Why Should Like Cases Be Decided Alike?
A Formal Model of Aristotelian Justice

Benjamin Johnson and Richard Jordan*
Princeton University
Baylor University

March 1, 2017

Abstract

The maxim "treat like cases alike" has been a bulwark against arbitrary rule and a core feature of almost every theory of justice since Aristotle. The maxim features in the largest debate in legal theory in the twentieth century, the response of legal positivists and Dworkin to legal realists. Yet, despite its prominence, it has received almost no sustained attention. From the very beginning, Aristotle treated the maxim as positive as well as normative. Yet the last formal theorist to investigate the maxim may have been Aristotle himself. We believe that a neglected but fruitful way forward is to reengage and advance Aristotle's positive approach. We undertake such an analysis using the framework of metric spaces. We find that, far from a core feature of justice, the maxim is either unhelpful or pernicious. Our findings should encourage political and legal theorists to reexamine some core assumptions of our normative theories.

*Department of Politics, Princeton University 130 Corwin Hall, Princeton, NJ 08544-1012 E-mail: bbj@princeton.edu; Department of Political Science, Baylor University E-mail: richard_ jordan@baylor.edu

> [J]ustice 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

A core concern in political and legal theory is avoiding arbitrariness when deciding cases. Deeply rooted in our notions of legitimacy and the rule of law, what distinguishes a judge from a despot is the substitution of generally applicable rules for arbitrary fiat. This concern has traditionally sounded in a maxim traced to Aristotle: *like cases should be treated alike*. On a plain reading, the like cases maxim (LCM) requires judgments to track intrinsic similarities between cases. If case A is objectively *like* case B, then the two cases should be decided in a similar way. By establishing a rule that defines similarities according to cases' objective characteristics as opposed to the subjective perceptions of the judge, the maxim forces decisions to be consistent with rules. The LCM is a concept enjoying nearly universal approval—and yet it has received almost no sustained inquiry.

The LCM is present in theories of justice going back thousands of years, and it 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 grounded in shared, objective features could hold ground against utilitarian theories that take into account 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 that judges can and will do as they please). These same concerns animate ongoing work about income inequality, racial discrimination, and global justice. All appeal to the basic belief that like persons should not be treated differently.

But it is not clear that the maxim is up to this task. Many have noted that the like-cases maxim risks being too weak for the mission, yet 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 seemingly common intuition that while "by itself" the maxim does not do much, given a set of substantive rules to build on, it can work justice. We disagree.

Instead, we show that if the LCM does augment substantive rules, it does so in a way that violates core moral intuitions; it is in fact pernicious. These interpretations of the maxim imply conclusions that would alarm most reasonable people. However, avoiding these consequences requires eliminating the maxim's ability to decide any cases at all, making it superfluous.

This does not mean that the LCM can serve no purpose: we suggest that it might be of some limited use as an ex-post check on theories to see whether or not 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 potentially very large herd.

If the LCM is superfluous or pernicious, then it is time to reopen some debates in legal and political theory, and it is imperative that theorists reexamine the foundations of major normative traditions. This is a strong claim, and so we need to clarify its scope. We limit our exploration to pure theory, by which we mean theories that provide a complete rule all at once. Since a pure theory can render judgment on any case at any time, the order cases arise should not matter. Of course in a common law tradition such as ours, we are conditioned to assume that the order of cases *does* matter. But we believe that as a matter of pure theory, whether the saintly widow or bitter miser brings the case first is irrelevant.

We are open to the possibility that the LCM may be useful *in practice*, especially as part of a judicial machinery deciding cases on an ongoing basis. But the LCM is presumed to have normative force of its own and to operate on the level of pure theory. Therefore, our results should redouble examination of the maxim in practice in light of its dubious value in pure theory. Perhaps there is practical value in the LCM even if there is no reason to maintain it as a matter of pure justice. That is a worthwhile study, but its value only becomes apparent after we have examined the LCM in a purely theoretical context.

Our argument proceeds by taking the notion of 'like' to imply two things: that there is a set of possible cases and that there is some measure of likeness between any cases, i.e. some metric. This allows us to formally define a theory using standard mathematical notation of metric spaces. Once we establish the possible conceptions of the maxim, we use these tools to consider Aristotle's own conception of the maxim alongside three modern theorists of law and justice: H.L.A. Hart, John Rawls, and Ronald Dworkin. Each of them uses the like-cases maxim in their theories, though each adopts a slightly different interpretation. We find that all four suffer the same limitations as the other interpretations we elucidate.

Insofar as possible, we would like to proceed as rigorously as we can. However, we recognize that this places some burdens on the reader. Accordingly, we have restricted the mathematical arguments to self-contained portions of the paper so that a reader can skip them if so desired.

Nonetheless, we believe our chief contribution is the development of a framework in which this type of analysis is made possible. In taking up the tools of formalism to engage an ancient question of political and legal theory, this article is a bridge between positive and normative theory. This type of work is valuable not only because it employs a different and rigorous methodology on important, timeless normative questions, but also because, as Knight and Johnson (2015) point out, the line between positive and normative theory is not always clear. Restoring the ancient bridges between formal and political theory sheds light on the maxim and holds out the hope that similar efforts on classic questions in political and legal theory 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 helps exclude kinds of arbitrary and unpredictable systems of rule (Rawls, 2009). For H.L.A. Hart, it joins with other principles of justice to exclude segregation and racial injustice. For Ronald Dworkin, the maxim is the foundation of political integrity, the indispensable virtue guiding how judges should interpret the law. Kant believed that violating the maxim could incur "blood guilt" (Wood, 2010, 112); Locke, that just governance requires "promulgated established laws, not to be varied in particular cases, but to have one rule for rich and poor, for the favourite at court, and the country man at plough" Locke (1980). And Aristotle, the maxim's progenitor, used 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. Certainly, we might all intuitively agree that similar cases should result in similar outcomes. But what makes two cases similar?

A famous instance of this question is *Powell v Pennsylvania* (1888).[1] There, manufacturers of margarine challenged laws applying to margarine but not to butter as discriminatory and therefore a violation of the 14$^{\text{th}}$ amendment. Their challenge was dismissed.[2] The law decided that margarine manufacturers were like each other but unlike butter producers, and the law treated them accordingly. Like cases were treated alike.

As a matter of pure theory, this approach is clearly unacceptable. Regardless whether margarine and butter makers should be treated alike, 'likeness' should not mean merely whatever the lawmaker decrees. Otherwise, any discriminatory practice could (in theory) not violate the maxim; it could even be required by the LCM.[3] Equal protection under the Fourteenth Amendment—upon which rests the American edifice of civil rights—would be rendered vacuous.[4]

We are not the first to note that too thin an interpretation of the maxim reduces it to nothing

[1] See also McColgan (2006), who uses the case to ground her discussion of international human rights law.

[2] In short, as a matter of constitutional law, the Supreme Court *can* believe it's not butter.

[3] For instance, many laws in the South once asserted that different races implied different cases and under that regime, treating like cases alike resulted in Jim Crow.

[4] Modern legal regimes have attempted to circumvent this difficulty by explicitly stating what cannot be grounds for discrimination. While satisfying in practice, this approach only begs the question.

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 must result 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 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.[5] In fact, legislators often try to enforce a uniformity of decisionmaking on judges by limiting their discretion (e.g. through sentencing guidelines) in order to guarantee that like cases are treated alike.[6] 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.[7] Marmor (2005) takes this same theme and strengthens it, arguing that the like-cases maxim might be used to decide cases when all other reasons for a decision in the case 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 usefulness of the LCM in practice is 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 decision-making in both economics and political science. We track most

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

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

[7]See Sunstein (1993).

closely the work of Gilboa and Schmeidler, who develop a theory of case-based decision-making under uncertainty (Gilboa and Schmeidler (1995); Gilboa and Schmeidler (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 in the formal literature include Kornhauser (1992a), Kornhauser (1992b), Lax (2007), Landa and Lax (2007), and Landa and Lax (2009).

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.). Theories of justice can therefore be defined by the set of objects they take as inputs, the set of possible outcomes, and the guidelines they use to map from the former to the latter. We call these guidelines *axioms*.

The like-cases maxim is an axiom. When deciding how to judge a particular case, a theory that values this axiom must consider how like cases are being or should be decided. The maxim is thus a normative command that requires a particular judgment.

The LCM contains several distinct parts. First, there are cases that can be compared. 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 cases depend on subjective criteria that would enable arbitrary decisions and undermine the normative value of the maxim. These objective factors may be either continuous or discrete. Continuous facts are things like money, share of income, level of violence, or anything that can be conceptualized along a continuum. By contrast, discrete factors are categorical: male or female; citizen or non-citizen; alive, dead, or only mostly dead.

Second, the LCM requires these cases to be alike or unalike in some measure. 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. Further, we assume that these differences can be captured by some metric over the set of cases that describes a weak notion of relative difference. We do not imagine some philosophical ruler that would tell us how objectively different cases are. Instead, we only assume that the relative differences between a pair of cases can be compared.

Third, the LCM requires cases to be decided. Judgments are the set of possible decisions consistent with a theory. They describe whether a case is just or unjust, equitable or unfair. Finally, to decide what judgment to render, we rely on rules, which map from cases to judgments. In a theory of justice, these rules are not arbitrary but are induced by the axioms of the theory.We examine how the LCM affects the rules of a theory of justice.

We introduce our different definitions in the context of pure theories of justice. While we do not claim that this is an exhaustive list of all possible definitions, we do claim that this list exhausts our ability to think of credible definitions. We invite the reader to propose alternatives and to undertake a similar analysis to what follows.

The key characteristics of pure theory that we wish to emphasize are that a pure theory can (potentially) decide any case and the order of cases does not matter. We will limit most of our examples to theories of justice that take an input and determine whether it is just or unjust. Different theories might recognize partial justice, a continuum of justice, or other categories, but for the sake of clarity we have confined our discussion to the outcomes just and unjust.

We argue that any interpretation of the like-cases maxim must be either superfluous or pernicious. In the following sections, we define each of these flaws, and we group different interpretations of the maxim by the flaw it suffers. We can neither find nor imagine any interpretation that does not fall prey to one of these objections.

3.1 Formal

We believe it is important to ground cases in objective facts. If cases depend not on objective, physical facts—about past, present, or possible futures—then cases depend on subjective criteria that would invite bias and caprice.

We argue all possible physical states of the world (including histories and futures) reside in a compact set Ω.[8] This Ω is obviously overly inclusive and redundant. For example, the world in which humans never evolved is in Ω, but it will not generate any cases of interest. Therefore we are only interested in a subset of all possible states, $X \subset \Omega$. We can easily ensure that X is closed (by adding additional worlds to X as necessary), and X is therefore compact. We let the surjective function $\Pi : \Omega \to X$ map states of the world into cases.

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

1. $\rho(x, y) = 0 \iff x = y$

2. $\rho(x, y) = \rho(y, x)$

[8] In brief, the location and velocity of every particle in the universe can be represented in a sufficiently large but closed and bounded subset of $\mathbb{R}^{6,2}$, 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 that 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 Ω.

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

Thus the set of possible cases is a metric space.

A case $x \in X$ is composed of a finite set of elements, $x = \{e_1, e_2, ..., e_n\}$, that may be continuous or discrete, i.e. for $i \in \{1, 2, \ldots n\}$, $e_i \in \Delta \cup \Gamma$, where $\Delta = \{D_1, D_2, ..., D_m\}$ is the set of discrete elements and $\Gamma = \{\gamma_{m+1}, \gamma_{m+2}, ..., \gamma_n\}$ is the set of continuous elements that together constitute the set of possible elements for any case in X. For instance, a case might consist of the speed at which a driver was going (a value of a continuous element γ_{m+1}) and whether or not he was texting while driving (a value of a discrete element D_1).

Discrete elements imply a strictly positive distance between cases. Formally, if two cases $x, y \in X$ have different realizations of a discrete factor, $\exists \epsilon > 0$ such that $\rho(x, y) > \epsilon$. We say a *D-profile*, $\bar{D}_j$, is a combination of discrete values $\bar{\Delta} = \{\bar{d_1}, \bar{d_2}, ...\bar{d_m}\}$. A *D-profile* is the set of cases that hold all discrete elements fixed. If $\bar{X}_j \in X$, the subset of X consistent with D-profile j, is not a singleton, then cases in $\bar{X}_j$ must depend on at least one continuous element.

Recall that Π maps from worlds to cases. It may be that $\Pi : \Omega \to X$ has the following property.

Property 1. $\forall \omega \in \Omega,\ \forall \delta > 0, \exists \omega' \in \Omega$ *such that* $0 < \tilde{\rho}(\omega, \omega') < \delta$
where $\tilde{\rho} : \Omega \times \Omega \to \mathbb{R}$ *is defined as* $\tilde{\rho}(\omega_1, \omega_2) = \rho(\Pi(\omega_1), \Pi(\omega_2))$.

If Π has Property 1, then there is always some incremental change in the natural world that will imply a new case arbitrarily close to the initial case. This does not, on its own, imply that every case depends on at least one continuous factor. For that, we need discrete factors to be finite. Proposition 1 says that if there are only a finite number of possible discrete values of interest, then Property 1 is sufficient to guarantee that each case depends on at least one continuous factor.

Proposition 1. *If* Δ *and every* $D_i \in \Delta$ *is finite and* Π *has Property 1 then every* $x \in X$ *relies in part on at least one continuous factor.*

If cases rely on some continuous factor, then we are guaranteed a set of D-profiles, which is critical to our analysis. We assume that these sets of cases that share common discrete elements all have the following property.

Property 2. *For any* $\bar{X}_j \subset X$ *defined by D-profile,* $\bar{D}_j$*, and at least one continuous element* Λ*, then* $\bar{X}_j$ *is strictly convex,*[9] *simply connected, and compact.*

Finally, define $\mathcal{C}$ as the set of possible judgments and let $r : X \to \mathcal{C}$ be the rule that renders judgment on any case $x \in X$. A rule may be induced by a set of axioms $\mathcal{A}$. If one of the axioms is some form of "like cases should be decided alike," we say the theory includes the like-cases

[9] We formally define strict convexity in the context of metric spaces below.

maxim (LCM). We define a theory as $T_\tau = \{X_\tau, \mathcal{A}_\tau, \mathcal{C}_\tau\}$. In other words, a theory takes a case as a primitive and then decides that case using a rule, r_T, derived from applying axioms $\mathcal{A}_T$ to determine the judgment, $c \in \mathcal{C}_T$.

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.'

2. Require that $xRy \implies r(x) = r(y)$.

As most of the interesting parts of the maxim reside in the first step, we now turn to possible interpretations of the relationship R.

4 Pernicious

We begin with pernicious interpretations of the maxim. These interpretations, while at first glace seemingly innocuous, must necessarily entangle any theory in undesirable consequences; they are also the interpretations that bear the strongest resemblance to how the maxim is deployed in legal practice.

We say that an interpretation is pernicious if it leads to outcomes that grossly violate our moral intuitions or contradict other axioms. That is, if an interpretation logically entails conclusions that a reasonable person would find morally repugnant, we call that interpretation pernicious. While violating moral intuitions is not a logical threat to the maxim, it may be a very good reason to discard it. If the theory leads to the conclusion that all self-defense is unjust or that all murder is just, then we may think that the theory cannot be correct. If we do so, it is because we have greater faith in those moral intuitions than we do in the like-cases maxim. In contrast, if the maxim conflicts with other axioms rendering the theory self-contradictory, that is a conceptual objection to the theory as a whole. At that point, the theorist must choose whether to discard the conflicting substantive axiom or the like-cases maxim.

We focus on three possible definitions of 'like cases' that we take to be obvious candidates. First, we consider the notion that two cases are alike if they are *close enough* to each other. The second possibility is that if two cases have already been decided in the same way, then any case between them must be decided accordingly. Finally, we examine the possibility that the LCM applies not to cases directly but to normative qualities that cases might share. We show that all three of these interpretations are deeply pernicious.

4.1 Close-Enough Cases

Example 1. *A regime where wealth is concentrated in the top 1% is unjust. So a regime where wealth is concentrated in the top 1.1% is unjust.*

Example 2. *You shouldn't steal a handbag. Therefore, you shouldn't steal a movie.*

We begin with the most intuitive and powerful definition of like-cases: the idea that there is some measure of 'close-enough' such that any case that is close enough to another case with a known outcome shares that outcome. This 'close-enough' rule seems attractive because it provides certainty of outcomes when there is a relevant precedent. We say that A is really similar to B, and since A was just, B should be just also. Implicitly, we are saying that B is close enough to A to require that they share the same outcome. For theoretical purposes, though, a close-enough rule has major disadvantages. It is extremely powerful and ends up forcing too many cases to the same outcome.

The close-enough interpretation makes it impossible for theories to make sharp distinctions along continuous factors. For instance, if an age-related axiom says that anyone under 18 should be treated as a minor, that axiom is sufficient to govern decisions based on age. But if we try to apply a close-enough rule, we are saying something like, "any two people born less than two hours apart are alike." But this obviously swallows everything: in the hour before one turns 18, one is still a minor, but the adult born 119 minutes before you is treated differently. The LCM conflicts with the age limit and is pernicious.

There is an obvious similarity between this phenomenon and the classic Sorites Paradox. There, 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. 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.[10]

[10] Also, the Sorites Paradox is circular. It argues as follows:

1. The Paradox assumes that n and $n+1$ share an outcome.
2. The two cases are then alike because they share an outcome.
3. They should be decided alike because they are alike.

4.2 Convexity

Example 3. *Sentencing a child to a life sentence with the possibility of parole or to the death penalty are both unjust, therefore life without parole for a child is also unjust.*

Example 4. *Lying to a stranger is wrong. Lying to your spouse is wrong. Therefore, lying to a friend is wrong.*

Our next definition of like-cases is *convexity*. It requires that when 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 two closely related reasons. First, convexity imposes a kind of consistency on decision-making. If we assume there are no other intervening elements to distinguish cases, it would seem wrong if a theory decides stealing \$30 or \$50 is unjust but stealing \$40 is just. The second benefit of convexity is its boundedness. 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 clean air and clean water. Then the unjust set is an L-shaped region (see figure 1 below). In this case, the just set is convex but the set of points that are unjust is not convex. In figure 1, cases 1 and 2 are clearly unjust. Under a principle of convexity, then case 3 must also be unjust. But clearly case 3 has sufficient clean air and clean water. 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.

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 1, the set of just cases is convex. There are several immediate problems that arise from such a move. First, it is unclear which set—just or unjust—should be convex. Nothing in the maxim or this interpretation can answer that question. Second, at least one point and often a continuum of points will lie between two cases that are arbitrarily close to each other and yet turn out differently. Finally, the LCM traditionally applies to all cases, not merely to just or unjust cases: it does not urge us to treat unjust cases alike; it demands that we treat all cases alike. Creating only one convex set would violate the plain language of the maxim.

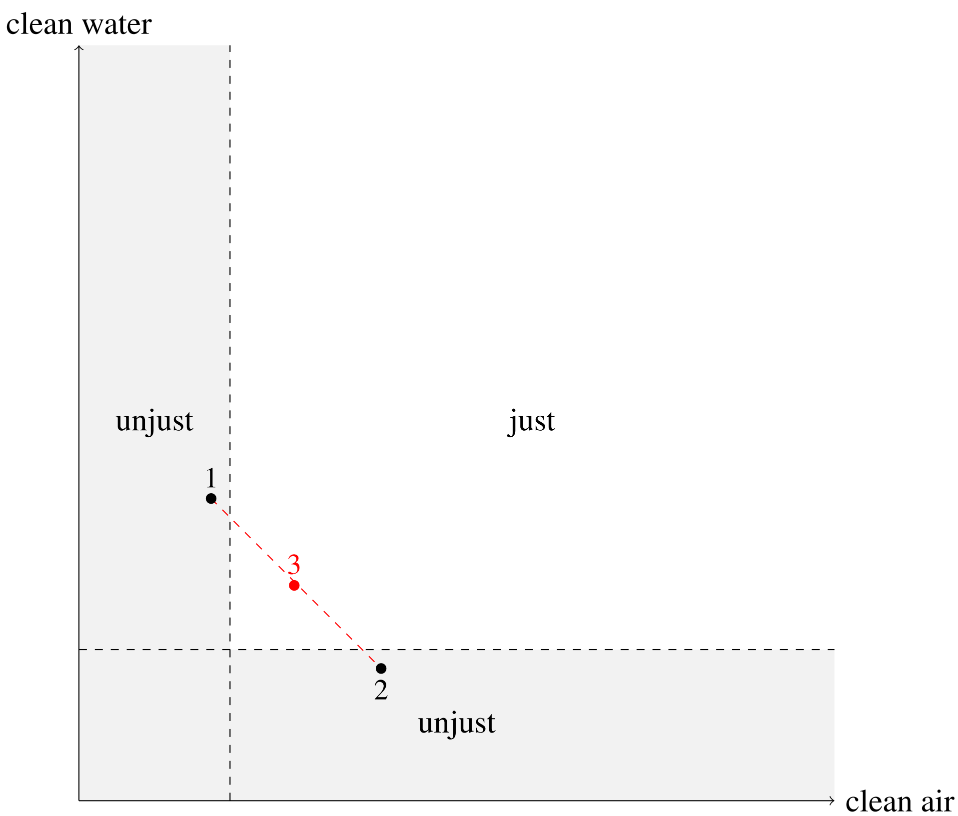

Figure 1: Convexity with Minimum Thresholds

4.3 Qualities

Example 5. *A case is unjust if a motorist drives too fast or if he drives too slow. If 25mph is too slow, and 15mph is too slow, then 20mph is also too slow.*

Example 6. *Entering an apartment without a warrant is just if the officer hears someone being harmed. Therefore entering an apartment without a warrant is also just if the officer believes the suspect is destroying evidence.*

A third way to operationalize the LCM is through the introduction of qualities. If justice is a matter of "fairness," for example, then it flows from abstract qualities or properties. One might also think of things as harm, pleasure, or other qualities such that the presence, absence, or level of the quality are constitutive of the case. There are two ways a theorist could bring qualities into the LCM analysis.

The first is simply to treat qualities as derived from the natural world. On this account, states of the world map to qualities. Qualities establish cases, and rules decide the case. This path does nothing to affect our analysis, as these qualities will be either continuous or discrete, and we may continue to compare them as before. Qualities would (on this interpretation) be logically identical to cases, and they would fall victim to the same problems as apply to cases.

For instance, consider the issue of search and seizure where intrusiveness and exigency are important qualities to consider. Figure 2 shows how one may think that justice demands an ever increasing change in the level of exigency to justify the marginal increase in intrusiveness. But notice that this does not divide the cases into two convex sets, as the set of unjust cases is not convex. So the convexity analysis from 4.2 holds, as would the remainder of our findings.

The other and more subtle way forward uses qualities not only to decide cases but to establish likeness. So one might say that two cases x and y are alike *because* they share quality P, and also that a case that has quality P is just. This guarantees that the like cases, x and y, are decided alike. But notice that this maneuver comes at the cost of grounding likeness in objective and intrinsic factors because P can be anything the theorist or judge desires. If cases x and y are just because a judge likes the outcomes, and cases x and y are alike because a judge likes both outcomes, that would count as treating like cases alike. This interpretation of the LCM would thus *justify* arbitrariness rather than contain it. As such, it is entirely inconsistent with the normative purpose of the LCM, even if it is logically coherent.

Thus, interpreting the LCM as qualities entangles a theorist in one of two difficulties: either qualities are grounded in cases' intrinsic features, and so bring with them all the problems detailed in this paper, or qualities are not grounded in cases' intrinsic features, and so justify arbitrary rule. A qualities interpretation is therefore either pernicious or, at best, unhelpful.

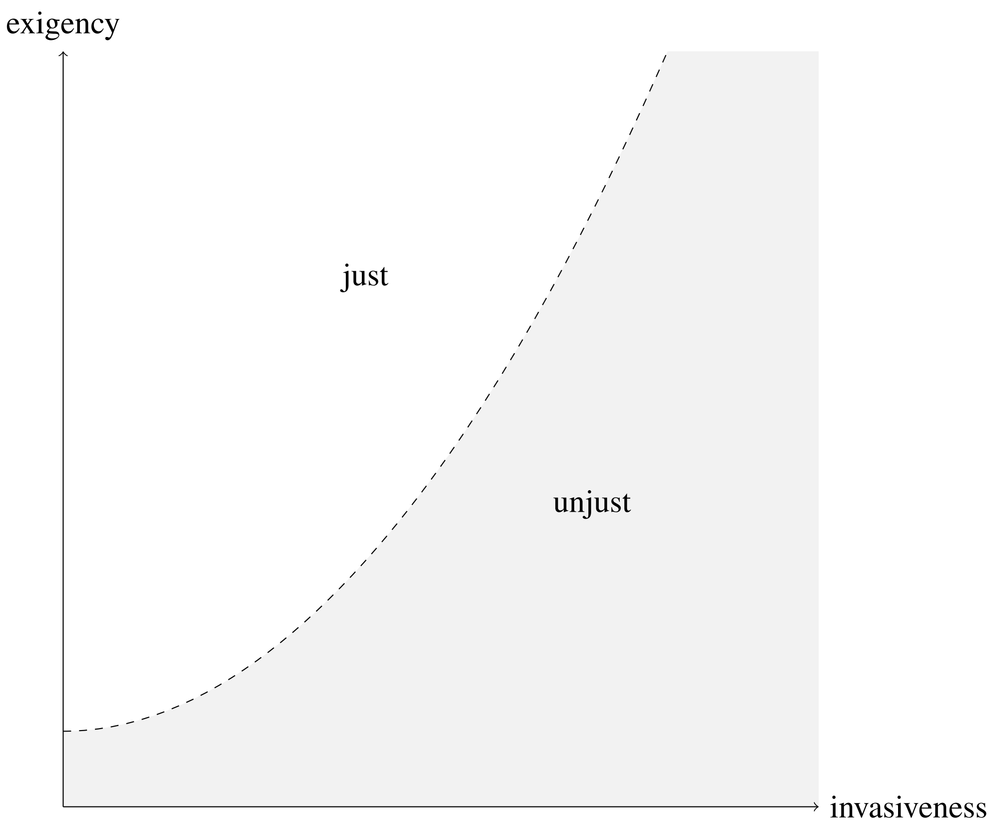

Figure 2: Search and Seizure

4.4 Formal

We now formalize the close-enough rule, convexity, qualities, and the dangers of each.

Interpretation 1. *Close-Enough Rule*
For any $x, y \in X$, xRy if $\rho(x, y) \leq k$, where $k > 0$ is some constant. $xRy \Rightarrow r(x) = r(y)$.

Definition 1. *k-chain*
A k-chain $Y \subset X$ has the following properties:

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

2. *The only partition of Y that satisfies 1 is $\{Y, \emptyset\}$.*

Proposition 2. *Properties of k-chains*

1. *If $\Pi : \Omega \rightarrow X$ has Property 1, the set of all k-chains is a partition of X.*

2. *If Property 1 holds, then every case is an element in some k-chain and no chain is a singleton.*

3. *If Y is a k-chain, then if $\exists y \in Y$ such that $r(y) = c \in \mathcal{C}$, then under a close enough rule $r(z) = c$ for all $z \in Y$.*

4. *If $\bar{X}_j$ is the subset of X consistent with D-profile j, then there exists some k-chain Y such that $\bar{X}_j \subset Y$.*

While Proposition 2 is obvious, the formalization provides more than just analytical rigor. The close-enough rule is even more powerful and pernicious than our initial casual treatment implied. The concept of k-chains is initially appealing, since it seems to do exactly what the like-cases maxim should do: it collects like cases into sets and guarantees an identical outcome. But the key point made clear intuitively and formally is that k-chains can only be distinguished from one another by discrete factors that impose a consistently large difference between cases. When a theory requires that discrete factors are irrelevant to the decision, the distance between cases in the metric space X will not be large enough to separate cases, and all cases will fall into the same chain.

Interpretation 2. *Convexity*
For any $x, y, z \in X$, if $\rho(x, z) + \rho(z, y) = \rho(x, y)$ and $r(x) = r(y)$, then xRy, yRz, and xRz. If xRz, then $r(x) = r(z)$.

First we establish the following definitions

Definition 2. *Strictly Convex Metric Space*

A metric space (M, d) is a convex metric space if there is a mapping $W : M \times M \times [0, 1] \to M$ such that $x, y \in M$ and $0 \leq \beta \leq 1$ where $W(x, y; \beta)$ is valued in M and satisfies

$$d(u,\ W(x, y; \beta)) \leq \beta d(u, x) + (1 - \beta)d(u, y)$$

for all $u \in M$. We say M is strictly convex if for any x, y, λ there exists a unique $z \in U$ such that $\lambda d(x, y) = d(z, y)$ and $(1 - \lambda)d(x, y) = d(x, z)$.

Recall $\bar{X}_j \subset X$ is a strictly convex, compact, and simply connected subset of X consistent with D-profile, $\bar{D}_j$. We begin by noting that if there are C possible outcomes in $\mathcal{C}$, there is no guarantee that $\bar{X}_j$ may be partitioned into C convex subsets or fewer. If it cannot be so divided, then the convexity definition is impossible to apply. However, suppose $U \subset \bar{X}_j$ is the convex subspace consistent with unjust outcomes under a theory and $J \subset \bar{X}_j$ is the convex subspace consistent with a just outcome. As both U and J are convex.Define the boundary between the subspaces as $B = cl(J) \cap cl(U)$ and note that for any point $x \in B$ and for any $\delta > 0$, the δ-ball centered on x contains points in both U and J. Assuming the theory is a pure theory and decides all cases, $\{int(U), int(J), B\}$ is a partition of $\bar{X}_j$. The next proposition shows that this boundary B must make cleave the set of cases very sharply.

Proposition 3. *Characterizing Boundaries*

$B = cl(J) \cap cl(U)$ is geodesically convex.

The proof is in the appendix. Proposition 3 relies on the strict convexity of $\bar{X}_j$. Since there is only one shortest path between any two cases, proposition 3 asserts that the shortest distance between any two points in the boundary must be contained in the boundary. We note in passing that if $\bar{X}_j$ is a vector space then the Hahn-Banach Theorem would guarantee that the boundaries between the sets are contained in separating hyperplanes. However, as we cannot be certain that the set of cases satisfies the conditions of a vector space, we generalize this condition in the proposition.

Interpretation 3. *Qualities*

There are two different ways to think of the LCM in terms of qualities:

1. *$q : X \to Q$ and $r^\star : Q \to \mathcal{C}$, or equivalently $\Pi^\star : \Omega \to Q = q(\Pi(\omega))$ where Q is a metric space with distance metric $\rho^\star$*

   OR

2. *$\hat{q} : X \to \hat{Q}$, where $\hat{Q}$ is a set of discrete qualities and $\hat{q}$ is defined for all $x \in X$. Then for any two cases $x, y \in X$, $\hat{q}(x) = \hat{q}(y) \implies xRy$. If $xRy,\ r(x) = r(y)$*

The first way to implement the idea of qualities is simply to generate cases through a composite function. The result is a metric space of cases as before, but now the cases hinge on qualities derived from the objective world rather than the objective features of that world. Nothing in the previous analysis changes.

The second possibility trivially guarantees the LCM since likeness and shared outcomes are both implications of sharing the same qualities. But now likeness is asserted directly by the theory (e.g. both cases share the quality of being liked by a judge) instead of being tied to the intrinsic characteristics of the cases. Thus, the second definition is coherent, but it does nothing to mitigate biased reasoning and arbitrary judgments.

We note that it is technically possible to specify a form of the Qualities interpretation which is neither unhelpful nor necessarily pernicious. In the following example, Interpretation 3 *is* helpful, and the theory does not obviously contradict any of our moral intuitions. Consider a simplified theory of justice that deals only with speeding violations on an interstate; $X = [0, \infty)$ is a person's speed; $\mathcal{C} = \{\text{issue a fine, issue no fine}\}$; $\hat{Q} = \{\text{too fast, permissible, too slow}\}$; $\hat{q} : X \to \hat{Q}$ and $\hat{q}$ is defined for all $x \in X$; and the axioms governing justice are

$$\mathcal{A} = \begin{cases} \text{if } \hat{q}(x) = \hat{q}(y),\ xRy;\ \text{if } xRy,\ r(x) = r(y)\ \forall x, y \in X \text{ (LCM)} \\ \hat{q}(x) = \text{too slow iff } x \in [0, 20] \\ \hat{q}(x) = \text{too fast iff } x \in [80, \infty) \\ r(3) = \text{issue a fine};\ r(85) = \text{issue a fine};\ r(55) = \text{issue no fine} \end{cases} \tag{1}$$

This example accords comfortably with our moral intuitions, and the LCM is indispensable to ensuring that the theory is complete. The example makes a problem with this approach obvious, as well. The qualities in this example seem reasonable, but they need not be. The cases in X can be arbitrarily assigned to qualities, and the LCM provides no check on such arbitrary assignment. This approach, while logically sound, reintroduces the very arbitrariness the LCM is designed to prevent.

5 Unhelpful

If the previous interpretations are too strong—if they result in pernicious implications—then perhaps less strict interpretations of the maxim are in order. Unfortunately, softening the maxim robs it of any force at all. We now turn to a class of definitions we take to be unhelpful.

By unhelpful we mean that these interpretations do not help decide cases. Importantly, these definitions are not therefore useless: they may allow us to examine the results of a theory, verifying that it decides cases consistent with our intuitions. Theories that fail such a test might be rejected.

Thus, while the following interpretations of the maxim do not decide any cases, they may offer a yardstick to choose between competing theories. Nonetheless, by this very virtue they should also be discarded as part of the theory itself.

5.1 Consistency Within and Across Cases

Example 7. *Regardless whether a person is rich or poor, it is wrong to steal.*

Perhaps the most basic interpretation of like-cases is that the same case must be decided the same way each time. This account is too minimalist for three reasons. First, the axiom plainly talks about *like* cases, not *same* cases. Second, it is not clear which cases count as the 'same.' For one theorist, a rich person and a poor person each stealing bread might count as two instances of the same case; for another theorist, they might count as two different cases. Finally, and most importantly, 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. This renders the maxim superfluous.

Another common interpretation of the maxim is consistency—a requirement that judges apply legal axioms the same way in each case. As a matter of practical law, this means that the law's application to subsequent cases must align with its application in previous cases. This practice promotes predictability, a key facet of the rule of law.

Nonetheless, as a matter of pure theory, consistency is strictly pointless since the order of cases does not matter. As an interpretation of the like-cases maxim, it only enjoins theorists to apply the rules when they apply. It is thus superfluous. The maxims of a pure theory already decide the outcome of a case, and so it is unnecessary for the theory to reiterate that they apply.

Consistency is closely related to the traditional philosophical concept of supervenience, which says there cannot be a difference in A-properties without a difference in B-properties (Kim, 1984). In the context of rules, supervenience requires that there cannot be a difference in outcomes without a difference in cases. The main contrast between consistency and supervenience is that the former is a normative command while the latter is descriptive. The LCM says there *should not* be a difference in outcomes without a difference in cases, while supervenience says there *cannot* be a difference in outcomes without a difference in cases.

Supervenience obviously requires the same case to be decided the same way. If two cases truly are the same, then there is no difference in the cases (B-properties), so there can be no difference in the judgments (A-properties). But if we weaken the definition of supervenience to require that there be no difference in outcomes without a meaningful difference in the cases, then we move away from the same-case interpretation toward a broader notion of consistency. But having done so, we are left with the difficult task of deciding which differences are relevant or meaningful. This task can only be accomplished by the other axioms, reducing supervenience to the statement that cases cannot be decided differently unless the theory says they should be decided differently. Since pure theories are complete by definition, the axioms must decide all cases already. Consistency and supervenience thus only require cases to be decided alike when they are decided alike and differently when they are decided differently. This is obviously consistent but just as obviously unhelpful.

5.2 ϵ-Rule

Example 8. *If it is wrong to steal $100, it is wrong to steal about $100.*

In section 4.1, we considered a fixed limit around a decision within which all future cases must be decided similarly. We now posit that the limit might vary by case. That is, some cases cast larger shadows than others. We call these shadows ϵ, and so we call this interpretation an ϵ-rule.

The difference between an ϵ-rule and a close-enough rule is subtle but significant. To see the difference, assume that we know that case A is just. A close-enough rule says that all cases within a fixed distance from A are also just. In contrast, an ϵ-rule says that if we narrow the scope of comparison enough—if we make ϵ small enough—then we can always find a range around A where the cases are also just. Importantly, while we know such an ϵ must exist, we do not know what it is.

The ϵ-rule does not provide any predictive leverage; it is unhelpful. We know that there is some shadow ϵ such that any case within that x's shadow will share x's outcome. But we do not know how long that shadow is. Saying the action y is very close to x does not tell us anything about whether or not y is unjust. Thus the second and larger deficiency of this interpretation is that it cannot be used to argue from one case to another.

But the final deficiency is that the ϵ-rule faces an impossible dilemma. On the one hand, it may be no different than the close-enough rule, sharing the latter's pernicious effect of forcing all cases to the same outcome. On the other hand, if it is to avoid this effect, then it cannot be part of a pure theory at all. The intuition here is that, if cases turn out differently than others, then there must be some sort of boundary between them—a line of cases where they have one outcome on one side and a different outcome on the other. But the ϵ-rule requires that any case turning out one way must be surrounded by similar cases on all sides. So any theory that employs the ϵ-rule cannot decide any case along a boundary. Thus, adopting an ϵ interpretation of like-cases results in a theory that is either either incomplete or self-contradictory.

5.3 Connectedness and Path Connectedness

The notion of connectedness is closely tied to reasoning. If we believe some case to be just and we want to know whether some different case is also just, connectedness recommends that we look for a path of related cases along which each case is also just. By tracing a 'line of argument' in this way, we can reason from A to Z. Clearly this is a nice property that lawyers, judges, and theorists all presume regularly.

There are two ways to think of like-cases as connectedness. First, we could take a strong view and assume both that like cases should be decided alike and that unlike cases should be decided differently. This suggests that when there are only finitely many outcomes available, there can only be an equivalent number of connected sets of like cases. If the relevant consideration is only one-dimensional, then the strong form of connectedness implies a strict form of monotonicity. Thus the strong form of connectedness is sometimes helpful in that it decides cases, but it is often a bit too strong and leads to pernicious results. For example, consider the question of whether or not a person driving on a highway is acting justly. Suppose it is unjust to drive at an unreasonable speed, perhaps because it puts others at risk. Intuitively, the unjust set should include both going too slow and going too fast. But notice that this generates types of cases divided into three connected sets: too slow, acceptable, too fast. We immediately observe that going too fast and too slow are not like cases because they are not connected. They are also different from each other, and on the strong view, should be decided differently. This means that if we decide driving too slow is unjust, we must also decide driving too fast is just (or we must multiply the possible outcomes a case can take).

The weak form of connectedness does not have this problem. The too slow cases are alike, the too fast cases are alike, and the acceptable speeds are alike, but none of the sets must be like the others. This is to the good, but it comes at the expense of deciding unalike cases as alike. More generally, the weak form provides essentially no structure. The weak form can always be trivially satisfied by treating each case as a separate set.

The more important problem is that both the strong and weak forms of connectedness are unhelpful. Connectedness says that there is some way to divide up the cases so that like cases are connected. It does not tell the theorist what the partition is.

5.4 Continuity

Example 9. *If driving 20mph over the speed limit results in a $100 fine, then driving a little bit faster should not result in a much larger fine.*

Often, we think of a theory as deciding not whether something is just or unjust (or partially just) but as measuring how just, legal, good, etc. some particular case is. Obviously, some decisions about justice do not fall nicely into discrete choices. For example, for how long should a judge sentence a criminal to prison? How large should a fine be? These decisions have continuous outcomes. Given their continuous nature, it makes sense that the relationship between cases and decisions should likewise be continuous.

However, continuity is still unhelpful. Knowing the outcome of one case does not guarantee anything about the outcome of another case. Continuity guarantees small moves for sufficiently small changes in the case, but it says nothing about either the direction of those moves or how small the changes have to be to guarantee the outcome has not changed significantly. Just like with the ϵ-rule framework, we cannot generate any predictive power from the continuity interpretation. It cannot guarantee that two cases x and y that are very close are actually decided the same. It only guarantees that there is a similar case that will be decided very similarly to an initial case.

5.5 Formal

Definition 3. *Superfluous*
Consider the set of axioms $\mathcal{A}_T$ that belong to theory $T = \{X_\tau, \mathcal{A}_\tau, \mathcal{C}_\tau\}$, and suppose that one of these axioms $a \in \mathcal{A}_\tau$ is the LCM. We say that the like-cases maxim a is superfluous if for $T' = \{X_\tau, (\mathcal{A}_\tau \setminus a), \mathcal{C}_\tau\}$, the rule $r_{T'}$ induced by T' is unchanged so that $r_T(x) = r_{T'}(x)\ \forall x \in X$.

Interpretation 4. *Circular*
xRy if $r(x) = r(y)$. If xRy, then $r(x) = r(y)$.

Obviously, a circular interpretation of the LCM is superfluous. While circularity is not a logical fallacy, it does leave the premises in need of independent justification. Notice that convexity could be interpreted as circular for border cases (but not for other cases).

Interpretation 5. *Consistency*
Let $a \in \mathcal{A}$ be Consistency. For all $x \in X$, if $\mathcal{A} \setminus a \Rightarrow r(x) = c$, then $r(x) = c$.

Obviously, Consistency is superfluous.

Interpretation 6. *Connectedness*

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

- *Strong Form — $Y_c = \{y \in X | r(y) = c\} \implies Y_c$ is connected.*

Interpretation 7. *Path Connectedness*
A set $Y \subset X$ is path connected if for any two points $y, y' \in Y$ there exists some continuous function $f : [0,1] \to Y$ such that $f(0) = y$ and $f(1) = y'$.

1. *Weak Form — For every $x \in X$, there exists a set $Y \subset X$ such that $x \in Y$, $r(y) = r(x)$ for all $y \in Y$, and Y is path connected.*

2. *Strong Form* — $Y_c = \{y \in X | r(y) = c\} \implies Y_c$ *is path-connected.*

Notice that the strong form is not necessarily superfluous. Consider the following example of a theory focusing on interstate speeding violations: cases are speeds, $X = [0, \infty]$; $\mathcal{C} = \{\text{just, unjust}\}$; and

$$\mathcal{A} = \begin{cases} r(x) = \text{unjust} & \text{if } x < 45\text{mph} \\ r(x) = \text{just} & \text{if } x = 45\text{mph} \\ \text{Connectedness (Strong Form)} & \end{cases}$$

This is a complete theory for which the maxim does real work to decide cases. It is functionally equivalent to the cutline rule in that it returns the same case outcomes for all cases, but here the maxim is not superfluous. The potential for perniciousness of the interpretation, however, is obvious (it says driving at 180 miles per hour is just).

Interpretation 8. *ϵ-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, then $r(x) = r(y)$.

Proposition 4. *If the LCM has an ϵ-rule interpretation, then for any D-profile $\bar{X}_j \subset X$, either*

i) $r(x) = c \,\forall\, x \in \bar{X}_j$, *where c is constant*

ii) $\exists\, x \in \bar{X}_j$ *such that $r(x)$ is undetermined*

For proof of Proposition 4, see appendix.

Interpretation 9. *Equivalent Definitions of Continuity*
Suppose now that $r : X \to \mathbb{R}$. Then the following are equivalent

- *For all $(x_n)_{n \in \mathbb{N}} \subset X$,*
$$\lim_{n \to \infty} x_n = c \implies \lim_{n \to \infty} r(x_n) = r(c)$$

- *For any $x \in X$ and $\epsilon > 0$, there exists some $\delta > 0$ such that for all $y \in X$ with $\rho(y, x) < \delta$, we have $r(y) - \epsilon < r(x) < r(y) + \epsilon$.*

We begin with the sequence-based definition of continuity because it is the most substantively appealing. It most directly says that as cases get closer to some limiting case, the outcomes of those cases must approach the outcome of that terminal case. However, the close similarity with the ϵ-rule is most plain in the epsilon-delta definition of continuity.

Continuity alone is unlikely to provide sufficient conditions to satisfy normative concerns. It may do a bit better if combined with a monotonicity requirement, so that, for example, as cases become more egregious, the punishment becomes monotonically harsher. Otherwise, punishments could wax and wane as the cases become more egregious. But even this would be unlikely to satisfy our notions of justice, because the punishment may ramp up arbitrarily quickly at certain points, flatten out over too long a range, or otherwise violate our moral intuitions. Even if one could pin down sufficient limitations on first, second, and third derivatives to restrain these problems, the continuity definition would almost never decide a case.

6 Four Theorists of Like-Cases

Before concluding, we briefly consider four famous philosophers who have employed the like-cases maxim in their theories: John Rawls, Ronald Dworkin, H.L.A. Hart, and Aristotle. We ask if any of these figures provides an interpretation that is neither pernicious nor superfluous. We find that they do not. Each offers a variation on one of the previous interpretations with its same attendant flaws.

6.1 John Rawls

> similar cases are treated similarly, the relevant similarities and differences being those identified by the existing norms...[this formal justice] excludes significant kinds of injustices...[because it] secures legitimate expectations(Rawls, 2009, 50-51)

In the above passage, John Rawls incorporates the like-cases maxim into the foundation of his *Theory of Justice*. While Rawls acknowledges that the maxim alone "is not a sufficient guarantee of substantive justice" (Rawls, 2009, 51), he 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, 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. Nonetheless, as others have noted, it is not obvious that one should prefer to live under a predictable rather than an unpredictable despotism. At least in the case of the Soviet Union, it seems preferable that the laws be only infrequently and arbitrarily enforced rather than consistently and predictably.

Setting aside the practical concerns, we have already discussed the objections to consistency—namely, that it is strictly unhelpful to a pure or ideal theory of justice. As Rawls explicitly identifies his interpretation of the maxim with consistency, we therefore conclude that the maxim is superfluous to his theory.

6.2 Ronald Dworkin

> 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. (Dworkin, 1986, 220)

Ronald Dworkin explicitly connects his idea of 'political integrity' to the maxim that like cases should be decided alike. He then carefully differentiates integrity from a narrow interpretation of the maxim as consistency.[11] In this section, we briefly study his broader interpretation of the maxim as integrity, and we ask if this interpretation falls victim to any of the flaws we have articulated.

As we understand him, Dworkin is attempting to offer both a pure theory of law (what justice demands in every case) 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. We conclude that it does not.

Dworkin's view of integrity has come under fire by Réaume (1989), Raz (1992), and others, who argue that such integrity is not morally desirable. Our goal in this paper is not to address the morality of Dworkin's view; we are only asking whether or not his view is, as a matter of theory, helpful.

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, 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.

[11]"Is integrity only consistency (deciding like cases alike) under a prouder name? That depends on what we mean by consistency or like cases" (Dworkin, 1986, 219). Dworkin argues that integrity requires judges decide cases not so much consistent with previous cases but with an overarching theory of justice. Moreover, integrity demands that cases should not (in theory or in practice) be treated differently unless that difference can be justified with reference to the axioms of that pure theory.

Integrity thus interprets like-cases to require that any differentiation of outcome must be explicitly justified by citing a principle of justice. In the example of British barristers, this requirement obviously excludes them from special treatment under the law. Clearly, in the example Dworkin offers, like-cases as integrity is essential to his argument; without the principle, it would not be clear why barristers might not be able to receive special treatment after all.

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 why they should not be. But clearly, there must be some principle of justice establishing the correct answer—otherwise, our British cousins would be caught in an impossible dilemma. But if such a principle exists, then we do not need integrity to decide whether barristers should be liable.

This example conveys a general truth. 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. Thus, while we believe Dworkin successfully distinguishes integrity from consistency as a matter of practical law, as a matter of pure theory the idea is logically identical.

6.2.1 Formalizing Dworkin

If Dworkin means only that $r(x)$ must equal $r(y)$ unless the principles articulated in $\mathcal{A}$ imply otherwise, then clearly integrity, as a matter of pure theory, is just consistency and is therefore superfluous. We therefore examine a more plausible interpretation.

Interpretation 10. *Integrity*
For any $x, y \in X$, then xRy unless some $a \in \mathcal{A} \Rightarrow r(x) \neq r(y)$. If xRy, then $r(x) = r(y)$.

This interpretation requires all cases be treated alike unless a particular principle a articulated in $\mathcal{A}$ implies otherwise. It thus guarantees the broadest possible equality between cases, so that any difference must explicitly be justified in principle.

From this definition, integrity seems attractive at first glance. It clearly would solve the example Dworkin uses, that of barristers in the UK: if $r(\text{all other professions}) = \text{liable}$ and no $a \in \mathcal{A} \Rightarrow r(\text{barristers}) \neq \text{liable}$, then $r(\text{barristers}) = \text{liable}$. The difficulty comes when we attempt to situate integrity within any theory beyond a single example.

Consider a simple world in which $\mathcal{C} = \{\text{just}, \text{unjust}\}$ and there exists at least one $x \in X$ such that $r(x) = \text{just}$ and another $y \in X$ such that $r(y) = \text{unjust}$. Let $I \in \mathcal{A}$ be Integrity. Consider the set of axioms $A = \mathcal{A} \setminus I$. Suppose that for some $z \in X$, $r(z)$ is undefined according to A alone. Then by I, $r(z)$ must be both just and unjust. Yet this is a contradiction. Therefore no such z can exist. But if no such z exists, then $A = \mathcal{A}$, and so I would be superfluous.

This example can be easily generalized to the following claim, the proof of which is obvious:

Proposition 5. *If there exist $x, y \in X$ such that $r(x) \neq r(y)$, then Integrity is superfluous.*

6.3 H.L.A. Hart

> 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 like-cases maxim. 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 and that we have previously called qualities. Cases sharing the same qualities have the same outcomes. For instance, driving twenty or thirty miles-per-hour above the speed limit would count as two different cases, but they would share the same quality of driving too fast. Since they share this 'too fast' quality, they should be decided the same.[12]

Notice that Hart has not yet shown how to decide whether or not cases share the same qualities. There are essentially two possibilities. First, we could say that like cases should have the same qualities. Second, we could say that cases that have the same qualities are alike. The first of these reintroduces all of the problems detailed in the previous sections. Because it still requires mapping from a possibly continuous set into a discrete set based on the similarities and differences between cases, all that this option does is push the problem back one stage. The second option introduces a new possibility. On this view, likeness is no longer a function of the intrinsic features of a case. Before, cases shared qualities because they were alike, but under this view cases are alike because they share the same qualities. This reversal of the causal arrow yields Hart's interpretation of the maxim: cases that have the same qualities are alike, and cases that have the same qualities 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.

[12]We believe this is an accurate representation of Hart's view. It matches the explication of Hart in Winston (1974).

Notice that intrinsically similar cases do not share qualities because they are alike; 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 qualities. Instead, it is the qualities that are at the root of the enterprise, and cases may be assigned to those qualities arbitrarily. Thus, Hart's approach cannot even solve his own example, since by his account it is only the law (and not their intrinsic qualities) that make people of different races alike.

6.3.1 Formalizing Hart

Hart's interpretation is the same as Interpretation 3 above.

6.4 Aristotle

> 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.13; see also *Nicomachean Ethics* V.3)
>
> We understand that mean [justice] by the proportion of things to persons, namely, in such a way that as one person surpasses another, so also the goods allotted to one person surpass the goods allotted to another. (St Thomas Aquinas *ST* II–II, Q.61)

The like-cases maxim originates with Aristotle, and so it is perhaps fitting that we should turn to his interpretation after exhausting his successors. His approach seems to be that justice *is* the like-cases maxim, and this maxim requires that desert is proportional to a person's qualities. Thus, he argues that the best flutes should be given, not to those of wealth or physical stature, but to the the best flautist.

Example 10. *Alice and Robert operate a lemonade stand that makes \$100 during the day. Alice worked for six hours and Robert worked for four hours. Proportional justice requires that Alice gets \$60 and Robert gets \$40.*

This understanding of proportional justice relies on a continuous outcome set. Proportions are by definition points along a continuum between zero and one. Proportional justice requires that the same proportion relate the relevant normative considerations to the final distribution. It is an obvious limitation of proportional justice that it will likely require the normative considerations be commensurable and measurable and the final good to be sufficiently divisible. While this limitation is real, it is not necessarily a threat to the interpretation.

We have already described the operation and limitations of the maxim when the outcome set is continuous, and that analysis applies here. We also wish to highlight a subtlety related to what question confronts what type of actor. The administrator must answer the question, "What is the just allocation of this good?" The answer to that question resides along a continuum. But once the decision is made, it may be asked of a theorist whether or not the allocation made by the administrator is just. That question has a dichotomous outcome. Notice that the possible outcomes of the first question are the continuous set of cases that make up the inputs for the second question that returns a simple yes or no answer. Applying the maxim to the second question encounters all of the problems explained above.

It is possible, of course, for justice to obey a variety of fixed proportions.[13] This might overcome potential indeterminacy. But then the objection to proportional justice is contained in its name: that justice must obey a fixed ratio. Justice would be linear—and so it would have to satisfy the same cutline requirements as convexity: the only way for both the just and unjust sets to be convex is for something like a straight line to divide the outcomes; proportionality requires a similar divide.

A final remark. The idea of proportional justice is clearly traceable to Aristotle, and many of his most famous successors (including St Thomas Aquinas, quoted above) articulate it. Nonetheless, it is not clear that Aristotle himself believes it. Instead, Aristotle may use proportionality against certain contemporary, naive theories of justice—ones that too quickly embrace equality as the goal of law. On this understanding, Aristotle's intends not to show that "like cases should be treated alike," but that unalike cases should be treated differently. We remain agnostic as to which view of Aristotle is correct.

6.4.1 Formalizing Aristotle

Interpretation 11. *Proportional Justice*

There are two different but related ways one may formalize proportional justice. The first is to assume that there is a single unit of some good G to be divided among n players $i \in \{1, 2, ..., n\}$ based on each player's relative holdings of some quality q. A case is then a set $\tilde{\mathbf{q}} = \{q_1, q_2, ..., q_n\}$ where q_i is the amount of q held by player i. Then player i should receive g_i of G, where

$$g_i = \frac{f(q_i)}{\sum_{j \leq n} f(q_j)} \tag{2}$$

[13] "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).

and $\sum_i g_i = 1$. The second possible understanding of proportional justice is to consider a case $x = \{\tilde{q}, \tilde{g}\}$, where $\tilde{\mathbf{g}} = \{g_1, g_2, ..., g_n\}$.

The first understanding resembles continuity (Interpretation 9). Justice is deciding what your share should be, and that is continuous. Setting aside the obvious objections (this interpretation would require all factors to be continuous and infinitely divisible), while Aristotle offers more structure than the simplistic continuity in Interpretation 9, he does not offer enough to be determinate. The function $f(\cdot)$ in equation 2 could be constructed to satisfy almost any arbitrarily desired outcome.

The second understanding looks more like discrete outcomes, because every pair will be mapped to $\{$*just*, *unjust*$\}$. The first definition tells the administrator how to distribute the good. The second judges that distribution as either just or unjust. This second case is understandable as the discrete metric over the set of cases. That is, if we define $Y = \{x \in X | \forall g_i \in \tilde{g}, g_i = \frac{f(q_i)}{\sum_{q_j \in \tilde{q}} f(q_j)}\}$, then we have a metric space (X, d) where $d(y, z) = 0$ if $y, z \in Y$ and 1 otherwise. This is conceptually useful, but since X is a metric space, it must be that y and z are the same case. Thus the like cases maxim is doing no work; it is simply saying that the same case should be decided the same way.

7 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 even 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 that it is a hallmark of justice. We show that this optimistic 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 articulate (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 maxim, when pursued to its logical extreme, must lead to unacceptable conclusions.

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 and independent 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 yet 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

Dworkin, Ronald. 1986. *Law's Empire*. Cambridge, Massachusetts: Belknap Press.

Eskridge, William N and John Ferejohn. 1992. "Making the deal stick: Enforcing the original constitutional structure of lawmaking in the modern regulatory state." *Journal of Law, Economics, & Organization* 8 (1):165–189.

Gaus, Gerald. 2016. *The Tyranny of the Ideal: Justice in a Diverse Society*. Princeton University Press.

Gilboa, Itzhak and David Schmeidler. 1995. "Case-based decision theory." *The Quarterly Journal of Economics* :605–639.

———. 2001. *A theory of case-based decisions*. Cambridge University Press.

Hart, H.L.A. 1961. *The Concept of Law*. Oxford University Press.

Khamsi, Mohamed A and William A Kirk. 2011. *An introduction to metric spaces and fixed point theory*, vol. 53. John Wiley & Sons.

Kim, Jaegwon. 1984. "Concept of Supervenience." *Philosophy and Phenomenological Research* .

Knight, Jack and James Johnson. 2015. "On Attempts to Gerrymander "Positive" and "Normative" Political Theory: Six Theses." *The Good Society* 24 (1):30–48.

Kogelmann, Brian and Stephen GW Stich. 2015. "The Irrelevance of the Impossibility of Pure Libertarianism." *The Journal of Philosophy* 112 (4):211–222.

———. 2016. "When Public Reason Fails Us: Convergence Discourse as Blood Oath." *Forthcoming in American Political Science Review* .

Kornhauser, Lewis A. 1992a. "Modeling collegial courts I: Path-dependence." *International Review of Law and Economics* 12 (2):169–185.

———. 1992b. "Modeling collegial courts. II. Legal doctrine." *Journal of Law, Economics, & Organization* 8 (3):441–470.

Landa, Dimitri and Jeffrey R Lax. 2007. "Disagreements on collegial Courts: a case-space approach." *U. Pa. J. Const. L.* 10:305.

———. 2009. "Legal Doctrine on Collegial Courts." *The Journal of Politics* 71 (03):946–963.

Lax, Jeffrey R. 2007. "Constructing legal rules on appellate courts." *American Political Science Review* 101 (03):591–604.

Locke, John. 1980. *Second Treatise of Government*. Hackett Publishing Company.

Marmor, Andrei. 2005. "Should Like Cases be Treated Alike?" *Legal Theory* 11 (1):27–38.

McColgan, Aileen. 2006. "Cracking the Comparator Problem: Discrimination, 'Equal' Treatment and the Role of Comparisons." *European Human Rights Law Review* .

Papadopoulos, Athanase. 2005. *Metric spaces, convexity and nonpositive curvature*, vol. 6. European Mathematical Society.

Rawls, John. 2009. *A theory of justice*. Harvard university press.

Raz, Joseph. 1992. "The Relevance of Coherence." *Boston University Law Review* 72 (2):273–324.

Réaume, Denise. 1989. "Is Integrity a Virtue? Dworkin's Theory of Legal Obligation." *The University of Toronto Law Journal* 39 (4):380–409. URL `http://www.jstor.org/stable/825866`.

Schauer, Frederick. 1987. "Precedent." *Stanford Law Review* 39 (3):571–605.

Strauss, David. 2002. "Must Like Cases be Treated Alike?" URL `http://ssrn.com/abstract_id=312180`. The Law School, University of Chicago.

Sunstein, Cass R. 1993. "On Analogical Reasoning." *Harvard Law Review* .

Walker, Robert. 2010. "Treating like cases alike and unlike cases differently." *Canterbury Law Review* .

Westen, Peter. 1982. "The Empty Idea of Equality." *Harvard Law Review* .

Winston, Kenneth. 1974. "On Treating Like Cases Alike." *California Law Review* .

Wood, Allen. 2010. "Punishment, Retribution, and the Coercive Enforcement of Right." In *Kant's Metaphysics of Morals: A Critical Guide*, edited by Lara Denis, chap. 6. Cambridge University Press.

Appendix

Proof. Proposition 1 Choose some $x \in X$. If x does not rely on a continuous factor, then $x = \bar{X}_j$ is the unique case consistent with D-profile j. Define $Y = \{y \in X | \ y$ differs by only one discrete factor from $x\}$. Let $Z = \{\rho(x, y) \forall y \in Y\}$ and define $z^\star = min\{z \in Z | z > 0\}$. There are two possibilities. Either every $z \in Z$ is zero or among the strictly positive values of z, there must be a minimum value (since by construction Z is finite).

If there is no such $z^\star$, then there is no case $\hat{Z}$ such that $\rho(\hat{z}, x) > 0$. If there is such a $z^\star$, define $\hat{\delta} = \rho(z^\star, x)$. Then there is no case $\tilde{z}$ such that $0 < \rho(\tilde{z}, x) < \hat{\delta}$ Either way Property 1 does not hold. □

Proof. Proposition 3

Every complete and convex metric space is a length space. (Khamsi and Kirk, 2011, 35). By assumption, $\bar{X}_j$ 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 $\bar{X}_j$ is a geodesic space. Further, since $\bar{X}_j$ is strictly convex, the geodesic connecting any two $x, y \in \bar{X}_j$ is unique, so $\bar{X}_j$ must be a uniquely geodesic space.

A subspace $S \subset M$ is *geodesically convex* if for every x and y in S, the geodesic segment $[x, y] \subset S$. Since U and J are assumed to be convex subsets of $\bar{X}_j$, it is obvious that both U and J are geodesically convex subspaces. Papadopoulos (2005, 76) proves that the closure of a geodesically convex subset of a uniquely geodsic space is also geodesically convex. So $cl(U)$ and $cl(J)$ must be geodesically convex. Obviously, the intersection of geodesically convex subsets of a uniquely geodesically convex space is convex. So $B = cl(J) \cap cl(U)$ is geodesically convex.

□

Proof. Proposition 4

Plainly if every case in a path-connected space yields the same outcome, the ϵ-rule is satisfied. Alternatively, suppose that for some path-connected subset $\tilde{X} \subset X$ there are two cases $x, y \in \tilde{X}$ such that $r(x) \neq r(y)$, and let $g : [0, 1] \to X$ be a path between x and y such that $g(0) = x$ and $g(1) = y$. We show that there must exist some $\alpha \in (0, 1)$ such that $r(g(\alpha)) \notin \mathcal{C}$.

Define the set $A = \{a \in [0, 1] | r(g(a)) = r(x)\}$ and let $\alpha = sup(A)$. Plainly $\alpha \in (0, 1]$. Further, since g is continuous, $\forall \ \epsilon > 0 \ \exists \delta > 0$ such that $b - a < \delta \implies g(b) - g(a) < \epsilon$. Suppose $r(g(\alpha)) = r(x)$. Then for any $\epsilon > 0 \ \exists \delta$ such that $r(g(\alpha + \delta)) \neq r(x)$ even though $\rho(g(\alpha), g(\alpha + \delta)) < \epsilon$, which violates the ϵ-rule. Similarly, if $r(g(\alpha)) = y \in \mathcal{C}$, then for any $\epsilon > 0 \ \exists \delta$ such that $r(g(\alpha - \delta)) \neq r(g(\alpha))$ even though $\rho(g(\alpha), g(\alpha - \delta)) < \epsilon$, which violates also the ϵ-rule. To avoid violating the ϵ-rule, it must be that $r(g(\alpha)) \notin \mathcal{C}$. But that means that the theory cannot decide the case at $g(\alpha)$, which means it is incomplete and thus not a pure theory. □