Constitutive rules as a logical problem in the origin of John Rawls' moral philosophy

Pablo Aguayo-Westwood (paguayo@derecho.uchile.cl) Faculty of Law, University of Chile (Santiago, Chile) https://orcid.org/0000-0003-3239-5441 Role: Conceptualization, Writing original draft.

Abstract

This article offers a critical review of John Rawls's 1955 essay Two Concepts of Rules and argues that a full understanding of his mature work requires examining his early philosophical development. The central hypothesis posits that Rawls's analysis of constitutive rules is essential for understanding the epistemic foundations of his later theory of justice. A key distinction is that drawn between justifying an action that falls within a rule (or practice) and justifying the rule itself. This distinction allowed Rawls to defend utilitarianism against its critics, highlighting their logical misunderstanding of how this theory evaluates actions. Furthermore, the article argues that Rawls's early concern with practices and social institutions directly influenced his subsequent interest in the basic structure of society, developed in A Theory of Justice. Ultimately, it reveals Rawls's nascent attention to the evaluation of systems of rules that shape social practices, rather than focusing solely on individual actions, representing a crucial step in the elaboration of his theory.

Keywords: Rawls, rules, social institutions, utilitarianism.

Introduction

It can be affirmed that understanding the mature work of most philosophers requires a critical review not only of the context in which their works appeared but also of the philosophical development reflected in their minor works. Just to give an example, the understanding of Kant's mature work is greatly enriched by a careful reading of his pre-critical works, which provide interesting conceptual insights for addressing the notions of space and time at the beginning of the transcendental aesthetic. This working hypothesis motivates the review of the article "Two Concepts of Rules" (hereinafter TCR) published by Rawls in 1955. This review aims to show the extent to which ideas such as institutions and social practices have their origin almost two decades before the publication of *A Theory of Justice* (hereinafter TJ).

It should be noted that the drafting of TCR was strongly influenced by Rawls's friendship with the Oxonian philosopher J. O. Urmson during the latter's stay at Princeton University during the years 1950-51. In this context, Thomas Pogge maintained that J. O. Urmson was the one who allowed Rawls to get to know "Oxford's most important philosophers" (Pogge 2007:16). Following Urmson's advice, Rawls applied for a Fulbright scholarship with which he was able to spend the academic year 1952-1953 at Christ Church College, Oxford, a city where he met experts in the utilitarian tradition such as the analytical legal philosopher H. L. A. Hart, and philosophers like Philippa Foot and Elizabeth Anscombe. In systematic terms, the analysis of the ideas Rawls offered in TCR allows us to understand the influence of methodological and epistemic issues on the development of his theory

of justice. In this sense, the hypothesis that will be discussed here is that the analysis of the notion of constitutive rules is central to understanding the epistemic foundations of the author's mature proposal.

The objectives of TCR

One of the main objectives Rawls set himself in TJ was to confront what had been, for much of modern moral philosophy, the predominant systematic theory, namely, utilitarianism. Since the publication of "Two Concepts of Rules", Rawls was aware that utilitarianism implied "a coherent view of society, and is not simply an ethical theory, much less an attempt at philosophical analysis in the modern sense" (Rawls 1955:19), in other words, he knew that confronting the main theses of utilitarianism implied recognising that this way of conceiving moral reflection transcended the evaluation of particular actions. This led him to continue his original project aimed at finding a reasonable way to validate moral rules for the correctness of our actions, an enterprise he had begun in his doctoral dissertation and developed in his "Outline of a Decision Procedure for Ethics" of 1951.

It should be noted that Rawls's concern in TCR is not only the possibility of finding a criterion for determining the moral value of particular actions, but rather the logical evaluation of certain rules that constitute practices of a general nature. Using the categories of TCR, we can maintain that the shift indicated above involves differentiating between *justifying an action that falls within a rule* (or practice) and the *justification of the rule itself*. Although this distinction is used throughout the essay to show the impropriety of certain criticisms of utilitarianism, particularly those by E. F. Carritt and G. E. Moore, the central issue that interests me is the identification of a field of discussion in moral philosophy focused on the justification of social practices, rather than on the determination of criteria for evaluating particular cases. Rawls's central thesis in TCR consists of demonstrating the inability of certain critics to observe the logical differences between justifying a rule and justifying a case within it; in other words, the criticism involves denouncing enumerative induction as a criterion for determining moral rules (Hamlin. The rule of rules). This transition from a concern with determining a criterion for valuing individual actions to determining criteria for judging social practices will accompany him until the elaboration of the central theses of TJ.

As I will try to show later, I believe that his concern for social institutions, as well as for the basic structure of society, are heirs to this change of perspective.

Analysis of the central ideas of TCR

Rawls's main objective in TCR was "to show the importance of distinguishing between justifying a practice and justifying a particular action that falls under it" (Rawls 1955:3). Once the logical bases of such a distinction were explained, Rawls defended utilitarianism from some objections that could be resolved from it. The central objections he faced concern the moral value of punishment and the keeping of promises. In the first part of the article (section I), Rawls discusses the problem of the justification of punishment and maintains that "various arguments for it have been given by moral philosophers, but so far none of them has won any sort of general acceptance; no justification is without those who detest it" (Rawls 1955:4). In this section, he critically analyses two ways of approaching the topic, namely, the retributive point of view and the utilitarian point of view.

From the retributive point of view, punishment would be justified on the basis that wrongdoings deserve punishment. For this conception, a person who has committed a wrongdoing must suffer in proportion to the wrong committed. Unlike this point of view, the utilitarian conception holds that punishment is justifiable by virtue of the favourable consequences of its application for the maintenance of social order. In this way, punishment would only be justifiable if it effectively promotes the interest of society. In this work, Rawls is not interested in the philosophical discussion about the morality of punishment, nor in issues related to human dignity and freedom. His objective is rather metaethical: it consists of examining the logical and epistemic bases of the distinction between the justification of a practice and the justification of an action that falls under it.

Given the above, the debate between retributivists and utilitarians is appropriate insofar as it allows us to observe that utilitarian arguments are valid (or significant) in relation to the justification of certain practices, while retributive arguments are confined to the application of particular cases. With this, Rawls uses a discussion about utilitarianism for the treatment of a second-order issue that is more significant to him, namely, the justification of social practices.

In TCR, Rawls maintains that a practice is any form of activity specified by a system of rules. He gives examples of games, rituals, and parliamentary trials; referring to them, he points out that they can be understood as a kind of technical term meaning "any form of activity specified by a system of rules which defines offices, roles, moves, penalties, defenses, and so on, and which gives the activity its structure. As examples one may think of games and rituals, trials and parliaments" (Rawls 1955:3, note 1). As is well known, the importance of the notion of practice in Rawls's mature philosophy is fundamental. The elaboration of ideas such as social institutions, as well as the basic structure of society demonstrate this. One need only consider that in his reflection on social institutions in TJ, Rawls uses the same examples he had offered almost twenty years earlier, as when he states: "As examples of institutions, or more generally social practices, we may think of games and rituals, trials and parliaments, markets and systems of property" (Rawls 1971:55).

Finally, in section III of TCR, Rawls examines the central issue of his work, that is, the two ways of understanding the notion of a rule. I will now review the central elements of this part of the article. Then I will establish some conclusions related to his way of understanding moral philosophy.

The summary conception of rules

The summary conception of rules reveals a tension that characterises much of his philosophical reflection. This tension is determined by two elements or ways of approaching philosophical questions that Rawls permanently combines. In the first place, there is the historical dimension. This involves defending utilitarianism (as an ethical theory that responds to problems of social interest) from some of its critics. In this sense, the discussion is not about mere conceptual definitions or problems of analytical metaphysics seeking the meaning of the notion of utility. Rawls is interested in utilitarianism as a doctrine that is part of a historical tradition that has reflected on morality. Secondly, there is the logical or epistemic dimension of the treatment of philosophical questions. Let us not forget that Rawls insists in several parts of TCR that the error made by some critics of utilitarianism is "misconceiving the logical status of the rules of practices" (Rawls 1955:19). In this sense, critics like Carritt or Moore would commit a *logical* error when understanding the way in which utilitarianism evaluates actions. For Rawls, critics of utilitarianism would commit a procedural error, an error in the way of understanding the forms of conceiving and applying (moral) rules.

Returning to the summary conception of rules, Rawls maintains that this way of understanding them assumes that each person could decide what to do in each case based on the application of the principle of utility. It is also assumed that different people would decide a particular case in the same way. In strictly logical terms, rules would be reached through inductive generalisation and would be applied in a deductive manner; only in this way would the possibility of obtaining the same results in different situations be comprehensible. It should be noted that Rawls does not have a favourable attitude towards deductivist conceptions within ethics, both concerning the justification of its principles and at the level of their application. So, what procedure should we adopt in ethics? And how should we understand that procedure? We know that in his early works of the 1950s, Rawls conceived of ethics as a discipline capable of offering a procedure for resolving moral conflicts. From this perspective, the summary view would in principle fit this parameter. The above would be justified insofar as this conception offers a procedure for the moral evaluation of particular cases in which rules are conceived as *summaries* of past decisions, rules that would be arrived at by applying the principle of utility. The scheme could be summarised in four parts:

- I have the principle of utility
- I apply that principle to particular cases
- I generalise that application (its results) and obtain a rule
- I use that rule as a framework for moral guidance for the evaluation of other cases

While it is true that the summary view of rules fits, in principle, the formal requirements of a procedure for ethics, for Rawls, such a procedure is insufficient. The author believes that this would not only be due to the logical problems underlying an understanding of rules as inductive generalisations but, above all, because it confuses the meaning of a moral rule with the notion of moral maxims or rules of thumb. In this sense, anyone could doubt the validity of the procedure in obtaining the rule that guides an area of action and, therefore, decide to use the principle of utility in each case. Thus, Rawls maintains: “On this view a society of rational utilitarians would be a society without rules in which each person applied the utilitarian principle directly and smoothly, and without error, case by case” (Rawls 1955:19). For him, this interpretation of rules distorts the central core of utilitarianism, understood as a doctrine that aspires to offer a criterion for defining social rules of behaviour. It is for this reason that Rawls proposes to advance in another way in the understanding of rules, a way that will not only offer fruits to defend utilitarianism from its critics but also to build his own notion of moral philosophy. I then proceed to critically examine the practice conception of rules.

The practice conception of rules

The central idea of the practice conception of rules is that they are not generalisations based on decisions in which the principle of utility is applied directly to particular cases. On the contrary, rules define and constitute a practice in themselves, and it is the rules themselves that are subject to the utilitarian principle. From the foregoing, it follows that under this way of understanding rules, it is not possible to conceive of cases prior to their stipulation, since it is the rules themselves that give rise to the possibility of cases of that rule existing. For example, there could not be an *offside* case without first having stipulated *the rules* of football, nor a *castling* case before defining *the rules* of chess. In the case of actions specified by certain practices, it is logically impossible to represent them outside the framework provided by those practices, because unless the practice exists, and unless the properties required by it are met, the actions will cease to count as such without this framework.

For Rawls: “The practice is logically prior to particular cases: unless there is the practice the terms referring to actions specified by it lack a sense” (Rawls 1955:25).

Recall that in the summary view, rules were obtained by summaries of cases. In contrast, under the practice conception of rules, the situation is reversed, not only in a logical-procedural sense but also in its philosophical background. Using Peirce’s categories, we could maintain that, for the practical conception, either cases would be instances of general rules, or *tokens* would be treated as *types*. In this sense, the task of the moral philosopher is the establishment and evaluation of social rules for their proper development. In other words, the moral philosopher would be responsible for the elaboration, evaluation, and organisation of a system of rules to guarantee the stability of society or, as Hage maintained, an assessment of the ontological priority of constitutive rules (Hage. Two concepts of constitutive rules). While it is true that Rawls argued in TCR that the criterion for the moral evaluation of these rules should be the principle of utility (an idea he would later abandon), what is relevant regarding the path his thought follows is the identification of this sphere of reflection.

Conclusion

In this work I have shown the relevance that TCR had for the formulation and significance of the idea of practice and its relationship with the justification of constitutive rules. In this sense, the central issue in TCR is the discovery of a field of moral reflection linked to systems of rules that define certain practices. While it is true that Rawls considers the distinction between justifying a rule and justifying an action that falls under it as a distinction of a logical nature, it is possible to maintain that this distinction transcends this formal framework. It is about considering the question that asks about the possibility of evaluating not only particular actions but also the ways in which a society organises itself. It seeks, to put it in Rawls’s mature language, criteria that guarantee social stability and allow us to achieve a well-ordered society.

It is for the reasons explained above that Rawls does not end his article once he has responded to the criticisms of utilitarianism. Although the author expresses an intuition that there is something more to be gained from the distinction outlined, section IV and the final part are too brief to delve into it. Rawls only manages to outline the importance of rules understood not only as mere generalisations but rather as human creations intended for the better functioning of society. This implies an important conclusion in the field of political liberalism that the author will discuss in the following decades, namely, that the capacity of particular agents to determine the morality of actions would be diminished under a summary conception of rules. The reason for this is because the principles of utilitarianism are no longer conceived as a power inherent to each subject to deliberate with their conscience about what the best decision to follow should be. In Rawls’s words, moral subjects would lose the fullness of moral authority that would characterise this summary view of rules. In contrast, under Rawls’s proposal, the principles of utilitarianism are conceived rather as criteria for evaluating rules that are part of the public life of moral agents, constitutive rules of certain practices. The point here is to show the need for the establishment of rules that are publicly known and understood as definitive. These rules would be constructed by moral subjects and would be available for evaluation in the public sphere. In this sense, the idea of social agreement achieved through constitutive rules that would serve as a framework for determining practices appears here as a seminal idea for his later works. Years later, Rawls would even reject this way of understanding the moral justification of rules. In fact, already in the first version of his article "Justice as Fairness"

this rejection becomes manifest. Slowly, this critical position towards utilitarianism would articulate, step by step, the central ideas of *A Theory of Justice*.

However, and despite the abandonment of utilitarian theses, two central notions explored in TCR will remain in his proposal: firstly, *the notion of practice* as "a sort of technical term" (Rawls 1995:3, note 1) capable of situating us at a level of philosophical reflection on general social issues (and not only the evaluation of particular cases as in the understanding of *act-utilitarianism*); secondly, his concern for *moral justification* will also be part of the philosophical framework with which he is to construct his theory of justice. It is this concern for the moral justification of rules that will guide his search for criteria to ensure the fairness of the principles of justice.

References

Pogge, T. (2007). *John Rawls. His life and theory of justice*. Oxford University Press.

Rawls, J. (1955). Two concepts of rules. *The Philosophical Review*, *64*(1), 3-32. https://doi.org/10.2307/2182230

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Received 28 May 2025
Accepted 1 Jul 2025