

## Xinghua Wang

Chinese Academy of Social Sciences, Institute of Philosophy, Jianguomennei Street, No. 5, CN-100732 Beijing  
wangxh-zxs@cass.org.cn

# Rousseauian Heritage of Rawls's Moral Psychology

### Abstract

*Some authors have noticed Rousseau's influence on Rawls's original position argument for justice as fairness, but few have argued for Rousseau's influence on his moral psychology, which constitutes the first part of his stability argument. I will argue that Rawls's account of moral development parallels, and is grounded in, Rousseau's thoughts on amour-propre. In particular, I argue that (1) Rawls's thesis that the sense of justice is derived from love and friendship is an illustration of Rousseau's thesis that moral sentiments are derived from natural sentiments, that (2) Rawls's explanation for how we acquire the reciprocity of disposition is grounded in Rousseau's explanation of how amour-propre can be extended into the sense of justice, and that (3) Rawls's thoughts on the principle of sympathy parallel Rousseau's thoughts on compassion.*

### Keywords

Jean-Jacques Rousseau, *amour-propre*, John Rawls, moral psychology, sense of justice

Some have noticed Rousseau's influence on Rawls's original argument for justice as fairness,<sup>1</sup> but few have argued for Rousseau's influence on his stability argument, which constitutes the second part of his argument for the two principles of justice. The importance of the stability in Rawls's overall project has been emphasised by many commentators. Nevertheless, only a few recognise the importance of Rousseau's concept of *amour-propre* in Rawls's stability argument, especially in his moral psychology.<sup>2</sup> Rawls acknowledges the influence of Rousseau on his moral psychology, as he says, his aim in "The Sense of Justice" (1963) is to reconstruct an account of moral development to

1

Bercuson argues for a Rousseauian interpretation of Rawls's justice as fairness. – Jeffrey Bercuson, *John Rawls and the History of Political Thought: The Rousseauian and Hegelian Heritage of Justice as Fairness*, Routledge, London 2014.

2

For the importance of Rawls's stability for his overall project, see: Samuel Freeman, "John Rawls – an Overview", in: Samuel Freeman (ed.), *The Cambridge Companion to Rawls*, Cambridge University Press, Cambridge 2003, pp. 1–59; Samuel Freeman, *Rawls*, Routledge, London 2007; Jon Garthoff,

"Rawlsian Stability", *Res Publica* 22 (2016) 3, pp. 285–299, doi: <https://doi.org/10.1007/s11158-015-9292-z>; Thomas E. Hill, "Stability, A Sense of Justice and Self-respect", in: Jon Mandle, David Reidy (eds.), *A Companion to Rawls*, Wiley-Blackwell, Hoboken 2014, pp. 200–215, doi: <https://doi.org/10.1002/9781118328460.ch11>; Paul J. Weithman, *Why Political Liberalism?: On John Rawls's Political Turn*, Oxford University Press, Oxford 2010; Thomas Pogge, *John Rawls: His Life and Theory of Justice*, Oxford University Press, Oxford 2007.

illustrate how “Rousseau’s thesis might be true” (CP 96).<sup>3</sup> What Rawls refers to by “Rousseau’s thesis” is the idea that “the sense of justice is no mere moral conception formed by the understanding alone, but a true sentiment of the heart enlightened by reason, the natural outcome of our primitive affections” (CP 96).<sup>4</sup> We have seen that *amour-propre* has a reciprocal feature. According to the reciprocity thesis, those who have self-respect tend to respect others. More generally, the reciprocity thesis says that those who receive benefits from others tend to benefit others. This reciprocity thesis originated in Rousseau but is developed by Rawls. In “The Sense of Justice” and Chapter VIII of *A Theory of Justice*, Rawls explains how the reciprocity of disposition is exhibited in each of the three stages of moral development. In *Restatement*, he re-emphasises the importance of the acquisition of the sense of justice through the three stages of moral development and argues that the reciprocity of disposition is an essential element in his reasonable moral psychology. Thus, it should come as no surprise that Rawls’s moral psychology has a Rousseauian interpretation.

In this paper, I will first briefly introduce Rawls’s moral psychology, and then I will argue that Rawls’s account of moral development parallels, and is grounded in, Rousseau’s thoughts on *amour-propre*. In particular, I argue that (1) Rawls’s thesis that the sense of justice is derived from love and friendship is an illustration of Rousseau’s thesis that moral sentiments are derived from natural sentiments, that (2) Rawls’s explanation for how we acquire the reciprocity of disposition is grounded in Rousseau’s explanation of how *amour-propre* can be extended into the sense of justice, and that (3) Rawls’s thoughts on the principle of sympathy parallel Rousseau’s thoughts on compassion. In arguing for the Rousseauian interpretation of Rawls’s moral psychology, I will also answer some of the objections against his moral psychology in a Rousseauian way.

## 1. Rawls’s Moral Psychology

Rawls says that his moral psychology is a psychological reconstruction which aims to illustrate how Rousseau’s thesis can be true. Rousseau’s thesis includes three aspects. First, he argues that moral sentiments are derived from natural sentiments. Second, he argues that *amour-propre* has a reciprocal feature. And third, he argues that self-love or self-esteem is the primary motivating force for human actions, not compassion. I will call the first aspect “Rousseau’s thesis” narrowly construed (as Rawls calls it), the second aspect “the reciprocity thesis”, and the third aspect “a realistic moral psychology”. I argue that these three aspects parallel and provide the basis for Rawls’s moral psychology. In this section, I will first briefly explain Rawls’s moral psychology. In the next section, I will explain the parallel relations between Rawls’s moral psychology and that of Rousseau.

Rawls’s moral psychology did not fundamentally change between his early and later works.<sup>5</sup> In “The Sense of Justice” he outlines the three stages of moral development through which a person acquires a sense of justice which motivates him to stabilise the basic structure of the well-ordered society. In *Theory*, this same account of moral development is restated. The role of moral psychology, however, is not limited to explaining the acquisition of a sense of justice, but also explains the congruence between the right and the thin conceptions of the good, and explains the absence of special psychology such as envy. In *Political Liberalism*, Rawls argues that there are two questions of

stability, the first one is “whether people who grow up under just institutions (as the political conception defines them) acquire a normally sufficient sense of justice so that they generally comply with those institutions” (PL 141).<sup>6</sup> The second question is whether the political conception of justice can be the focus of an overlapping consensus. He argues that the first question is answered by his account of moral development in *Theory*. The second question is answered by the fact that the political conception of justice is acceptable for all reasonable comprehensive views. In *Restatement*, in response to the objection that the consensus is utopian, he argues that it is a reasonable moral psychology, which includes the reciprocity of disposition that is exhibited in each stage of moral development (JF 196).<sup>7</sup> Thus, even though the role of moral psychology shifts slightly in his later writings, the content of moral psychology remains unchanged.<sup>8</sup>

The content of Rawls's moral psychology includes an account of moral persons and an account of moral development to specify the account of moral persons. Rawls's moral persons are free and equal, rational and reasonable. They are free in the sense that they have two moral powers, the capacity for a sense of justice and the capacity for a conception of the good (TJ 11; JF 21; PL 19).<sup>9</sup> They are equal in the sense that they equally have these moral powers to a minimal degree to be fully cooperating members of the society (TJ 17; JF 20; PL 19). They are rational because they can choose and alter their ends, and they seek efficient and probable means to achieve their ends (TJ 10; JF 6–7; PL 50). They are reasonable in the sense that they are willing to honour the fair terms of cooperation, provided that others are willing to do the same. The idea of reasonableness thus is a part of the idea of reciprocity (JF 6–7; PL 40–50). Thus, to honour the two principles of justice is to honour the principle of reciprocity. They are reasonable also in the sense that they are willing to recognise the burdens of judgment and to use public reason to resolve fundamental political problems (PL 54).

3

CP refers to John Rawls, *Collected Papers*, edited by Samuel Freeman, Harvard University Press, Cambridge (MA) 1999.

4

This thesis is explicit in *Emile* where Rousseau says: “Our first duties are to ourselves: our primary sentiments are centered on ourselves; all our natural movements relate in the first instance to our preservation and our well-being. Thus, the first sentiment of justice does not come to us from the justice we owe but from that which is owed us; and it is again one of the mistakes of ordinary education that, speaking at first to children of their duties, never of their rights, one begins by telling them the opposite of what is necessary, what they cannot understand, and what cannot interest them.” (E 97). E refers to: Jean-Jacques Rousseau, *Emile: Or, on Education*, translated by Alan Bloom, Basic Books, New York 1979.

5

This point is made by Wenar and Estlund. See: Leif Wenar, “The Unity of Rawls's Work”, in: Thom Brooks, Fabian Freyenhagen (eds.), *The Legacy of John Rawls*, Continuum, Lon-

don 2005; David Estlund, “The Survival of Egalitarian Justice in John Rawls's *Political Liberalism*”, *Journal of Political Philosophy* 4 (1996) 1, pp. 68–78, doi: <https://doi.org/10.1111/j.1467-9760.1996.tb00042.x>.

6

PL refers to John Rawls, *Political Liberalism*, Columbia University Press, New York 1993.

7

JF refers to John Rawls, *Justice as Fairness*, Harvard University Press, Cambridge (MA) 2001.

8

For the roles of moral psychology in Rawls, see: Thomas Baldwin, “Rawls and Moral Psychology”, in: Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Oxford University Press, 2008, p. 251.

9

TJ refers to John Rawls, *A Theory of Justice*, Harvard University Press, Cambridge (MA) 1971.

The principle of reciprocity is specified in Rawls's account of moral development. His three-stage development of the morality of principles is initially stated in "The Sense of Justice" and Chapter VIII of *Theory*. The origin of moral sentiments is the natural love in the infant stage. Family is the first institution of moral education. Thus Rawls says that the basic structure of the well-ordered society includes family (TJ 405). In the first stage, parents love their child, and they express their love for their child evidently through their care for him, their intention to do what the child's rational self-love would incline and the fulfilment of this intention. When the child recognizes parents' love of them, the child loves them back. This is the first psychological law. The child's returning sentiment is not immediate after they see "the evident intention" of their parents' love. It needs to go through several steps: first, the child is assured of their self-worth when they see "the evident intention" of their parents' love; second, the child trusts their parents and have confidence in their surrounding environment, as they experience their parents', unconditional love; third, in developing various skills, the child gradually acquires a sense of self-confidence, which affirms their self-esteem; and fourth, the child has tightened bonds with and deep affection for their parents as they connect them with their self-development and self-esteem. The child's returning sentiment is the first appearance of the reciprocity of disposition. This stage is called "the morality of authority", for the child sees their parents as the sources of authority. And the child would have a sense of guilt, "authority guilt", which is a proto-moral sentiment, if they violate their commands, provided that they have the natural sentiments of love and trust. Since the natural sentiments of love and trust are necessary for the moral sentiment of guilt, the absence of guilt indicates the absence of love and trust (TJ 425). But this stage is not a moral stage, but a proto-moral stage of moral development, for the child does not have moral sentiments in reference to moral concepts.<sup>10</sup>

In the proto-moral stage, the child evaluates the goodness or badness of action according to a person's external behaviour regardless of their motives. But in the second stage of moral development, as they assume more demanding roles in more complex schemes of rights and duties in various associations, the examples of which are family, school, games and play, and national association, they will have a more robust moral point of view. They gradually learn the standard of being a good son or daughter, a good student, a good sport and companion, and a good citizen. They learn to identify themselves with others and see things from their perspective. The development of their intellectual abilities is necessary for their acquisition of moral sensibility. But this intellectual development is not sufficient for the acquisition of moral sensibility, for a person who possesses these intellectual skills may be manipulative and exploitative. Thus, a necessary condition for moral sensibility is the recognition of others as equals.<sup>11</sup> By the recognition that other people are equally doing their part in the association that they belong to, the child acquires the feelings of trust and friendship toward them, provided that they love their parents during Stage 1. If the parents fail to do their part, the child will have a sense of guilt, "association guilt", the absence of which indicates the absence of the natural feelings of trust and friendship. In other words, the natural feelings of trust and friendship are necessary for the moral sentiment of association guilt. And so the presence of the moral sentiment of guilt indicates the presence of the natural feelings of trust and friendship. The returning sentiments of trust and friendship also need to go through several steps: the first step is to recognise that others' doing their part is beneficial to each and all in the association. Secondly, it is regarded as a form of good will that others are doing

their part, and they want to, and this arouses feelings of trust and friendship. Thirdly, everyone who participates in the same activity reciprocally does their part in the association.

In a certain form of the morality of association, once a person has the recognition that their fellow citizens willingly contribute to the cooperative scheme of a well-ordered society, they will have friendly feelings and trust toward them, and they will reciprocally do his share. They will have a sense of guilt if they fail to contribute to social cooperation, provided that they have feelings of trust and friendship toward them. But since the natural feelings of trust and friendship are necessary for the moral sentiment of guilt, their moral sentiments are contingent upon their natural feelings of trust and friendship, without which they would not have “association guilt”. In the final stage of moral development, the morality of principles, a person acquires a sense of justice if they recognise that the just institutions benefit them and those whom they care for, provided that they acquired feelings of love and friendship in the first two stages. Unlike in the second stage, now their moral sentiments are not contingent upon their natural feelings toward their fellow citizens, but upon the acceptance of the principles of justice. While a citizen is attached to several other citizens, nobody is attached to all. And so citizens cannot be bound with each through the ties of friendship; rather, they are bound by the allegiance to the principle of justice. Note that the acquisition of love and friendship in the first two stages is still necessary for the sense of justice. But once a person has a sense of justice, they will act upon the principles of justice regardless if they have feelings of love or friendship toward others or not. If they violate the principles of justice, they will have a sense of guilt, “principle guilt”, regardless of the contingencies of the world. But if they, at the same time, have fellow feelings toward their countrymen, their moral sentiments will be intensified. Rawls thus says that the sense of justice and the love of humankind are two forms of the morality of principles. But he argues that the later is supererogatory while the former is a natural duty. He concludes that justice as fairness have a more realistic moral psychology than other conceptions of justice. It is more realistic than intuitionism because intuitionism fails to explain the sense of justice in a psychologically understandable way (TJ 418). And even though utilitarianism and perfectionism naturally explain the sense of justice, Rawls suggests that they ground justice in the love of humankind, a supererogatory duty.<sup>12</sup> What makes Rawls’s moral psychology realistic is that it recognises the primary motivating force of self-respect, which inclines a person to respect others. The principle of reciprocity is neither egoistic nor altruistic.

10

Rawls argues that “it is a necessary feature of moral feelings, and part of what distinguishes them from the natural attitudes, that the person’s explanation of his experience invokes a moral concept and its associated principles. His account of his feelings makes reference to an acknowledged right or wrong” (TJ 421). Since children in the first stage of moral development do not have a clear idea of moral right or wrong, they do not have the strict sense of moral feelings.

11

As we have seen that in the first stage, children are affirmed of self-respect through their

parents’ manifest love and the development of their own intellectual skills. The sense of self-respect inclines them to respect others and recognize them as equals.

12

Rawls connects utilitarianism with the principle of sympathy or the love of humankind in several places, for example, see chapter 5, section 30 in *Theory*. For Rawls, both perfectionism and utilitarianism may sacrifice freedom for the sake of the greater good (perfection or utility), and so their moral psychologies are similarly unrealistic.

To sum up, the content of Rawls's moral psychology has three themes. First, the sense of justice is derived from love and friendship. In other words, to acquire the sense of justice in the final stage, a person must love their parents back in the first stage and have friendly feelings toward their associates in the second stage. Second, the reciprocity of disposition is exhibited in each stage of moral development. In the first stage, once the child recognises their parents' love with evident intention, they will love their parents back. In the second stage, given that the child loves their parents back in the first stage, once they see that their associates do their parts in the associations they belong to, they will have feelings of trust and friendship toward them. In the third stage, given that they have returning sentiments of love and friendship in the first two stages, once they see that the just institutions are beneficial to them and those around them, they will acquire a sense of justice. Third, Rawls's moral psychology is realistic moral psychology. It is more realistic than utilitarianism and perfectionism because in Rawls' theory the love of humankind, which is their psychological basis is a supererogatory duty. Of course, only in an ideal society, everyone will go through these three stages of moral development. In the real life, many people will deviate from these stages. People in the later stage of moral development have the moral (and natural) sentiments developed in the previous stages.

## 2. The Parallel Relations Between Rousseau's Concept of *Amour-propre* and Rawls's Moral Psychology

I argue that Rawls's moral psychology parallels, and is grounded in, Rousseau's moral education in *Emile*.<sup>13</sup> First, Rawls's thesis that the sense of justice is derived from love and friendship is an illustration of Rousseau's thesis that moral sentiments are derived from natural feelings. Second, Rawls's explanation of how we acquire the reciprocity of disposition is grounded in Rousseau's explanation of how *amour-propre* can be extended to the sense of justice. And third, Rawls's thoughts on the principle of sympathy or the love of humankind parallel Rousseau's realistic moral psychology.

### 2.1. Rousseau's Thesis

Rawls acknowledges that his moral psychology is an illustration of how Rousseau's thesis might be true (CP 96). He says that Rousseau's thesis is "the sense of justice is no mere moral conception formed by the understanding alone, but a true sentiment of the heart enlightened by reason, the natural outcome of our primitive affections" (CP 96). His account of moral development tries to show that the understanding or recognition do not simply form a person's sense of justice, that the just institutions are beneficial to them and those around them, but also the outcome of the natural feelings of love and friendship. It is true that in the final stage of moral development, the sense of justice is not contingent on the natural feelings. But without the acquisition of the natural feelings of love and friendship in the first two stages, it is impossible to acquire a sense of justice. Or we may say, the acquisition of the natural feelings of love and friendship is a necessary condition for the acquisition of the sense of justice.

John Deigh (1996) argues that Rawls fails to explain how the sense of justice is derived from love and friendship. This failure, he argues, comes from Rawls's confusion between guilt and remorse. Deigh argues that the moral feeling, necessarily connected with the natural feelings of love and friendship,

is remorse, and that Rawls mistakenly confuses this feeling with guilt. He argues that this confusion leads Rawls to claim that a sense of guilt connects the sense of justice with the natural feelings of love and friendship. But since guilt is not necessarily connected with the natural feelings of love and friendship, he argues, the sense of justice is not derived from love and friendship.

“To explain how the child comes to love its parents is not yet to explain how it comes to respect them. The former establishes that the child identifies with its parents but not that it accepts their authority, and acceptance of authority is the facet of moral development Rawls means to explain in giving his account of the first stage.”<sup>14</sup>

Deigh correctly points out that love is not sufficient for “authority guilt”. It is not sufficient to explain how children come to respect their parents’ authority. It is also not sufficient to explain how citizens in the well-ordered society of justice as fairness acquire the sense of justice. Rawls acknowledges that love is merely a necessary but not sufficient condition for “authority guilt”.

“In the light of this sketch of the development of the morality of authority, it seems that the conditions favoring its being learned by the child are these. First, the parents must love the child and be worthy objects of his admiration. In this way they arouse in him a sense of his own value and the desire to become the sort of person that they are. Secondly, they must enunciate clear and intelligible (and of course justifiable) rules adapted to the child’s level of comprehension. In addition they should set out the reasons for these injunctions so far as these can be understood, and they must also follow these precepts insofar as they apply to them as well.” (TJ 408)

Thus, for Rawls, love is merely a necessary but not sufficient condition for moral sentiments. For a child to acquire “authority guilt”, they not only need to have parental love but also understand the rules that are set up and followed by their parents. But Deigh not only denies that love is sufficient for moral sentiments, he also denies that love is necessary for moral sentiments. He endorses Freud’s pessimistic view of moral education and explains guilt in terms of fear.

“At this stage, fear is the emotion that they in their role as authorities evoke in the child. Then, as their authority becomes internal, as the child comes to see itself as part of a moral order over which its parents preside, its fear of their power is transformed into respect for their authority. Correspondingly, its fear of punishment – or of whatever evil the child senses will be the consequence of disobedience – becomes a liability to guilt.”<sup>15</sup>

Here Deigh argues that a person has a feeling of guilt only if they are afraid of punishment. But Deigh confuses the acceptance of authority with the acquisition of moral sentiment. To explain why a person accepts parental (or anyone’s) authority does not yet explain why this person has the moral sentiment of guilt when they violate their parents’ (or anyone’s) demands. In fact, if they have a gun to their head, then a person may accept their parents’ (or anyone’s) authority, but it does not follow that they have any feelings of guilt when they violates their demands. In any case, Deigh does not explain why and how a person internalises an external authority. Fear of punishment may explain

13

Samuel Freeman points out that Rawls largely agrees with Rousseau’s description of human nature and that Rawls’s moral psychology is largely influenced by Rousseau’s moral education in *Emile*. – S. Freeman, *Rawls*, p. 19.

*Theory*, Cambridge University Press, Cambridge 1996, p. 57.

15

*Ibid.*, p. 62.

14

John Deigh, *The Sources of Moral Agency: Essays in Moral Psychology and Freudian*

why a person accepts an external authority, but it does not explain why and how this person internalises such authority.<sup>16</sup> Rawls's and Rousseau's moral psychologies fill in the blank. Rawls agrees with Rousseau that love is the origin of moral sentiments. Without the natural sentiment of love, the moral sentiment of guilt is inexplicable. Fear cannot explain the moral sentiment of guilt. A person may do as someone commands out of fear, but he will not have the moral sentiments of guilt if they violate this person's commands. Rousseau warns us about the consequences of punishing children:

"Firstly, by imposing on them a duty they do not feel, you set them against your tyranny and turn them away from loving you. Secondly, you teach them to become dissemblers, fakers, and liars in order to extort rewards or escape punishments. Finally, by accustoming them always to cover a secret motive with an apparent motive, you yourselves give them the means of deceiving you ceaselessly, of depriving you of the knowledge of their true character, and of fobbing you and others off with vain words when the occasion serves." (E 91)

Rousseau writes:

"It is quite strange that since people first became involved with raising children, no instruments for guiding them has been imagined other than emulation, jealousy, envy, vanity, avidity, and *vile fear* – all the most dangerous passions, the quickest to ferment and the most appropriate to corrupt the soul, even before the body has been formed. With each lesson that one wants to put into their heads before its proper time, a vice is planted in the depth of their heart." (E 92)<sup>17</sup>

Thus, for Rousseau, fear is not the appropriate means of making a child a good person. On the contrary, fear is "most appropriate to corrupt the soul". A child educated in this way would be either servile or arrogant. This child, in Rousseau's terms, would have an inflamed *amour-propre*. If in the early stage of their moral development, a person's *amour-propre* is inflamed, it will be impossible for them to acquire the sense of justice in their later stages. Rousseau argues that the sense of justice is derived from primitive sentiments.

"Our first duties are to ourselves; our primary sentiments are centered on ourselves; all our natural movements relate in the first instance to our preservation and our well-being. Thus, the first sentiment of justice does not come to us from the justice we owe but from that which is owed us." (E 97)

For Rousseau, since self-love is reciprocal (E 214), a person who is affirmed of his self-love by his parents' manifest love tends to love his parents back. If this person "sees that everything approaching him is inclined to assist him" (E 213), they will have friendly feelings toward those around them. Eventually, they will acquire a sense of justice.

To sum up, Rawls's moral psychology is an illustration of Rousseau's thesis that moral sentiments are derived from natural sentiments, or, in particular, the sense of justice is derived from love and friendship. Deigh argues that Rawls's moral psychology mistakenly grounds justice in love. He argues that the sense of justice is derived from fear. But as we have seen, in Rousseau, fear is not the appropriate way to nurture the sense of justice. On the contrary, it gives rise to an inflamed *amour-propre*, which, as we have seen, is the root of all evils.

## 2.2. The Reciprocity Thesis

Rousseau suggests that our natural and healthy form of *amour-propre* makes a rightful claim on our conduct, a claim to grant others equal standing with us provided they do the same in return. And so those who have self-respect are more likely to respect others. This explains why the natural and healthy



form of *amour-propre* is the psychological foundation of a society of mutual respect and self-esteem for both Rousseau and Rawls. But it remains to be seen how citizens in the well-ordered society of justice as fairness acquire the reciprocity of disposition which inclines them to accept the principle of reciprocity or the principles of justice.<sup>18</sup> Rawls's moral psychology provides an answer to this question. Rawls says that "In the account in *Theory* of the three-stage development of the morality of principles [as it is called there], the psychological laws for each stage exhibit this reciprocity of disposition" (JF 196). The reciprocity of disposition is a "tendency to answer in kind, to answer others' being fair to us with our being fair to them, and the like" (JF 196). In the first stage of moral development, children acquire the reciprocity of disposition for the first time. They are disposed to love their parents back once they recognise that their parents love and want to love them. In the second stage, as they play more complex roles in the society, they reciprocally contribute to the cooperative arrangement of the association once they recognise that others willingly contribute to it. And in Stage 3, they willingly comply with the two principles of justice once they recognize that the just institutions are beneficial to them and those around them. It is important to note that the reciprocity of disposition is acquired for the first time when the children are disposed to love their parents back on the condition that their parents love them "with evident intention". As Rawls says:

"The parents, we may suppose, love the child and in time the child comes to love and to trust his parents. How does this change in the child come about? To answer this question I assume the following psychological principle: the child comes to love the parents only if they manifestly first love him." (TJ 406)

Rawls notes that this psychological principle is drawn from Rousseau's *Emile* (TJ 406, fn. 9). In explaining the second psychological law that people acquire feelings of trust and friendship toward their associates on the condition that their associates do their part "with evident intention", he notes that "the idea of evident intention as used here is from Rousseau's *Emile*" (JF 196, fn. 16). In the *Emile*, Rousseau says:

"A child's first sentiment is to love himself; and the second, which derives from the first, is to love those who come near him, for in the state of weakness that he is in, he does not recognize anyone except by the assistance and cares he receives. At first the attachment he has for his nurse and his governess is only habit. He seeks them because he needs them and is well-off in having them; it is recognition rather than benevolence. He needs much time to understand that not only are they useful to him but they want to be; and it is then he begins to love them." (E 213)

Rousseau points out that children are disposed to return love for love if the caregivers manifestly love them. This is the first step of the extension of self-love. Human beings naturally have self-love, and they naturally love those who manifestly love them. Love is extended from themselves to those around them (provided that they love them with evident intention). Parental love planted the seed of humanity into *Emile* so that he could become a good

16

In certain cases, fear of punishment or power cannot explain why a person accepts an external authority. Even children have self-esteem, and once their self-esteem is hurt by the punishments imposed on them, they will not obey the external authority.

17

Italics added.

18

In "Justice as Reciprocity" (1971), Rawls says: "The principle of reciprocity requires of a practice that it satisfy those principles which the persons who participate in it could reasonably propose for mutual acceptance under the circumstances and conditions of the hypothetical account." (CP 208) Also, he argues that the principles of justice satisfy the principle of reciprocity while utilitarianism does not.

person in his later life. Once Emile acquired the returning sentiment of love, in reaching the adulthood, after he became interested in his surroundings (or after he began to compare himself with others around him), he acquired the first sentiment of friendship toward others. In this step, self-love becomes relative; *amour de soi* turns into *amour-propre*. But Emile is brought up “(1) To have a sense of the true relations of man, with respect to the species as well as the individual. (2) To order all the affections of the soul according to these relations” (E 219). Thus Emile’s *amour-propre* is not inflamed. He recognises that human beings have equal status and that they are not in the relations of subjugation and domination, alienation and exploitation. By recognising others as equals, Emile’s *amour-propre* extends to the love of human species or the love of justice.

For Rousseau, there are two sources of the sense of justice: the recognition of human beings as equals and the natural feelings of love and friendship. And the natural feelings of love and friendship are aroused in their heart because of their recognition of the manifest love of their parents and those around them. As Rousseau says:

“A child is therefore naturally inclined to benevolence, because he sees that everything approaching him is inclined to assist him; and from this observation he gets the habit of a sentiment favorable to his species.” (E 213)

Thus, not only is Rawls’s reciprocity thesis grounded in Rousseau’s reciprocal feature of *amour-propre*, but also Rawls’s explanation of how citizens in the well-ordered society of justice as fairness acquire the reciprocity of disposition, which inclines them to accept the principle of reciprocity or the principles of justice, is grounded in Rousseau’s explanation of how *amour-propre* can be extended to the sense of justice.

### 2.3. A Realistic Moral Psychology

Rawls suggests that justice as fairness has a more realistic moral psychology than other conceptions of justice, such as utilitarianism, perfectionism and intuitionism. The flaw of intuitionism is that it fails to describe the sense of justice as “psychologically understandable”. Rawls says that for intuitionism, the sense of justice is simply a desire for what is right, but the right is unanalysable. And thus, intuitionism fails to explain the sense of justice understandably. Rawls argues that utilitarianism and perfectionism can naturally interpret the sense of justice, and thus, they can describe the sense of justice as “psychologically understandable” (TJ 417). But he suggests that the love of humankind, which is the psychological foundation for utilitarianism and perfectionism, is supererogatory for human beings – utilitarianism’s (and perfectionism’s) moral psychology is the psychology of the saints (and the psychology of heroes), not realistic moral psychology for ordinary human beings. For Rawls, justice as fairness provides a more realistic moral psychology than other conceptions of justice.

Rawls’s realistic moral psychology parallels Rousseau’s thoughts on *amour-propre*. In particular, Rousseau argues that *amour-propre*, rather than compassion, is the primary motivating force for human actions. In the second *Discourse*, Rousseau suggests that even though *amour de soi* and compassion are the two principles of action in the state of nature, *amour de soi*, rather than compassion, is the primary motivating force. He also says that in the civil state, compassion is too weak to motivate a person to act compared with *amour-propre*. In *Emile*, Rousseau suggests that *amour-propre* has a natu-

ral and healthy form if guided by reason and that it is the starting point for justice. He argues that compassion plays a secondary, supporting role in the transformation of *amour-propre*, but it is not the starting point for justice. We have also seen that in the *Social Contract* and other of Rousseau's political works, he suggests that once *amour-propre* is guided by the general will (public reason), it will be the primary motivating force to form and retain the just society. Compassion, thus, is not the primary motivating force to stabilise the just society.

Michael Pritchard (1977) argues that Rawls's moral psychology faces a dilemma: either his principle of equal respect is simply based on the principle of reciprocity, as he claims it to be, in which case there is a gap between the principle of equal respect and the principle of reciprocity; or, his principle of equal respect also relies on the principle of sympathy, as utilitarianism does, in which case his objection against utilitarianism fails.<sup>19</sup> To explain the first horn of the dilemma, Pritchard argues that the principle of reciprocity cannot generate the sense of justice. In other words, he argues that moved by the principle of reciprocity, people will not have the sense of justice toward those who cannot contribute to the social cooperation of the well-ordered society.<sup>20</sup> For example, he says that the disabled are incapable of contributing to the cooperative scheme of the well-ordered society, and, thus, according to Rawls they do not deserve equal respect.<sup>21</sup> To explain the second horn of the dilemma, he argues that the principle of reciprocity cannot generate the principle of sympathy. He argues that the principle of reciprocity cannot generate the love of humankind, which, he maintains, is the rightful basis of the just society because it explains the equal respect for each and all.

But Pritchard fails to see the distinction between love and respect. Rawls points out that the love of humankind is supererogatory for ordinary human beings. As Rawls says, "morality [the love of humankind] is not one for ordinary persons, and its peculiar virtues are those of benevolence, a heightened sensitivity to the feelings and wants of others, and a proper humility and unconcern with self" (TJ 419). Thus it is unrealistic to think that all human beings have "a heightened sensitivity". Furthermore, the love of humankind

19

Cf. Michael S. Pritchard, "Rawls's Moral Psychology", *Southwestern Journal of Philosophy* 8 (1977) 1, pp. 59–72, doi: <https://doi.org/10.5840/swjphil1977817>.

20

For similar objections, see: Allen E. Buchanan, "Justice as Reciprocity versus Subject-centered Justice", *Philosophy and Public Affairs* 19 (1990) 3, pp. 228–252, pp. 228–230; Brian M. Barry, *Justice as Impartiality*, Oxford University Press, Oxford 1995, p. 31; Gerald Allan Cohen, *Self-Ownership, Freedom, and Equality*, Cambridge University Press, Cambridge 1995; Martha Nussbaum, *Hiding From Humanity: Disgust, Shame, and the Law*, Princeton University Press, Oxford 2006, p. 123, 224.

21

Pritchard says that "suppose that reliable methods were developed for determining which among newborns would create the

greatest problems for the society – e.g., those with serious intellectual or physical disabilities, those who will contract crippling or disabling diseases relatively early in life, or those who will end up being the dissenters in the society and try to bring down its institutions. If the society adopted the practice of removing such newborns from the society (either by infanticide or by sending them to less well-ordered societies), this would seem to be consistent with principles of reciprocity, assuming the newborns had made no significant contribution to the society". – M. S. Pritchard, "Rawls's Moral Psychology", p. 65. But on the one hand, Rawls might say that those with serious intellectual or physical disabilities can make significant contribution to the society; on the other hand, he might say that even though they cannot make contribution to the society due to the misfortune in the natural lottery, they still deserve equal respect.

is not necessary for a society of equal mutual respect. What is needed is the reciprocity of disposition.

Pritchard also fails to understand Rawls's principle of reciprocity fully. He fails to understand why the reciprocity of disposition will generate equal mutual respect. He thinks that the principle of reciprocity is the morality of exchange. On this reading, since those who are disabled do not contribute to the cooperative scheme of the well-ordered society, they are not the objects of equal respect. But Rawls never merely understands the reciprocity of disposition in this way. Indeed, in the first stage of moral development, the morality of authority may be the morality of exchange because the returning sentiments of love are contingent on the love they receive from their parents. But in the second stage, the morality of association may not be the morality of exchange; for we may still do our parts in the associations we belong to, even though we know that there are free-riders in the associations. And in the final stage of moral development, the morality of principles, the sense of justice is not contingent on the benefits they receive from others.

“Once a morality of principles is accepted, however, moral attitudes are no longer connected solely with the well-being and approval of particular individuals and groups, but are shaped by a conception of right chosen irrespective of these contingencies.” (TJ 416)

In the final stage of moral development, recognising that the just institutions are beneficial to them and those around them, citizens will have the sense of justice, provided that they have feelings of love and friendship in the first two stages. Their sense of justice is not toward those they love and befriend, for, if so, they would still be in the first two stages of moral development. The mark of their being in the final stage is that their sense of justice is not dependent on the benefits they receive from particular persons. Pritchard is wrong, therefore, to say that for Rawls those with a sense of justice will not have equal respect to those who cannot contribute to the cooperative scheme of the well-ordered society.

But perhaps Pritchard's question is how those who have a sense of justice may have equal respect with those who cannot contribute to the cooperative scheme of the well-ordered society. In other words, his question might be about how it is possible that human beings can proceed into the final stage of moral development. The empirical psychologies of Piaget and Kohlberg are also subject to the same question. But, unlike their empirical psychology, Rawls's moral psychology is not a science of human nature (PL 86). His reasonable moral psychology is compatible with human nature, and even more so than the moral psychologies of utilitarianism and perfectionism. Rawls does not need to show that his moral psychology is the most realistic one; he only needs to show that it is more realistic than the moral psychologies of other conceptions of justice. Furthermore, if Rawls's moral psychology is plausible, it will provide support for Piaget's and Kohlberg's empirical psychologies and refute the objections against their moral psychologies. It is simply too early to say that it is impossible that human beings can develop into the final stage of moral development. After all, we do not live in the just society that Rawls depicts as a realistic utopia. And that most of the people in an unjust society cannot develop into the final stage of moral development says nothing about the moral psychology of those who live in the just society. For Rawls and Rousseau, institutional factors have a significant impact on what human beings are like, and so we can reasonably expect that in the just society people will have the sense of justice in the final stage of moral development.

### 3. Conclusion and Some After-Thoughts

To conclude, I have argued that there is a Rousseauian interpretation of Rawls's moral psychology which constitutes the first part of his stability argument and that on this interpretation, some of the objections against his moral psychology can be answered. In response to these objections, I hope that I have shown that Rawls's moral psychology is Rousseauian in a deeper and better way.

However, on the one hand, one might argue that my interpretation does not represent Rawls as a distinctive thinker, a political liberal who is not like the republican Rousseau; On the other hand, one might argue that if there are Republican thoughts in Rawls, it does not show Rawls's moral psychology is deeper or better.

In response to this objection, while I acknowledge that there are Republican thoughts in Rawls, he is still a distinctive thinker, a political liberal. In his *Political Liberalism*, he argues that the stability of a well-ordered society is secured by the overlapping consensus among different comprehensive doctrines. And the overlapping consensus cannot be achieved if citizens are not reasonable. My interpretation tries to lend support to Rawls's idea of reasonableness. In making an analogy between Rousseau and Rawls's moral psychology, I try to show how Rousseau's theory of moral education in *Emile* influence Rawls. Rousseau and Rawls's moral psychology see human beings as having the capacity of perfectibility. By going through several developmental stages, they can have a sense of justice, which makes Rawls's political conception of justice possible. This realistic moral psychology, as Rawls says, makes his conception of justice more defensible than other conceptions of justice.

Xinghua Wang

#### Rousseauovsko nasljeđe u Rawlsovoj psihologiji morala

##### Sažetak

*Rousseauov utjecaj na Rawlsov izvoran argument za pravdu kao poštenje neki su autori primijetili, ali malo je njih raspravljalo o Rousseauovu utjecaju na Rawlsovu psihologiju morala koja sačinjava prvi dio njegova argumenta o stabilnosti. Argumentirat ću da Rawlsov iskaz o moralnom razvoju podudara i zasniva se na Rousseauovu konceptu amour-propre. Napose tvrdim da je (1) Rawlsova teza da osjećaj za pravdu proizlazi iz ljubavi i prijateljstva prikaz Rousseauove teze da moralni sentimentali proizlaze iz prirodnih sentimentata, da se (2) Rawlsovo objašnjenje o tome kako zadobivamo reciprocitet dispozicija zasniva na Rousseauovu objašnjenju kako amour-propre može biti produžen u smisao pravde, te da (3) Rawlsova misao o principu simpatije odgovara Rousseauovim mislima o suosjećajnosti.*

##### Ključne riječi

Jean-Jacques Rousseau, *amour-propre*, John Rawls, psihologija morala, smisao za pravdu

Xinghua Wang

#### Rousseausches Erbe in Rawls' Moralphysikologie

##### Zusammenfassung

*Rousseaus Einflussnahme auf Rawls' ursprüngliches Argument für Gerechtigkeit als Fairness wurde zwar von einigen Autoren bemerkt, jedoch nur wenige haben sich für Rousseaus Einfluss auf Rawls' Moralphysikologie ausgesprochen, die den ersten Teil seines Stabilitätsarguments*

*ausmacht. Ich werde substantzieren, dass Rawls' Bericht über die moralische Entwicklung mit Rousseaus Gedanken über die amour-propre übereinstimmt und darauf aufbaut. Insbesondere argumentiere ich, dass (1) Rawls' These, der Sinn für Gerechtigkeit rühre von Liebe und Freundschaft her, eine Illustration von Rousseaus These ist, dass moralische Empfindungen auf natürlichen Empfindungen beruhen, dass (2) Rawls' Auslegung, wie wir die Reziprozität der Disposition erlangen, auf Rousseaus Erklärung basiert, wie amour-propre auf den Gerechtigkeitsinn ausgedehnt werden kann und dass (3) Rawls' Gedanken zum Prinzip der Sympathie mit Rousseaus Gedanken zum Mitgefühl korrespondieren.*

**Schlüsselwörter**

Jean-Jacques Rousseau, *amour-propre*, John Rawls, Moralpsychologie, Gerechtigkeitsinn

**Xinghua Wang**

**Le patrimoine de Rousseau dans la psychologie morale de Rawls**

**Résumé**

*Certains auteurs ont remarqué l'influence de Rousseau sur l'argument de la position originelle de Rawls en faveur de la justice comme équité, mais peu ont argumenté en faveur de l'influence de Rousseau sur sa psychologie morale, qui constitue la première partie de son argument de la stabilité. Je soutiendrai que le compte-rendu de Rawls sur le développement moral est un parallèle et il se fonde sur les réflexions de Rousseau sur l'amour-propre. En particulier, je soutiens que (1) la thèse de Rawls selon laquelle le sens de la justice procède de l'amour et de l'amitié est une illustration de la thèse de Rousseau selon laquelle les sentiments moraux procèdent de sentiments naturels, que (2) l'explication de Rawls sur la manière d'acquérir la réciprocité des dispositions est fondée sur l'explication de Rousseau sur la manière dont l'amour-propre peut être étendu au sens de la justice et que (3) les réflexions de Rawls sur le principe de sympathie correspondent aux pensées de Rousseau sur la compassion.*

**Mots-clés**

Jean-Jacques Rousseau, *amour-propre*, John Rawls, psychologie morale, sens de la justice