Should Like Cases be Decided Alike?
A Formal Analysis of Four Theories of Justice

Benjamin Johnson  Richard Jordan
Princeton University  Baylor University

Abstract

The maxim ‘treat like cases alike’ has underpinned theories of justice since Aristotle. It is commonly wielded as a shield against arbitrary rule and as a sword for civil rights. But it is not clear what the maxim actually means. In this paper, we formalize and evaluate possible definitions of the maxim, including those of John Rawls, Ronald Dworkin, HLA Hart, and Aristotle. We demonstrate that, as a theoretical principle, the maxim is always either unhelpful or pernicious. It provides no moral reason to make any particular decision. Importantly, we limit our inquiry to pure theory, and so there may still be room for using the maxim in practice. But such pragmatic use could only be justified with reference to other values, since the maxim contains no moral force of its own. Our findings should encourage political and legal theorists to reexamine some core assumptions of our normative theories.

9973 Words
Justice demands, wherever that concept is found, that like men be treated alike in like conditions. Why, I do not know; the fact is given.

– Karl Llewellyn, *The Bramble Bush*

# 1 Introduction

About 2,400 years ago, Aristotle argued that *like cases should be treated alike*. This Like Cases Maxim (LCM) has been a core feature of political philosophy ever since. The maxim was at the heart of some of the most important theoretical disputes of the last several decades. For political theorists in the liberal tradition like John Rawls, a maxim rooted in shared, objective features could hold ground against utilitarian theories that rest on subjective individual preferences. For legal theorists like Fuller, Hart, and Dworkin, the maxim is ground zero in reclaiming legal practice and theory from legal realists (who argued all cases are really different and judges would do as they please). Today, the LCM undergirds arguments about income inequality, racial discrimination, and global justice, as scholars and advocates appeal to the basic conviction that like persons should not be treated differently.

The maxim is a seemingly simple, normative command. It asserts that there is a moral imperative to treat like cases alike. On a plain reading, it requires judgments to track intrinsic similarities between cases. If case $A$ is objectively like case $B$, then the two cases should be decided similarly. By defining similarity according to cases’ objective characteristics rather than a judge’s subjective perceptions, the maxim forces decisions to be consistent with rules.

Unfortunately, its perceived simplicity and universal acceptance have shielded the maxim from sustained inquiry. Though many have noted the like-cases maxim risks being too weak for its mission, its hold on theorists is too strong to be easily abandoned. H.L.A. Hart, for instance, while acknowledging that the like-cases maxim “is by itself incomplete,” nonetheless insists it is “a central element in the idea of justice” (Hart, 1961, 159). Hart shares the common intuition that while “by itself” the maxim does not do much, given a set of substantive rules to build on, it can
Our central claim is that the Like-Cases Maxim is either superfluous or pernicious, and so it carries no normative weight. By superfluous, we mean the maxim is not useful to decide cases. This is the most common concern regarding the LCM, and we formalize it. But we also show that some interpretations of the LCM are not superfluous and are useful when deciding cases. Unfortunately, such interpretations are pernicious in at least one of two ways: they might violate our core moral intuitions, or they might lead to logical contradictions.

Nonetheless, we do not claim the LCM serves no purpose. The LCM might offer an *ex post* check on theories to see whether cases *once decided* are decided consistently with the maxim. That is, the LCM might not help a theorist decide a case, but it can help us reject some potential theories of justice. This possibility is, so far as we know, novel to this article, but we note that it is a very weak tool to thin a very large herd.

We also believe the LCM may be a good rule-of-thumb in practice, especially as part of a judicial machinery deciding ongoing cases in a world of uncertainty. When a judge faces a case where the law is unclear, if she can decide the case in a way consistent with an earlier decision, that seems a plausible reason to choose one option over another. The LCM might be bad philosophy but good advice.

Still, the LCM is framed as a normative command, not advice. For it to have moral force, it must hold together and mean something at a theoretical level. We argue it does not. It either leads to unacceptable outcomes, or it does nothing at all. If the LCM is superfluous or pernicious, then it is time to reopen some debates in legal and political theory, and we may need to reexamine some foundations of our normative traditions.

Our argument proceeds by taking the notion of like cases to imply two things: that there is a set of possible cases and that there is some measure of likeness between them, i.e. some metric. This allows us to define a theory formally using standard mathematical notation of metric spaces. We then use these tools to consider Aristotle’s conception of the maxim alongside three modern theorists of law and justice: H.L.A. Hart, John Rawls, and Ronald Dworkin. Each uses the like-
cases maxim in their theories, though each adopts a different interpretation. We find that all four are either pernicious or superfluous. Even if the reader does not find our subsequent analysis convincing, we hope this paper does make plain that a maxim enjoying unanimous approval lacks a common definition. It is not clear what everyone is agreeing about.

Insofar as possible, we would like to proceed as rigorously as we can. However, we recognize that formalism places burdens on a reader. Accordingly, we have restricted the mathematical arguments to self-contained portions of the paper so that a reader can skip them if desired. Nonetheless, we believe the development of a mathematical framework is important. By using formalism to engage an ancient question of political and legal scholarship, this article is a bridge between positive and normative theory. Such work is valuable not only because it employs a different and rigorous methodology on important normative questions, but also because, as Knight and Johnson (2015) point out, the line between positive and normative theory is not always clear. We thus see this project in the tradition of Amartya Sen’s critique of the rights literature and Restoring the ancient ties between formal and political theory sheds light on the maxim and offers hope that similar efforts on classic questions might be similarly susceptible to formal analysis (Kogelmann and Stich 2016, Gaus 2016, Kogelmann and Stich 2015, Eskridge and Ferejohn 1992).

2 Literature Review

The like-cases maxim recurs throughout political, philosophical, and practical writings on justice. For John Rawls, it relates to formal justice, and it excludes certain arbitrary and unpredictable systems of rule (Rawls 2009). For H.L.A. Hart, it joins with other principles to exclude segregation and racial injustice. For Ronald Dworkin, the maxim establishes political integrity, the indispensable virtue guiding how judges should interpret the law. Kant believes that violating the maxim can incur “blood guilt” (Wood 2010, 112); Locke, that justice requires “promulgated established laws, not to be varied in particular cases, but to have one rule for rich and poor, for the favourite at

---

1We have in mind his work on equality and his formulation of the liberal paradox. For representative works, see Sen (1970), (1980), and (1990). We also draw on Sen (1996) in section 7.
court, and the country man at plough” (Locke [1980]). And Aristotle, the maxim’s progenitor, uses it to guarantee that desert should always be proportional to merit. On its surface, the maxim seems uncontroversial. The trouble comes when we ask what it actually means. Intuitively, similar cases should result in similar outcomes. But what makes two cases similar? A famous instance of this question is Powell v. Pennsylvania (1888), a case that challenged the legal distinction between butter and margarine. Margarine producers said that margarine is like butter, and so laws treating them differently were unjustly discriminatory and violated the 14th Amendment. The Court ruled that since margarine producers were all treated the same and butter producers were all treated the same, there was no violation. For the Court, like cases were treated alike. But this essentially says that discrimination written into a statute cannot be unjustly discriminatory; likeness is whatever the lawmaker says. If maintained, this notion would undermine the modern jurisprudence of Equal Protection.

We are not the first to note that too thin an interpretation of the maxim reduces it to nothing more than consistency. In a seminal article, Westen (1982) argues that the maxim could not have any content. His sentiment echoes in many quarters. For instance, Winston (1974) quotes C. Perelman as observing that justice and the like-cases maxim might simply mean logical consistency. Winston then rejects this idea, since the maxim would have “lost its peculiarly moral importance” (Winston [1974] 9). We agree: the like-cases maxim is intended to carry normative weight.

Thus, while we concur with Westen and others that the maxim is often superfluous, we do not find that it is necessarily so. As we show, it is possible to construct interpretations of the maxim that are not empty. These interpretations just have undesirable implications. In this paper, we show that both the maxim’s defenders and its detractors are wrong: the like-cases maxim is not always devoid of content, but any attempt to imbue it with content results in consequences its defenders would probably be forced to reject.

In the realm of practical justice, the maxim’s application is more apparent. Valuing precedent

---

So, the Supreme Court can believe it’s not butter. The opinion reads thus (678 1888 127): See also McCollan (2006), who uses the case to explore international human rights law.
or legal consistency, if only to ensure the law’s predictability, should lead us to treat similar cases similarly, even if doing so offends some of our moral sensibilities. In fact, legislators often try to enforce uniform decisionmaking by limiting judges’ discretion (e.g. through sentencing guidelines) in order to guarantee like cases are treated alike. Some notion of likeness seems to guide the evolution of law, as well: in the United Kingdom, Walker (2010) notes that married couples are increasingly seen as more similar to cohabiting couples, and so the UK is beginning to extend the same rights to both. Finally, even if we do not value precedent or consistency, the maxim might be helpful if a judge is simply uncertain about how to decide a particular case. In this instance, she will employ analogical reasoning to deduce from a known, settled outcome the decision she should reach in the present circumstances. Marmor (2005) takes this theme further, arguing the LCM could decide cases when all other reasons for a decision are unclear.

But none of these practical interpretations speak to the maxim’s most common applications: normative questions of equality, especially racial, gender, and sexual equality. In this usage, which is that at least of Dworkin and Hart, the maxim is not just a rule of thumb but a fundamental principle of justice: it carries moral weight, and in conjunction with other principles it decides crucial cases about which we care deeply. Moreover, the use of the LCM in practice is often legitimated by the belief that the LCM is a core feature of justice. The first-order question is whether it is, in fact, fundamental.

Our framework mirrors that found in Gaus (2016). Methodologically, our work relates to theories of case-based decisionmaking in economics and political science, especially that of Gilboa and Schmeidler (1995, 2001). Responding to the vast literature on subjective expected utility, they develop a theory where past decisions in similar cases inform decisions in a current case. In political science, our work echoes the influential case-space model underlying formal and empirical studies in judicial politics. Examples include Kornhauser (1992a), Kornhauser (1992b), Lax (2007), Landa and Lax (2007), and Landa and Lax (2009).

\[3\] For a classic work on the subject, see Schauer (1987).

\[4\] See a working paper by David Strauss. While Strauss focuses more on the maxim in practice, our conclusions about the maxim in theory bear a striking resemblance to his.

\[5\] See Sunstein (1993).
3 Conceptual Framework

Theories of justice share a common objective: to examine some input or case (e.g. an action, an institution, a history, a distribution of goods, etc.) and then to categorize that object in some way (e.g. as just, unjust, partially just, etc.) according to a set of axioms. Theories of justice can therefore be defined by the set of objects they take as inputs, the set of possible outcomes, and the rule used to map from the former to the latter. The like-cases maxim is an axiom. When deciding how to judge a particular case, \( x \), it requires a theorist to first identify other cases like \( x \) and to decide \( x \) so that the outcomes are alike.

The LCM contains several parts. First, it compares cases. We believe it is important to ground cases in objective facts. If cases depend not on objective, physical facts—about the past, present, or possible futures—then they depend on subjective criteria that would enable arbitrary decisions and undermine the maxim’s normative intent. These objective factors may be either continuous or discrete. Next, the LCM regards cases as like or unalike. At a minimum, ‘likeness’ implies a measure of difference: if case \( A \) is like \( B \) but unlike \( C \), then \( A \) is more different from \( C \) than it is from \( B \). The more different two cases are, the (weakly) less alike they are. We assume that these differences can be captured by some metric describing relative difference. Last, the LCM decides cases. Judgments are the set of possible decisions consistent with a theory. They describe whether a case is just or equitable or fair. We allow a judgment to take any number of descriptions (not merely just or unjust).

We emphasize that defining a measure of likeness precedes the application of the LCM. This must be true for two reasons. First, the LCM plainly presupposes a metric. “Treat like cases alike” assumes there is some extant definition of likeness. The LCM cannot merely assert which cases are alike without regard for objective reality, else it would justify arbitrariness rather than limit it. Second, the LCM must be more than a bid for moral superiority. Consider an argument between an animal rights activist and a butcher. The activist might tell the butcher that his willingness to carve beef but not children violates the LCM, because the activist believes there is no moral difference between animals and people. They are, to his mind, like cases. To the butcher, the cases
are different and should be treated differently. Accordingly, the activist’s invocation of the maxim is a bid for moral superiority—a declaration that he has truer knowledge of justice and the proper likeness of cases. Whether he is correct or not has nothing to do with the LCM, and the argument is unintelligible to anyone who does not share his worldview. Appealing to the LCM offers no moral direction to either.

We claim that every interpretation of the LCM we have found or could imagine is either pernicious, superfluous, or arbitrary. By pernicious, we mean that an interpretation would cause a logical contradiction or entail conclusions that grossly violate our moral intuitions. Although violating moral intuitions is not a logical threat to the maxim, it may be a very good reason to discard it. By superfluous, we mean that an interpretation of the LCM is useless for deciding cases. Perhaps the best example comes from The Andy Griffith Show. In one episode, the inept deputy Barney Fife locks two criminals in a cell and then explains the rules of Mayberry jail: “Here at the Rock we have two basic rules...Rule Number One: Obey all rules! Secondly, do not write on the walls, as it takes a lot of work—to erase writing—off of walls.” In our view, there is really only one rule here. The first rule is superfluous. As we show, many interpretations of the LCM boil down to the command to obey all rules!

To better explain what we mean by pernicious and superfluous, consider a stylized example. Suppose acts of interpersonal violence exist on the unit interval. Self-defense or defense of another person is at or near zero and premeditated murder close to one. Other acts of violence fill in the points between the two. Suppose we start with the assumption that self-defense that causes little harm is just and apply a close-enough rule that says that any two cases within 0.01 of each other are like one another and should be decided alike. When we say that self-defense is just (located at 0 on the interval), we are also saying that the action at 0.01 is also just. But when we say that, we must recognize that the case at 0.02 is just. From there, we must acknowledge that the case

---

6By arbitrary, we mean that an interpretation does not treat cases as alike or different based on criteria that reflect objective features about the world. For example, suppose one were to say that case A is like case B because A and B are decided alike. Then it is easy to make sure that cases that are like each other are decided alike, but the reasoning is entirely circular.

7We have in mind something akin to Rawls' process of reaching a reflective equilibrium.
at 0.03 is just and so on. This continues until one must conclude that premeditated murder is also just. This violates our deeply held moral beliefs and is pernicious. So, though the close-enough rule seems attractive initially because it decides outcomes when there is a relevant precedent, it is clearly an unsatisfactory way to operationalize the maxim.

The obvious solution is to draw a line. For example, we might say that cases to the left of 0.5 are justified acts of violence, and cases to the right of 0.5 are unjustified. Such line-drawing is a legitimate, even necessary operation for a just and functioning society. But notice that it makes the LCM superfluous. Drawing the cutpoint decides all of the cases. The cutpoint says, cases up to 0.5 are just and cases to the right of 0.5 are unjust. The LCM adds nothing. At most, it simply says that since cases to the left of 0.5 are justified, cases to the left of 0.5 should be justified. That is, the LCM merely says to “apply the rule when the rule applies”—or as Barney would say, “Obey all rules!” Again, we are not saying that such a cutpoint is unjust, illegitimate, or unreasonable. We are simply saying that the cutpoint is a great way to decide cases and does such a good job, it makes the LCM superfluous.8

In the remainder of the paper, we take up more plausible interpretations of the LCM, including four from philosophers of justice: Rawls, Dworkin, Hart, and Aristotle. But before proceeding, we want to stress two points. First, the plethora of different interpretations of the maxim makes it clear that theorists need to specify carefully what they mean when establishing likeness. Second, while we generally focus on decisions where the outcome is either “just” or “unjust,” this is merely for expositional convenience. As we show in the formal proofs, the same logic applies if we use more discrete categories or a continuum. We reiterate that our results in no way limit the outcome set to merely just or unjust.

8There is an obvious similarity between this phenomenon and the classic Sorites Paradox. We compare the two in the appendix.
4 John Rawls and Consistency

One common interpretation of the LCM is consistency. This appears to be the view of John Rawls. Rawls posits a system of axioms that would “satisfy a certain conception of justice” and institutions that are “im impartially and consistently administered by judges and other officials. That is, similar cases are treated similarly, the relevant similarities and differences being those identified by the existing norms” (Rawls 2009, 50-51). Here, Rawls incorporates the like-cases maxim into the foundation of his Theory of Justice. While he acknowledges the maxim alone “is not a sufficient guarantee of substantive justice,” Rawls does believe it can prevent some forms of injustice: later, he argues that the maxim and other basic principles “impose rather weak constraints on the basic structure, but ones that are not by any means negligible” (Rawls 2009, 51, 209). The injustices avoided are somewhat unclear, but Rawls seems to have in mind primarily practical injustices: an example he gives is of someone in an unjust regime being able to protect themselves because the unjust regime at least acts predictably.

Rawls’ interpretation could take two forms. First, it might merely require the same cases to be treated the same. If ‘different cases’ are identical across all meaningful dimensions, then they are just different instances of the same case, and the rule that decides one decides all. This interpretation of the LCM is too minimalist for several reasons. Most immediately, the maxim plainly talks about like cases, not same cases. Moreover, it is not clear which cases count as the same. Finally, if the remaining axioms are sufficient to decide a case the first time, they should be sufficient to decide a case in each repeated instance. So the LCM would not decide any new cases, it would merely assert that one should “obey all rules!”

A second and more charitable view of consistency runs as follows. Rawls posits a system where prevailing norms decide which cases are alike. Then, institutions consistently and impartially apply axioms to decide those cases. If \( S \) is a set of cases that prevailing norms regard as similar, but the axioms do not decide every case in \( S \) alike, then Rawls’ interpretation of the LCM would reject

---

\[ ^9 \text{As others have noted, it is not obvious that one should prefer to live under a predictable rather than an unpredictable despotism. This, though, is not our main objection to consistency.} \]
this system as unjust. Nonetheless, if 90% of cases in $S$ are judged to be just and 10% to be unjust, Rawls’ LCM can only say that this is inconsistent; it cannot tell us whether any particular case in $S$ is just or unjust. Because his LCM only operates as an ex-post check on a theory, it does not decide any cases, and so it is superfluous to the theory itself. It does not impose any obligation on a person or institution within a system to decide cases in a certain way. At best, it provides grounds to criticize a system as a whole, but it does not offer any particular guidance.

5 Ronald Dworkin and Integrity

For some time British judges declared that although members of other professions were liable for damage caused by their carelessness, barristers were immune from such liability...integrity condemns the special treatment of barristers unless it can be justified in principle, which seems unlikely. The House of Lords has now curtailed the exemption: to that extent it has preferred integrity to narrow consistency. Integrity will not be satisfied, however, until the exemption is entirely erased. \cite{Dworkin1986, 220}

Ronald Dworkin offers a second interpretation of the LCM. Under political integrity, cases should be decided alike unless there is a principle of justice that differentiates them. As we understand him, Dworkin offers both a pure theory of law (what justice demands) and an ideal-type description of practical law in a particular kind of society. We set the latter aside. Our goal is rather to ask if political integrity has any place in a pure theory of justice.

As a principle of interpretation, political integrity requires us to regard the law as if conceived and formulated by a single, coherent author. It insists that, whenever possible, a jurist should seek the broadest possible set of axioms (within certain limits) from which the law might follow. Thus,

\footnote{Dworkin works hard to distinguish his idea of integrity from consistency, which he sees as too narrow. “Is integrity only consistency (deciding like cases alike) under a prouder name? That depends on what we mean by consistency or like cases” \cite{Dworkin1986, 219}.}

\footnote{Dworkin’s view of integrity has come under fire by Réaume \cite{1989}, Raz \cite{1992}, and others, who argue that such integrity is not morally desirable.}
in the preceding example, Dworkin objects to the special treatment of barristers because there is no comprehensive conception of the law that might have led to their special exemption from liability. While preserving the exemption might be consistent with past practice, it violates the spirit from which the law flows. The exemption, he argues, ought therefore to be eliminated. Integrity thus interprets like-cases to require that any differentiation of outcome must be explicitly justified by a principle of justice.

The weakness with this interpretation appears once we try to extend it beyond a single example and incorporate it into a theory of justice. The principle excludes special treatment for barristers because there is no group in Dworkin’s example receiving special treatment. But suppose that our theory of justice allowed for such a group; for instance, suppose that British law also excused Oxford dons from liability for their students’ failing marks, and let us suppose further that this excuse is justified in principle. Integrity would require that barristers be treated the same both as Oxford dons (not liable) and as other professions (liable) unless a principle of justice could establish otherwise. But clearly, there must be some principle of justice establishing the correct answer—otherwise, our British cousins would be caught in an impossible dilemma. And if such a principle exists, then we do not need integrity to decide whether barristers should be liable.

This example conveys a general fact. Invoking integrity can never do any actual work in a pure theory. If it would exclude an outcome, it can do so only by referring to a set of principles that would have excluded the outcome already.

6 H.L.A. Hart and Relevant Aspects

its leading precept is often formulated as ‘Treat like cases alike’...So when, in the name of justice, we protest against a law forbidding coloured people the use of the public parks, the point of such criticism is that such a law is bad, because in distributing the benefits of public amenities among the population it discriminates between persons who are, in all relevant aspects, alike. (Hart [1961], 159)
H.L.A. Hart provides a third view of the LCM. Like Rawls and Dworkin, Hart begins with the intrinsic and objective features of a case. But instead of deciding the justice or injustice of the case directly, he first imposes an intermediate step that he calls ‘aspects.’ These aspects, or qualities, of a case determine whether the cases are alike. Cases sharing the same relevant aspects should have the same outcomes. For instance, driving twenty or thirty miles-per-hour above the speed limit are two different states of the world, but they would share the same quality of driving too fast. Since they share this ‘too fast’ quality, they should be decided the same.\(^\text{12}\)

Hart’s is the subtlest and perhaps most common view of the LCM. Before analyzing it, we note that his account implicates two points raised before. First, suppose the rules require that all people are welcome in public parks and that race is not relevant to park access. Further, suppose the state forbids minorities from parks. This would violate the LCM. But notice that it also violates the rules. A state that will violate the rules is unlikely to be moved by the addition of a rule that says “Obey all rules.” If this is all Hart’s LCM is after, it is superfluous.

Second, one must be careful not to abuse the LCM as a bid for moral superiority. If a person asserts that race is not a relevant characteristic in how to treat other people and then treats people differently based on race, Hart could appeal to that person on the grounds of the maxim. He could alert the person that their actions are not treating people alike whom that person believes to be like one another. But such an argument will not work with a racist who believes that racial differences matter. We would join Hart in saying the racist is wrong, but Hart must also recognize that, according to his own theory, the racist is not violating the LCM.

That said, we now attempt to explicate how reasoning through aspects/qualities would work. Hart is rather vague on this score, so we attempt to be charitable. The key point is that for Hart, cases are not collections of objective facts about the world; instead, they are collections of qualities. Thus, judgment is a two-step process: a theorist first maps from objective facts into aspects, and then from aspects to outcomes. To prevent arbitrariness sneaking in through these (likely subjective) aspects/qualities, there must be some mapping from objective states of the world to

\(^{12}\)We believe this is an accurate representation of Hart’s view. It matches the explication of Hart in [Winston (1974)].
qualities.

Qualities can be either discrete or continuous. For example, drivers could have the characteristic of driving ‘too fast’ (a discrete variable), and police searches could be more or less invasive (a continuous variable). Suppose the aspects are discrete, for instance a case where the relevant quality is driving too fast. Drivers going 100mph and those moving at 110mph both have the quality of driving too fast. As such, they are the same case for Hart, since cases are bundles of relevant aspects, and these two cases share the common relevant, discrete aspect of going too fast. At best, this would reduce the LCM to “the same cases should be decided the same.”

Now, suppose relevant aspects are continuous. We assume for the moment that there is an easy mapping from the continuous set of cases to the continuous set of qualities and turn our attention to the second stage, the mapping from aspects to outcomes. It is immediately apparent that Hart does not tell us how to handle cases that are slightly different in some continuous factor. The LCM would require us to figure out which are like others and treat them alike, but the definition of likeness is unspecified. Presumably, we must adopt some interpretation that defines likeness.

For instance, suppose we adopted a convexity rule when determining whether a search is justified. Under such a rule, if cases $A$ and $B$ are both just, then any case $C$ on the line between them must also be just. Convexity would require the sets of just and unjust cases be divided into convex sets. Figure 1 shows how justice might demand an ever increasing change in the level of exigency to justify a marginal increase in intrusiveness. But notice that this does not divide the cases into convex sets, as the set of unjust cases is not convex. This example shows the deep problems with a convexity rule, but more importantly, it shows that processing objective cases through a set of continuous but subjective qualities will not fix the LCM. All the problems associated with close-enough rules, consistency, integrity, or any other interpretation will still apply.

To escape these difficulties, a theorist might be tempted to use aspects not only to decide cases but to establish likeness: cases that have the same qualities are alike. But then likeness would

---

This is an oversimplification. We develop convexity and explain its benefits and limitations in the appendix.
Figure 1: Search and Seizure
no longer be a function of the intrinsic features of a case. Before, cases shared qualities because they were alike, but under this view cases would be alike because they share the same qualities. This reversal would interpret the LCM as follows: cases that have the same aspects are alike, and cases that have the same aspects are decided in the same way, so cases that are alike are decided alike. This is a logically coherent understanding of the maxim, but it comes at a high cost. Notice that intrinsically similar cases do not share qualities because they are objectively similar; rather, cases that may or may not be intrinsically similar are alike because they share the same qualities. Likeness is no longer derivative of intrinsic or objective criteria, and likeness is not used to map from these characteristics to relevant aspects. Instead, it is the qualities that are at the root of the enterprise, and cases may be assigned to those qualities arbitrarily. This approach would justify arbitrariness rather than contain it. As such, it is inconsistent with the normative purpose of the LCM, even if it is logically coherent. By this understanding, Hart’s interpretation could not solve his own example, since it could only be the law (and not their intrinsic qualities) that made people of different races alike.

7 Aristotle and Proportionality

This, then, is what the just is—the proportional; the unjust is what violates the proportion. (Nicomachean Ethics V.3)

The like-cases maxim originates with Aristotle, and so it is perhaps fitting that we should conclude with his interpretation. His approach seems to be that justice is the like-cases maxim, and this maxim requires that desert be proportional to certain facts. Thus, he argues that the best flutes should be given, not to those of wealth or physical stature, but to the best flautist.

An enormous literature dwells on the idea of proportional justice, and so we discuss it with some trepidation. Our objective is not to argue for or against this approach, but rather to ask what

---

14“Justice is held by all to be a certain equality...justice is a certain thing for certain persons, and should be equal for equal persons. But equality in what sort of things and inequality in what sort of things—this should not be overlooked” (Aristotle, Politics III.12).
role in it the LCM can logically play. As with the other interpretations, we conclude that it must be
either pernicious, superfluous, or arbitrary. However, we also believe Aristotle intends the LCM
more as an *ex post* check on theories of justice, a role for which it is well-suited. But before we
reach that conclusion, we must make four points.

First, a theorist must be very careful to establish the proportionality rule at the outset. While
Aristotle seems to suggest that justice should follow an arithmetic proportion, Aquinas later sug-
ests that a geometric proportion is best in some circumstances.\(^{15}\) In general, there are infinite
ways to define proportions to achieve an infinite number of distributions. In order to actually de-
cide cases, the LCM would have to define the type of proportion to be used. This decision could
easily be arbitrary.

Still, proportional accounts of justice avoid most of our previous concerns as they rely on a
continuous outcome set. Proportional justice requires that the same proportion relate the relevant
normative considerations to the final distribution, and by definition, proportions lie along a con-
tinuum between zero and one.\(^{16}\) The earlier theories fell largely because they attempted to decide
cases in discrete ways, for example as just or unjust. Aristotle does not follow this path.\(^{17}\) He is
concerned with administrative acts that apportion goods rather than moral or legal judgments on
acts or institutions.

Second, what is particularly interesting about this interpretation is that, for Aristotle, the LCM
may be less about promoting equality based on similarities and more about honoring the differences
between cases. Aristotle’s LCM says that it is unjust for a person who works five hours at a task to

\(^{15}\)“The general form of justice is equality, in which distributive and commutative justice agree. But there is equality
in distributive justice by geometric proportionality, and equality in commutative justice by arithmetic proportionality”
(St Thomas Aquinas *ST* II-II Q.61). Of course, once we permit a variety of proportions, the maxim risks becoming
arbitrary, since many rules could then satisfy the LCM.

\(^{16}\)It is an obvious limitation of proportional justice that it will require the normative considerations be commensu-
ral and measurable and the final good to be sufficiently divisible.

\(^{17}\)Proportionality also suggests another interpretation of the LCM: continuity. Continuity is the belief that small
changes in cases should produce small changes in outcomes. While a bit oversimplified, the idea is that a marginal
change in a case should not lead to a large ‘jump’ in their outcomes. As cases change continuously, so should outcomes.
Continuity is a nice property for outcomes that do not fall into discrete categories. Continuity might be a good thing
for determining the length of prison sentences or the size of fines. But continuity is superfluous in that it cannot
determine how fast fines can increase or even whether prison terms should move up or down from a base case. It
can only guarantee that outcomes move continuously as cases change. We address this possible interpretation in the
appendix.
make the same as a person who works ten hours. Aristotle may thus be using the maxim to attack equality of outcome. On this understanding, Aristotle intends not to show that “like cases should be treated alike,” but that unalike cases should be treated differently.

This leads to our third point. To put it bluntly, Aristotle’s proportionality leads to conclusions that violate our deeply held beliefs. Most standards of justice require that even those unable to contribute to the production of any good (e.g. infants, the seriously ill, or the insane) are entitled to a minimum standard of living. Under Aristotle’s proportional rule, it is unjust for those who contribute nothing to receive anything. One need not be a Rawlsian to believe that infants have a moral claim to sustenance and the insane to the protection of the state. Proportional justice—at least as defined in Aristotle—denies these claims. As such, it is pernicious.

It is possible to avoid these consequences. A rule could specify that goods should be proportional to need, or that proportionality kicks in only after some threshold is met. As we are not trying to launch a fusillade against the entire literature on proportional justice, we take no stand on whether these efforts succeed. But we do insist that they are no longer based on treating like cases alike. Proportionality may be a very good way to think about justice, but if so, it is only because it is not derived from the LCM.

And so we conclude with a final point. We do not believe Aristotle intends the LCM to decide any case. Instead, he seems to intend the LCM as an *ex post* check on theories of justice. He argues that any theory of justice must take likeness (proportion) into account; he is specifying the form a theory should take. If a theory does not treat like cases alike and unalike cases differently, then it is not a good theory. On this much, we agree, but once again the LCM is superfluous to the theory itself; it decides nothing. To the extent that later theorists have seized upon the maxim to derive claims about justice, they have departed from Aristotle’s vision. As we hope to have shown, this departure is untenable.

---

18See e.g. the Universal Declaration of Human Rights Articles 25-26.
8  Formal Analysis

Our formal argument proceeds as follows. Any theory must first identify the relevant dimensions of different states of the world (cases). Once it has done this, it considers how different the cases are from one another along these relevant dimensions. Having identified the set of different cases and what makes them differ, the theory has effectively created a metric space, \((X, \rho)\), where \(X\) is the set of cases and \(\rho\) is a measure of the differences between cases. Since \(X\) is a metric space, if two cases \(x, y\) share all relevant characteristics in common, then they are the same case and \(\rho(x, y) = 0\). If \(\rho(x, y) > 0\) then there is some difference between the two cases on some relevant dimension.

Since the LCM aims to eliminate arbitrariness, we assume that the likeness and difference between cases for purposes of the maxim must be grounded in objective factors. Hence, the judgment that case \(x\) is like case \(y\) must relate to the distance metric.

Define \(C\) as the set of possible judgments and let \(r : X \rightarrow C\) be the rule that renders judgment on any case \(x \in X\). A rule may be (partially) induced by a set of axioms \(A\). If one of the axioms is some form of “like cases should be decided alike,” we say the theory includes the like-cases maxim (LCM). We define a theory as \(T = \{X_t, A_t, C_t\}\). In other words, a theory takes a case as a primitive and then decides that case using a rule, \(r_T\), that is consistent with the axioms \(A_T\) to determine the judgment, \(c \in C_T\).

We assume that within the set of all cases, there is a subset of cases satisfying the following assumptions:

**Assumption 1. An Interesting Subset**

1. There is a set \(S \subset X\) that is strictly convex, simply connected, and compact. We refer to this set as \(S\) throughout the proofs.\(^{19}\)

2. There exist \(x, y \in S\) such that \(r(x)\) and \(r(y)\) are defined but \(r(x) \neq r(y)\).

\(^{19}\)Case space and policy space models both assume vector space frameworks that easily satisfy these conditions. Strict convexity is borrowed from Takahashi (1970) and formally defined in the appendix.
3. Let \( V \) be the set of all rules \( r' \) which are consistent with the axioms \( A \) and for which \( r'(x) \) is defined for all \( x \in X \). \( V \) is not empty.

The first assumes that justice will sometimes depend on one or more factors that are connected (such as a driver’s speed)\(^{20}\) The second, that for some such factor, not all outcomes are the same (e.g. driving 25mph through a residential subdivision is permissible, but driving 180mph is not). And the last, that all cases can in principle be decided, even if a particular theory does not decide them all (e.g. a theorist might be unsure how to deal with a particular case—like whether it is permissible to drive 60mph—but his theory does not prevent that case from having a decision).

Put differently, we allow a theory to be incomplete so long as any attempt to decide the rest of the cases would not necessarily cause a logical contradiction.

We believe that any interpretation of the LCM must do two things. First, it must establish which cases count as alike. Second, it guarantees that all like cases are decided alike. Conceptually, this means that every interpretation will have the following form:

1. Define some reflexive relationship \( R \) on elements of \( X \), where \( xRy \) means that ‘\( x \) is like \( y \).’ (Note that the relationship need not be symmetric or transitive.)

2. Require that if \( x \) is like \( y \), then \( r(x) \) and \( r(y) \) are alike. That is, \( xRy \implies r(x) = r(y) \) or, in the case of continuous outcomes, \( r(x) \approx r(y) \).

**Different Interpretations**

We now formalize several possible ways to define the relationship \( R \).

**Interpretation 1. Close-Enough**

*For any \( x, y \in X \), \( xRy \) if \( \rho(x, y) \leq k \), where \( k > 0 \) is some constant. If \( xRy \), then \( r(x) = r(y) \).*

**Interpretation 2. \( \epsilon \)-Rule**

*For any case \( x \in X \) there exists some \( \epsilon > 0 \) such that for any case \( y \in X \), if \( \rho(x, y) < \epsilon \), then \( xRy \). If \( xRy \), \( r(x) = r(y) \).*

\(^{20}\)Further motivation for assumption is provided in the appendix.
Though similar, interpretations [1] and [2] differ in an important respect. The first asserts that for any case $x$, any case $y$ such that the distance between $x$ and $y$ is less than or equal to $k$ must be like $x$. That is, cases that are close enough to each other (based on a pre-specified distance) must be like one another. The second does not specify the distance guaranteeing likeness.

**Interpretation 3. Convexity**

For any $x, y, z \in X$, if $\rho(x, z) + \rho(z, y) = \rho(x, y)$ and $r(x) = r(y)$, then $yRz$ and $xRz$. If $xRz$, then $r(x) = r(z)$.

**Interpretation 4. Connectedness**

For every $x \in X$, there exists a connected set $Y \subset X$ such that $x \in Y$ and $r(y) = r(x)$ for all $y \in Y$.

**Interpretation 5. Circular**

$xRy$ if $r(x) = r(y)$.

**Interpretation 6. Basic Consistency**

Let $a \in A$ be Consistency. For all $x \in X$, if $A \setminus a \Rightarrow r(x) = c$, then $r(x) = c$.

**Interpretation 7. Same Case**

For any $x, y \in X$, if $\rho(x, y) = 0$ then $xRy$. If $xRy$, then $r(x) = r(y)$.

**Interpretation 8. Consistency (Rawls)**

Define $A_x = \{ y \in X | xRy \}$ and $B_x = \{ y \in X | r(y) = r(x) \}$. Consistency holds if $A_x \subset B_x$.

**Interpretation 9. Integrity (Dworkin)**

Let $a \in A$ be Integrity. For any $x, y \in X$, then $xRy$ unless $A \setminus a \Rightarrow r(x) \neq r(y)$. If $xRy$, then $r(x) = r(y)$.

**Interpretation 10. Aspects (Hart)**

Define some function $q : X \rightarrow Q$ where $Q$ is capturing the ‘qualities’ of a case. The Aspects interpretation may take one of two forms:
1. If \( xRy \), then \( q(x) = q(y) \) and \( r^*(q(x)) = r^*(q(y)) \), where \( r^*: Q \rightarrow C \).

2. If \( q(x) = q(y) \) then \( xRy \) and \( r(x) = r(y) \).

**Interpretation 11. Continuity**

For all \((x_n)_{n \in \mathbb{N}} \subseteq X\),

\[
\lim_{n \to \infty} x_n = c \implies \lim_{n \to \infty} r(x_n) = r(c)
\]

**Proportionality (Aristotle)**

Suppose there is an infinitely divisible good \( G \) to be divided between \( n \) persons \( i \in \{1, 2, \ldots, n\} \). Suppose \( G \) is in part produced by the combined efforts of all players contributing up to \( m \) different inputs, \( j \in \{1, 2, \ldots, m\} \). Let the input matrix be \( Z \), which contains the amount of each input contributed by each individual. We say \( z_{ij} \in \mathbb{R} \) is individual \( i \)'s contributed amount of input \( j \) and \( Z_i = \{z_{i1}, z_{i2}, \ldots, z_{im}\} \) is the vector that contains the amount of each input contributed by \( i \). We say \( Z_i = 0 \) if \( z_{ij} = 0 \) for all \( j \).

Let \( G = \{g_1, g_2, \ldots, g_n\} \) be a vector of the amount of good \( G \) provided to each individual. We say that \( \hat{G} = \{\hat{g}_1, \hat{g}_2, \ldots, \hat{g}_n\} \) is the observed distribution of \( G \). We now define two functions: \( f: Z \rightarrow G \) and \( f_i: Z \rightarrow \mathbb{R} \). These functions simply take the inputs provided by all individuals and return the amount of good \( G \) that each individual should receive. Interpretations of the LCM that emphasize proportionality place restrictions on \( f \) and \( f_i \). Possible restrictions include (but are not limited to):

1. If \( Z_i = Z_k \), then \( f_i(Z) = f_k(Z) \) for all \( i, k \).

2. If \( z_{ij} \geq z_{kj} \) for all \( j \) then \( g_i \geq g_k \). The relationship is strict if \( z_{ij} \geq z_{kj} \) for all \( j \) and \( z_{ij} > z_{kj} \) for at least one input, \( j \).

3. \( \hat{g}_i = f_i(Z) \)

4. \( f_i(0) = 0 \)

\[\text{This approach largely tracks how Sen (1996) discusses proportional justice.}\]
We can distinguish a strict from a weak interpretation of proportional justice:

**Interpretation 12. Strict Proportionality**

Restrictions 1-4 obtain

**Interpretation 13. Weak Proportionality**

Restrictions 1-3 obtain

**Objections to the Interpretations**

We identify three potential problems an interpretation may face. An interpretation could fail because it is pernicious, superfluous, or arbitrary. A pernicious interpretation is one that violates core moral intuitions or leads to a contradiction. It is difficult to formalize exactly what we mean by ‘core moral intuition,’ but we have in mind statements such as “premeditated murder is unjust,” “infants should not be starved to death,” or “there should be at least some minimum amount of clean air.” An arbitrary interpretation is one that determines likeness without regard to the objective similarities of cases, i.e. without reference to $\rho$. An interpretation is superfluous if it does not help decide cases.

Formally, consider two theories $T = \{X, \mathcal{A}, \mathcal{C}\}$ and $T' = \{X, (\mathcal{A} \setminus a), \mathcal{C}\}$ where $a \in \mathcal{A}$ is the LCM. The rules induced by the theories are $r : X \rightarrow \mathcal{C}$ and $r' : X \rightarrow \mathcal{C}$, respectively.

**Definition 1. Superfluous**

An interpretation of the LCM is superfluous if any of the following hold:

1. For all $x \in X_T$, $r(x) \in \mathcal{C}$, $r'(x) \in \mathcal{C}$, and $r(x) = r'(x)$

2. There is no $x \in X_T$ such that $r'(x)$ is undefined but $r(x) \in \mathcal{C}$.

3. Compared to $X$, the set of cases decided by the LCM has measure zero.\(^{22}\)

\(^{22}\)It is technically necessary to belabor the definition of (3). We say a theory is superfluous if there is a theory $T^*$ such that, if the LCM is replaced by an axiom $a^*$ which is a function over some $Z \subset X$ and the rule induced by $T^*$ is $r^* = r$, then the set $Z$ has measure zero compared to $X$. 

23
We now state our major finding. The complete proof is in the appendix, but we include several parts here.

**Proposition 1.** Each of interpretations 1-13 is either Pernicious, Superfluous, or Arbitrary.

We prove this proposition through a series of lemmas under assumption 1 above. Since each interpretation of the LCM applies to all cases (the full set \(X\)), it must also apply to all subsets of cases, including \(S\). In many of the lemmas below, the strategy is to prove that the proposed interpretation contradicts the assumptions that define subsets like \(S\)—i.e. we analyze sets of cases that can be decided, vary along some continuous variable, and contain at least two possible outcomes.

The LCM must function when Assumption 1.2 applies; if it only applies when all cases share the same outcome, the LCM is essentially trivial. Similarly, we believe the LCM must apply when cases vary along some continuous variable, since the world is generally continuous. We do not assume that a theory decides all cases, though we do assume that a theory must not prevent some case from having a decision. (For example, we consider cases where the speed of a car is relevant. Assumption 1 requires that the speeds are continuous, as are the cases, but some speeds are legal and others are not, and every speed is susceptible to judgment.)

**Lemma 1.** Close-Enough Is Pernicious

**Proof.** We define a k-chain \(Y \subset X\) as follows:

1. \(\forall y \in Y\), there exists a \(z \in Y\) s.t. \(\rho(y, z) \leq k\), and \(\exists x \in Y^c\) s.t. \(\rho(y, x) > k\) \}

2. The only subsets of \(Y\) that satisfy 1 is \(\{Y, \emptyset\}\).

Two things follow immediately. First, since \(S\) is simply connected, there is some k-chain \(Y\) such that \(S \subset Y \subset X\). Second, if \(Y\) is a k-chain, then if \(\exists y \in Y\) such that \(r(y) = c \in C\), then under a close-enough rule \(r(z) = c\) for all \(z \in Y\), which in turn means \(r(z) = c\) for all \(z \in S\). But this contradicts Assumption 2. So the close-enough rule is pernicious.

**Lemma 2.** The following interpretations are Superfluous: Circular, Basic Consistency, Same Case, and Rawlsian Consistency.
For Circular, Basic Consistency, and Same Case, this is obvious. For Rawlsian Consistency, notice that it simply restates the generic form of the LCM. Recall that any interpretation of the LCM must first define the relation \( R \) and then ensure that \( xRy \implies r(x) = r(y) \). We show that \( xRy \implies r(x) = r(y) \) iff \( A_x \subseteq B_x \) where \( A_x = \{ y \in X | xRy \} \) and \( B_x = \{ y \in X | r(y) = r(x) \} \).

First, suppose \( xRy \implies r(x) = r(y) \). If there was some \( y \in A_x \) but \( y \notin B_x \), then \( xRy \) but \( r(x) \neq r(y) \). Now suppose \( A_x \subseteq B_x \). Then for any \( y \in A_x \), we have \( xRy \) by the definition of \( A_x \), \( y \in B_x \) by assumption, and thus \( r(y) = r(x) \) by the definition of \( B_x \). Accordingly, Rawls simply restates the LCM without giving it any content.

**Lemma 3. Integrity (Dworkin) is Superfluous or Pernicious**

*Proof.* Suppose for any \( x \in S \) we have \( r'(x) \in C \). Then plainly Integrity would be superfluous. Now suppose there is some \( x \in S \) such that \( r(x) \) is undetermined. By assumption, there are at least two cases \( y, z \in S \) such that \( r(y) \neq r(z) \). By integrity, \( r(x) = r(y) \) and \( r(x) = r(z) \), but by assumption \( r(y) \neq r(z) \). So there is a contradiction. \( \square \)

**Lemma 4. Convexity is Pernicious**

Let \( x, y \in S \) be cases such that \( r(x) \neq r(y) \) and define \( J = \{ z \in S | r(z) = r(x) \} \) and \( U = \{ z \in S | r(z) = r(y) \} \). Then \( B = cl(J) \cap cl(U) \) is geodesically convex.

Roughly, the lemma claims that for any two points on the border, the shortest path between those two paths lies on the border. In the traditional vectors spaces of the case-space and policy-space models, this is obvious, because collecting cases into convex sets requires separating hyperplanes. The Separating Hyperplane Theorem would do the job there. However, we prove the rough equivalent for metric spaces. Convexity makes it essentially impossible to place upper or lower bounds or to require increasing amounts of one thing to offset the loss of another.

*Proof.* Every complete and convex metric space is a length space. ([Khamsi and Kirk](2011), 35). By assumption, \( S \) is compact and convex, and is therefore a proper length space. The Hopf-Rinow theorem guarantees that for any two points \( x, y \) in a proper length space, there exists a geodesic...
path that joins \( x \) and \( y \). Therefore \( S \) is a geodesic space. Further, since \( S \) is strictly convex, the geodesic connecting any two \( x, y \in S \) is unique, so \( S \) must be a uniquely geodesic space.

Consistent with the assumptions, pick \( x, y \in S \) such that \( r(x) \neq r(y) \). Let \( U = \{ z \in S | r(z) = r(x) \} \) and \( J = \{ z \in S | r(z) = r(y) \} \). If the LCM (Convexity) holds, then \( U \) and \( J \) are convex subsets of \( S \) and further are both geodesically convex. Papadopoulos (2005, 76) proves that the closure of a geodesically convex subset of a uniquely geodesic space is also geodesically convex. So \( \text{cl}(U) \) and \( \text{cl}(J) \) must be geodesically convex. Obviously, the intersection of geodesically convex subsets of a uniquely geodesically convex space is convex. So \( B = \text{cl}(J) \cap \text{cl}(U) \) is geodesically convex.

**Lemma 5. Aspects (Hart) is Superfluous, Pernicious, or Arbitrary**

Interpretation 10.1 is obviously incomplete, since it hasn’t defined likeness. Presumably, this interpretation would have to be coupled with another interpretation from this paper, which would make it either superfluous, pernicious, or arbitrary. (In section 6, we couple it with convexity and show that it is pernicious.)

Interpretation 10.2 is clearly arbitrary. It makes no reference to \( \rho \), the metric over cases. Consequently, a philosopher employing the LCM could decide any two cases he pleased were alike—the LCM would provide no check at all. While this is not a formalized objection to theory, we believe it is clearly untenable.

**Lemma 6. Strict Proportional Justice (Aristotle) is Pernicious. Weak Proportional Justice Is Superfluous**

Since strict proportionality requires that those unable to contribute to the production of a good be denied a distribution of that good, it violates our core beliefs and is pernicious, just as convexity was. A weaker form of proportional justice could drop the last requirement. The remaining two would simply require that the individual identities should not influence outcomes—only the individuals’ contributions should—and that those who contribute more should receive more. Notice
that these are merely restrictions on \( f \); they do not define \( f \). These restrictions function as a check on a moral theory, but they do not actually decide cases. Hence, this interpretation is superfluous.

9 Conclusion

The like cases maxim is supposed to guard against arbitrariness in theory and in practice. By requiring objectively similar cases be decided alike, it is meant to prevent judges and legislators from discriminating unfairly. Some theorists have suggested the maxim cannot accomplish much of this mission on its own; nonetheless, philosophers from Aristotle to Rawls have assumed it is a hallmark of justice. This confidence is misplaced. At best, the maxim may help adjudicate between a few contending theories of justice, even if theoretically it only reiterates what other principles already say (e.g. that a person’s race should not affect the severity of his punishment). At worst, the maxim logically entails conclusions that contradict serious moral intuitions. Thus, invoking the like-cases maxim is not just a waste of paper; it may also be deeply insidious.

We could not have reached this conclusion without the assistance of formal reasoning. Other thinkers have intuited some of our findings, but they often take them too far: Westen (1982) and other critics tend to believe the maxim is necessarily vacuous, while Hart (1961) and other proponents tend to believe the maxim can helpfully decide cases when joined with other principles. By formalizing both sides, we discover that neither is correct. Naive criticism fails to see that the maxim, including its original interpretation in Aristotle, can in fact meaningfully decide cases. Naive optimism fails to see that the maxim, when pursued to its logical conclusion, must lead to unacceptable results.

This result has direct implications for longstanding debates in political and legal theory. Legal realists like Llewellyn have a strong point about the difficulty of applying the maxim; responses to their critique from Hart, Dworkin, and others seem to fall apart upon close inspection. Similarly, insofar as Rawls relies on the maxim to develop general rules to govern a just society, rebutting the utilitarian preference for individualized and subjective reasoning, his effort will falter unless
alternative grounds can be found for his claims. Finally, these conclusions speak to ongoing work about the nature and normative implications of inequality. We worry that people who are alike in all relevant ways are still treated differently in their income, wealth, or opportunities. Framed in this way, concerns about inequality rely implicitly on the LCM as a principle of justice, which we have shown it cannot be.

Our formal analysis also indicates how we might recover a role for the like-cases maxim. We have shown how the LCM should not play a role in pure theories of justice. But in the course of this inquiry, we also showed its promise in adjudicating between theories. Further, we suspect that the maxim might be usefully employed as a rule of thumb by judges and lawmakers attempting to put justice into practice. Nonetheless, since it is not a principle of justice, it cannot justify overturning longstanding traditions or precedents. The maxim has no moral force of its own, and so it cannot justify sweeping change of any kind. It therefore cannot bear the burdens most often laid upon it, namely, questions of discrimination and equal treatment. To these, it can give no answer.
References


URL: http://www.jstor.org/stable/825866


URL: http://www.jstor.org/stable/1829633


URL: http://www.jstor.org/stable/2265406


Strauss, David. 2002. “Must Like Cases be Treated Alike?” The Law School, University of Chicago.

URL: http://ssrn.com/abstract_id=312180


Universal Declaration of Human Rights. 1948.


Appendix

Sorites Paradox

In the Sorites Paradox, the argument runs that if one grain of wheat is not a heap, then two grains do not make a heap. If two grains do not make a heap, then three grains do not, and so on. Essentially, assuming \( n \) grains do not make a heap, and supposing there is no discernible difference between \( n \) and \( n + 1 \) grains, then \( n + 1 \) grains do not make a heap either.

A few key differences distinguish the paradox from the close-enough rule (Interpretation [1]). First, while the Paradox explicitly relies on sequential decisions, a pure theory of justice decides all cases simultaneously. In the Sorites Paradox, whether \( n \) grains constitutes a heap depends on whether the examiner begins with a single grain and adds grains one at a time, or whether she begins with a million grains and begins removing them one by one. Second, the paradox depends on vagueness while the close-enough rule operates on clarity. The Sorites Paradox defines \( A \) and \( B \) as like if the observer cannot tell them apart. The close-enough rule is based on concrete differences. Finally, the Sorites Paradox is something of the converse of the LCM. The LCM says that cases should be decided alike (both heaps) if they are alike (if \( n \) is like \( n + 1 \)). The Paradox says they are alike (\( n \) is like \( n + 1 \)) because they are decided alike (both heaps).

Convexity

Suppose sentencing a child to a life sentence with the possibility of parole or to the death penalty are both unjust. It would seemingly follow that sentencing a child to life without parole is also unjust. Such analysis relies on understanding the LCM to imply convexity as a principle of justice—i.e. if two cases \( A \) and \( B \) share the same outcome, then any case \( C \) between \( A \) and \( B \) should also share their outcome.

Convexity is an attractive property for several reasons. First, convexity imposes the kind of consistency on decision-making that Rawls and Dworkin desire. If we assume there are no other intervening elements to distinguish cases, it would seem wrong if a theory decides stealing $30
and stealing $50 are unjust but stealing $40 is just. But unlike mere consistency, convexity does
decide cases. Knowing that stealing $30 and $50 are both unjust and adding the LCM decides the
case of stealing $40. Thus, the second attractive property of convexity is that it provides a type
of consistency while remaining useful. The third benefit is its boundedness. Recall that the close-
enoough rule tended to metastasize and swallow the entire space. Where the close-enough rule leads
to a potentially unconstrained chain reaction, convexity stops the chain reaction at a boundary.

Nonetheless, convexity is a very demanding assumption, and any theory that complies with it
must make very sharp distinctions. Consider a world where justice requires a minimum threshold
of education and food. Then the unjust set is an L-shaped region (see figure 2). In this case, the
just set is convex but the set of points that are unjust is not convex. In figure 2, cases $A$ and $B$ are
clearly unjust. Under a principle of convexity, then case $C$ must also be unjust. But clearly case
$C$ has sufficient education and sufficient food. This problem is easily generalized: it is impossible
for a theory of justice both to set minimum thresholds and to understand like-cases as convexity.
(Note that this result does not require limiting the analysis to only two potential outcomes.)

Thus, while it is possible to divide cases into convex sets, one must make very stark choices to
do so. Either entire connected sets must share an outcome as under the close-enough rule, or the
axioms that split the space must comply with a very sharp, essentially linear, divide between the
cases.

It is tempting to try to salvage convexity and avoid making such stark choices by limiting
convexity to one set. For example, in figure 2, the set of just cases is convex. Several immediate
problems arise from such a move. First, the maxim plainly says that like cases should be decided
alike, not that just cases should be decided alike. Second, it implies that two points extremely
close together might both be just (or unjust) but a point between them might be the opposite (thus
undermining the intuition behind the interpretation in the first place). Finally, it is not obvious that
the just set will always be convex. Take for instance a ceiling on air and water pollution as shown
in figure 3, where we are sometimes willing to trade a little more smog for a little less sewage.
Here, the unjust set is convex while the just set is not.
Figure 2: Convexity with Minimum Thresholds
Figure 3: An Example of a Convex Unjust Set
In short, convexity offers some real improvements over consistency and the close-enough interpretations. However, these benefits come at a high cost. To ensure convexity, axioms must make incredibly sharp distinctions across a host of competing factors. Basic tools such as minimum or maximum thresholds are nearly impossible to use. Indeed, any theory of justice that uses multiple thresholds or even a single boundary that curves (as in figure 3) will conflict with convexity, thus making this interpretation pernicious.

Motivating the Assumptions

We begin by motivating the assumptions on subset $S$. To prevent bias and caprice, we ground cases in objective, physical facts. Recognize that the location and velocity of every particle in the universe can be represented in a sufficiently large but closed and bounded subset of $\mathbb{R}^{N,M}$, which is therefore compact. The product of all of the compact metric spaces for each particle is the set of all possible worlds and is compact by the Tychonoff Theorem. Histories and futures are continuous paths through the set of possible states of the world and are each compact. The product of all histories, futures, and current states—which are thus all compact—is also compact again by Tychonoff. We call this set $\Omega$. Thus all possible physical states of the world (including histories and futures) reside in a compact set $\Omega$. This $\Omega$ is obviously overly inclusive and redundant. For example, the world in which humans never evolved is in $\Omega$, but it will not generate any cases of interest. Therefore we are only interested in a subset of all possible states, $\Omega^* \subset \Omega$. We can easily take the closure of these morally relevant states, which ensures that $\overline{\Omega^*}$ is closed and therefore compact. We let a surjective function $\Pi : \overline{\Omega^*} \to \tilde{X}$ map states of the world into the set of cases.

Individual cases are elements $x, y, z \in \tilde{X}$. Comparing cases implies a distance metric, $\rho^*$ defined on $\tilde{X}$ with a distance function $\rho^*(x, y)$, where $\rho^* : \tilde{X} \times \tilde{X} \to \mathbb{R}^+$. So for any three points $x, y, z$ in $\tilde{X}$:

1. $\rho^*(x, y) \geq 0$, if $x = y$, then $\rho^*(x, y) = 0$
2. \( \rho^*(x, y) = \rho^*(y, x) \)

3. \( \rho^*(x, z) \leq \rho^*(x, y) + \rho^*(y, z) \)

Thus the set of possible cases is a pseudometric space\(^{23}\) We now say \( x \sim y \) if \( \rho^*(x, y) = 0 \). Let \( X \) be the quotient space \( \tilde{X}/\sim \) and \( p : \tilde{X} \to X \) be the canonical projection that maps each point of \( \tilde{X} \) onto its equivalence class\(^{24}\). Further, let \( \rho \) be the metric on \( X \) defined as \( \rho(a, b) = \rho^*(p^{-1}(a), p^{-1}(b)) \) for \( a, b \in X \). Now \((X, \rho)\) is a traditional metric space.

Most—if not all—moral theories, \( T = \{X_\tau, A_\tau, C_\tau\} \), have the following property:

**Property 1.** There exists some compact subset \( W \subset \Omega^*, S \subset X_\tau \), and a continuous, surjective function \( \Pi^* : W \to S \) such that for every \( \omega \in W \) and for every \( \delta > 0 \), there exists \( \omega' \in W \) such that \( 0 < \tilde{\rho}(\omega, \omega') < \delta \) where \( \tilde{\rho} : W \times W \to \mathbb{R} \) is defined as \( \tilde{\rho}(\omega_1, \omega_2) = \rho^*(\Pi^*(\omega_1), \Pi^*(\omega_2)) \).

This property simply says that there are some continuous changes in the natural world that lead to continuous changes in the set of cases. For example, suppose one is worried about water pollution. The amount of pollutants in the stream varies continuously both in the physical world and in the set of cases concerned with pollution. Similarly, if the speed of a car is relevant, that speed varies continuously in the physical world and a moral theory cares about continuous speed as well. The only two ways a theory would not have Property 1 are if the theory did not care about any continuous features of the world or if it did not base likeness on objective, physical characteristics.

If a theory satisfies Property 1 then by construction \( S \) is a compact, simply connected subset of \( X_\tau \). We will assert that \( S \) is strictly convex, defined as follows:

**Definition 2.** **Strictly Convex Metric Space**

A metric space \((M, d)\) is a convex metric space if there is a mapping \( W : M \times M \times [0, 1] \to M \) (i.e. \( W(x, y; \lambda) \) defined for all pairs \( x, y \in X \) and \( \lambda \in [0, 1] \)) and valued in \( X \) satisfying

\[
d(u, W(x, y; \lambda)) \leq \beta d(u, x) + (1 - \lambda)d(u, y)
\]

\(^{23}\)The difference between a pseudometric space and a metric space is that in the former, one cannot use the distance measure to distinguish between two points. That is, in a metric space \((M, d)\), if \( x \neq y \) then \( d(x, y) > 0 \). In a pseudometric, the inequality is not strict.

\(^{24}\)Howes (2012).
for all \( u \in M \). We say \( M \) is strictly convex if for any \( x, y \in M \) and \( \lambda \in [0, 1] \) there exists a unique \( z \in M \) such that \( \lambda d(x, y) = d(z, y) \) and \( (1 - \lambda)d(x, y) = d(x, z) \).

We note that case-space models used in the judicial politics literature and policy-space models used throughout the social sciences assume a vector space structure. Vector spaces are similarly simply connected and strictly convex, but they also have further restrictions. Our assumptions are more general than current state of the art models.

Further Lemmas

Lemma 7. An \( \epsilon \)-rule Is Pernicious

Proof. If every point in \( S \) is in the domain of \( r(\cdot) \), the rule induced by the axioms of the theory, then the \( \epsilon \)-rule will be pernicious. To see this, assume that \( r(x) \in C \) for all \( x \in S \). Further, by assumption 2 there exist two points \( y, z \in S \) such that \( r(y) \neq r(z) \). Since \( S \) is simply connected, there exists a path \( g : [0, 1] \to S \) such that \( g(0) = y \) and \( g(1) = z \). Define the set \( A = \{a \in [0, 1] | r(g(a)) = r(y)\} \) and let \( \alpha = \sup(A) \). Plainly \( \alpha \in (0, 1] \). Further, since \( g \) is continuous, \( \forall \epsilon > 0 \exists \delta > 0 \) such that \( b - a < \delta \implies g(b) - g(a) < \epsilon \).

If \( r(g(\alpha)) = r(y) \), then for any \( \epsilon > 0 \exists \delta \) such that \( r(g(\alpha + \delta)) \neq r(y) \) even though \( \rho(g(\alpha), g(\alpha + \delta)) < \epsilon \), which violates the \( \epsilon \)-rule. Similarly, if \( r(g(\alpha)) \neq r(y) \), then for any \( \epsilon > 0 \exists \delta \) such that \( r(g(\alpha - \delta)) = r(y) \) even though \( \rho(g(\alpha), g(\alpha - \delta)) < \epsilon \).

Lemma 8. Connectedness Is Superfluous

Proof. Notice that a single point is connected, so any rule \( r(\cdot) \) satisfies connectedness. Hence connectivity is superfluous.

Lemma 9. Continuity Is Pernicious or Superfluous

Proof. Suppose \( C \) is finite. Then continuity reduces to the \( \epsilon \)-rule, and it is pernicious. Suppose \( C \) is continuous. It is obvious that the only way continuity could establish the outcome of a case would be if the case in question is the limit of a convergent sequence. However, the set of such cases in comparison to \( X \) has measure zero.
Additional Concerns with Proportional Justice

There are two additional ways in which proportionality falls short. Since proportional justice requires a continuous set of possible outcomes, it is unable (by design) to answer discrete moral questions, e.g. whether an action or institution is just, illegal, etc. While a proportionality rule could theoretically determine how much wealth every individual should have, it could not tell us whether or not lethal self-defense was or was not justified. This seems like a serious shortcoming.

Second, Aristotle insists that like cases be decided alike and that different cases be decided differently. This would imply that \( f \) should be injective. But this might be nearly impossible when \( m > 1 \). For example, suppose \( G \) is the result of labor and capital with each factor equally important. Consider two different situations: 1) \( A \) and \( B \) each provide half of the labor and half of the capital; and 2) \( A \) provides all of the capital and \( B \) all of the labor. Under a reasonable proportionality rule, \( A \) and \( B \) should split the gains equally in both situations. However, they appear to be different cases, so \( f \) would not be injective. These two cases seem to be ‘like’ each other only in that they lead to the same outcome. A proportional interpretation would thus risk becoming circular.