ORIGINAL ARTICLE

A new problem for rules

Jeffrey Kaplan ⬤

The University of North Carolina at
Greensboro

Correspondence
Jeffrey Kaplan, The University of North
Carolina at Greensboro.
Email: jikaplan@uncg.edu

Abstract

This paper presents a series of arguments aimed at show-
ing that, for an important subclass of social rules—non-
summary rules—no adequate metaphysical account has
been given, and it tentatively suggests that no such
account *can* be given. The category of non-summary
rules is an important one, as it includes the rules of eti-
quette, fashion, chess, basketball, California state law,
descriptive English grammar, and so on. This paper
begins with behavioristic accounts of the conditions for
the existence of such rules, and proceeds through pro-
gressively more complex accounts, all the way to what
are labeled 'normative attitude accounts,' which are
prima facie plausible and particularly popular. In each
case it is argued that the type of account under consid-
eration cannot explain the existence of the relevant type
of social rule.

Consider the following rules of fashion: *do not wear flats with an evening gown*, *match one's shoes and belt*, and *do not wear socks with sandals*. The first is a traditional rule of women's formal wear. The second—originally a rule of men's semi-formal attire, but now often applied more gender-neutrally—dictates that brown shoes should be worn with a brown belt and that black shoes should be worn with a black belt. The third was fairly widely accepted, but may now be on the way out, at least in some circles. All of these rules, even when and where they are in force, are occasion-ally violated. These are *social* rules because their existence and content are determined, at least in large part, by social facts—facts about people. But they are *not* what John Rawls (1955) would

call "summary" rules—i.e., rules that, so to speak, attempt to *summarize* other rules or reasons. For an example of a summary rule, consider a rule of what we might call "conventional morality," a society's accepted moral code: *do not have premarital sex*. Conventional morality can be understood as a society's collective attempt to represent what independent moral reasons they take to exist. If that is right, then a rule prohibiting premarital sex can be an expression of a soeciety's view that premarital sex is morally impermissible *independent of whether or not they have a rule of conventional morality deeming it impermissible*. The rules of fashion mentioned above are not like this. The rule requiring matching one's shoes and belt, for instance, may be non-arbitrary—there may be good reason for a group of people to create a rule with that particular content—but it is *not* an attempt to represent the fact that, *independent of the existence of this social rule*, people ought not wear brown shoes with a black belt. Therefore, this rule of fashion falls on the non-summary side of Rawls' distinction.[1]

This paper presents a skeptical argument in regards to this kind of rule—non-summary social rules. The skeptical conclusion is that it is impossible to give a satisfactory metaphysical account of rules of etiquette, fashion, chess, basketball, California state law, descriptive English grammar, and many other rules of this kind.

The resulting rule skepticism—if we can call it that—differs from previous prominent forms of rule skepticism in many ways, two of which are particularly noteworthy. First, unlike Wittgensteinian or Kripkensteinian rule skepticism, it is not directed at *all* rules or rule-following, but applies only to non-summary social rules.[2] Second, the rule skepticism developed here does not, strictly speaking, concern whether these social rules *exist*. Even those who would object to my classification of rules of fashion as non-summary rules, will accept that *some* non-summary rules exist. Indeed, such rules are ubiquitous—and nothing said here calls that into question. Rather, this paper calls into question the claim that it is possible to give an *account* of their existence.[3]

The kind of account we are talking about is a properly metaphysical one. Therefore, this paper is *not* concerned with issues such as: which social rules there ought to be, which members of society get to create them, how social rules erode, how they are maintained, etc. Rather, the accounts discussed are ones that attempt to present, in the most general sense possible, the conditions for the existence of social rules.[4]

This paper does not present an *in principle* argument that no such account is possible. Nor does it argue against each item on an exhaustive list of possible accounts. The argument of this paper is somewhat less ambitious than that. It considers the most plausible general types of metaphysical

---

[1] More on the summary/non-summary distinction in sections 4.1 and 5.

[2] Wittgenstein 1953; Kripke 1982. If Quine (1960) is understood as presenting rule skepticism, then his rule skepticism also applies to all rules or rule-following, and therefore also differs from the rule skepticism presented here in that the rule skepticism presented here does not apply as widely. Additionally, the skepticism developed here differs from Kripkensteinian skepticism in that it does not assume the controversial normativity of meaning thesis, and it differs from Quinean skepticism in that it does not assume naturalism.

[3] The difference between skepticism about the existence of some phenomenon and skepticism about giving an account of that phenomenon is discussed more thoroughly in section 4.2. It may be helpful to think of the kind of skepticism defended here by way of a limited analogy with what is sometimes called "meta-epistemological skepticism," of the kind espoused by Stroud (1984). The view is that we are unable to give a satisfactory account of knowledge that doesn't fall prey to first-order epistemological skepticism. The result is not skepticism about knowledge, but about our ability to give an account of knowledge that meets certain explanatory desiderata.

[4] It is worth noting a distinction occasionally drawn between individualistic and holistic accounts of social rules. Holistic accounts maintain that rules can only be explained by some appeal to features of groups of people which cannot be reduced to features of the individual people. Following others, such as Brennan et al. 2013, 4, my focus is on individualistic accounts.

accounts of social rules, one at a time, and argues that accounts of each type fail. Though this style of argument inevitably leaves open the possibility of the success of some other, unconsidered type of account, there is a systemicity to the way that the accounts are considered here that reduces, but does not eliminate, this possibility. The accounts of social rules considered here are considered in ascending order of richness of explanans (or, what comes to the same, descending order of ontological minimalism).[5] Behaviorist accounts—those that attempt to explain the existence of social rules in terms of behavioral patterns and dispositions—are considered first. I argue that these explanatory resources are insufficient for explaining social rules. Next, commonplace folk-psychological mental states, like belief and desire, are added into the pool of explanatory resources. I call the resulting accounts 'Humean' since they make central appeal to the characteristic mental states of Humean belief-desire psychology. Accounts of social rules based on David Lewis's account of conventions (1969) are Humean in this sense, as are the more contemporary accounts defended by H. Peyton Young (2003) and Christina Bicchieri (2006). I argue that, even with the additional explanatory resources of belief and desire, accounts of this type cannot explain the existence of social rules. Finally, evaluative or deontic attitudes are added into the mix.[6] Most of the resulting 'normative attitude accounts,' as they are sometimes called, draw on HLA Hart's practice theory of rules (1961), including the most prominent contemporary example presented by Nicholas Southwood and Lina Eriksson (2011).[7] These accounts are explanatorily substantive, *descriptive* accounts in the sense that they attempt to explain the existence of social rules by appeal not to the existence of other social rules, but to non-rule-having descriptive states of affairs. The characteristic descriptive state of affairs that these accounts appeal to include normative attitudes that are more than just belief and desire. What exactly these attitudes are will be a significant part of our focus. But it is crucial to note that *the fact that some people have a normative attitude* is itself a descriptive fact. Normative attitude accounts of this type are popular and seemingly plausible. But, in the last few years, a distinction between robust and non-robust normativity has come to prominence, mostly in the philosophy of law literature. I argue that when we bring this distinction to bear on normative attitude accounts, we find that those accounts cannot capture non-summary social rules.[8]

This paper makes several original contributions to the philosophical literature, but here are the three most important. First, it presents a new argument (in section 3) that neither behaviorist nor Humean accounts can be supplemented with underlying robust normative reasons so as to

---

[5] The types of accounts considered in this paper are construed very broadly, but within those broadly construed types all of the diverse, well-known accounts of social rules are encompassed. This includes Ullmann-Margalit 1977; Gilbert 1989; Coleman 1990; J. R. Searle 1995; J. Searle 2010; Young 2003; Bicchieri 2006; Southwood and Eriksson 2011; Brennan et al. 2013.

[6] There are those who take desire to entail normative attitudes (see Scanlon 1998; Dancy 2000; Oddie 2005). If those views are correct, then we can group Humean accounts together with normative attitude accounts and understand the former as facing the same problems as the latter.

[7] They further develop the same account in Brennan et al. 2013.

[8] It is natural to wonder whether the accounts considered here are all "reductive" accounts. The answer depends on what we mean by "reduction." All of the accounts considered attempt to explain the existence of social rules without appeal to the existence of social rules. So they are reductive in that sense. But I do allow for accounts that appeal to robust normative reasons, if there are such things, such as moral or prudential reasons for action. And, moreover, I consider accounts that make appeal to normative attitudes. This differs, in nomenclature, from the approach of Southwood & Eriksson (2011), which is later revised in Brennan et al (2013), as their view appeals to normative attitudes, but is classified as "non-reductive" in the sense that the attitudes are normative, even though the state of affairs that are relevant to their account (i.e., the fact that some people have these attitudes) are themselves descriptive.

adequately explain social rules. Second, it presents (in section 4.2) a wholly new problem for normative attitude accounts by way of the robust/non-robust distinction just mentioned. And third, it puts these original arguments together with disparate existing points in the philosophical literature to point out that no successful account of non-summary social rules has been given, and to make the tentative suggestion that, perhaps, none can be given.[9]

The Rawlsian (1955) label, "non-summary," deserves a more thorough explanation than the brief one given above. But it is difficult to give that more thorough explanation before reaching the final step of this paper's skeptical argument. So I proceed through that argument by using a commonplace type of social rule as a running example, and clarify the full range of social rules to which the argument applies at the end. The running example will be the rules of fashion, such as one of the rules with which we began: match one's shoes and belt.[10]

1 | BEHAVIORIST ACCOUNTS

The behaviorist accounts discussed in this section are specifically behaviorist *in regards to social rules*. It is possible for a behaviorist about, for instance, mental states to also espouse behaviorism about social rules. But it is not necessary, and it will not be assumed that the two views go together.

Behaviorist accounts of social rules can be divided into two broad categories, based on the complexity of the behavioral and dispositional regularities that they employ.

1.1 | Simple behaviorist accounts

The most explanatorily parsimonious way of accounting for social rules is to explain them only with appeal to behavioral regularities. But such accounts of the rules of fashion cannot be made to work.

Consider the following state of affairs: people do not wear blue on the moon. As far as I am aware, this is a real regularity, instantiated in the actual world. All clothing ever worn on the moon has been non-blue. Contrast this with the following: do not wear brown shoes with a black belt. In some societies, or sub-societies, this is also a regularity, albeit a less absolute one. But it is

---

[9] For some of these disparate existing points, see Gilbert 1989; Davis 2003; Millikan 2005; Guala 2007; Southwood and Eriksson 2011; Brennan et al. 2013. [*Redacted*]

[10] It is possible to object to the characterization of the rules of fashion as non-summary. This objection is natural if one supposes that non-summary rules are necessarily arbitrary. But non-summary rules are not necessarily arbitrary. A rule of chess empowering a player with the authority to move a bishop diagonally is not particularly arbitrary. Plausibly, having pieces on the board that can move diagonally adds interesting elements to the game, making it more stimulating and fun. What makes this rule a non-summary rule is the fact that there are no reasons related to the movement of a wooden icon of that type, which exist prior to the existence of the bishop-movement rule, and which the bishop-movement rule attempts to summarize. Things are similar with rules of fashion. Though they involve some degree of arbitrariness, they are not entirely arbitrary. There may be good reasons to create a rule of fashion requiring matching one's shoes and belt. Regardless, as elaborated in sections 4.1 and 5, the rule requiring matching one's shoes and belt is not an attempt to summarize what reasons individuals *independently* have to wear clothes of various colors. Moreover, those who remain convinced that rules of fashion are summary rules can replace those examples with any rules that they accept to be non-summary throughout the paper and the overall skeptical argument should be unaffected.

more than that: it is a *rule*. Wearing blue on the moon is *uncommon*.[11] Wearing brown shoes with a black belt is *forbidden*.[12]

As far as philosophical distinctions go, this one—between rules and mere regularities—is relatively straightforward and intuitive.[13] And we can make explicit the feature that characteristically distinguishes rules from mere regularities: the existence of a rule entails that some evaluative or deontic concept applies to the behavior that accords or discords with the rule. That is, the existence of a rule of fashion requiring people to match their shoes and belt entails that, when *S* wears brown shoes with a black belt, *S* has done something *impermissible*.

It is worth noting that these evaluative concepts need not be *moral* concepts (though they can be). Indeed, the two evaluative concepts mentioned so far—*forbidden* and *impermissible*—are broadly normative or evaluative, but not necessarily moral.[14] However, concepts the application of which is entailed by mere behavioral regularities—such as *typical*, *common*, *atypical*, *uncommon*, etc.—are entirely descriptive concepts.

We have said enough, I think, to show that simple behavioral regularities are insufficient for the existence of social rules. Will adding simple behavioral dispositions make a difference?[15] No. Suppose that not only has no one ever worn blue on the moon, but also that everyone is disposed not to wear blue on the moon were they to travel there (perhaps because none of the available space suits are blue, or because everyone has been hypnotized not to pick blue when given their space suit options, or whatever). Still, there would be no rule prohibiting anyone from wearing blue on the moon. Simple regularities of behavior and dispositions to behave—no matter how robust—are not enough to explain the existence of social rules requiring or prohibiting that behavior.[16] If so, then simple behaviorist accounts of social rules do not work.

---

[11] I am using "uncommon" here to include both (a) events that have occurred sometimes but infrequently and (b) events that have never occurred. If, indeed, no one has ever worn blue on the moon, then wearing blue on the moon is *maximally* uncommon.

[12] If the word "forbidden" appears too strong to some readers (as some such terms are occasionally reserved for violations of only the most important rules), then it can be replaced with potentially weaker terms such as "improper," "impermissible," or "in violation of an evaluative standard."

[13] See Sellars 1954.

[14] It is important not to confuse the existence of a rule with the existence of robust normative reasons for action. It may or may not be the case that the existence of a rule of fashion necessarily generates genuine, robust reasons for action. Either way, fashion regularities are distinct from fashion rules. If a rule of fashion requiring matching one's shoes and belt entails that one has reason not to wear brown shoes with a black belt, then the existence of a regularity of brown-shoes-with-black-belt-wearing is insufficient for the existence of the relevant rule of fashion. But also, if a rule of fashion requiring matching one's shoes and belt merely entails that wearing brown shoes with a black belt is fashion-wise impermissible, but does not entail that one has reason not to wear brown shoes with a black belt, then the existence of a regularity of brown-shoes-with-black-belt-wearing is *still* insufficient for the existence of the relevant rule of fashion.

[15] According to Haugeland (1982), Heidegger holds that behavioral dispositions are sufficient for the existence of social rules.

[16] The purpose of this section is to show that a simple regularity of behavior is not sufficient for the existence of a social rule, such as a rule of fashion. There is also a compelling case to be made that a simple regularity of behavior is not necessary for the existence of a social rule. See Brennan et al. 2013, 20; Southwood and Eriksson 2011.

1.2 | Complex behaviorist accounts

As we have seen, for a rule of fashion to exist it is not enough for there to be a regularity of *the behavior that the rule governs*. But what about regularities of other behavior? For example, criticism. Is a combination of first-order behavior *plus* criticism behavior sufficient for the existence of social rules, such as the rules of fashion?

The first thing to say before answering this question is that in order for an account like this to be truly behavioristic, we must interpret "criticism" purely behavioristically. Most criticism behavior is accompanied by certain attitudes or mental states. But we will consider attitude-based accounts of the rules of fashion in the following sections. So, for the purpose of assessing a behaviorist account of social rules here, the attitudes of the speaker are irrelevant to whether some behavior counts as 'criticism.' The question is: is first-order behavior (e.g., matching one's shoes and belt) plus criticism behavior (e.g., uttering criticism of those who wear brown shoes with a black belt) sufficient for the existence of a rule (e.g., one requiring the matching of one's shoes and belt)?[17] I do not think so.

Since criticism behavior is almost always accompanied by certain attitudes, a good deal of imaginative discipline is required to construct and evaluate circumstances in which the two come apart, allowing us to isolate the criticism behavior itself. A fanciful thought experiment is called for. Consider an eccentric billionaire who holds an exclusive get together. The billionaire likes to hear criticism of other people's fashion choices. So she lets it be known that she expects her guests to criticize one another for wearing orange. But she does not demand that the guests refrain from wearing orange. In fact, she prefers that they wear orange. That way she gets to hear them criticize one another. The guests are all willing to play some part in pleasing their eccentric host. Only a small minority of them are disposed to wear orange themselves, but they all arrive prepared to criticize those who do wear orange. On the rare occasions when a guest arrives wearing orange, public criticism behavior follows.[18] If the sophisticated behaviorist account of rules of fashion is correct, then the circumstances just described are sufficient for the existence of a rule of fashion prohibiting orange wearing (at least at this get together).[19] But plausibly there is no such rule. And we can further modify the example to make the non-existence of such a rule even clearer.

So long as the guests continue to exhibit the relevant behavior—both first-order and criticism behavior—then the conditions set out by the sophisticated behaviorist account are satisfied. So, according to the sophisticated behaviorist, nothing we stipulate about the attitudes of the guests whatsoever should make a difference to whether or not a rule is present. But now suppose that the guests resent their billionaire host. In this modified case, the guests hope that they see others wearing orange, so that they can, at least some of the time, *fail* to criticize orange-wearing. Suppose that they still have to voice criticism most of the time in order to avoid retribution from their eccentric host. But they know they can get away with failing to criticize it some of the time, and they revel in these opportunities. If all of this is the case, it very much seems like there is no rule against wearing orange. It is important to note that the question is not whether there is a rule requiring some kind of *criticism behavior*. Perhaps there is such a rule. The question, rather, is whether there is a rule against wearing orange. It is hard to see how there could be, considering that, in the case we are discussing, *everyone* approves of wearing orange clothing. The host

---

[17] This question could just as easily be constructed vis-à-vis a rule requiring behavior, rather than one prohibiting behavior.

[18] It is worth emphasizing, if it is not obvious already, that in all of these scenarios wearing orange is, in fact, uncommon.

[19] If we like, we can extend the example *mutatis mutandis* for a larger society over the span of many years.

approves of it (because it presents opportunities for criticism) and the guests approve of it (because it increases the number of opportunities to fail to criticize). If we like, we can add that everyone knows that everyone else approves of wearing orange. The sophisticated behaviorist conditions are met, but there is no rule against wearing orange.[20]

2 | HUMEAN ACCOUNTS

A Humean account of the rules of fashion, as I am using the term, is one that makes appeal not only to patterns of first- and higher-order behavior, but also to the characteristic mental states of Humean belief-desire psychology: belief and desire. Can the existence of rules of fashion be explained by appeal to behavior, beliefs, and desires?[21]

David Lewis's account of conventions is Humean in this sense.[22] Lewis's main explanatory resources are what he calls "expectations," which are really a specific kind of belief—beliefs about the future behavior of others—and what he calls "preferences," which are really a specific kind of desire—desires about how one behaves given the behavior of others.[23] Since Lewis's account is so well-known, I consider it first. But the problems that doom a Lewisian account of the rules of fashion apply only to that particular account given some of its rather idiosyncratic features. So it is natural to wonder whether a non-Lewisian but still broadly Humean account might succeed. This possibility is discussed and rejected at the end of this section.

2.1 | Lewisian convention

Here is a fashion-related example from Lewis himself: there is a party and guests can wear either a tuxedo or a clown suit.[24] Some partygoers prefer wearing one or the other, but they all have a greater preference for wearing the same thing as everyone else. In this sense, their overall preferences are *conditional*: they prefer wearing a tuxedo, if everyone else wears a tuxedo; and they prefer wearing a clown suit, if everyone else wears a clown suit. In this circumstance, if everyone expects that others will wear clown suits, and is rational, then they will wear clown suits

---

[20] Other, even more complex, forms of behaviorism about social rules are possible. One such behaviorist might insist that in order for a social rule to exist criticism behavior must be exhibited in an even greater set of contexts. So, for example, people must criticize those who wear orange even privately, or when they are at home with their family. To respond to this form of behaviorism, the example of the eccentric host can be modified accordingly, such as, for example, to allow her to have the threat of surveillance of her guests even when they return to their homes.There is also the question of whether Peter Railton's (2006) view of normative guidance can be built up into a theory of social rules. Railton claims that the internalization of a norm can be understood in terms of dispositions to notice violations of the norm in combination with dispositions to feel affective pressure toward conformity. How exactly these dispositions are understood determines whether a view of social rules developed out of them would fall under the umbrella of behaviorist accounts or, more likely, fall under the umbrella of normative attitude accounts, as discussed in section 4.

[21] Humean accounts have some initial appeal, especially to philosophers familiar with Lewis, but less steeped in the contemporary literature on norms/rules/social institutions. So it is important to include discussion of them. But those who find this old hat are encouraged to skip this section.

[22] Lewis 1969.

[23] For an important distinction between types of expectations, see Bicchieri 2006, 15.

[24] Lewis (1969), p. 6, 45-6.

themselves. When this process iterates over time, there develops a convention, in the Lewisian sense.[25]

To charitably apply this to fashion, some changes must be made to Lewis's approach. For one, it is obvious that not every single person in a society has the requisite expectations and conditional preferences. Yet there are conventions in place in these societies.[26] Lewis does not discuss this explicitly, but we can weaken the requirement that *everyone* have the right preferences and expectations. In order for there to be a rule of fashion in place among a group of people, we can say, all that is required is that *enough* of the members of that group prefer to conform their fashion behavior with the fashion behavior of others and also expect that others will exhibit fashion behavior of a certain type. In weakening Lewis's view in this way, we must not only stop requiring that all people have the preferences and expectations, but we must also adjust our requirements on the content of the preferences and expectations themselves.[27] That is, even those who still have the conditional preferences do not prefer to wear clown suits provided that *everyone* else wears clown suits. Rather, they prefer to wear clown suits provided that *enough* of everyone else wears clown suits. And similarly for the expectations about the behavior of others.

In giving the most charitable application of Lewis's theory to fashion, we should also respond on his behalf to the objection that rules of fashion apply to those who do not have the required conditional preferences (or expectations). For instance, there are those who do not prefer to match the fashion behavior of others. Yet, when they wear brown shoes with a black belt they nonetheless violate a rule of fashion. Rules of fashion often apply to those who do not meet the conditions for a Lewisian convention. This objection assumes that the group to which a Lewisian convention applies is the group of people who have the requisite preferences and expectations. But we can avoid this objection by reinterpreting Lewis's view such that the relevant group is independently identified, and that the convention applies to all those—including those who do not themselves have the requisite preferences and expectations—who are members of a group of which enough members have the requisite preferences and expectations. That is, so long as a convention is in place in my society, the convention applies to me, even if I am not one of the people with the beliefs and desires in virtue of which the convention applies to my society. We analyze conventions at the group level, not the individual level.

Having made these minor adjustments to Lewis's theory, we are left with a laudably parsimonious account. However, it fails to capture rules of fashion for two reasons. First, even at the group level, there can be rules of fashion in place among groups of people who do not have the very specific preferences required for a Lewisian convention. Consider a case where breaking a rule of fashion comes with a severe sanction. Perhaps, for instance, those who wear brown shoes with a black belt face a prison sentence. Matching one's shoes and belt is no longer a Lewisian convention because the preferences are no longer conditional in the required sense of depending on the behavior of others. If the sanction is severe enough, then many members of the society will prefer

---

[25] Lewis's account is explicitly modeled on Hume's account of "artificial virtue". Hume (1738, 477) thinks that justice—the practice of repaying loans and respecting property rights—is only a virtue because of an artifice, a convention in place in our society. Justice arises, Hume thinks, from the following form of self-interest: "I observe, that it will be for my interest to leave another in the possession of his goods, provided he will act in the same manner with regard to me." (1738, 490). When these interests are "mutually express'd, and…known" to all parties, then people leave others in possession of their goods. Hume stresses that this occurs without a promise being made. We leave everyone with his or her possessions because it is in our best interest, given that they do the same.

[26] Lewis has been criticized on this point. See Gilbert 1989; Davis 2003; Millikan 2005.

[27] Lewis 1969, 76–80.

to match their shoes and belt *unconditionally*—i.e., regardless of what other people wear. In fact, Lewis says that law cannot be understood as a system of conventions in his sense for precisely this reason.[28] The rules of fashion are not typically associated with formal sanctions such as imprisonment. But the sanctions need not be formal. Social sanctions, like those commonly attached to rules of fashion, are sufficient, so long as they are severe enough to motivate individuals to make their preferences no longer conditional in the relevant sense. When the rules of fashion are associated with sufficiently serious sanctions they cease to be Lewisian conventions, but they do not cease to be rules of fashion.[29]

Lewis himself also mentions another way in which members of societies with rules of fashion lack conditional preferences and therefore do not have Lewisian conventions.[30] Suppose that everyone matches their shoes and belts, everyone is aware of this behavioral regularity, and they all prefer to continue behaving this way. Suppose also that a significant subclass of people in this society like matching their shoes and belts so much that they would continue to do so even if everyone else was purposefully *mis*matching their shoes and belts.[31] This society lacks an alternative behavioral equilibrium.[32] Simply because enough people prefer matching their shoes and belts *unconditionally*, there is no Lewisian convention. But there may very well still be a rule of fashion requiring people to match their shoes and belts. If people's dress preferences were conditional in the relevant sense, then fashion might be a Lewisian convention. But, as Lewis himself notes, that is not a plausible account of the preferences that people really do have.[33] Unfortunately, it is common that people prefer that they themselves conform with the popular style of dress, but they also prefer that some minority fails to conform. This allows them to feel even better about fitting in, because they can exclude some others. Because preferences of this kind are had, fashion is not a Lewisian convention.

---

[28] Lewis 1969, 103.

[29] One might attempt to save the Lewisian account with a disjunctive approach. The idea here is that those rules of fashion that are backed by serious social sanction are given one account and those rules of fashion that are not backed by serious social sanction are given a Lewisian account. Put very briefly, I see two problems with this approach. First, a disjunctive account does not explain what unifies these two types of rules of fashion in such a way so as to explain why they both count as rules of fashion. One might say that what they have in common is that they govern how one can dress, but that will not do because many rules governing dress, such those of school or work dress codes, are not rules of fashion. Second, the disjunctive account seems to poorly match up with the fact that the phenomenon of fashion itself involves not two ways that rules are related to sanction, but rather a spectrum along which rules are related to sanction. It is not as though some rules of fashion come with sanctions and others do not. Rather, all rules of fashion are associated with social sanctions, and those sanctions fall along a spectrum with the most severe at one end and the least severe at the other. Indeed, there is a threshold of severity at which the sanctions become so severe that the relevant preferences are no longer conditional in the sense relevant to Lewis's account of convention. But this threshold only marks the point at which Lewis's account stops being plausible, and not a point at which there is some genuine division amongst rules of fashion.

[30] Lewis 1969, 45–6.

[31] Of course, they will never be confronted with this situation because many others share their dislike.

[32] With these preferences, each possible way of dressing is merely an equilibrium, not a coordination equilibrium, to use Lewis's jargon.

[33] Lewis 1969, 45–6.

2.2 | **Non-Lewisian Humean accounts**

The Lewisian view just considered requires rather specific conditional preferences and behavioral expectations in order for a convention to exist. And the problems just presented for an application of that view to rules of fashion have force only against a view requiring those narrowly construed attitudes. So the question is: can social rules be explained in broadly Humean terms—with appeal to beliefs, desires, or similar non-normative folk-psychological attitudes? Young (2003, 390) and Bicchieri (2006) present (very different) accounts that meet this description.[34] Both are impressive accounts that build on Lewisian insights while maintaining parsimony. If any Lewisian account can be successfully put to the task of explaining social rules, it is one of these. But it seems to me that thanks to Southwood and Eriksson (2011, 200) there is reason to think that no approach of this kind can succeed.[35]

Imagine a group of people who all wear clown suits to a party. They enjoy wearing clown suits, enjoy seeing one another wearing clown suits, believe (or, if we like, have knowledge or common knowledge) that all parties involved wear (or will wear) clown suits, etc.[36] But they very much dislike the idea that there are any rules *requiring* people to dress in certain ways. This came about, let us suppose, because they observed the ill effects of rules of fashion that are associated with gender and as a result they condemn *all* rules of fashion. (It is plausibly *misguided* for them to, so to speak, throw out the fashion-rules baby with the gender-rules bathwater. But rational or not, this is what they have done.) These partygoers abhor both (a) the existence of rules of fashion and (b) any judgemental or evaluative attitudes toward people's fashion choices. So they believe all sorts of things about what people will wear, and they desire that people wear certain things, but they would *never* allow themselves to judge another person for not wearing what even they themselves believe and desire that that person wear.

The distinction between desire and evaluation here is fairly familiar. I may enjoy the taste of olives and have a desire for them. But it does not follow from that that I think there is anything wrong with those who dislike olives or who refrain from eating them. My neighbor may have an olive tree, and I may desire that she offer me some free olives. But it does not follow that I think she has an obligation to, or in any sense ought to, offer me free olives. Similarly, the group of people under consideration may desire clown-suit-wearing, but they need not, and in this case *do* not, regard non-clown-suit-wearing as incorrect in any sense.

Even though these people take such a sweeping, hardline view of the rules of fashion, they are sophisticated and self-reflective enough to immediately recognize a rule of fashion if one were to arise among them or a judgemental or evaluative attitude if one were to arise within themselves. As a result, if there ever arose a rule of requiring partygoers to wear clown suits, then they would immediately recognize it and disband the party.

---

[34] The following example may be less effective against Bicchieri's view, which is rather sophisticated. However, in the interest of not repeating points that have been made elsewhere, I refer readers to Brennan et al. 2013, 25; Guala 2007. Another view that fits this description well is that of Vanderschraaf 2018, which is a theory of justice, but could plausibly be applied to rules of fashion as well.

[35] See Southwood and Eriksson 2011, 200; Brennan et al. 2013, 18. Those familiar with their *Imelda's Inn* example are encouraged to skip ahead, as the example presented below is a modified version of that.

[36] One can add in more-or-less whatever non-normative, folk-psychological attitudes one likes. For an important distinction relating to whether an "expectation" counts as a normative attitude or not, see Bicchieri 2006, 15.

This example—modified from Southwood and Eriksson (2011, 200)—builds in that there is no rule of fashion.[37] But it does not, it seems to me, problematically beg the question. Of course, it has to do the rather artificial work of teasing apart beliefs and desires from the normative attitudes with which they are normally associated. But what it shows, if we accept that a circumstance like this is possible, is that there is more to a social rule than patterns of behavior, beliefs, and desires.[38]

3 | MORAL AND PRUDENTIAL TRIGGERING ACCOUNTS

We are not done with behaviorist and Humean accounts quite yet. These accounts could be saved, perhaps, if only there were a way for them to explain the one thing that might seem to many to be missing: genuine reasons for action. The goal of this section, therefore, is to see if we can produce an account of the rules of fashion that accommodates two claims: (1) Descriptive, social facts play a large part in determining the content of the rules of fashion of a given society or group of people. That is, descriptive, social facts are among the explanatory inputs. And (2) rules of fashion constitute genuine reasons for action. That is, there are reasons among the outputs. As far as I am aware, there is only one mechanism that allows both of these claims to be true in a given domain: triggering. I borrow the term "triggering" from David Enoch.[39] The aim of this section is to explain what this mechanism is and why it cannot be used to adequately explain rules of fashion.[40]

To understand what is meant here by "triggering," consider the possibility of explaining rules of fashion in terms of prudential reasons or norms.[41] Suppose that there is some underlying prudential reason or norm: e.g., *do not dress in a way that causes widespread disapproval*. And also suppose that the following descriptive, social state of affairs obtains: wearing brown shoes with a black belt causes widespread disapproval. The result is that the descriptive, social fact *triggers* the underlying reason so as to generate a new, more specific prudential reason. One now has prudential reason not to wear brown shoes with a black belt. The suggestion is that the rule of fashion prohibiting wearing brown shoes with a black belt just *is* this resulting *pro tanto* prudential norm or reason, which resulted from a social fact, so to speak, tapping into an independent prudential reason.

We can consider the same suggestion for moral, or any other variety of genuine underlying, reasons. If there are moral facts, then they can be triggered by descriptive, social facts in the same way that prudential facts can. And insofar as other varieties of genuine normative reasons exist, they can be triggered as well.

The most basic problem with any triggering-based approach to the rules of fashion is simply that there are possible and actual cases where rules of fashion require or prohibit actions that one has no prudential, moral, or other genuine reason to perform or refrain from performing. When one is going to attend a gathering of people who admire fashion transgressions, and one will not

---

[37] Those who think that this sort of example fails to undermine Bicchieri's view as applied to rules of fashion may still be persuaded by a different example presented by Brennan et al. (2013, 25).

[38] "What [this sort of example] shows is that, even though, *typically*, a social practice will be accompanied by a corresponding form, this is not due to the practice as such, but something in addition to the practice: namely, our background *normative attitudes*." (Brennan et al. 2013, 19).

[39] Enoch 2011.

[40] The discussion of this section allows for the possibility of non-naturalistic accounts of social rules, as many ways of understanding genuine reasons for action are non-naturalistic.

[41] I use "reason" and "norm" synonymously since I am talking only about practical reasons.

encounter anyone who disapproves of fashion transgressions, and one will not feel negatively about a fashion transgression oneself, it is plausible that one has no prudential reason whatsoever to refrain from wearing brown shoes with a black belt. But it is still true that wearing brown shoes with a black belt violates the relevant rule of fashion. The rule continues to exist and it does not contain an exception for such circumstances. It is not hard to imagine similar circumstances with regard to moral or any other genuine reasons.

Aside from this general suggestion, I offer the following argument, meant to show that rules of fashion cannot be understood as triggered underlying norms or reasons. In order for a triggering-based theory of the rules of fashion to be true, it must be the case not merely that rules of fashion *can* generate genuine reasons for action, but that they *necessarily* generate genuine reasons for action. That is why such a theory is a theory *of* the rules of fashion, as opposed to a mundane fact about the rules of fashion. Almost anything *can* trigger underlying norms or reasons and thereby generate genuine reasons for action. Rain falling from the sky triggers a general underlying reason and makes it the case that I now have reason to pick up an umbrella on my way out the door. But if the the appeal to triggering is meant not merely to explain the utterly unremarkable fact that rules of fashion, like anything else, can generate reasons in this way, but is, instead, meant to explain the *very nature* (or conditions for the existence) of rules of fashion, then rules of fashion must necessarily generate reasons for action.

But it seems to me that there are only two ways that such a necessary connection could exist between underlying reasons and descriptive social facts. The first way that there could be such a connection is if it is built into the underlying reasons or norms. It is natural to think—though this is certainly controversial—that this is how promising works.[42] There is, one might think, an underlying moral norm, such as: if one promises to φ, then, barring exceptional circumstances, one ought to φ. Mention of the practice of promising is built into the moral norm itself, guaranteeing that promissory obligations have moral force. Though it is not certain, it is plausible that there is an underlying moral norm that encodes a necessary connection to the practice of promising. This is not the case for any underlying norms and the practice of fashion.[43]

The second way that there could be a necessary connection between underlying norms and descriptive social facts is for that connection to be built into the descriptive social facts. It is hard to think of an example, so we will invent one. Consider a collection of social rules called *financial prudence*. These are social rules requiring individuals to do what is in their own financial best interest—diversifying one's investments, avoiding vacation timeshares and Ponzi schemes, etc. But these are rules of thumb in the following sense. When circumstances are such that prudence requires one to go against a straightforward reading of the rule, the rule does not require what a straightforward reading suggests. So, for instance, if a particular vacation timeshare is an excellent value, then the rule of financial prudence that normally prohibits purchasing timeshares simply does not apply in this case. The rules of financial prudence are Rawlsian "summary rules."[44] They attempt to capture, or 'summarize,' the reasons that one already has for behaving in a certain way.

---

[42] For an alternative view of promising see Prichard 2002.

[43] That is, promising seems to be such a universal, fundamental part of human social life that it is plausible that objective morality itself, if there is such a thing, encodes an obligation to keep promises. Fashion, though ubiquitous, simply doesn't seem worthy of this kind of moral encoding.

[44] Rawls 1955. To get a sense of what Rawls means, note that one really does have reason to diversify one's investments. "Financial prudence," as we are understanding it here, is the name for a collection of social rules that exist and have the content that they do because of social facts. However, the purpose of these social rules is to capture what ways are really financially, prudentially best. So these social rules aim to align with, so to speak, the non-social rules of prudence itself.

However, we are stipulating, it is built into the concept of *financial prudence* that these rules are infallible. In the cases where the reasons being summarized do not match up with the summary, the summary rule is, by definition, not recommending what it appears to be recommending. It is irrelevant whether or not the practice of financial prudence exists in the actual world. The point is just that social rules of this kind, if they were to exist, would have a necessary connection to some underlying genuine reasons for action *because that connection is built into the rules themselves.* And, relevant for our case, rules of fashion are not like this. The concept of fashion imposes no such constraint on the content of its rules. It is not a part of the concept *rule of fashion* that such rules only require what one has prudential, moral, or some other genuine reason to do.[45]

Since no underlying norms or reasons build in a necessary connection with the rules of fashion, and since the rules of fashion do not build in a necessary connection with any underlying norms or reasons, therefore there is no necessary connection between the rules of fashion and underlying norms or reasons. So triggering accounts of the rules of fashion fail.[46]

4 | NORMATIVE ATTITUDE ACCOUNTS

If we cannot explain the existence of social rules in terms of behavior, beliefs, desires, or underlying reasons triggered by any of these things, then the natural next step is to expand our set of explanatory resources to include normative attitudes. At the moment, this is the most popular approach to the metaphysics of social rules. On this view, wearing brown shoes with a black belt is unfashionable—and I mean nothing by "unfashionable" other than *violates a rule of fashion*—because enough of the right people have the right kind of fashion-disapproval for wearing brown shoes with a black belt. On the face of it, this kind of view is highly plausible. What makes it the case that wearing brown shoes with a black belt is unfashionable is just that the right people *regard it as unfashionable.* Matters of fashion just are, at some level, matters of popularity. So it can seem—and, at times, it seems this way even to me—that some version of this view *must* be correct.

The question is: how do we flesh out the notion of 'regarding some behavior as unfashionable' or 'fashion-disapproval,' as I have tentatively labeled it here? Unlike belief and desire, which are familiar folk-psychological attitudes, 'fashion-disapproval' is vague and mysterious. In this section, I argue that there is no satisfactory way of making sense of this normative attitude so as to allow the attitude to play the role that it needs to play within the normative attitude approach to

---

[45] One obvious way to see this is to note how we would respond in a case where we discover, or think we discover, some underlying prudential facts. If it turned out that, in shocking contrast to almost all known cases, purchasing a certain timeshare was a wise investment, then we would say that the rules of financial prudence never prohibited purchasing that timeshare. (This is just a stipulated feature of financial prudence as we are understanding it.) But if we discovered that some unfashionable way of dressing turned out to be very attractive (or to have whatever value we might think rules of fashion strive toward, if they strive toward anything), then we would not say that the rules of fashion never prohibited that way of dressing. In other words, it is impossible, by definition, for a rule of financial prudence to recommend or require something that is financially unwise. But it is far from impossible—indeed, it happens all the time—for a rule of fashion to recommend or require something that is ugly.

[46] Everything said in this section applies just as well to the specifically Lewisian view when it is understood as a triggering account. Lewis himself was not particularly interested in explaining the existence of rules or reasons of any kind. He was interested in explaining certain patterns of behavior. But it is not hard to see how his view could be transformed into a triggering view: prudential reasons would be triggered. The idea is simply that there are, at least in some cases, reasons to satisfy one's own preferences. See Lewis 1969, 41, 45.

the explanation of social rules. This problem has gone entirely unnoticed because it only becomes apparent when we apply a somewhat underappreciated distinction to our attempt to specify the normative attitude involved in an account of this kind.

As a starting point, consider the normative attitude approach of Southwood and Eriksson, which, like many normative attitude accounts, is inspired by H.L.A. Hart's appeal to 'the internal point of view' in his theory of law.[47] Southwood and Eriksson's view is that a social rule exists when enough of the right people take normative attitudes and when they know that they and others take those normative attitudes.[48] These normative attitudes "include at least the following: (a) normative beliefs, judgements, and other cognitive states, (b) normative expectations, (c) reactive attitudes and dispositions to have such attitudes, and (d) any other attitudes that entail (a), (b), or (c)."[49] This is an expansive conception of 'normative attitude.' And there is nothing wrong with a broad category as such. The problem, as I explain in the following paragraphs, is that there is one dimension along which this category needs to be more precisely specified.[50]

Take, as an example, a negative normative assessment that one might make of someone's choice of clothing. In an extreme case, the clothing choice may be regarded as *immoral* or *foolish*. When one takes this attitude one regards the other person as having some genuine reason—either moral or prudential, in this case—not to wear the clothing in question. In keeping with the terminology that is currently arising for this type of attitude, one makes a "robust" normative assessment.[51] One regards the clothing choice as going against what there is *real reason* to do. Every type of normative attitude mentioned by Southwood and Eriksson can come in this robust variety.

Alternatively, one could regard some choice of clothing as merely *unfashionable* or *in violation of a rule of fashion*. Of course, *sometimes* one has genuine reason to dress in accordance with the rules of fashion, but not always. So some behavior's status vis-à-vis a rule of fashion does not by itself entail anything about whether one has robust reason to exhibit or not exhibit that behavior. And, similarly, to judge that some choice of clothing is unfashionable is not itself to judge that there is any reason not to wear it. We can use the label "non-robust" for this type of normative assessment. All of the normative attitudes mentioned by Southwood and Eriksson can also come in this non-robust variety, whereby they involve merely assessing some behavior as according or failing to accord with a standard, and not as going with or against some genuine reason for action.[52]

So there are two types of normative attitudes: robust ones (e.g., regarding behavior as immoral or foolish) and non-robust ones (e.g., regarding behavior merely as impermissible or unfashionable). The question is whether the normative attitudes that are supposed to explain the existence of rules of fashion are robust or not. That is, when we are trying to explain the existence of a rule of fashion in terms of enough of the right people making normative assessments of fashion behavior,

---

[47] The citations for this view will be from the first chapter of Brennan et al. 2013, which is an updated and more-detailed version of the view originally published as Southwood and Eriksson 2011. Hart 1961, 55

[48] Southwood and Eriksson 2011; Brennan et al. 2013, 35, 57–58.

[49] Brennan et al. 2013, 29.

[50] This is not to suggest that there is not detailed and worthwhile discussion of the relevant normative attitudes (see Brennan et al. 2013, 29–31). The point is simply that there is one crucial detail that remains underspecified.

[51] See Copp 2005; Parfit 2011, 144; McPherson 2011; McPherson and Plunkett 2017; McPherson 2018; Plunkett and Shapiro 2017, 48; Enoch 2011, 2019; Plunkett 2019, 113–115; Berman 2019.

[52] This distinction is now becoming quite commonplace in the literature (see the previous footnote), but there are details that need to be worked out. [*Redacted*]

are they assessing that behavior as (a) something that one has or lacks genuine reason to exhibit or (b) merely as something that violates a standard?

This is a question that can be asked not just of the normative attitudes that Southwood and Eriksson think are at the heart of the existence of social rules, but of the normative attitudes appealed to by anyone who endorses a normative attitude account of social rules.[53] For instance, on John Searle's suggestive approach to institutional reality, instances of behavior have rule-conforming or rule-transgressing status because people take normative attitudes, which he variously calls "acceptance," "recognition," and "endorsement."[54] The question is: does this attitude necessarily involve taking there to be genuine reasons for or against that behavior, or does it merely involve acknowledgement that the behavior is incorrect according to some standard?

The problem, as I will argue in the remainder of this section, is that either answer leads to a dead end. Either way, central appeal to normative attitudes leave us unable to explain the existence of rules of fashion. This problem has gone entirely unnoticed in the philosophical literature on social rules because those discussing normative attitude accounts—though such accounts are the most sophisticated and plausible accounts on offer—have failed to ask this robust/non-robust question, and that is primarily because that question is based on a distinction that has only risen to prominence (mostly in the philosophy of law literature) somewhat recently.[55]

4.1 | Robust normative attitude accounts

Let us start by considering the robust version of normative attitude accounts. On views of this type, the normative attitude that centrally explains the existence of a rule of fashion requiring matching one's shoes and belt is some way or other of regarding people as having real reason to match their shoes and belt. For instance, if enough of the right people think that wearing brown shoes with a black belt is immoral, then there exists a rule of fashion against wearing brown shoes with a black belt.[56] Or, similarly, if enough of the right people think that matching one's shoes and belt is prudent, then there exists a rule of fashion requiring matching one's shoes and belt.

These attitudes, let us suppose, represent what reasons agents independently have for behaving in certain ways, and the rules of fashion result from these attitudes. The resulting rules are summary rules, which are, as Rawls puts it, "reports that cases of a certain sort have been found on other grounds to be properly decided in a certain way."[57] Summary rules are like maps—they continue to exist and continue to have the representational content that they do even when they are inaccurate. When a map gets the terrain wrong, it—the map—continues to exist and, until it is altered, it continues to represent the terrain as it previously has.

This picture of social rules as summary rules constituted by robust normative attitudes of individuals is fairly plausible, as mentioned above, when applied to rules of conventional morality. The believed, conventional moral code of a given society represents moral reasons that members

---

[53] See H.L.A. Hart's "reflective critical attitude" known as the "internal point of view" (1961, 57), which forms the foundation for many normative attitude accounts. For attitude-based accounts of language, see Brandom 1994; Lance and O'Leary-Hawthorne 1997, 62. See also Raz 1982; Taylor 2015.

[54] J. Searle 2010, 8, 42–60; J. R. Searle 1995.

[55] See Plunkett, Shapiro, and Toh 2019 for a collection of new essays, many of which acknowledge or develop the crucial distinction.

[56] This is just meant to illustrate the nature of this type of view, not to claim that such views are particularly plausible.

[57] Rawls 1955, 19.

of that society take there to be. This picture is also plausible when applied to accepted systems of strategy or positive prudence. A society may, for instance, have a well-established rule of financial prudence that recommends investment diversification. This rule exists, plausibly, because enough members of the society take robust normative attitudes, wherein they take there to be independent prudential reason to diversify their investments. The question is whether this picture can also be plausibly applied to rules of fashion. I argue that it cannot.

As mentioned in section 3, not all rules of fashion generate or constitute genuine reasons for action. However, this does not yet show that the robust normative attitude account of rules of fashion is a failure. After all, rules of fashion may be explainable in terms of people *taking there to be* genuine reasons to behave in certain ways without it always being the case that there *are* such reasons. That is, the robust normative attitude account might be an error theory of the rules of fashion. They may be inaccurate summary rules, which miss the mark, as it were, but continue to exist despite their inaccuracy.

This kind of error theory is plausibly correct for several other systems of social rules. For instance, a system of conventional morality that requires human sacrifice is mistaken. It is mistaken, I will suppose, because genuine morality itself does not require—indeed, it does not even permit—the slaughter of innocent people for use in religious ceremonies. But this rule of conventional morality still exists. It is still true that this system of conventional morality requires human sacrifice. It is just that, in this respect, the system of conventional morality is *mistaken*.[58] It is appropriately explained, therefore, by an error theory—one centered around mistaken normative attitudes taken by members of the relevant group. Human sacrifice is required because enough members of the relevant group take a robust normative attitude. They take there to be independent moral reason to perform human sacrifice.

But this sort of approach is implausible when applied to rules of fashion. A rule of fashion requiring matching one's shoes and belt is not *mistaken*. At least, it is not mistaken in the way that a rule of conventional morality permitting human sacrifice is mistaken. Rules of fashion are not the kind of thing that can be mistaken in this way. And that is because, unlike rules of conventional morality, they are not summary rules. They are simply not in the business of representing independent reasons for action. So although a robust version of the normative attitude type of account is plausible for summary rules, such as rules of conventional morality, it is not plausible for non-summary rules, such as rules of fashion.

4.2 | Non-robust normative attitude accounts

This brings us to our final option: explaining the existence of rules of fashion in terms of people taking non-robust normative attitudes. Though this approach is highly attractive, it is circular.

On accounts of this type, what makes it the case that there is a rule of fashion prohibiting wearing brown shoes with a black belt is that enough of the right people regard wearing brown shoes

---

[58] I do not mean that this sort of rule of conventional morality is *altogether misguided*. It may be that there are *some* reasons for adopting a rule of conventional morality even though that rule does not accurately represent the independent moral reasons that agents have. Perhaps, for instance, the putative rule of conventional morality keeps society stable and prevents disorder. That may be. But there is one clear sense in which summary rules of this kind can be misguided, and I am supposing that this rule of conventional morality is misguided in precisely that sense: it inaccurately represents the independent reasons that agents have. This is a feature of summary rules. Because they are a kind of 'summary,' they can be accurate or inaccurate. The rule of conventional morality requiring human sacrifice is meant to be an example of a summary rule that is inaccurate.

with a black belt as *unfashionable*. Naïvely, something like this is *surely* correct. Think about the newest fashionable clothing item. As it becomes regarded as fashionable, and *because* it becomes regarded as fashionable, it becomes, in fact, fashionable.

The problem is as follows. We are trying to explain the existence of a type of social rule. The explanation being offered makes mention of a certain attitude or mental state. But our specification of that mental state makes mention of the very thing we initially set out to explain: the social rule. What is it for there to be a fashion rule against φ-ing? It is for enough people to take φ-ing to violate a rule of a fashion. This is a striking circularity.

How might an advocate of non-robust normative attitude accounts respond to this circularity worry? Broadly, they have two options: (a) further develop the account so as to eliminate the circularity or (b) embrace the circularity and argue that it is not vicious. Since circularity problems of this kind are not uncommon in philosophy, we can explore both options by considering some precedents for them in other areas of philosophy.

Beginning with the second option (i.e., embracing the circularity and arguing that it is not vicious), consider the institutionalist definition of art most prominently espoused by George Dickie, building off of the work of Arthur Danto. This account holds, roughly, that works of art are artifacts created with the intention of being presented to "the artworld." Danto introduced the term "artworld" and defined it in terms of appreciation of, and theorizing about, art.[59] Dickie maintained that membership in the artworld is determined by self-identification.[60] The result is an explanation of art in terms of individuals regarding themselves as related, in a certain relevant way, to art.

Here we have a circularity wherein the concept in need of explanation—*art*—appears within the explanans, but, at least on some ways of developing the institutional definition of art, that appearance happens within a referentially opaque context.[61] It is not merely that art is defined in terms of an institution focused on the appreciation of art. Rather, art is defined, ultimately, by reference to people who *regard themselves* as part of an institution focused on the appreciation of art.

Broadly, this is the same kind of circularity that is revealed within some normative attitude accounts of social rules when the robust/non-robust distinction is applied to them. The question is whether this circularity is vicious. Or, better put: *how* vicious is it? First of all, the fact that the concept in need of explanation appears within a referentially opaque context, wherein one cannot substitute co-referential terms without opening up the possibility of changing the truth value of the sentence as a whole, makes the circularity, at minimum, less direct. And there are several things that this kind of indirect circularity does *not* do. It does not call into question the possibility of the relevant phenomenon (i.e., social rules). It also does not call into question the possibility of saying informative things about the phenomenon. What it *does* do is present a nontrivial flaw with this variety of metaphysical account of social rules. The flaw is that metaphysical accounts of this kind, because they seek to provide explanation by way of appeal to an attitude that is only identified by reference to the phenomenon that was in need of explanation in the first place, will remain, at minimum, imperfect or incomplete.

The circularity built into non-robust normative attitude accounts of social rules is not a metaphysical problem, but, rather, a meta-metaphysical one. It constitutes a limitation on the

---

[59] Danto 1964, 1981.

[60] Dickie 1974, 1984.

[61] Davies 2003.

possibility of *explaining* or *accounting for* rules of fashion. The non-robust normative attitude approach attempts to explain the existence of a rule of fashion in terms of some attitude. What attitude? The attitude of taking some behavior to be unfashionable. And what attitude is *that*? When we take some behavior to be unfashionable, what are we taking that behavior to be? The answer given by accounts of this type is that when we take behavior to be unfashionable we are taking it to be the kind of behavior that is taken to be unfashionable.

Should we conclude, therefore, that non-robust normative attitude accounts of social rules are *doomed*—that they are *failures*? The answer depends on how we understand the ambitions of those accounts. To the extent that such accounts merely aim to say something informative and accurate about the metaphysical underpinnings of social rules, I think these accounts still have something to offer. The claim that social rules depend—not merely in a contingent or historical sense, but in a metaphysical sense—for their existence on certain people having certain normative mental states is, if true, a genuine, nontrivial insight. But to the extent that such accounts aim to explicate the fundamental nature of social rules without taking for granted, pre-theoretically, what counts as a social rule, then non-robust normative attitude accounts are, in relation to this aim, deficient. Because accounts of this type are circular, there is a limit to their informativeness.

There remains, however, the first option for how an advocate of non-robust normative attitude accounts might respond to the circularity worry (i.e., by further analyzing the relevant attitude so that the circularity disappears). A clear example of this route has been mapped by Alan Gibbard.[62] Gibbard attempts to explain the existence of practical reasons in terms of someone taking there to be practical reasons. When T.M. Scanlon points out that this analysis is circular in that the specification of the attitude makes mention of precisely the phenomenon that was originally meant to be explained, Gibbard responds by proposing that the attitude can itself be further reduced to mere dispositional features of the agent with the attitude, thereby avoiding mention of 'reason' and exiting the circularity.[63] That is, Gibbard proposes that we understand 'taking there to be practical reasons' simply as a way of "calculating what to do on a certain pattern, a pattern we could program a robot to mimic."[64] If this response works, then the circularity problem goes away. Gibbard is no longer explaining the existence of practical reasons in terms of someone taking there to be practical reasons. Now he is explaining the existence of practical reasons in terms of someone calculating what kinds of behavior fit a certain pattern. Saying that someone has determined that some behavior does or does not fit with a certain pattern does not make use of the concept of *practical reason*. So no circularity is involved.

Whether or not this type of approach works in Gibbard's case, we can see that it does not work when it comes to social rules, like rules of fashion. We have already established that a regularity of behavior is insufficient for the existence of a social rule. But we are now asking: is the mental activity of determining whether some instance of behavior fits with a regularity—even when this activity is performed by many people and repeated over time—sufficient for the existence of a social rule? The answer is 'no.' Suppose that you travel to the moon and I 'calculate' that your fashion behavior fits with the pattern of not wearing blue on the moon. Moreover, many other people, including you, make the same calculation. We all notice that you are wearing clothing that conforms to a pattern that we take to be typical of people on the moon. Do our attitudes make it the case that your outfit is *fashionable* or in any way *correct*? No. And the same would hold,

---

[62] Gibbard 1990; Scanlon 1998, 58–59; Gibbard 2003, 188–191.

[63] Scanlon 1998, 58–59; Gibbard 2003, 188–191.

[64] Gibbard 2003, 190.

mutatis mutandis, if you had deviated from the pattern rather than conformed with it. What is needed, we might think, to make it that your clothing choice violates a rule of fashion is not for enough of the right people to regard it going against some regularity, but rather for enough of the right people to regard it as going against some rule of fashion. But, as just discussed, if we are interested in explaining the existence of such social rules, it is not ideal to refer to those rules in our explanation.[65]

5 | CONCLUSION

I conclude by (a) clarifying which rules this skeptical argument applies to and (b) explaining the unavoidable limitations of an argument of this kind.

Obviously, the argument presented here does not apply to non-social, mind-independent rules, if there are any. It is controversial whether there are moral or prudential rules that apply to human behavior independent of what anyone thinks about moral or prudential matters. But if there are such rules, our ability to account for them is untouched by what has been said here. But also, as we saw in the previous section, the argument presented in this paper does not apply to summary social rules, such as the rules of conventional morality and conventional prudence. These are social rules, but they can very plausibly be accounted for with some form of robust normative attitude account. For example, it is plausible that we can explain the existence of a society's rule of conventional morality prohibiting premarital sex by appealing centrally to the fact that enough members of that society take the robust normative attitude of regarding premarial sex as immoral. Whether this account constitutes an error theory depends on whether one thinks that premarital sex really is immoral or not. But, crucially, this account avoids the circularity problem from section 4.2 because, even though the explanation makes appeal to an attitude the content of which mentions morality, this account is not itself an account of morality, but of conventional morality. This is in contrast to non-robust normative attitude accounts, which attempt to explain the rules of fashion, for instance, by appeal to an attitude the content of which includes the concept of fashionableness. So we can see that the skeptical argument of this paper applies to accounts of *non-summary* social rules—that is, rules that do not attempt to capture other, independent, non-social evaluative standards. The problem, therefore, plausibly applies, in addition to fashion, to rules of etiquette, games, and all manner of social institutions, including clubs, universities, condominiums, etc.

Finally, it should be noted that I am merely presenting a problem that I believe has never before been appreciated and suggesting that it is difficult to solve. I have no argument that it is *impossible*

---

[65] This section discussed normative attitude accounts and argued that some such accounts involve a circularity. But what about attitude-based accounts, other than the robust normative attitude accounts, that do not—at least not explicitly—involve this circularity? As I see it, there are two possibilities. First, it may be that although those accounts do not appear to involve the relevant circularity, this appearance is only a result of the fact that they have not yet been subjected to the robust/non-robust normativity distinction, which has recently become prominent in the philosophy of law literature. The idea is that when we ask of attitude-based accounts, 'is the normative status involved in the relevant attitude robust or non-robust?' those accounts will turn out to either (a) only be plausible accounts of summary rules or (b) involve the circularity discussed in this section. The second possibility is that the normative attitude accounts that do not explicitly involve this circularity might turn out, after being subjected to the robust/non-robust distinction, to not *implicitly* involve the circularity either. In that case, they would evade this objection and thereby undermine the central project of this paper. A demonstration that there is such an account—one that avoids the circularity problem and is still plausible when applied to non-summary rules—would be a philosophically illuminating and productive reply to this paper.

to solve. Like many other skeptical arguments, the argument of this paper proceeds by canvassing a finite number of potential accounts and arguing that they each fail. So, like other skeptical arguments of this form, the argument of this paper is not challenged by the mere suggestion that there are other ways to account for rules of fashion. There are, indeed, *countless* potential ways. What would constitute a challenge would be a somewhat worked-out account of non-summary social rules, perhaps along with an argument that the account avoids the problems for other accounts discussed here. Indeed, if such an account were successfully presented, I would consider it a victory, of sorts, for this paper. Encouraging such an account to emerge in the philosophical literature would itself be a significant contribution.

ORCID
Jeffrey Kaplan https://orcid.org/0000-0003-1011-0126

REFERENCES

Berman, Mitchell. 2019. "Of Law and Other Artificial Normative Systems." In *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, edited by David Plunkett, Scott J. Shapiro, and Kevin Toh, 137–64. Oxford University Press.

Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.

Brandom, Robert B. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

Brennan, Geoffrey, Lina Eriksson, Robert E. Goodin, and Nicholas Southwood. 2013. *Explaining Norms*. Oxford University Press UK.

Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, Mass: Harvard University Press.

Copp, David. 2005. "Moral Naturalism and Three Grades of Normativity." In *Normativity and Naturalism*, edited by P. Schaber. Ontos-Verlag.

Dancy, Jonathan. 2000. *Practical Reality*. Oxford University Press.

Danto, Arthur. 1964. "The Artworld." *The Journal of Philosophy* 61 (19).

Danto, Arthur. 1981. *The Transfiguration of the Commonplace: A Philosophy of Art*. Cambridge: Harvard University Press.

Davies, David. 2003. *Art as Performance*. Oxford: Blackwell Publishing.

Davis, Wayne A. 2003. "Psychologism and Humeanism: Review of Dancy's Practical Reality." *Philosophy and Phenomenological Research* 67 (2). Wiley-Blackwell: 452–59.

Dickie, George. 1974. *Art and the Aesthetic: An Institutional Analysis*. Cornell University Press.

Dickie, George. 1984. *The Art Circle: A Theory of Art*. New York: Haven.

Enoch, David. 2011. "Reason-Giving and the Law." In *Oxford Studies in Philosophy of Law*, edited by Leslie Green and Brian Leiter. Oxford University Press.

Enoch, David. 2019. "Is General Jurisprudence Interesting?" In *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, edited by David Plunkett, Scott J. Shapiro, and Kevin Toh, 66–86. Oxford University Press.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Harvard University Press.

Gibbard, Allan. 2003. *Thinking How to Live*. Harvard University Press.

Gilbert, Margaret. 1989. *On Social Facts*. Routledge.

Guala, F. 2007. "CRISTINA BICCHIERI : The Grammar of Society: The Nature and Dynamics of Social Norms." *The British Journal for the Philosophy of Science* 58 (3): 613–18.

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

Haugeland, John. 1982. "Heidegger on Being a Person." *Noûs* 16 (1). Wiley-Blackwell: 15–26.

Hume, David. 1738. *A Treatise of Human Nature*. Oxford University Press.

Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language*. Harvard University Press.

Lance, Mark Norris, and John O'Leary-Hawthorne. 1997. *The Grammar of Meaning: Normativity and Semantic Discourse*. Cambridge University Press.

Lewis, David K. 1969. *Convention: A Philosophical Study*. Wiley-Blackwell.

McPherson, Tristram. 2011. "Against Quietist Normative Realism." *Philosophical Studies* 154 (2): 223–40.

McPherson, Tristram. 2018. "Authoritatively Normative Concepts." In *Oxford Studies in Metaethics Vol*. 13, edited by R. Shafer-Landau.

McPherson, Tristram, and David Plunkett. 2017. "The Nature and Explanatory Ambitions of Metaethics." In *TheRoutledgeHandbook of Metaethics*, edited by Tristram McPherson and David Plunkett, 1–28. Routledge.

Millikan, Ruth. 2005. *Language: A Biological Model*. Oxford: Clarendon Press.

Oddie, Graham. 2005. *Value, Reality, and Desire*. Clarendon Press.

Parfit, Derek. 2011. *On What Matters: Two-Volume Set*. OUP Oxford.

Plunkett, David. 2019. "Robust Normativity, Morality, and Legal Positivism." In *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, edited by David Plunkett, Shapiro Scott J, and Kevin Toh, 105–36. Oxford University Press.

Plunkett, David, and Scott Shapiro. 2017. "Law, Morality, and Everything Else: General Jurisprudence as a Branch of Metanormative Inquiry." *Ethics* 128 (1). The University of Chicago Press: 37–68.

Plunkett, David, Scott J. Shapiro, and Kevin Toh, eds. 2019. *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*. Oxford University Press.

Prichard, H. A. 2002. Moral Writings*. Edited by Jim MacAdam*. Oxford University Press.

Quine, Willard Van Orman. 1960. *Word and Object*. MIT Press.

Railton, Peter. 2006. "Normative Guidance." In *Oxford Studies in Metaethics, Volume* 1, edited by Russ Shafer-Landau, 3–34. Oxford University Press.

Rawls, John. 1955. "Two Concepts of Rules." *The Philosophical Review* 64 (1). Jstor: 3–32.

Raz, Joseph. 1982. "Liberalism, Autonomy, and the Politics of Neutral Concern." *Midwest Studies in Philosophy* 7 (1): 89–120.

Scanlon, Thomas. 1998. *What We Owe to Each Other*. Belknap Press of Harvard University Press.

Searle, John. 2010. *Making the Social World: The Structure of Human Civilization*. Oxford University Press.

Searle, John R. 1995. *The Construction of Social Reality*. Simon and Schuster.

Sellars, Wilfrid. 1954. "Some Reflections on Language Games." *Philosophy of Science* 21 (3). Association for Symbolic Logic: 204–28.

Southwood, Nicholas, and Lina Eriksson. 2011. "Norms and Conventions." *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* 14 (2). Taylor & Francis: 195–217.

Stroud, Barry. 1984. *The Significance of Philosophical Scepticism*. OUP Oxford.

Taylor, Kenneth A. 2015. "How to Hume a Hegel-Kant: A Program for Naturalizing Normative Consciousness." *Philosophical Issues. A Supplement to Nous* 25 (1): 1–40.

Ullmann-Margalit, Edna. 1977. *The Emergence of Norms*. Oxford University Press.

Vanderschraaf, Peter. 2018. *Strategic Justice: Convention and Problems of Balancing Divergent Interests*. Oxford University Press.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Vol. 2001. Blackwell Publishing.

Young, H. Peyton. 2003. "The Power of Norms." In *Genetic and Cultural Evolution of Cooperation*, edited by P. Hammerstein, 389–99. MIT Press.