

# A New Approach of Wide Reflective Equilibrium-from the Perspective of Morality as Cooperation

**Yanzhou Chen**<sup>1, \*</sup>

<sup>1</sup>YKPao School, Shanghai, China

\*Corresponding author:  
centianzhou@ldy.edu.rs

## Abstract:

This review discussed a possible combination of Justice as Fairness and Morality as Cooperation. The concept of Justice as Fairness, which was introduced by the renowned philosopher John Rawls, advocates for the use of wide reflective equilibrium as a tool to derive and establish the fundamental principles that should govern a just society. In this process, Rawls suggests that we should not only consider our own moral intuitions but also engage with the perspectives and reasoning of others, including various philosophical theories and arguments. The theory of Justice as Fairness, with its emphasis on wide reflective equilibrium, thus provides a structured yet flexible approach to moral and political philosophy. There are, however, sociological and epistemic difficulties in implementing its social blueprint, thereby giving evolutionary psychology and game theory chances to participate in formulating justice, as Oliver Curry had proposed in his theory of Morality as Cooperation. This review relates Curry's moral molecule model and evolutionary game theory as substitutes for Rawls' Fairness principles and original position in wide reflective equilibrium. It justifies their practical advantage, and by replacing the disadvantaged concepts with their counterparts, this review aims to produce a method with the characterization of being instinctively highly accepted with simpler methodologies; thus, this review looks to inspire future directions of moral psychology and philosophy.

**Keywords:** Justice as Fairness; wide reflective equilibrium; morality as cooperation; evolutionary Psychology.

## 1. Introduction

### 1.1 Importance of Wide Reflective Equilib-

### rium

Wide Reflective Equilibrium (wRE) is a philosophical justification method created by John Rawls and is useful when assessing certain moral intuitions. This theory is renowned as the prevalent approach in

conversations in the context of politics and ethical philosophy [1, 2]. wRE is a derived concept from the social contractarian belief of “Justice as fairness (JAF)” proposed in *A Theory of Justice* [3]. Thought as one of the most important philosophies of the 20th century, JAF proposes a social system that endorses fairness as the sole value important in society, that ethical improvement would be achieved only through broader recognition of this value [4]. This review looks to improve this social blueprint by extending wRE to a broader means and, therefore, improving the justice it produces.

## 1.2 Missing criteria: beyond the veil of ignorance

Rawls has argued that stability and efficiency are only good when it is just [3]—when it is fair and justified by wRE—but had few considerations on the difficulty implementing “Justice” as a society caused by the conflict of his doctrines with human nature. He also faced problems making “the original position” realistic, as intuitions had long been considered divergent, and this relativism made legislation nearly impossible.

This article aims to extend the theory from a new perspective of Morality as Cooperation (MAC), perfecting his theory, adding an incentive to obedience, and providing tangible justification tools, coping with more complications in decision-making.

## 1.3 Connecting MAC to JAF: a perfection

Another purpose of this review is to connect MAC with JAF. MAC is a theory developed by Oliver Curry, who took Naturalists’ perspectives further and proposes that humans are naturally social animals adapted for it and show a tendency to cooperate by which they modulate intuitions and values to facilitate novel forms of cooperation [5]. Curry made the following logic: Naturalists such as Darwin argued that humans are bundles of adaptations that ensure the survival of human beings and that reason and belief are legacies of evolution, and for this purpose, people generate thoughts and intuitions for the common good, which would betray their selfishness and create cooperation altruistically under certain conditions, and thus he argues maximum mutual benefit (or cooperation) is the natural moral goal of human beings, and the most cooperative decision is, therefore, the most moral decision. Curry also endorsed Game theory as the meter for how cooperative a thing is, contrasting the intuition-reliant theory with Rawls [5, 6].

The extension of Curry to Rawls involves arguing for more aspects of Justice, decided by evolutionary psychology, than fairness; he invented seven types of cooperation:

Kinship, Mutualism, Reciprocity, Hawk, Dove, Possession, and Division (ranked in no particular order) [7]. Furthermore, he argues that a combination of these types of cooperation attached to different weights would lead to infinitely many scenarios that would account for every situation possible on earth and each situation available for calculations of Game Theory. Curry proposes that in the future, it is possible that Game Theory will develop into a state where it can calculate the effectiveness of such decision-making, and that would then be a universal truth for ethics [5].

It is not the ambition today to completely facilitate Justice as legislation or to universalize it to the world. The only purpose is to extend these important works of Rawls with the psychology and science basics of MAC, amending its shortcomings in humanism to a more applicable state to the general public and to provide a more readily available future methodology for deciding true Justice.

## 2. Method

The literature review is based on wRE, JAF, and MAC. The reference materials first pursue first-hand works from related key scholars, such as John Rawls and Oliver Curry, then resort to databases Jstor and Google Scholar. A few materials also come from the Stanford Encyclopedia of Philosophy. Keywords used in the general search include wide reflective equilibrium, justice as fairness, morality as cooperation, A Theory of Justice, Moral foundation theory, free will, individual autonomy, liberty, obligation, and duty.

## 3. Results

### 3.1 Theoretical Frameworks of JAF and MAC

#### 3.1.2 JAF and wRE

The theory of JAF was established by John Rawls, a Harvard faculty member for more than 30 years. As a survivor of World War Two, the Vietnam War made Rawls “prosecute so ruthlessly what he saw as an unjust war” and triggered his thoughts on justice in society and justifiable government actions [8]. He condensed these beliefs into his 1971 book *A Theory of Justice* [3] and later publications such as *The Law of Peoples* [9]. They are concluded in his final book, *Justice as Fairness*, written in 2001, the year before his death.

Concerning wRE, the starting point is a number of “considered judgments,” or convictions, as referred to by Rawls [3]. These are judgments we have confidence in and are empirically and logically justified. After this process, peo-

ple must filter and reflect on moral descriptions containing all the related philosophical principles and arguments for or against them [10]. Finally, people would need to consider the “devices for justification,” most importantly, be in their original position, a device that Rawls used to ensure fairness in wRE. He proposes that if people do not acknowledge their social status, they are freed from the constrictions of their interests. Therefore, their judgments are fair. Rawls argued, then, that objective justice is the most accepted concept of justice (given a differentiation between concept and conception of justice) by individuals in the original position [3].

### 3.1.2 MAC

The theory of MAC comes from a disagreement with the Moral Foundations Theory (MFT). MFT is an account of morality, much like virtue ethics. Jonathan Heidt argues that cooperative social existence is made possible by moral systems, which are interconnected “sets of values, virtues, norms, practices, identities, institutions, technologies,” and evolutionary psychological mechanisms that restrict or govern selfishness, specifically, foundations including Care, Fairness, Loyalty, Authority, and Purity [11, 12]. Curry rebukes MFT by challenging two Moral Foundations: (1) Care. Care is a foundation that consists of too much complexity and massive variation of the reason to care and the divergent forms of care; this ambiguity casts a shadow on whether care is purely cooperative. (2) Purity. Purity is a foundation that consists of no cooperation. Maintaining one’s own purity involves a sole individual. Therefore, it is not a corporation [13]. Curry also argued that there are empirical flaws in MFT’s key empirical studies; some of the data gathered by the Moral Foundations Questionnaire denies the five-factor model of MFT [14]. By contrast, Curry did his empirical research on better strategies.

## 3.2 Comparison of MAC and JAF

The principal difference between MAC and JAF comes from their aim in means-end reasoning: Rawls focused on answering the question, “What reasonable conditions should ‘free and equal citizens’ have for social cooperation?” [8]. The original position is often argued as a device that helps to answer this question by experimenting: “What conditions would ‘free and equal’ agree?” Curry takes a different path; he questions what—as a human race—people would want from an evolutionary basis, seeking the means of cooperation that benefits this aim.

The contrasting principles deduce operational differences in the judging process for MAC and JAF. More concerned with fairness, JAF emphasizes its original position, which cuts vested interests of decision makers, as a trusted de-

vice for justification; MAC focused merely on humans’ evolutionary trend, marking the importance of mutual benefit, relies on increasing total cooperation, and remarked game theory with seven non-zero-sum games as its sole protocol (as stated in 1.3).

### 3.2.1 Limitations of JAF

JAF potentially faces severe problems in practical implementations. However, Rawls argued that people could conduct wRE in a small, independent society and then apply the doctrines in other dealings of justice; he ignored the cost of universalising formulated to a bigger and more complex society [3]. Rawls is aware that he is assuming that society being “well-ordered” [3]. Moreover, in a more common saying, his Justice “are not everywhere satisfactory” [3]. This review would later follow this path to discuss a more comprehensive and applicable extension of his theory.

In addition, Rawls had few guidelines on how to conduct the ideal wRE practically. First, the original position originated from the “state of nature,” which, as admitted by Kant, is hypothetical [3, 15], and Rawls is fully aware of this situation. The ambiguity made Rawls concede that the “course of reflection itself” would lead to many equilibriums [3], causes could take from a list of factors, such as framing effect [16], the emotional tendency caused by diction or language, emotions, such as the deviation of footbridge version of the trolley problem and its original version [17]. Such objections made reflective equilibrium to be described as only “a potential source of justification” [4].

### 3.2.2 MAC provides features perfection to JAF

On actual implementations, psychology and game theory-based deliberation do a better job than JAF on social recognition, MAC shows. First, evolutionary psychology has made MAC more acceptable through its close relation to human evolution, which directly affects our cognitive behavior and learning process and indirectly affects our behavior by affecting its determinant norms and education prevalent in societies [18-20]. That said, it is likely more naturally accepted than JAF, which is only considering individuals in the original position. Second, on an empirical basis, Curry had done research to justify the psychological acceptance of MAC [3]. Curry described his study, which was conducted in 60 societies randomly distributed around the globe, as an anthropology-based moral valency test with the societies’ ethnographic records. The results entailed that there was a uniform positive agreement with his theory, which led to his belief that MAC has cross-culture recognition and that it is universally applicable through legislation. To enhance the conclusion, Curry

conducted another study with Natural Language Processing (NLP) algorithms, analyzed using the means of Linguistic Inquiry and Word Count, with a bigger sample of 196 societies recorded in the Human Relations Area Files [21]. Through his findings, we can locate the advantage of MAC over JAF: Though Rawls' justification is sound in a coherent sense, the product of wRE has only a scientific sample number of one, or it is an individualist. This does not incentivize the general public to obey the doctrines deduced. By contrast, MAC (as a delicate combination of psychology and game theory) is not only theoretically and empirically more accepted, as shown above, but also generates maximum profit based on scientific calculations. This theory naturally incentivizes humans to act accordingly and face less pressure as a social justice.

MAC is a lot more tangible in practice than JAF. The studies on moral molecules demonstrate its vast account for moral values, and features of game theory facilitate empirical research on morality. In his 2021 article [22], Curry proposed that morality could be combinatorial in the periodic table in natural sciences. With his seven types of cooperation as basic elements combined in different proportions, there are infinite possible "moral molecules" that account for any situation on earth. To test the claim, Curry reviewed various literature to seek examples of primary molecules (with two elements combined only) and succeeded in every case [22]. This discovery derives two conclusions from this theory: (1) all possible moral values can be predicted; (2) all actual moral values can be analyzed as a moral molecule. The key benefit of both features is that it facilitates a scientific study of morality: as moral values are ways of cooperation in a non-zero-sum game, the best balance of the elements is an empirical question, potentially a ground of study for game theory, which is a readily available subject yet in progress [23]. This benefit outruns JAF, which fails to provide basics for posterior judgment in actual implementations and demonstrates moral molecules as a reliable model for morality or, at least, seeking justice.

## 4. Implications

Finally, is there a way to assimilate MAC with wRE? Two key elements of MAC are discussed above: the moral molecule theory and the utilization of game theory. Is it possible to replace JAF and original position with the more beneficial theory and formulate another path to justice principles while keeping the coherence of judgments and principles?

A potential path to achieving this is similar to wRE. The collection of considered judgments should be kept, and the careful logical validation of the judgment should still

be on the bench. To further analyze and consider judgments, they must be carefully positioned into moral molecules, identifying the elements they consists of. Then, one must consider arguments for and against the judgments, reflect on the judgment and arguments, and find the moral elements common to both sides; similar to the process of wRE, this step confirms the harms or elements of morality suppressed the judgment. Game theory would now help predict the most mutually beneficial balance of these elements and formulate a deviated moral molecule that ought to be accepted as part of justice.

At this point, the review argued for the benefits of game theory and the combinatorial system's replacement of JAF and its original position. The problems is that claims still need empirical justification, and the analysis would also have to extend to various stakeholders. Though this article is aimed at making wRE more practical, more research is still anticipated to facilitate the suggested method. For example, to what extent can we break down judgments into moral elements, and what scaler standard should we create to measure the proportion of elements in a moral molecule? To bring the theory further, scientists must develop a systematic branch of evolutionary game theory that provides an account for MAC. In the end, this review would have done its job if it had proved the existence of ways to perfect moral theories through scientific research.

## 5. Conclusion

In conclusion, this review looked into the theories of JAF and MAC. It located potential improvements needed in wRE when seeking justice as a society. MAC was proven to provide amendments to this limitation by considering people's psychology, thus attracting natural supporters of the generated justice and facilitating scientific research on morality by constructing the model of moral molecules. Nevertheless, this review lacks direct empirical support and anticipates quantification of the moral molecule model soon. Overall, this review considers the existence of a potentially better version of wRE awaiting development in evolutionary game theory and provides thinking for future ethics.

## References

- [1] McPherson, Tristram. *The methodological irrelevance of reflective equilibrium*. Palgrave Macmillan UK eBooks, 2015, 652–674.
- [2] De Maagt, Sem. "Reflective equilibrium and moral objectivity." *Inquiry*, 2016m 60(5): 443–465.
- [3] Harvard College, editor. *A Theory of Justice*. Harvard UP, 1999.

- [4] Knight. Reflective Equilibrium.” Stanford Encyclopedia of Philosophy/Winter 2023 Edition, Metaphysics Research Lab, Stanford University, 2023.
- [5] Curry, O. S. Morality as Natural History: An adaptationist account of ethics. (PhD), London School of Economics, London. 2005.
- [6] Curry, O. S. Morality as cooperation: A problem-centred approach. In T. K. Shackelford & R. D. Hansen (Eds.), *The evolution of morality* (pp. 27–51). Springer International Publishing AG. 2016.
- [7] Curry, Oliver Scott, et al. Is It Good to Cooperate? Testing the Theory of Morality-as-Cooperation in 60 Societies. *Current Anthropology*, 2019, 60(1): 47–69.
- [8] Wenar, Leif. John Rawls. *The Stanford Encyclopedia of Philosophy/Summer 2021 Edition*, Edward N. Zalta (ed.) 2021.
- [9] Rawls, J. The law of peoples. *Critical inquiry*, 1993, 20(1): 36-68.
- [10] Daniels, Norman. Justice and Justification: Reflective Equilibrium in Theory and Practice.” *The Philosophical Review*, 1998, 107(3): 33-39.
- [11] Haidt, J. & Kesebir, S. *Handbook of Social Psychology*. Chapter 22, Morality, 2010.
- [12] Haidt, J. Moral psychology for the twenty-first century. *Journal of Moral Education*, 2013, 42(3): 281–97.
- [13] Curry. What’s Wrong with Moral Foundations Theory, and How to get Moral Psychology Right. *Behavioral Scientist*. *Behavioral Scientist*, 27, 2019.
- [14] Graham, Jesse, et al. Mapping the moral domain. *Journal of Personality and Social Psychology*, 2011, 101(2): 366–385.
- [15] Kant’s Political Writings. The UP, 1970, 73-81.
- [16] List, Christian, & Laura Valentini. *The Methodology of Political Theory*. The Oxford Handbook of Philosophical Methodology, 2016, 525–553.
- [17] Greene, Joshua D., et al. An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, 2001, 293(5537): 2105–2108.
- [18] Boyd, R., Richerson, P. J., & Henrich, J. The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108: 10918–10925.
- [19] Pinker, S. The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107: 8993–8999.
- [20] Boyd, R., and P. J. Richerson. Culture and the Evolution of the Human social Instincts. *Roots of Human Sociality*, 2020, 453–77.
- [21] Alfano, Mark, et al. Moral Universals: A Machine-reading Analysis of 256 Societies. *Heliyon*, 2024, 10(6): e25940.
- [22] Curry, O. S., Alfano, M., Brandt, M. J., & Pelican, C. Moral molecules: Morality as a combinatorial system. *Review of Philosophy and Psychology*, 2022, 13(4): 1039-1058.
- [23] Traulsen, A., & Glynatsi, N. E. The future of theoretical evolutionary game theory. *Philosophical Transactions of the Royal Society B*, 2021, 378(1876): 20210508.