In his paper, "Two Concepts of Rules" Rawls reexamines utilitarianism, and makes the important distinction between justifying a practice and justifying a particular action that falls under that broader practice in order to then defend utilitarianism against objections that have been made against it with regards to punishment. The utilitarian view on punishment in terms of maximizing total welfare runs into a contradiction between the fact that we think the state ought never to punish a person it knows to be innocent, but sometimes maximizing total welfare may require the punishment of an innocent person. In order to reconcile this problem, Rawls proposes that we examine rules as falling under two different categories, summary rules and practice rules. What differentiates practice rules from summary rules is that practice rules can be thought of as all jointly contributing or pertaining to a single goal or purpose, which is the practice. Rawls believes that punishment can be thought of as exactly such a practice, and argues that we can still use the utilitarian approach to analyze the practice of the institution of punishment, but once we have done this, it is no longer appropriate to judge individual acts within the practice outside of the context of the practice. Finally, I will expand on Rawls' ideas by making the utilitarian claim for the practice of punishment where only the guilty can be punished through the framework of the time-inconsistency problem.

In the traditional utilitarian view, punishment is only justified if it leads to an increase in total welfare within society. If the choice of punishing a person results in lower social welfare than the choice not to punish, then the punishment would not be justified. This definition of just punishment runs into some difficulties when the question arises of whether it can ever be just to punish an innocent person. Our intuitive feeling is that punishing an innocent man or woman would not be just, but the traditional utilitarian argument for why this is the case is not very strong. Bentham argues that the harm felt by the innocent person being punished would be so great, that it would always outweigh any potential benefits. He believes that the principle of utility would never justify punishing a person who has committed no harm.[1] But this argument isn't very satisfying, because it doesn't create an absolute rule against punishing innocents, but instead leaves it up an empirical calculation.

There are hypothetical examples we can create in which punishing an innocent person would actually lead to higher levels of total welfare, the classic one being the case of punishing someone to make an example of them in order to prevent a riot, which would cause much greater violence and destruction than the act of punishment. The traditional utilitarian explanation does not give us an absolute rule against punishing innocents, and it does not tell us how to deal with a situation in which maximizing total welfare might require the state to punish a person it knows to be innocent. Rawls uses the distinction between practices and actions within those practices to argue on behalf of utilitarianism against these types of criticisms. Rawls makes the argument that utilitarianism can in fact be consistent with the position that the state should never punish a person it knows to be innocent by appealing to the separation between the practice of punishment and the act of punishment.

---

[1] Bentham, 675

In doing so, he begins by defining the difference between practice rules and summary rules. Summary rules can considered as "rules of thumb" or rules of convenience; they represent a type of summary of past decisions arrived at the direct application of basic reasoning. An example of a summary rule would be the idea that we should look both ways before crossing the street or that we should knock before entering a room. In the case of summary rules, we can always reconsider the correctness of applicability of the rule in particular cases. We adopt summary rules because, on the whole, following the rule leads us to be better off than we would be if we were to rely on basic reasoning every time. This does not preclude us from using our own reasoning prior to acting, and deciding that, in this particular case, it might be better to not apply the rule.

Rawls also defines a second type of rules, and these are practice rules. Practice rules are logically prior to the particular actions governed by the rules. Practice rules set up offices, define specific actions, and specify and establish appropriateness and consequences of actions. To explain this relationship between practice rules and actions, Rawls uses the metaphor of baseball. Baseball involves certain actions: throwing the ball, swinging the bat, or running to the base, but in order for to say that we are playing the game of baseball, the rules of baseball must exist prior to us engaging in these actions. We can imagine that before baseball was invented, people could still perform the actions of throwing a ball and hitting it with a stick, but they wouldn't be playing baseball until the rules were created which set up and defined the relationships in the game.[2] In addition, particular actions within the practice can only make sense in the context of that practice. For example, in baseball, it is impossible to strike out or steal a base outside of the context of following the rules of baseball. These actions are specific to the practice, and do not exist outside of the game of baseball.

Furthermore, because the rules are logically prior to the actions, unlike in the case of summary rules, it is impossible logically ask whether or not the rule should apply in a particular case or instance. In baseball, the practice rules dictate that every batter get three strikes. If someone were to ask for a fourth strike, the other players would say: no, we are playing baseball; the rules of baseball only allow three strikes. It would not make sense for the first person to then say: yes I agree that we are playing baseball, and that there should be 3 strikes, but I think that in this case, I should get a fourth. Once we have established and accepted the rules of a practice, it is irrelevant to talk about the merits of an action or rule outside of the context of its role within the practice, since once we no longer follow the rule, we are no longer participating in the practice. If we begin giving some batters three strikes, and some batters four, we are no longer playing baseball. We cannot simultaneously participate in the practice of baseball and also give batters different numbers of strikes.[3]

In some cases, however, we may feel that it would be acceptable to make individual or case specific exceptions. What if you are playing baseball with young children and one of them hasn't been able to get a hit the whole game. You know she will be very upset if she strikes out again, so you make an exception and allow her to take a couple more tries. You know that in baseball, the rules say that there are only three strikes allowed, but in this case your ultimate goal is that all the kids have fun rather

[2] Rawls, 25
[3] Rawls, 26

than to adhere strictly to all the rules of baseball, so it would actually be better not to follow the rule. However, given this difference in the ultimate goal here in this example, a supporter of Rawls could respond that the practice of baseball could actually be divided into two sub-practices. There is official baseball, where the rules must be followed strictly in order to ensure a fair game and winner, and there is also recreational baseball, where the main objective is to have fun; and in recreational baseball, the rules may actually allow for a change in the number of strikes for the sake of making sure that everyone is enjoying themselves. If prior to beginning the game, it is clearly established that recreational baseball is being played as opposed to official baseball, the decision to allow the child to take an extra strike would no longer be problematic.

The above claim could imply that in any case where you think there may be a legitimate exception to a practice rule, it shouldn't be considered as breaking the rule, but instead as creating a sub-practice where the rules are actually slightly different. But isn't this problematic for cheating, someone might ask. Every time you want to break the rules, you could simply say that you are not actually participating in the original practice, but in a new sub-practice where your actions are not actually against the rules and that you have no obligation to the rules of the original practice. Here it is important to remember the definition of practice rules: practice rules are set up prior to actions.

So in order for a cheater to make an appeal to a sub-practice, the sub-practice must already exist and have rules that correspond exactly to the way that the cheater is behaving. If this is indeed that case, then cheating can be prevented by clearly establishing consent for the practice in which everyone is participating before anyone has committed to any particular actions. If, before the game, the players come to a consensus that they are playing official baseball, thereby giving their implied consent to be bound by the rules of official baseball, then any attempt to sneak in an additional strike cannot be attributed to the belief of the cheater that they were actually playing recreational baseball, and even if it was truly a misunderstanding, it still would not be an exculpatory defense for breaking the rules. Knowingly participating in a particular practice implies that you have consented to following the rules of that practice. This is akin to establishing jurisdictions for various laws if we were compare the different sub-practices to different geographical areas with different laws. Maybe jaywalking is illegal in New York, but not in New Jersey. However, once you enter New York, you have implicitly consented to follow the New York state laws, and arguing a jay walking fine by saying that jaywalking is perfectly legal in New Jersey will not be an acceptable defense.

So, if we can't question whether to perform a particular action in the case of practice rules, what is the correct way to exercise our judgment about whether we should be following a rule? Rawls argues that the questioning whether to follow a practice rule must take the form of questioning the rule itself, not its implications in a particular circumstance. In order to question a practice rule, you must question whether the practice itself is effectively designed.[4] Going back to the baseball example, if you wanted to challenge the practice rule about strikes, you would have to challenge the entire practice of allowing only 3 strikes. You can argue that the rules of official baseball should be changed to allow four strikes, but you cannot argue that the rules of baseball say there are only three strikes, but in this particular

---

[4] Rawls, 16

circumstance, you think you should be allowed a fourth. If you want to defend or appeal a particular action within a practice, the argument must take the form of appeals or defenses of the entire practice as a whole. It does not make sense to argue about the correctness of an action outside of the context of its role or function within a practice.

By now I'm sure we are all wondering how this lengthy discussion of practices and rules relates back to the original issue we wanted to address about how it can be possible to make a utilitarian argument for never punishing a person that is known to be innocent. Well, Rawls believes that punishment can be thought of as a practice, and all the rules regarding the appropriateness, nature, and logistics of punishment can be considered practice rules. In order to create a practice or institution of punishment, we must clearly define the common goal of punishment which all of its practice rules contribute to. This common goal, or definition, of punishment is: to legally deprive someone of their typical rights in the way permitted or described by the law due to that fact that he or she has violated the law when that violation has been determined to have occurred by a fair and legal trial.[5] This implies that punishment is defined by the guilt of the person who is being punished, and that the practice rules of punishment require that the person who is being punished by the state has been found guilty of a crime. This means that the practice laws of the institution of punishment require proof of guilt in order to punish someone.

When we consider punishment as an institution, and accept its existence as practice in the form of the definition above, it is no longer relevant to question whether it is appropriate to punish an innocent person in a particular instance. In the same way that the rules of baseball require three strikes and it is illogical to argue for four strikes under particular circumstances, the rules of punishment require that the person being punished be guilty of breaking the law, and it is irrelevant to argue that the person can be innocent under particular circumstances. Punishment is defined as a consequence of law-breaking, so once you punish an innocent person, you are no longer punishing them; you are participating in a completely different practice, which Rawls names "telishment"[6].

The creation of practices and the enforcement of practice rules can be superimposed relatively closely onto the roles of legislature and judiciary. The legislature has the job of determining whether a certain practice is worth integrating into law, but once a practice has been accepted into law, the judiciary cannot make a decision regarding the correctness of an individual action outside of the context of the practice. We may usefully compare the roles of judge and legislature can be compared to the roles of the creators of baseball and to the umpire.

Although there is indeed a major difference between the government and the baseball team, I think they can be reconciled by examining the judiciary-legislature relationship in more detail. Unlike in case of umpire, the checks and balances system of our government does allow for some discretion and creation of new laws on the part of judges. However, in instances where judges can use discretion, this ability to use discretion is built into the practice.

---

[5] Rawls, 10
[6] Rawls, 11

For example, the rules of punishment often say something like: if the defendant is guilty of X, the punishment is a fine of between 100 and 500 dollars. In this way, judicial discretion is directly built into the practice of punishment. The judge can say: yes, this person committed X, but she was simply very careless and did not commit the wrongdoing with scienter, so I will give her a only the $100 fine. But, he only has the power to use that discretion to the extent that is allowed by the practice. He cannot say that he thinks that we should not punish the defendant for her actions at all, nor can he say that he thinks the defendant acted so badly that she should pay 600 dollars in fines since neither of these actions is within the range of discretion permitted by the practice.

The second way in which discretion is built into punishment is instances where the law is silent. In these cases the practice laws do not address whether punishment is acceptable in a particular circumstance or not, so it is unavoidable that the judge make a discretionary decision. If the judge/umpire chooses not to enforce, then it sets one type of example, and if she does choose to punish, it sets the opposite example. Given the efficiency constraints, it is not practical for the judge or umpire to appeal to the legislature or creators of baseball for a decision. The practice dictates that these individuals are trained in the intentions and purposes of the rule creators so that they can interpret the rules in the best unbiased and scientific way. Just as rules are not rewritten by umpires, judges do not rewrite rules, they only interpret them.

The final way that the judiciary can exercise discretion is through an appeal to the Supreme Court. In this case, I think it is helpful to keep in mind the idea of practices and sub-practices explained earlier in the paper. The ultimate practice to which the courts are beholden to is upholding the constitution. In this way, the courts may appeal the correctness of a certain practice by saying that it violates the goal of this ultimate practice. The practice of upholding the constitution may require additional sub-practices with their own rules, but ultimately even they are answerable to this ultimate practice. We have the practice of baseball, which so far includes the sub-practices of official baseball and recreational baseball. Someone may try to create a new sub-practice of baseball called British Baseball in which the players use a flat bat and run back and forth instead of around the bases. It can then be decided that this new sub-practice is sufficiently different from the original category of baseball that it can no longer be called baseball, but falls into a different practice called cricket. In the same way, it can be decided that a sub-practice within the law is sufficiently different from the objectives of our legal system that it no longer falls within the practice agreed to in the constitution.

Alright, so we have an idea of practices, and we have established two important conclusions: first, that a practice which allows for punishing innocent as well as the guilty is a different from a practice in which punishment is only applied to the guilty, and second, that we cannot question the righteousness of actions within practices without questioning the practice itself. But, the question we really want to answer is: why do we find that the practice which includes a rule allowing us to punish the innocent is unjust while the practice which only allows us to punish the guilty is just? Well, in this case, utilitarianism can be applied relatively straightfowardly in order to answer this question.

Utilitarian considerations would not allow for a practice which allows us to punish the innocent since such a practice would undermine the entire deterrent benefit of punishment. Decisions about

punishment have an inherent time-inconsistency problem, which is characterized by circumstances under which the best course of action at a particular time may not be the best course of action over all.[7]

In the case of punishment, the best course of action over all is an institution of punishment where those that are guilty of breaking the law face certain consequences for their disobedience, so that people are deterred from breaking the law in the future. However, in every particular instance, there is an incentive for the state to punish an innocent person. Punishing an innocent person would be less costly than finding the guilty party and going through the trouble of proving their guilt, but would serve the same purpose of setting an example for those considering breaking the law. However, because people are forward looking, once they can anticipate that the state is no longer adhering to this promise to punish only the guilty, they no longer have any incentive to follow the law since their likelihood of punishment will be same regardless of whether they are guilty or innocent.

The utility gain of punishing an innocent can only be realized if the people still believe that punishment is only for the guilty. So once you have an established practice of punishing the innocent, the arbitrage opportunity created by the discrepancy between the state's actions and the public's beliefs about those actions no longer exists, and the system of punishment will collapse.

The classic example of this type of problem is when you have a professor with a class full of students. The professor really wants everyone in the class to study and learn the material, but he doesn't necessary want to give students exams since exams take away class time and he doesn't want to grade them. So the professor tells the class that at the end of the semester there will be a lengthy, challenging, cumulative final exam. However, once the end of the semester approaches, the professor realizes that he really doesn't want to spend the first few weeks of June grading all of these exams; so the logical thing for him to do would be to cancel the exam. The system was based on the promise that there would be this tough exam at the end of the semester, and as long as the students believe this promise and study for the exam, then the professor does not need to keep his promise. But, just like people evaluating the consequences of criminal actions, the students are smart and forward looking, they can anticipate that it's not really in the professor's interest to have to proctor and grade all of the exams. And once they no longer believe that the professor will keep his promise, then they no longer have an incentive to study, and the system ceases to function.

The final exam example contains the same problem as the punishment example since the summation of individual utility maximizing decisions over a course of time does not actually produce the overall utility maximizing decision over the course of multiple years or repetitions. In each individual instance, it will be in the interest of the state to punish the innocent man as opposed to the guilty man, but over time, this practice will produce a state where people no longer deterred by punishment and no longer trust their government. In order to preserve a practice which will ultimately be the most welfare maximizing action in the long run, we may have to consistently make a series of individual decisions in which the highest welfare option for each particular instance is not chosen.

---

[7] Strotz, 165

I think Rawls' use of a distinction between practices and individual actions characterized by those practices was very enlightening for the understanding of the justification for never punishing the innocent. Separating the practice of punishment from the individual actions which make up the process of punishing someone allows us to contain utilitarian considerations to only the practice in general, rather than applying them to individual cases. Furthermore, I think the theory of the time-inconsistency problem provides additional insight into how this distinction can be understood and supported from the utilitarian perspective. Utilitarianism calls for welfare maximization, but it does not often specify a time frame. In cases where certain decisions produce different welfare outcomes in the short and long run, practices can be created precisely to preserve certain long run utilitarian benefits in the presence of short run incentives to defect.

Works Cited

Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation*. Chapter I (Sections I-II) and Chapters XIII-XIV.

Rawls, John. *Two Concepts of Rules*. The Philosophical Review. Duke University Press, 1955. Vol. 64, No. 1, pp. 3-32

Strotz, R. H.. *Myopia and Inconsistency in Dynamic Utility Maximization*. The Review of Economic Studies. Oxford University Press, 1955. Vol. 23, No. 3, pp. 165-180

24.235J / 17.021J Philosophy of Law
Spring 2012