

Rules, Norms and Principles: A Conceptual Framework

1.1 Introduction

In looking at the literature on rules one is struck by two related observations: one is that different notions are often conflated; the other is that it is often hard to see when a dispute is merely *verbal* and when it is substantive. Part of the problem here is that ordinary language is not precise when it comes to the word ‘rule’—one can legitimately talk in different ways and so there is a danger of people talking past one another.

In this paper, I will try to suggest a framework for talking about rules that will allow me to clearly pose the issues that interest me, and that will help us distinguish between merely verbal issues about rules and substantive ones.

1.2 What Is a Rule?

1.2.1 Rules as Abstract Objects

Let me start with the following basic question: *What is a rule?*

We should start with the fact that, whatever rules are, it must be possible for them to be the sorts of things that are *accepted*, either by a person or by a community. It must be *possible* for rules to be accepted, even if some are not. That suggests that rules themselves should be thought of as *contents*, as the possible objects of intentional states of acceptance. (An intentional state is a mental state that has an intentional object. This is not to be confused with the notion of an intention, although intentions are, of course, instances of intentional states).

Now, contents are best thought of as *abstract objects*. Exactly what an abstract object is, as we will presently see, is controversial. But the idea is that, on this conception, rules belong with such objects as numbers, properties and propositions, rather than with concreta like tables or (token) books. If we say that rules are abstract objects, what kinds of abstract objects should we take them to be?

I’ll come back to this question below, but already we might be thought to face a puzzle: *How could rules be abstract objects?* Abstract objects, after all, are said to exist outside of time and space. That means that they can’t come in and out of existence. But, surely, we do often want to say that *a rule came to exist* when it didn’t exist before. For example, those who think that (legal) *laws* are rules will want to say that rules come to exist as a result of certain kinds of legislative activity and that those rules did not exist before. Another kind of example would be the rules of chess. Surely, an objector might insist, the rules of chess did not exist prior to the game’s being invented.

However, on an abstract conception of rules, we seem committed to rules’ existing atemporally and necessarily. How can we reconcile these two observations? One possible response to this problem might be to say that we *do* after all need to make sense of abstract objects coming in and out of existence. Some philosophers see the need to say this even in connection with abstracta that don’t concern the topic of rules.

For example, a number of philosophers who hold that *concepts* are abstract objects have expressed sympathy for the idea that some of these concepts could not have existed since the dawn of time, but must have been brought into being by the creative activity of human beings. While knowing no good alternative to thinking of concepts as abstract objects, such philosophers find it odd to think that a

distinctively *cultural* concept, such as that of a *piano*, for example, or a *high school prom*, could have been around before there were humans. One might feel a similar discomfort with the idea of a certain *type* of musical composition, for example, the symphony, pre-existing its invention by human beings.¹

This sort of temporal thinking about abstract objects does not appeal to me. I lose my grip on the distinction between abstract and concrete objects if I think of abstracta as having temporal properties. Moreover, I believe that we can do justice to our intuitions about the role of conceptual and artistic creativity within a framework in which we talk not about *creating* abstracta, but about *selecting* them, or *discovering* them. But no matter, we will see below that our puzzle about rule existence can be solved on either the temporal or the atemporal view.

1.2.2 Imperative vs. Normative Content

Let me turn next to the question: If rules are abstracta, what kinds of abstracta are they?

One kind of abstract object that one might naturally think a rule might be, is a certain sort of *normative proposition*—a proposition that specifies a permission or a requirement. These, for example, are among what we call the ‘rules of chess’:

a.

White *must* move first.

b.

If one’s king and rook have not previously moved, and if the king is not currently in check, and if there are no pieces between the king and said rook, then, under those conditions and only under those conditions, one *may* castle.

Another kind of abstract object that is also often called a ‘rule’ is an *imperative* content, or an *instruction*:

c.

If C, write a 0, move to the next square and go into state S2!

d.

If x is an email that calls for an answer and you have just received x, answer it immediately!

Both of these types of content have been called ‘rules.’ Yet they are quite different from one another. For example, normative propositions look to be truth-evaluable, whereas instructions or imperatives are not.

So, are rules abstract objects with normative contents, or imperative contents? Which is it to be? I believe that it would be all right to say that a rule might be *either* an imperative content *or* a normative proposition.² This may come as a shock to some in the philosophy of law, because in that literature it is very common to insist that there is an important three-fold distinction between imperatives, rules and normative propositions, and to claim that there is a big issue whether laws should be thought of as consisting in the one sort of thing or the other.

To put things very roughly, John Austin argued that laws are *imperatives* with a certain pedigree; H. L.

A. Hart argued that laws are *rules* with a certain pedigree; and Ronald Dworkin argued that laws essentially involve certain sorts of normative propositions, *principles*, in addition to rules.³

How could we make sense of this central debate in the philosophy of law if we adopt the framework I am suggesting?

The answer to both this puzzle and the one about existence lies in seeing that the substantive debate is not about *whether a rule exists* or *what a rule is*, but about *what it takes for a rule to be norm on behavior*. Assuming that a rule is an abstract object, under what conditions will such a rule come to be a *standard* that behavior must conform to on pain of being open to a certain kind of criticism.

On the view I am recommending, we can say that a rule exists before it becomes law. What the legislator does is to take an antecedently existing abstract object and turn it into a certain kind of *legal norm on behavior*; a standard on behavior violations of which open one to a certain sort of criticism. Analogously, what the inventor of chess does is take a certain number of antecedently existing rules and put them together in such a way that they become norms on the behavior of a person with the right sorts of *intention*—namely, the person intending to play a game of chess.

In terms of this framework, we can pose Dworkin's question whether there is more to law than imperatives and rules by seeing it as the question not whether law involves principles in addition to rules, but rather as the question:

Are the rules that are involved in law norms on behavior *independently of whether they are accepted* or does the law involve *only* rules that have been somehow or other accepted (either directly or indirectly, via a rule of recognition)?

Dworkin's view, of course, was that certain normative propositions—namely, moral principles, could be norms on legal reasoning, even if they had not been explicitly accepted.

On the view I'm recommending, then, rules are just certain types of abstract content. Their essential feature is *generality*. Rules are about types of states of affairs, and types of action, not particular states of affairs and particular actions. The key question about them is whether some of them can be norms on behavior independently of acceptance, or whether acceptance is always a precondition on their being norms on behavior.

Intuitively, moral rules are norms on behavior independently of whether they are accepted—this amounts to saying that, intuitively, a relativistic view of morality is incorrect. Furthermore, rules of etiquette are not norms on behavior in a given community independently of whether they have been accepted by that community. In the way I am proposing we think about things, what Dworkin did is make a real issue out of the question whether legal laws always depend on explicit acceptance if they are to be regarded as normatively constraining one's behavior.

1.2.3 Rule-Following

What about the question what it is to *follow a rule* or to *obey a rule*? This is one of the most-talked about issues in connection with rules.

This notion has obvious application in the philosophy of law. Under what conditions is someone following the law? Let us ask first: what is the *intuitive* idea of someone's following a given rule R? Clearly, this is not just the idea of a rule's being *applicable* to one's behavior, being a norm on one's behavior. As we have just seen, a natural view about moral norms is that they can be applicable to one, even if one has not accepted them and so is not following them (see the ACER model of rule following below).

The idea of following rule R is also not just the idea of doing something and thereby *conforming* to R. Conformity to R is neither necessary nor sufficient for following R. It is not necessary because one could be following R but, for one reason or another, one fails to conform—for example, because of a performance error. And it is not sufficient because in conforming to R one simultaneously conforms to an infinite number of rules. So, what is it?

The intuitive account of rule-following is given by what we may call the ACER model: A, C, E, R (see Boghossian [2008](#)):

- (A) Acceptance: The agent must somehow or other accept the rule.
- (C) Correctness: If an agent is following rule R, then there is a sense of correctness, according to which what he does is correct iff he conforms to R.
- (E) Explanation: Acceptance of R explains what the agent does.
- (R) Rationalization: Acceptance of R rationalizes or makes sense of what the agent does.

On this view, for an agent to be following R, she had to have accepted R. This marks a contrast with the notion of R's applying to her behavior as a standard, since R may apply to her even if she has not accepted it. A close approximation to the idea that an agent is following rule R is that, in following R, she is *trying* to conform to R. Now, as we all know, there is supposed to be a question, made famous by Wittgenstein ([1953](#)), about how rule following is so much as possible. Given the gloss I have just given, this raises another puzzle. How could there be any such problem? Isn't 'trying to conform to a rule' the most mundane of phenomena? For example, don't we all, typically, try to conform to the rule: *If the traffic light is red, stop!* How could there be a problem about this?

According to Saul Kripke's famous exposition, Wittgenstein is drawing our attention to the problem of explaining how there could be a determinate fact of the matter about *which* rule has been *accepted* by someone. Maybe the rule is: '*If the traffic light is red and it is before the year 3000, then stop! Otherwise, keep going!*'⁴ After all, all the behavior we will ever observe will conform to that rule as much as to the normal one. But it looks as though there is a simple answer to Kripke: I am following the one rule and not the other because it is the one rule and not the other that I *intend* to conform to. Surely, there can be a fact of the matter about whether my intention involved anything about the year 3000 or not. Kripke thinks that this answer won't do because we need to explain how there could be a fact of the matter about what the determinate content of my intention is. But this threatens to look like just a generalized skepticism about determinate intentional content, and there are two things to be said about that.

First, this has nothing specifically to do with rule-following. It's a very general skepticism about intentional content and, if it were to be accepted, we would be in trouble in a number of fundamental ways, and not just because we couldn't be said to follow determinate rules. Second, the very generality of the argument makes it vulnerable. At most, Kripke could hope to have shown that *naturalistic reductions* of intentional content don't work.

While I am inclined to agree that such naturalistic reductions don't work, I also think that he could not have refuted all possible anti-reductionist conceptions of intentional content. Surely, the conclusion that there is no meaning or content is bound to be less plausible than any premise leading up to that conclusion.

Given these considerations, it looks as though we may just think of rule-following as involving an intention to conform to a certain pattern, where that intention is non-reductively understood. And that state can then explain and rationalize my behavior. So is that all there is to the famous rule-following problem? Unfortunately, no.

1.2.4 The Problem of Rule-Following Recast

The further problem I have in mind is not about *vagueness*. Rules, like any contents, can be vague. They can have borderline cases. How should they be applied to borderline cases? What constitutes correct following of the rule in a borderline case? These issues are of great relevance to the study of law, but they are not my topic here. Once again, vagueness is a very *general* problem about contents and does not have to do with rules specifically.

The rule-specific problem I have in mind is that in addition to cases of rule-following where *there is an explicit intention*, there are cases of rule-following where there is *no explicit intention*. And the difficulty is to explain how we are to think about those cases. How do we know that there are such cases? We know this because we know that there are phenomena that cry out for saying that the agent is following a rule, but in which the agent *couldn't formulate the rule if asked*, without doing a great deal of empirical work. Why is the agent's ability to formulate the rule a good test for whether the rule is carried by explicit intention of his? Because with intentions, as we encounter them in ordinary action, we know their contents relatively effortlessly. If we didn't, they couldn't play the role in guiding our actions that they clearly play.

What are examples of phenomena in which rules are clearly involved but in which the agent could not formulate the rules without doing a lot of empirical work? Two central such phenomena involve the use of language and the fixation of belief in response to experience. In both of these central cases, our productive and general competence appears to require attributing to us the acceptance, or internalization, of rules that guide our behavior. Yet we would, of course, be hard pressed to give an account of what those rules are on a purely introspective basis.

So, there clearly look to be many cases of rule-following unaccompanied by explicit intention. Indeed, there is an argument, laid out in Boghossian (2008), to the effect that there *have* to be some such cases, that not all rule-following could be explicit.

The argument goes something like this. If all rule-following involved an explicit intention, rule-following would always involve inference from that intention. But inference itself is a form of rule-following. It can't be, though, that rule-following always involves inference and inference always involves rule-following (For details see Boghossian (2008)). So, there has to be some rule-following that does not involve inference, and that in turn seems to imply that there must be rule-following that is not anchored in an explicit intention. But, in what sense could a rule be said to be *guiding* our behavior, if that does not amount to an intention's guiding our behavior?

One idea here that has appealed to many is that there is a representation of the rule guiding the behavior, but this representation is *sub-personal*, rather than being the sort of conscious mental state that an intention is. We can't know about this sub-personal state by introspection; but we can get at in various other more indirect ways—for example, by what we are inclined to say about possible scenarios (the method of cases), and by means of other analytic techniques.

One of the enduring difficulties with this line of thought is to say in what way such sub-personal regulation by rules can be properly said to rationally guide us in our use of language or in our reasoning. Many behaviors of ours, such as our digestion and breathing could be said to be regulated by homeostatic sub-personal systems, but we clearly don't think of them as rationally guiding those behaviors. Why would regulation by sub-personal systems be any different? I think we are still far from properly understanding these issues. I outline the matter here so we can properly distinguish those substantive issues that remain vexed, from those that we are relatively well understood.

1.2.5 The Normativity of Rules

Let me turn next to the question of the *normativity* of rules, a topic that has exercised quite a few scholars.

As we can see from the proposed framework, in asking about the normativity of rules one could be asking at least two rather different questions:

a.

Are rules themselves normative?

b.

Is following a rule normative?

Before we tackle these, let us briefly address a large question: what is it for something to be normative? Let us agree that something is normative when it provides *reasons* for doing something. This is, of course, very rough and terse, but it will do for now.

Question (a), then, is the question: are rules themselves normative? Do rules provide reasons? Well, rules are just abstract objects of a certain sort, we have agreed: among them we may find the following imperative: ‘kill the first-born child of every family!’ Or the normative proposition: ‘you ought to kill the first-born child of every family.’

Clearly, this rule does not give anyone a reason to kill anyone. So, no rule, merely *qua* rule, gives one a reason to do anything. Some rules can be bad—and no one would have any reason for accepting them or following them. This ought not to be conflated with the fact that a *true* normative proposition gives one a reason to do something. For example, among the normative propositions that exists is the following: ‘one ought to educate girls and young women.’ That true normative proposition gives one a reason to do something—namely, to see to it that girls and women are given an education along with boys and men. But it’s trivial that a *true* normative proposition gives one a reason to do something.

What I have just said about normative propositions also applies to imperatival contents, perhaps more obviously so. No imperative, *qua* imperative, gives one a reason to do something. A *correct* imperative, on the other hand, namely one that corresponds to a true normative proposition, does do so—but again, that’s trivial.

What about question (b)? Is there normativity that flows from the mere fact that one is following a given rule? Kripke famously claimed that there was. He said:

(...)Suppose I [am following the rule for] addition by ‘+’. What is the relation of this supposition to the question how I will respond to the problem ‘ $68 + 57 = ?$ ’? The dispositionalist gives a descriptive account of this relation: if ‘+’ meant addition, then I will answer ‘125’. But this is not the proper account of the relation, which is *normative*, not descriptive. The point is not that, if I meant addition by ‘+’ I *will* answer ‘125’, but that, if I intend to accord with my past meaning of ‘+’ I *should* answer ‘125’. Computational error, finiteness of my capacity, and other disturbing factors may lead me not to be *disposed* to respond as I *should*, but if so, I have not acted in accordance with my intentions. The relation of meaning and intention to future action is *normative*, not *descriptive*. (...).

Kripke uses this alleged fact to argue against naturalistic accounts of rule-following. But I think Kripke made a mistake here—not in being against naturalistic accounts of rule-following, but in endorsing an unqualified conception of rule-following as normative. Suppose I am following the invalid rule known

as affirming the consequent.

- From q and ‘if p , then q ’, infer p !

Does it follow that I have *reason* to conclude p from q and ‘if p , then q ’? From a subjective point of view, it might *make sense* for me to draw that conclusion, given that I am following that rule. Drawing that conclusion is *rationalized*. But it would be odd to hold that it’s *justified*. Similarly, if (like most people) I find plausible the fallacious probabilistic rule that says that if red has come up a number of times in a row in a game of roulette, it is much more likely than not that the ball will land on black the next time around, it might make sense for me to bet on black, but it wouldn’t be justified. (The rule is fallacious if each spin of the ball is probabilistically independent of the other spins).

So, there is a *quasi*-normative notion that enters into the analysis of following a rule, but it is not a genuinely normative notion. That notion is *rationalization*. This is different from *justification*. It is one thing to say that a certain behavior makes sense from a subjective point of view. It is another thing to say that it is objectively justifiable. In general, nothing normative follows from the mere fact that one is following a rule. It all depends on the content of the rule. So, where does the normativity of obeying the law come from?

If what I’ve been saying is on the right track, it doesn’t come from the mere fact that the law is a rule, nor does it come from the fact that the rule has been accepted. Where, then, could it come from? It looks as though it has to come from the truth of some underlying moral proposition roughly to the effect that:

When one is born into a society that has accepted certain norms and lives by them, and if one continues to live with and benefit from that society, then, other things being equal, one is obligated to live by the norms that are accepted in that society.

To formulate the moral proposition in question correctly would require a great deal of work. The main point right now is that it would be a mistake to look for a source of normativity either in the rule itself or in the mere fact that a rule has been accepted. If there is an obligation to obey a rule it cannot come from any source other than from the requirements of morality, which, as I previously emphasized, provide a norm on behavior independently of whether they have been accepted.

Conclusion

To sum up, then, these are the main elements of the framework that I am proposing for a talk about rules and norms.

Rules are themselves abstract objects: either normative propositions or instructions.

Their status as norms on behavior can be explained in some cases without anything—as in the case of true moral propositions—or, in other cases, via their acceptance, either directly or indirectly.

Following a rule is not in general a problem. What is a problem is explaining rule-following in cases where there is no explicit intention to conform one’s behavior to a rule.

Finally, rules and rule following facts are not normative in themselves. They derive what normativity they may on occasion have from the holding of some underlying moral truth.

Footnotes

1.

For a recent endorsement of a temporal conception of abstract see Mark Sainsbury and Michael

Tye (2012).

2.

For some considerations that bear on this issue see Boghossian (2008).

3.

See Austin (1832), Hart (1961), and Dworkin (1977).

4.

See Kripke (1982).

References

Austin, John. 1832. *The province of jurisprudence determined*. London: John Murray.

Boghossian, Paul. 2008. Epistemic rules. *Journal of Philosophy* 105 (9): 472–500.

Dworkin, Ronald. 1977. *Taking rights seriously*. Cambridge: Harvard University Press.

Hart, Herbert L. A. 1961. *The concept of law*. Oxford: Oxford University Press.

Kripke, Saul. 1982. *Wittgenstein on rules and private language*. Cambridge: Harvard University Press.

Sainsbury, Mark, and Michael Tye. 2012. *Seven puzzles of thought*. Oxford: Oxford University Press.

Wittgenstein, Ludwig. 1953. *Philosophical investigations*. New York: Macmillan.