation based on ideals which are part of a wider moral conception. In the case of the political conception of justice as fairness, this moral psychology means that members of a liberal democratic society *may* develop a desire to be regarded as free and equal citizens. This is an ideal embedded in the public political culture of their society.

> Once this condition is imposed, a political conception assumes a wide role as part of public culture. Not only are its first principles embodied in political and social institutions and in public traditions of their interpretation, but the derivation of citizens' rights, liberties and opportunities also contains a conception of citizens as free and equal. In this way citizens are made aware of and educated to this conception. They are presented with a way of regarding themselves that otherwise they would most likely never be able to entertain. To realize the full publicity condition is to realize a social world within which the ideal of citizenship can be learned and may elicit an effective desire to be that kind of person. (Rawls 2005, 71)

The suggestion that members of liberal democratic society *may* acquire a desire to live up to a certain ideal of citizenship is analogous to the concept of expression as seen with the lens of the ideal-based view but it is far more modest and narrower in scope.[37] This trimming, so to speak, of the concept of expression in *PL* could be regarded as evidence that in *TJ* Rawls used this concept along the lines of the essentialist view and he later retreated to a more restricted proxy along the lines of the ideal-based view. This is a plausible assumption if someone adopts the widespread belief that the so-called political turn signified a retrenchment from the universalistic, Kantian liberalism of *TJ*. According to this reasoning, if Rawls relied in *TJ* on an ideal of the person and not on a metaphysical conception he wouldn't have to abandon the concept of expression. But it is also plausible that Rawls abandoned it not because of

---

36 In *PL* Rawls makes a distinction between object-based desires, ideal-based desires and concept-based desires. The first kind of desire includes bodily desires, the desire to engage in pleasurable activities as well as desires for status, power and glory and attachments, affections, loyalties and devotions. In this category falls also the desire to pursue a vocation. In contrast, specifying the objects of principle-based desires involves rational and reasonable principles. Similarly, conception-based desires are desires to act from the principles we see "as belonging to, and as helping to articulate, a certain rational or reasonable conception, or a political ideal."

37 The ideal of citizenship is a strictly political ideal which applies only to our conduct as citizens, the desire to live up to this ideal is not connected with the Aristotelian principle and this desire is no longer described as a fact but as a *possibility*. Rawls remarks in *PL* that "the exercise of the two moral powers is experienced as a good" and this is a "consequence of the moral psychology used in justice as fairness" and he notes that in *TJ* "this psychology uses the so-called Aristotelian principle" (Rawls 2005, 203).

its metaphysical foundation but because it was founded on a comprehensive ethical ideal which exceeded the limits of a political conception.

8 Methodological Coherence

In light of what I have already said there isn't any conclusive argument in favor of one or the other view of the concept of expression. There are, however, reasons to think of the ideal-based view as more plausible on grounds of methodological coherence. As I will argue, the essentialist view is less coherent with the overall theoretical framework of justice as fairness and its methodological presuppositions. One apparent difficulty with the essentialist view is that it conflicts with Rawls's thesis in *TJ* that the principles of justice are not derived from necessary truths and self-evident premises but from general assumptions about human life and "widely accepted but weak premises" (Rawls 1971 [1999], 18/16). Rawls says that the conditions imposed on the adoption of principles are simply reasonable stipulations and it can't be claimed that "they are necessary or definitive of morality" (Rawls 1971 [1999], 578/506). They "set up an Archimedian point for assessing the social system without invoking *a priori* considerations" (Rawls 1971 [1999], 261/231), which are "too slender" as a basis (Rawls 1971 [1999], 51/44). He also stresses that "even the judgments we take provisionally as fixed points are liable to revision" (Rawls 1971 [1999], 20/18) since "there is nothing *a priori* about moral philosophy" (Rawls 1971 [1999], 438/384). The assumption that we are by nature free, equal and rational beings could hardly be considered a weak premise and it takes the status of a self-evident principle or necessary truth if associated with a Kantian justification. And if Rawls ultimately relied on such a premise, one would expect that he would offer a justification for its validity. The chapter "The Kantian Interpretation of Justice as Fairness" presents a standpoint under which justice as fairness *could* be seen as a Kantian moral theory. But one shouldn't overstate the extent of autonomy Rawls wanted to achieve from the historical contexts in which moral judgments are made.[38] Rawls doesn't explicitly invoke any Kantian-style justification of freedom and rationality but rather states that justice as fairness detaches Kant's theory from its "metaphysical surroundings" (Rawls 1971 [1999], 264/233). These surroundings should be taken to include Kant's dualism between noumena and phenomena and the conception of freedom of the will as a necessary presupposition of our practical agency.

Even so, one could argue like Robert S. Taylor that we can interpret the Kantian conception of the person as a "self-evident principle" in the form of a "necessary

38 On how Rawls attempts to embed Kant's practical philosophy in a historical context, see Berguson (2016); Dombrowski (2022).

presupposition or a postulate of practical reason"(Taylor 2011). He finds support for this belief in a remark Rawls made in "The Independence of Moral Theory."

> I note in passing that one's moral conception may turn out to be based on self-evident first principles. The procedure of reflective equilibrium does not, by itself, exclude this possibility, however unlikely it may be. For in the course of achieving this state, it is possible that first principles should be formulated that seem so compelling that they lead us to revise all previous and subsequent judgment inconsistent with them. (Rawls 1999, 289)

But although in this passage Rawls seems indeed to leave open the option of basing a moral conception in self-evident principles, he points out that this is probably a rather unlikely possibility and that even these principles must be examined opposite other judgments in a process of reflection. They are not accepted beforehand as necessary truths but rather acquire the status of first principles because they play an overriding role in leading our judgments in reflective equilibrium. In any case, a reading of the Kantian conception of the person as a necessary presupposition or a postulate of practical reason is not consistent with Rawls's explicit assertion in the final chapter of *TJ* that, regarding the conditions imposed in the original position, "it was not claimed that they are self-evident, or required by the analysis of moral concepts or the meaning of ethical terms" (Rawls 1971 [1999], 579/507). In *TJ* the Kantian conception of the person is not presented as self-evident or necessary, but it is in principle liable to revision, even if Rawls was confident that it matched or organized our considered judgments correctly.

One possible reply in favor of the coherence of the essentialist view with the methodological presuppositions of justice as fairness could be that the Kantian interpretation is not the definitive justification of justice as fairness but only one possible interpretation. Rawls makes clear that the restrictions imposed in the original position "may be viewed" as "a procedural interpretation of Kant's conception of autonomy and the categorical imperative" (Rawls 1971 [1999], 256/226). The original position is not justified on Kantian premises. It is justified because "the conditions embodied in the description of this situation are ones that we do in fact accept" (Rawls 1971 [1999], 587/514) and "are in fact widely recognized" (Rawls 1971 [1999], 585/512) and they "lead to a match with our considered judgments" (Rawls 1971 [1999], 352/310). We can still accept the conditions in the original position even if we don't accept a Kantian conception of the person (Cohen 1989).[39] As Anthony Taylor (2022, 100) says "when it comes to the argument from the original position, the

39 Cohen (1989) argues that even the original position is not essential for the derivation of the two principles of justice and that they could be defended by the notion of acceptable social position.

Kantian interpretation is presented by Rawls as having the status of an optional extra – he does not suggest that readers must find this conception of autonomy compelling to accept the argument." So the justification of the principles of justice is not based on *a priori* premises. But the Kantian conception of the person ceases to be an optional extra when the reader of *TJ* reaches the stability argument in the third part of the book. The congruence argument rests on the premise that the citizens of a well-ordered society would have a desire to express their nature as free, rational and equal beings. Without this premise the congruence argument fails and the stability of a well-ordered society can't be established. Given the fact that justification of justice as fairness depends on the stability argument, the Kantian conception of the person and the concept of expression are essential parts of the central argument of *TJ*.

Another possible – and more plausible – reply is that Rawls sees our free and rational nature as a description of a universal human essence and not as an ideal, but this description should be viewed as a general empirical claim and not as an *a priori* judgment or postulate of practical reason. And so should the desire to express this nature be viewed. As Guyer (2013, 558) says, for Rawls "it is an empirically known fact that humans generally desire to express their nature (or potential) as free and equal beings," even if it is "a deep and widespread human desire." But this empirical claim remains a controversial candidate for a weak and widely accepted premise. It is anything but an obvious fact that human beings want to express their nature as free, rational and equal irrespective of the society they live in. And it is certainly not widely accepted or shared if we expand the scope of reflective equilibrium beyond the considered judgments of the citizens of liberal democratic societies. It is an empirical claim that needs some kind of philosophical argument to support it. This argument can be found in the Aristotelian principle and the history of progress implicit in justice as fairness. But still the application of this argument beyond the sphere of liberal democratic societies remains a challenge which can be surpassed only if Rawls adopts a developmental account of human nature. Even so, the acceptance of this argument poses a limit in *TJ's* universalism, a limit which is more evident in the ideal-based view. It is possible that Rawls was well-aware of this limit when he wrote *TJ* but chose not to reflect on it.

It is of course possible that Rawls aimed to construct his theory with a light metaphysical weight, but he was, so to speak, carried away into essentialist waters (or comprehensive in political liberalism's terminology) by Kantian "sirens." By this account, Rawls transgressed the nonfoundational justificatory apparatus of *TJ* in his attempt to elaborate the Kantian interpretation. Although he organized the overall justification of the two principles of justice around an ideal conception of the person,

he framed the concept of expression from an essentialist point of view.[40] This hypothesis can find some support in the fact that this "very long book" was twenty years in the making and conflates Rawls's early nonfoundationalism with his later engagement with Kant and the social contract tradition. Even if *TJ* is a remarkably coherent book, it may be the case that some of the arguments formulated during Rawls's more Kantian phase, like the congruence argument for example, contain elements which are in tension with its justificatory method. It may even be the case that Rawls believed that he could somehow remove this tension by further explicating his methodological and metaethical approach at some later point.

It is telling in this respect that in *PL*, while he omitted any mention of a desire to express our nature as free and equal rational beings, Rawls might have made an implicit attempt to give a more essentialist grounding to the ideal of the person, only to apologize for being too comprehensive. In particular, in a passage in *PL* Rawls states that his justificatory method "embodies all the relevant requirements of practical reason and shows how the principles of justice follow from the principles of practical reason in union with conceptions of society and person, themselves ideas of practical reason" (Rawls 2005, 90). The statement that the conception of the person is "an idea of practical reason" which is connected with "principles of practical reason" can be read as in line with the essentialist view since it implies that this conception is a self-evident principle or postulate and is not contingent on historical facts. This is probably the reason Rawls called it a "serious mistake" in a letter to his editor from 1998, reprinted in a later edition of PL, where he admits that "throughout the original text the idea of principles of practical reason" gave the impression "that Kant's ideas of practical reason were being used" (Rawls 2005, 438) One motivation behind this attempt to infuse, as it were, a kind of essentialist grounding in an otherwise ideal-based conception of justice could be that Rawls wanted to stave off the danger of historical relativism.[41] If we see the conception of the person as an idea of practical reason which is connected with principles of practical reason we open the way for the justification of a conception of justice to humanity in general and not just to a particular historical society. This justification is in line with the universalistic aspirations of justice as fairness but is in tension with its non-foundationalism. It might be the case that Rawls was aware of this tension and struggled to accommodate it even after the publication of *PL*.

40 Essentialist in the terms I explained above.

41 As many commentators have remarked, one implication of Rawls's method of justification is that different persons might be justified to hold different conceptions of justice depending on the society in which they live. (e.g., Singer 1974).

9 Conclusions

There are two plausible interpretations of Rawls's claim that we have a desire to express our nature as free and equal rational beings. The first one, the essentialist view, is based on a conception of this nature as a universal human essence invariant in all cultures and places. The second one, the ideal-based view, conceives this nature as an ideal that we strive to realize. The essentialist view is universal in character while the ideal-based view appeals to liberal democratic societies. As I have shown, there is strong textual support for both views. But the essentialist view is less coherent with the methodological presuppositions of *TJ*. It acknowledges as a logical or empirical fact that we want to express ourselves as free, equal and rational beings. In the first case, it is in conflict with Rawls's thesis that the conditions imposed on the original position are not derived from necessary truths and self-evident premises. In the second case, it is in tension with Rawl's thesis that the conditions in the original position rely on weak but widely accepted premises. And although it can survive this tension, the conception of the person as an ideal qualifies as a weaker and more widely accepted premise. In this respect, the ideal-based view fits better with the methodological framework of justice as fairness and the nonfoundational character of reflective equilibrium.

The ideal-based view entails, however, a more qualified universalism. This qualified universalism is a consequence of the method of reflective equilibrium itself and must have created a dilemma for Rawls during the writing of *TJ*. This method is based on *our* considered judgments but the broader the group of people whose judgments are taken into account the fewer the chances of actually achieving an equilibrium. A theory that rests on the existence of shared judgments is most vulnerable to objections that these judgments are not universally shared. Despite the fact that *TJ* was limited to a set of decisive political questions and aimed to be a guiding framework and not a deductive schema, it would be difficult to guide *our* judgments into agreement if it was applied outside the range of liberal democratic societies. Only a justification which relied on some objective truth – metaphysical or anthropological in the case of the conception of the person – could guarantee a universal application. This limitation, and dilemma, was implicit in *TJ* but was never explicitly clarified. When *TJ* came into print in 1971 it was a grandiose work with universal aspirations, but in reality its universalism was more qualified than its aspirations. In the years that followed, Rawls reformulated his theory and in *PL* he made explicit that it applied only to liberal democratic societies. Yet his self-canceled attempt to ground the conception of the person as an idea of public reason might be a sign that he still struggled with this dilemma.[42]

42 Some proponents of Rawls's philosophical project have responded to this dilemma by founding the original position on a more comprehensive Kantian conception of the person grounded on a practical postulate of freedom (e.g., Taylor 2011).

Apart from the challenges it poses to the global advancement of liberal values, the ideal-based view entails a more demanding design of civic education for the resilience of these values in liberal democratic societies. If the desire to express our nature as free, equal and rational beings is not a given fact or a given fundamental preference, at least for those of us living in liberal democratic societies, as implied by the essentialist view, but is acquired by our exposure to attractive role models and ideals, as implied by the ideal-based view, then the channels through which these ideals are transmitted acquire a more prominent role in the design of the institutional framework of a well-ordered society. And this brings to the fore not only questions about liberal neutrality but also about the role and scope of the basic structure, since these channels are not necessarily institutional.[43] This is even more so if we take into account that the pervasiveness of these ideals is critical for the stability of justice as fairness. As I noted above, although Rawls abandoned in *PL* the concept of expression and the congruence argument, he made the more limited case that members of a well-ordered society may develop a desire to be regarded as free and equal citizens and that the development of this desire is crucial for stability. Rawls also claimed that the public political culture of a well-ordered society plays an educative role in helping citizens form a conception of themselves as free, equal and rational. As Rawls (2001, 146) writes "citizens acquire an understanding of the public political culture and its traditions of interpreting basic constitutional values" and they do so by "attending to how these values are interpreted by judges in important constitutional cases and reaffirmed by political parties." But the confinement of this educational role to political and judicial institutions may not be enough to create support for the ideal of free, equal and rational citizens,[44] especially if we take into account that the associations of civil society and families may bar their members from being exposed to this ideal.[45]

References

Berguson, J. 2016. *John Rawls and the History of Political Thought: The Rousseauvian and Hegelian Heritage of Justice as Fairness*. New York: Routledge.

43 I would like to thank an anonymous referee for constructive comments regarding this issue.

44 Some commentators (e.g., Costa 2010) have argued that we should assign a significant educational role to public schools.

45 Schouten (2019) argues that gendered division of labor restricts access to ideals of comprehensive and political autonomy and this creates a threat to stability. This justifies according to Schouten gender egalitarian policies which in incentivize the promotion of these ideals.

Button, M. 2010. *Contract, Culture, and Citizenship Transformative Liberalism from Hobbes to Rawls*. University Park, PA: Pennsylvania State University Press.
Cohen, J. 1989. "Democratic Equality." *Ethics* 99: 727–51.
Costa, M. V. 2010. *Rawls, Citizenship and Education*. New York: Routledge.
Dierksmeier, C. 2021. "Drop Rawls?" *Business Ethics, the Environment and Responsibility* 31: 281–92.
Dombrowski, D. A. 2022. *Pre-liberal Political Philosophy: Rawls and Plato, Aristotle, Augustine, Aquinas*. Boston: Brill.
Freeman, S. 2007. *Justice and the Social Contract: Essays on Rawlsian Political Philosophy*. New York: Oxford University Press.
Galisanka, A. 2019. *John Rawls: The Path to a Theory of Justice*. Cambridge, MA: Harvard University Press.
Guyer, P. 2013. "Rawls and the History of Moral Philosophy: The Cases of Smith and Kant." In *A Companion to Rawls*, edited by J. Mandle, and D. A. Reidy, 546–66. Chichester: Wiley Blackwell.
Guyer, P. 2018. "Principles of Justice, Primary Goods and Categories of Right: Rawls and Kant." *Kantian Review* 23 (4): 581–613.
Korsgaard, C. M. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.
Lefebvre, A. 2022. "The Spiritual Exercises of John Rawls." *Political Theory* 50 (3): 405–27.
Müller, J.-W. 2006. "Rawls, Historian: Remarks on Political Liberalism's 'Historicism." *Revue Internationale de Philosophie* 237 (3): 327–39.
Nielsen, Kai. 1977. "Rawls and Classist Amoralism." *Mind* 86: 19–30.
Rawls, J. 1971 [1999]. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Rawls, J. 1973. "Reply to Dworkin." John Rawls Faculty Papers, Harvard University Archives, HUM 48, Box 33, folder 8.
Rawls, J. 1999. *Collected Papers*. Cambridge, MA: Harvard University Press.
Rawls, J. 2001. *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.
Rawls, J. 2005 [1993]. *Political Liberalism*. New York: Columbia University Press.
Rawls, John. 2007. *Lectures on the History of Political Philosophy*, 207. Cambridge, MA: Harvard University Press.
Schouten, G. 2019. *Liberalism, Neutrality and the Gendered Division of Labor*. Oxford: Oxford University Press.
Singer, P. 1974. "Sidgwick and Reflective Equilibrium." *The Monist* 58: 490–517.
Stern, P. 1986. "The Problem of History and Temporality in Kantian Ethics." *The Review of Metaphysics* 39: 505–45.
Taylor, R. S. 2011. *Reconstructing Rawls: The Kantian Foundations of Justice as Fairness*. University Park, PA: Pennsylvania State University Press.
Taylor, A. 2022. "Rawls's Conception of Autonomy." In *The Routledge Handbook of Autonomy*, edited by B. Colbern, 96–110. London: Routledge.
Weithman, P. 2010. *Why Political Liberalism? On John Rawls's Political Turn*. New York: Oxford University Press.
Yack, B. 1993. "The Problem with Kantian Liberalism." In *Kant and Political Philosophy: The Contemporary Legacy*, edited by R. Beiner, and W. J. Booth, 224–45. New Haven: Yale University Press.