RATIONAL Realm

Exploring a rational approach to knowledge and life

John Rawls: Two Concepts of Rules

Leslie Allan

Published online: 6 December 2020

Copyright © 2020 Leslie Allan

In his seminal essay, ‘Two Concepts of Rules’, John Rawls defends utilitarianism against a number of key objections. Using the examples of unjust punishment and promise-breaking, he draws a central distinction between justifying a practice and justifying a particular action falling under it. In this essay, Leslie Allan walks through Rawls’s essay, highlighting his key arguments for a strengthened version of rule utilitarianism and reflecting on the lasting influence of his analysis.

To cite this essay:

Allan, Leslie 2020. John Rawls: Two Concepts of Rules, URL = <https://www.RationalRealm.com/philosophy/reviews/john-rawls-two-concepts-of-rules.html>

To link to this essay:

www.RationalRealm.com/philosophy/reviews/john-rawls-two-concepts-of-rules.html

Follow this and additional essays at: www.RationalRealm.com

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sublicensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms and conditions of access and use can be found at www.RationalRealm.com/policies/tos.html

1. Introduction

John Rawls was a leading moral and political philosopher in the latter part of the twentieth century. Before his death in 2002, he left behind key writings that remain highly influential to this day, including his major work, *A Theory of Justice*. An early essay of Rawls that had a profound impact on my own thinking about ethics is his 1955 piece, 'Two Concepts of Rules'. In this essay, Rawls demonstrates how utilitarianism can be defended from some debilitating objections. Rawls, himself, did not subscribe to utilitarianism, being a leading proponent of social contractualism.

First, what exactly is 'utilitarianism', the moral theory that Rawls is trying to defend against objectors in this essay? Utilitarianism is one of a type of theories labeled 'consequentialism'. Perhaps the best way to understand consequentialism is by seeing how it is contrasted with opposing deontological theories. This class of opposing theories considers the moral rightness or wrongness of an action or rule as being dependent on not just the consequences of the action or rule. Unlike these duty-based theories, consequentialist theories consider the consequences of the act or rule being evaluated as the sole determinant of its moral rightness or wrongness.

Utilitarianism is a particular kind of consequentialist theory. Its distinguishing feature is that it regards only one kind of consequence as being intrinsically valuable. This could be 'well-being', 'happiness', 'pleasure' or 'preference satisfaction', depending on the variant of utilitarianism. Rawls's essay focuses our attention on another aspect of utilitarian thinking. And that is whether, when we evaluate rightness and wrongness, we ought primarily to consider the consequences of individual actions or of practices and rules. The former kind of utilitarianism is known as 'act utilitarianism' (or 'direct utilitarianism'), while the latter kind is termed 'rule utilitarianism' (or 'indirect utilitarianism').

Now, act utilitarians do value the efficacy of rules as general guides to action. However, for act utilitarians, this use is secondary. Where all of the consequences of a particular action can be determined infallibly, then the consequences of the individual act must hold primacy, with any general rule being amended to take account of exact circumstances. However, as Rawls shows, this way of conceiving of rules does not avoid all of the woes of the act utilitarian.

2. Problems with Utilitarianism

In moral theory, I'm very partial to utilitarianism. However, I find two aspects of act utilitarianism highly problematic. And this is where I found Rawls's essay, 'Two Concepts of Rules', very influential in my own thinking. The two problems that I think Rawls's analysis solves are that, first, act utilitarianism seems to countenance some actions that common-sense morality and our reflective intuitions consider highly unjust and unfair. And, second, that act utilitarianism seems to require utilitarians to be two-faced in their public pronouncements. Let me outline each problem in turn.

Act utilitarianism seems to require in some special circumstances that the moral agent trade off the misery of one person for the greater happiness of many others in a way that leads to a great injustice. This case is just one such example:

> Imagine that each of five patients in a hospital will die without an organ transplant. The patient in Room 1 needs a heart, the patient in Room 2 needs a liver, the patient in Room 3 needs a kidney, and so on. The person in Room 6 is in the hospital for routine tests. Luckily (for them, not for him!), his tissue is compatible with the other five patients, and a specialist is available to transplant his organs into the other five. This operation would save all five of their lives, while killing the "donor". There is no other way to save any of the other five patients
>
> [Sinnott-Armstrong 2019]

Not only is the action one of committing a grave injustice, it also requires that the surgeon keep her intentions and reasoning secret. This illustrates the second problem with act utilitarianism. It requires moral agents to keep their motivations private. What I mean by this is that adopting act utilitarianism as a guide to action seems to require a utilitarian to:

1. advocate publicly some moral principles that they themselves do not accept as guiding their own actions
2. praise people publicly for some actions that they do not regard as morally right and punish people publicly for some actions that they do not regard as morally wrong

The classical utilitarian, Henry Sidgwick, argued for this 'esoteric morality' of utilitarianism to be kept secret from the general public lest it lead to bad consequences. As he put it:

> Thus, on Utilitarian principles, it may be right to do and privately recommend, under certain circumstances, what it would not be right to advocate openly; it may be right to teach openly to one set of persons what it would be wrong to teach to others; it may be conceivably right to do, if it can be done with comparative secrecy, what it would be wrong to do in the face of the world; and even, if perfect secrecy can be reasonably expected, what it would be wrong to recommend by private advice or example.
>
> [Sidgwick 1874 (1907): 489]

In relating his defence of a professor who deliberately misgrades a student, contemporary act utilitarian, Peter Singer, defends Sidgwick and this duplicity:

> It may be right for a professor to give a student a higher grade than his work merits, on the grounds that the student is so depressed over his work that one more poor grade will lead him to abandon his studies altogether, whereas if he can pull out of his depression he will be capable of reaching a satisfactory standard. It would not, however, be right for a professor to advocate this publicly, since then the student would know that the higher grade was undeserved and—quite apart from encouraging other students to feign depression—the higher grade might cheer the student only if he believes that it is merited.
>
> [Singer 1981: 166]

A second example is Jack Smart's response to Richard Brandt's case of a utilitarian Frenchman living in wartime England who must decide whether to comply with a government-directed rationing of gas and electricity. As Smart recognizes, with the vast majority of citizens complying, a few citizens using more gas and electricity to raise their comfort level will result in an increase in the general level of happiness [Smart and Williams 1973: 57]. So, in the usual case where the other citizens are predominantly non-utilitarians, Smart concludes that:

> The act-utilitarian will have to agree that *if the Frenchman's behaviour could be kept secret* then he ought in this case to use the electricity and gas. But the Frenchman should also agree that he should be condemned and punished if he were found out.
>
> [Smart and Williams 1973: 58]

These two cases (misgrading a student and breaking a ration) not only offend our basic sense of justice, they require us to be duplicitous in our public-facing statements. The great worth of Rawls's essay, I think, is in articulating clearly how a more insightful version of utilitarianism avoids this kind of duplicity and the stark clash with common-sense morality on questions of justice. In the next section, I will walk you through Rawls's essay, and then in the final section I will review what we have learned from Rawls's analysis and the benefits it brings to utilitarian theory.

3. Two Concepts of Rules

In this section, I will summarise Rawls's essay, 'Two Concepts of Rules'. In Part I, he illustrates his thesis by considering a case of unjust punishment. In Part II, he examines the practice of promising. In Part III, he distils the lessons from the first two parts to formulate more formally the difference between the two ways of looking at rules. In his final part, he sums up with some concluding remarks.

Rawls [1955: 3] begins his essay by outlining its purpose:

> I want to show the importance of the distinction between justifying a practice and justifying a particular action falling under it, and I want to explain the logical basis of this distinction and how it is possible to miss its significance.

Rawls brings out the distinction by considering two conceptions of rules. This distinction is not wholly new, though. He gives examples of how it is presumed by other major authors, including Hume, Austin, Mill, Mabbot and Toulmin. The importance of the distinction is illustrated in how it can be used to defend utilitarianism against objections centring on punishment and promise-keeping.

I – Punishment

In this first part of his essay, Rawls deals with supposed counterexamples to utilitarianism based on the unjust punishment of innocents. He has in mind here cases of the sort in which, in order to prevent future crimes of a very cruel kind, it becomes highly expedient to frame and hang an innocent man when none of the real criminals can be caught and the injustice can be hidden successfully from the public [1955: 10].

First, Rawls identifies two attempted justifications for punishment of legal wrongdoing. The retributive view justifies punishment on the grounds that it is morally fitting independently of any good consequences following from the punishment [1955: 4–5]. The utilitarian view, on the other hand, justifies punishment by looking to the future benefits to society through maintaining social order. In sum, utilitarianism justifies the practice of punishment while retributivism fits particular cases of wrongdoing to the general rules.

To illustrate the point, Rawls distinguishes between two types of questions:

1. Why was a particular person put in jail?
   Answer: They committed a crime (this answer is backward looking).
2. Why do people put other people in jail?
   Answer: To maintain peace (this answer is forward looking).

The apparent conflict of views disappears as the two questions are about two different and essential types of roles. In this case, these roles are:

1. judge (looking back)
2. legislator (looking forward)

Utilitarianism justifies the institution of criminal law as a system of rules that enforces retributive punishment in particular cases of law-breaking [1955: 6–7]. Historically, the classical utilitarians advocated for the rule of law and objected to capricious legal judgments [1955: 8].

The retributionist can mount two objections here.

Objection 1: Utilitarianism may justify punishing the wrong behaviours.

Rawls's Answer: The retributionist agrees with the utilitarian that behaviours that spread terror and alarm throughout society deserved to be punished [1955: 8–9].

Objection 2: Utilitarianism justifies too much; that is, it justifies the framing and punishing of an innocent when it benefits society [1955: 9–10].

Rawls's Answer: Not if the utilitarian draws on the distinction between the justification of an institution and the justification of a particular action falling under it.

So, the utilitarian can define the institution of punishment as:

1. "punishment" $=_{df}$ the legal deprivation of normal rights because of a violation of a rule of law
2. the violation is established by trial according to due process
3. the deprivation is carried out by the recognized legal authorities
4. the rule of law clearly specifies both the offense and the attached penalty
5. the courts construe statutes strictly and the statute was on the books prior to the offence [1955: 10]

Utilitarianism justifies the institution of punishment with specific roles. However, within this system, there is no role whose responsibilities include framing and punishing an innocent person that would survive utilitarian criticism and receive societal ascent.

Rawls [1955: 11–12] asks us to imagine an institution (called "telishment") designed and set up to frame and punish an innocent whenever it will benefit society, assigning such powers to various roles in parliament, law enforcement and the judiciary. It is unlikely such an institution will have utilitarian justification.

II – Promises

In this part, Rawls deals with cases of promise-keeping and promise-breaking. The objection to utilitarianism here is that it justifies the keeping of a promise *only* when it will lead to the best consequences [1955: 13].

A standard utilitarian defence to this objection is that breaking a promise diminishes the practice of promising and the general societal benefits that come with the practice.

David Ross has three objections to this kind of defence:

> *Objection 1*: Even taking the diminishing of the utility of the practice into account, some promises would still be justified on utilitarian grounds.
>
> Rawls considers Ross's argument unconvincing because it is low on specifics. However, Ross is right in that: 'For a general utilitarian defense is not open to the promisor: it is not one of the defenses allowed by the practice of making promises' [1955: 15].
>
> *Objection 2*: The utilitarian appeal overestimates the damage to the practice of promise-keeping.
>
> Rawls gives some *prima facie* weight to this objection.
>
> *Objection 3*: The utilitarian appeal fails to account for the obligation where the promise is not public (for example, 'a promise a son makes to his dying father concerning the handling of the estate' [1955: 15]).

Rawls [1955: 16] points out that Ross's objections fail to make the distinction between the justification of the practice of promise-keeping and the justification of a particular act of promise-keeping. The utility of the practice of promise-keeping derives precisely from the promisor's abdication of utilitarian excuses. As Rawls [1955: 16] puts it:

> But if one considers what the practice of promising is one will see, I think, that it is such as not to allow this sort of general discretion to the promisor. Indeed, the point of the practice is to abdicate one's title to act in accordance with utilitarian and prudential considerations in order that the future may be tied down and plans coordinated in advance. There are obvious utilitarian advantages in having a practice which denies to the promisor, as a defense, any general appeal to the utilitarian principle in accordance with which the practice itself may be justified.

Rawls likens the barring of utilitarian defences for breaking a promise to the barring of excuses for invalid game-play made by players in games such as chess and baseball. However, he does allow for promises to be broken in order to avoid *extremely* severe consequences. This excuse is allowed within the socially-mandated practice of promise-keeping.

A person breaking a promise with a general utilitarian excuse fails to understand what 'I promise' means. Pointing to this barring of general utilitarian excuses is how we teach children what it means to make a 'promise'. Nonetheless, every practice admits of defences for severe circumstances and these are not fully specified. However, the acceptance of these defences does not allow for general utilitarian calculations in ordinary cases.

III – Logical Status of Rules

In this part, Rawls expands on the difference between two concepts of rules that will illuminate the distinction between the justification of a practice and the justification of a particular action falling under it. He deals with the summary concept and the practice concept in turn.

1. Summary View: Rules are summaries of past direct applications of the utilitarian principle (induction from particulars).

Here, Rawls reviews the writings of Austin, Mill and Moore to conclude that they overtly espoused the summary view [1955: 19–21, f. n. 22]. However, in their writings and social advocacy, the classical utilitarians, in practice, applied the utilitarian principle to social institutions.

Rawls [1955: 22] distinguishes four key features of the summary view:

1. The point of using summary rules is to make decisions more quickly.
2. Decisions on particular cases are 'logically prior' to the rule, meaning:
    a. a particular case 'may exist whether or not there is a rule covering that case'
    b. a particular case can be described as a particular case of a sort whether or not there is a rule regarding that sort of case
    c. *A* and *B* in the rule 'Whenever *A* do *B*' can be described whether or not there is the rule or practice

    Here, Rawls [1955: 22–3] gives the example of the summary rule of telling a lie to a fatally ill person to hide the seriousness of the illness.
3. As rules are guides or 'rules of thumb' only, they are revisable in each instance.
4. As a *general* rule, it is arrived at by estimating probabilities of outcomes in particular cases; that is, 'as a generalization from experience' and therefore given to exception [1955: 23–4].

2. Practice View: Rules define a practice set up to reduce confusion and coordinate social behaviour.

Rawls [1955: 25] explains the differences between this view and the summary view as these:

1. The rules of a practice are logically prior to particular cases, meaning that there cannot be a particular case falling under a rule without a pre-existing practice.

    A practice defines offices, moves and offences (that is, penalties for a breach of rules).

That a practice is 'logically prior' to particular cases means:

a. an action/move falling under a rule cannot be *described* as that kind of action without the practice that defines it

Here, Rawls gives the example of moves within the game of baseball. A player cannot 'steal base', 'strike out', 'draw a walk', and so on, without the rules of the game that define these moves.

Rawls deals with the objection that a practice is not 'logically prior' as a practice presumes actual instances of actions falling under the practice by pointing out that an actual instance of an action still presupposes the practice [1955: 25–6, f. n. 23].

2. To engage in a practice is to give up one's authority to question whether the rule applies in one's particular case. To raise a question is to misunderstand the practice.

Here, Rawls [1955: 26] gives the example of the baseball batter asking, 'Can I have four strikes?', when the rules of baseball only allow three.

3. and 4. The structure of a practice specifies its generality to particular cases. There is no 'exception' to a practice; rather a 'qualification' or 'further specification' of the rule [1955: 27].

For Rawls, an action under a practice is explained or defended by referring to the rules of the practice and *not* by direct appeal to actual consequences of the action. (Question: 'Why are you in a hurry to pay him?' Answer: 'I promised to pay him today' [1955: 27]).

The **summary view** ignores the distinction between justifying a practice (defender) and justifying an action under a practice (office holder). On this view, there is only one 'office'; that of a 'rational person seeking case by case to realize the best on the whole' [1955: 28]. With no distinction between 'offices' on the summary view, there are no real 'offices'.

On the **practice view**, one's actions are justified by the specific 'office' one holds defined by the rules of the practice. Reformers question the practice from a different 'office' (for example, that of legislator).

Rawls revisits the case in which the promise is not public (a son's promise to his dying father on how to handle his estate). Here, the promiser steps out of his 'office' as promisor to consider utilitarian arguments for changing the rules for the practice of promising as they apply to promises made in private [1955: 28–9, f. n. 25].

Rawls [1955: 29] ends this section by stating three qualifications to what he has written in this essay:

1. Not all rules are rules of practices. Some are maxims or 'rules of thumb'.
2. There are yet other types of rules not considered in this essay.
3. There are borderline cases between rules as summary and rules as practice.

IV – Summing Up

In this final part, Rawls [1955: 30–2] draws the various threads of his arguments together and makes some concluding remarks. He observes that philosophers have assumed uncritically the summary view of rules and missed the logical import of the distinction between justifying a practice and justifying an action governed by a practice.

Most importantly, a practice provides defined discretions by office holders (for example, a judge determining penalties), but it does *not* provide a general discretion to act on direct utilitarian grounds.

In particular, promising is a practice. Saying, ‘I promise’, is a performative act that commits the promiser to *not* use general utilitarian excuses to break the promise. Rules governing the practice of promising are not strictly codified, so there are variations in allowable defences. However, no practice governing promising allows for the general direct utilitarian excuse to break a promise.

For the practice of punishment, this practice defines the rights of citizens, laws, due process, trials, courts, and so on, ‘none of which can exist outside the elaborate stage-setting of a legal system’. ‘Punishment is a move in an elaborate legal game and presupposes the complex of practices which make up the legal order’ [1955: 31].

Finally, Rawls emphasizes that his distinction between the two types of rules does not lead to social and political conservatism. Rawls’s point is logical, leaving open the possibility of radical reformers challenging social practices. The import of the distinction he draws between the two concepts of rules is that it saves utilitarianism from several traditional objections.

4. In Retrospect

In this final section, I want to consider the legacy of Rawls's essay and its lasting impacts on the defensibility of utilitarianism as a moral theory. To begin with, I think Rawls's choice of practices to discuss was very prudent. His two choices span the entire spectrum of practices from the highly codified and complex practice of punishment within our legal systems to the relatively loose and unspecified rules governing promises. Lying parallel to, and entangled with, the uncodified end of the spectrum are those rules that Rawls points to as difficult to categorize as either a summary rule or a rule under a practice. Like all things in life and philosophy, nothing is easy.

A further complexity arising is that our social practices are a network of interlocking and overlapping systems, some of which can perhaps be organized into a hierarchy. So, for example, our criminal justice system consists in several sub-practices. These include law enforcement, legal representation, the courts and penal and rehabilitative systems. What characterizes each of these systems, drawing on Rawls's components, are the following:

1. rules for behaviour (maximum driving speeds, treatment of the accused)
2. roles (license holder, police, judge, jury member, witness)
3. responsibilities (obtain licence, collect evidence, represent client)
4. rewards and punishments (promotion, fine, imprisonment)

Rawls's insight is that the most defensible form of utilitarianism warrants these components of a social practice as a bundle. What I find most beneficial about Rawls's analysis is that it captures the deontological aspects of our moral rules. It explains concisely and precisely our intuitions condemning unjust acts that, nonetheless, on balance lead to more good than bad consequences, while at the same time explaining our allowance for excuses in extreme circumstances.

Even while explaining the deontological character of many of our rules, Rawls shows how this version of utilitarianism allows for, and even encourages, a reformist zeal against unjust social and political systems. And this reformist mindset has played out historically in the work of both classical and contemporary utilitarians. Campaigns driven by the utilitarian's fundamental principle of impartiality—that no one's interests ought to be discounted simply on the basis of their unchosen identity—have led to the substantial dismantling of racist, sexist, and homophobic practices in modern times. Currently, utilitarians are leading the charge begun by the classical utilitarians against the industrial level of cruelty found in our systems of food production.

A second benefit of utilitarian moral theory is that it explains the importance we place on moral character. Mill, for example, placed great emphasis on how repressive family and work institutions stunt the development of moral virtue and, consequently, the capacity for human happiness. (For an overview of Mill's approach, see, for example, Homiak [2019].) Turning back to Rawls, the institutions and practices that govern the education of our children, political enfranchisement and family relationships can either promote social goods or impoverish them. Applying Rawls's two concepts of rules, those social practices that build morally virtuous characters can also now be seen as obtaining their justification on utilitarian grounds.

Rawls identified a practice view of rules not recognized by act utilitarians. The particular form of rule utilitarianism that capitalizes on Rawls's insight has been labeled by Sinnott-Armstrong [2019] as *public acceptance rule consequentialism*. As this is a mouthful, I have been referring to this view as *rules in practice utilitarianism*. In places, Mill seems to have espoused this form of rule utilitarianism himself. In one place, he writes: 'For the truth is, that the idea of penal sanction, which is the essence of law, enters not only into the conception of injustice, but into that of any kind of wrong' [1861 (1962): ch. 5, 303]. Mill continues: 'When we call anything a person's right, we mean that he has a valid claim on society to protect him in the possession of it, either by the force of law, or by that of education and opinion' [1861 (1962): ch. 5, 309].

On a final note, I think it useful to consider how divergent the rules in practice utilitarianism that Rawls defended against objectors and his own social contractualism. I'm quite partial to Rawls's version of the social contract, with perhaps my only substantive concern being that it relegates our obligations to non-human sentient animals and to children to the status of indirect duties. I think there is a case to be made, though, that the rules agreed to by rational and informed persons deliberating behind Rawls's 'veil of ignorance' are co-extensive with that agreed to by a community of utilitarians. Whatever the case may be, we owe a great debt to Rawls originality and his essay remains as a landmark piece of work in moral philosophy.

References

Allan, Leslie 2015. A Defence of Emotivism, URL = <https://www.rationalrealm.com/philosophy/ethics/defence-emotivism.html>.

Brink, David 2018. Mill's Moral and Political Philosophy, *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), ed. Edward N. Zalta, URL = <https://plato.stanford.edu/archives/win2018/entries/mill-moral-political/>.

Driver, Julia 2014. The History of Utilitarianism, *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), ed. Edward N. Zalta, URL = <https://plato.stanford.edu/archives/win2014/entries/utilitarianism-history/>.

Homiak, Marcia 2019. Moral Character, *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), ed. Edward N. Zalta, URL = <https://plato.stanford.edu/archives/sum2019/entries/moral-character/>.

Hooker, Brad 2016. Rule Consequentialism, *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), ed. Edward N. Zalta, URL = <https://plato.stanford.edu/archives/win2016/entries/consequentialism-rule/>.

Lazari-Radek, K. de and P. Singer 2014. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*, Oxford: Oxford University Press.

Mill, John Stuart 1861 (1962). *Utilitarianism*, ed. M. Warnock, Glasgow: Collins.

Rawls, John 1955. Two Concepts of Rules, *Philosophical Review* 64/1: 3–32.

Rawls, John 1972. *A Theory of Justice*, Oxford: Oxford University Press.

Sidgwick, Henry 1874 (1907). *The Methods of Ethics*, 7th edn, London: Macmillan.

Singer, Peter 1981a. *Practical Ethics*, Cambridge: Cambridge University Press.

Singer, Peter 1981b. *The Expanding Circle: Ethics, Evolution, and Moral Progress*, Oxford: Clarendon Press.

Sinnott-Armstrong, Walter 2019. Consequentialism, The Stanford Encyclopedia of Philosophy (Summer 2019 Edition), ed. Edward N. Zalta, URL = <https://plato.stanford.edu/archives/sum2019/entries/consequentialism/>.

Smart, J. J. C. and B. Williams 1973. *Utilitarianism: For and Against*, Cambridge: Cambridge University Press.

Wenar, Leif 2018. John Rawls, *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), ed. Edward N. Zalta, URL = <https://plato.stanford.edu/archives/spr2017/entries/rawls/>.