The problem of the social foundations of normativity can be illuminated by discussing the narrower question whether rule-following is necessarily a social matter. The problems with individualistic theories of rule-following seem to make such a conclusion unavoidable. Social theories of rule-following, however, seem to only push back one level the dilemma of having to choose either an infinite regress of interpretations or a collapse into non-normative descriptions. The most plausible of these models, Haugeland’s conformism, can avoid these objections if it is supplemented with an ontologically reasonable concept of the collective attitude of a group. Groups of individuals who are bound to shared norms by recognizing each other as equipped with a standard authority of criticism have the necessary properties for ascribing to those groups such collective attitudes. Given such a weak notion of a collective attitude, there is hope for a plausible collectivist theory of rule-following.

Keywords: Rule-following; practice; recognition; collective attitude.

Introduction

Among the diverse forms of norms, rules have always attracted special philosophical interest. Since the notion of a rule, formally conceived, is especially suitable for philosophical analysis, and, since the question of whether rule-following has a social foundation is widely discussed, it seems advisable for a theory of social norms to focus on rules as an exemplary sub-set of normative entities. If it could be made plausible that rule-following necessarily exhibits social properties, there is at least some reason to think that all forms of norms could do so. In the Wittgensteinian tradition of rule-following theory, such a social foundation is widely believed to be necessary. Many interpreters of Wittgenstein attribute to him the belief that the capability of persons or of other systems to follow rules depends on social conditions. But Wittgenstein unfortunately only hinted at what these conditions might be by referring to them as “customs”, “traditions” or “ways of life” which should ground social practices, so there is ample room for disagreement as to how these social conditions should appropriately be understood.

In any case, the introduction of these social conditions is held to be necessary in the light of some specific problems with rule-following. Thus, in order to justify this introduction, it would be useful to suggest why these problems cannot be
solved with a non-social solution and to make clear that there is at least one social solution which does solve them successfully. Before discussing non-successful social solutions and before presenting a more promising candidate, it is necessary to identify these problems.

The Rule-following Problems

The first distinction which is necessary in order to identify the problems is the well-known distinction between merely conforming to a rule, that is exhibiting behaviour describable by an abstract rule, and following a rule. The behaviour of any system can be described by a number of abstract regularities, because any sequence of behaviour instantiates an infinite number of more or less complex rules. Any of these rules is as “true” as the other, as long as they are seen as objectively describing the behaviour of the system. Conforming to a rule, therefore, is possible even for non-intentional systems.

Genuine rule-following, however, constitutes a normative relation: the rule is a normative constraint on an agent’s behaviour over an indefinite variety of situations (cf. Pettit 2003, 65). The behaviour of an agent is thought to be fallibly oriented towards following the rule. Therefore, the direction of fit is such, that the rule should determine the actual behaviour and not just retrospectively describe it. Therefore rule-following entails the possibility of the agent’s acting contrary to the rule. It is a norm rather than a description, and non-conformity must be possible at least in principle. Because, if non-conformity was impossible, the distinction between a norm and a description would collapse.

If there are rule-following agents, it therefore must be possible for such an agent’s behaviour to be oriented towards the rule. That means, as a first condition, that there must be something towards which one can refer when following a rule (for these conditions cf. Pettit 2003, 82). The second condition is that it must be explainable exactly how agents can try to follow a rule. That is, in any specific case of a decision between two ways of action, there must be a way, albeit it might be a fallible one, to somehow determine which option could be the one that the rule requires. A pseudo-rule like “Always do the right thing” fails in this regard because it is of no help in making specific decisions.

Here, the first problem emerges. If there must be a way to find out what the rule requires one to do, it must—at least in some basic cases—not be a way which itself requires rule-following. This is the well-known regress problem of Wittgenstein (Wittgenstein 1958; Sellars 1963). There must be at least some rules which allow us to follow them without resorting to interpretation, if interpretation itself is a rule-governed activity (which most philosophers think it to be). This seems to rule out the idea that we can refer to the rule in all cases as if it were an abstract, law-like formulation, since abstract formulations seem to need interpretation before being applied to specific cases.
But what if we just could refer to the rule as an abstract object \textit{without} interpreting it? One is tempted to say that there are some cases where the rule needs no further interpretation because it exerts some kind of non-propositional mental power (“I somehow feel that I should continue this way”) which just enforces the correct kind of action (“the rule extends itself”). This argument is prominently refuted by Kripke (1982) following Wittgenstein’s discussion of a private language. While rules indeed determine independently from our individual interpretation what is correct (“the hardness of the logical must”, Wittgenstein 1958, 129), they do not require that anyone \textit{knows} the correct application of the rule. If the rule could just be read off from experiences of mental force, the difference between the way this mental compulsion happens and the way it \textit{should} happen according to the rule would vanish (Wittgenstein 1958, 133). In this case, the content of a norm would be whatever interpretation our mental make-up would give it; that is, it would assimilate rule-governed to rule-conforming behaviour. Therefore, a theory of rules and norms must take rules as incapable of being reduced to a level at which there are no norms (McDowell 1984, 348f.). They are normative all way down if they are held to have a definite meaning (Brandom 1994, 44).

If, however, according to the first condition, agents must be oriented towards something, and this something cannot be a Platonic abstract object like the “meaning” of the rule, then the second problem emerges: since all rules may be applied to an infinite number of cases, it is not possible for finite beings to be oriented towards the extension of a rule.

Therefore, one can now construct a skeptical argument: Neither the rule-in-intension nor the rule-in-extension can guide an agent’s behaviour, therefore there is no rule-following at all.

However, there is a flaw in this argument. It is implicitly assumed that the way in which an agent should determine what the rule demands is somehow connected with the “rule itself”. Often, the rule is pictured as an abstract “object” which can be examined or “read” (cf. Pettit 1993, 82). Such accounts therefore picture rule-following as a sort of epistemic activity in which agents are able to determine beforehand the significance of an abstract object for their deliberations (“rule-Platonism”, cf. Liptow 2004, 123).

But it is neither necessary that there is an epistemic relation to the rule “itself” in determining what constitutes following the rule, nor is it necessary that an agent should know beforehand what the correct course of action is. It is logically conceivable that one can try to follow a rule by just doing anything and hoping that one, by chance, did the right thing. It is questionable whether we could sustain rule-following as a real practice if it were so, but even given this additional constraint it would be sufficient if we just had an ability to guess correctly more often than not without identifying this capability with any epistemic access to the rule “itself”.

12
Social Solutions to the Rule-Following Problems

The upshot of the preceding argument is therefore that rules to which agents are able to refer explicitly by means of interpretation are only comprehensible if there are also some other rules which agents are able to follow implicitly. That is mysterious to some degree. On the one hand, we are warned by John McDowell (1984, 341) that we should steer clear of the Charybdis of non-normative foundations. On the other hand, when we are told to avoid Scylla (the concept of rule-following as interpretation), we cannot draw on the normativity of intentional action, motivated by the desire to follow an explicitly represented formulation of a rule.

If we abandon the idea that an agent should be able to “read” a rule directly, however, some new options open up. If an agent is trying to follow a specific rule, we could allow that it is not necessary for her to interpret this rule in order to determine the correct course of action, but that it may be sufficient to consult a community of agents and do whatever the common opinion is what the rule would require.

This general line of reasoning is in the background of Kripke’s skeptical solution to the rule-following problem as well as of the more explicit social solution of Crispin Wright, and, more recently, of John Haugeland (1990). The attractiveness of Wittgenstein’s idea of a “custom” from which these solutions take their inspiration consists precisely in that it seems to fit both of our conditions. A “custom” is not only a description, but a normative standard; and at the same time it is not some abstract proposition which stands in need of interpretation but is some way of life which is just embedded in the doings and sayings of persons.

Therefore, the social accounts share a common strategy. In this view, when agents follow a rule, it is the attitude of the community towards which they are oriented. Whatever behaviour is accepted by the community as correct is indeed correct according to the rule. The group cannot get its own customs wrong. On the one hand, this would satisfy the condition of normativity: the rule is not identified with the behaviour of the individuals who follow the rule. Their behaviour may always be judged and corrected by the community. On the other hand, there is no regress of interpretation: One does not need to interpret the rule in order to follow it, but it seems enough to act in a way that gains approval by the community.

But these solutions are confronted, in turn, with an obvious objection: If we are following a rule, we are bound to do whatever the rule requires us to do, not what the community thinks the rule would require us to do (McDowell 1984, 335). For this reason, Kripke (1982, 66) admits that this is indeed not a true solution but at most a sceptical solution which only explains how we come to think that there is something like rule-following. By contrast, Crispin Wright, for example, openly embraces the idea that there is no normativity at the level of the community.
Here, the correct sort of action is whatever the community takes to be the correct sort of action.

This objection seems to entail that the social account fails to satisfy the normativity condition, as well. For if rule-following behaviour collapses into whatever a community takes as correct, this cannot ground any normative significance. While this may be true for most simple models of group norms, there are several strategies to avoid this conclusion, depending on how one conceptualizes “the attitude of a group”.

Haugeland’s Model

John Haugeland has developed a model of rule-following, and hence also of normativity, which is explicitly designed to withstand this sort of criticism. Its basic assumptions are that of behaviourism: That means that the orientation of the agents towards the rule is not accomplished by intentions but by conditioning. He starts from minimal assumptions: There are agents who are disposed to imitate each other and who are able to react to each other’s behaviour differentially. Some of these reactions have the effect of positively reinforcing or negatively sanctioning so that the future occurrence of the respective performance becomes more or less likely. Furthermore, they have a “second-order disposition” (Haugeland 1990, 404) to react to each other’s behaviour such that they reinforce behaviour which conforms to their own and discourage behaviour which is correspondingly deviant.

Given such a community, it is inevitable that some “clusters” of behaviour will emerge to which individuals will adapt. These might be called, collectively, the “way of life” of the community. And, whenever there is such an aggregation of individual adaptions, it becomes a communal standard or, in other words, a social norm.

This account has the advantages of the simple social accounts in satisfying the conditions named above but seems to lack the disadvantages: It provides agents with something to orient their behaviour towards, namely their past conditioning which will cause them to choose a way of action which is unlikely to be punished by others; and it provides a way an agent can fallibly try to conform with the rule by acting on its dispositions, which are highly likely but not guaranteed to be a product of a successful conformist training. But does this account satisfy the normativity condition as well, that is, is it able to avoid the conclusion that the community can’t go wrong? First of all, it seems that it is. In this highly individualistic social model, there is no such thing as the position of a community. Therefore, it is useless to ask whether the community could go wrong. One may ask and give an affirmative answer to the question, whether an individual reaction of a specific agent could be incorrect. Since any reaction on any level is possibly subject to further conformist censorship by further agents, evaluation of individual performance is all there is.
Nevertheless, Haugeland’s model has attracted criticism from Robert Brandom in his *Making it Explicit*. Here Brandom argues that Haugeland’s theory is a way of failing to take sufficiently serious Kant’s distinction between acting according to a rule and acting according to a conception of a rule. Sanction theories fund this crucial distinction by means of the distinction between producing a performance and assessing it. But assessing, sanctioning, is itself something that can be done correctly or incorrectly. If the normative status of being incorrect is to be understood in terms of the normative attitude of treating as incorrect by punishing, it seems that the identification required is not with the status of actually being punished but with that of deserving punishment, that is, being correctly punished. Of course, sanctioners can be in turn sanctioned for their sanctioning, which is thereby treated as itself correctly or incorrectly done. Nonetheless, if actual reinforcement of dispositional regularities is all that is available to appeal to in making sense of this regress, it may still be claimed that what is instituted by this hierarchy of regularities of responses to regularities of responses ought not to count as genuinely normative (Brandom 1994, 36).

As Jasper Liptow noted (Liptow 2004, 140), this criticism does misrepresent Haugeland in an important respect. Haugeland has no concept of a hierarchy of reactions or of sanctions in the sense that the correctness of a specific sanction is dependent on the correctness of the sanction by which the sanctioner is sanctioned, so that it would form a regress which could only have a “dogmatic” termination. Rather, Haugeland’s normative element is to be found in the conformist disposition, which is not a specific reason for sanctioning this or that way, but which regulates on a meta-level the choice between the options in the light of the social context.

But in my opinion, Brandom’s criticism is justified if one slightly reformulates it. Even if the reactions of Haugeland’s creatures are not simply reducible to the non-normative, there still is a regress: In order to exert the sanctioning capability toward others, an agent must be able to discern conforming from non-conforming behaviour. Since it is not possible that the sanctioner and the sanctioned are exactly in the same situation at the same time, the sanctioner must choose between the possible reactions in the light of past performances by herself and other community members and decide if the behaviour displayed by the sanctioned individual constitutes a case of “going on in the same way”. And this obviously leads back into the all too well-known problems of rule-following.

The same problem arises at another point. In Haugeland’s description, agents are both governed by the desire to be imitative and the desire to censor others into conformity. Relevant situations for these desires are situations in which the performances of two (or more) individuals come apart. How should an agent now decide between adapting her own behaviour and punishing that of the other individual? She could only do so by referring to the way the community does act, and this is something Haugeland wants to avoid, since it is the cause of the problem for the traditional social accounts.
Therefore, Brandom’s criticism seems to be valid, taken in this way. Brandom therefore proposes to combine Haugeland’s account of normative attitudes with the classical, stronger social account. Such a theory would have as its central element the communal assessment of a performance (Brandom 1994, 37). At the same time, Brandom is concerned that such an account could be incomprehensible since “the idea of communal performances, assessments or verdicts on which it relies, is a fiction” (ibid., 38). Therefore, he does not fully endorse such a theory.

The question whether a community as a whole can have a certain attitude towards a performance, or if this is indeed a fiction, is therefore of central importance for a social theory of rule-following. In my view, it is possible to construe this in a way which is not as ontologically problematic as it seems on first glance.

**Collective Attitudes and Recognition**

What would it be for a group to have a collective attitude towards a specific performance? One misunderstanding which must be avoided is the view that a collective attitude is in every aspect identical to an individual belief or some other sort of individual intention. These forms of relation towards propositions might involve properties which can only be exhibited by minds. If one does not feel comfortable with a strong ontological commitment to group minds, one will want to avoid this claim.

Nevertheless, collective attitudes are similar to individual beliefs in other, essential respects: Individual beliefs have inferential properties, that is, if one endorses a specific belief, one at the same time acquires a specific normative status in terms of what further attitudes or actions one is rationally committed to in specific situations. Collective attitudes, too, go along with such a normative status, only in respect of the community.

The notion of a normative attitude would not be very helpful in enlightening the rule-following problems if it required conscious belief on the part of all individuals in the relevant community. Firstly, the possibility of a single individual going wrong would be ruled out by definition. Secondly, explicit belief is to be avoided on the danger of the interpretation regress. Therefore, it cannot be necessary to ascribe to all group members identical beliefs about what the rule demands. On the other hand, it cannot be that all members are completely and explicitly disagreeing about what the rule demands, either. Since then it would be hard to argue that they are all referring to the same rule.

We are entitled, however, to claim that a group has a certain attitude towards a specific type of performance, thereby classifying it as correct or incorrect, if all members are normatively committed to accept the inferential properties of the corresponding judgement. We could, for example, say that a group considers a certain action as being immoral if they are committed to accepting their obligation of punishing this sort of action, even if they do not have any explicit beliefs of the
relevant sort. This commitment is, in turn, possibly a purely practical disposition: if they are accepting punishment as a legitimate sanction for their own non-punishing reactions to this type of performance, they are implicitly accepting their obligation to sanction it and thereby collectively accepting it as being incorrect.

What now, in comparison to Haugeland’s account, is the specific collective element about this? It is the special role played by the concept of obligation, specifically the obligation of group members to accept sanctions—that is, practical criticism or praise towards their reactions—from each other, without demanding any special justification for this. This sort of obligation is what constitutes a group or simple community in the relevant sense. Individuals become group members by implicitly ascribing to each other an authority of judgement about communal matters. They are obliged to accept normative evaluations of other members in normal cases without demanding any evidence for their legitimacy. At this point, however, the capability of agents to represent norms explicitly is no more optional. If you are able to justify your reaction to my behaviour by pointing to a rule, it makes a difference whether you must prove why your interpretation should be binding for me or whether I am obliged to accept you as “one of us” and thereby as an authoritative interpreter of our shared rules.

Therefore, the line between “us” and “them” in such simple communities is drawn by the reactions of individuals towards normative criticism of their own reaction: if you are one of “us”, then you are entitled to criticize my reaction without giving a reason for it, just by referring to our community and the way “we” act. If, however, you are not one of “us”, I am entitled to demand explicit justification for your criticism.

This mutual ascription of authority of judgement is the concept which is called recognition in moral and political philosophy; it is the acknowledgement of the other agent as one source of the normative power of common rules.

Recognition is the act of granting one’s fellow creatures a special sort of authority in relation to some abstract objects, called shared norms or rules. This authority provides normative sanctions with an a priori, default justification. Whenever one of “us” sanctions a performance, it is not the sanctioner who has to cite a meta-norm to prove her or his authority, it is the sanctioned who has to do so, in order to deny this authority. The borders of a community of rule-followers are defined by relations in which this standard justification does not hold, that is, in which the sanctioned individual is not obliged to silently accept the sanction in favourable circumstances.

There still seems to be, however, a regress in regard to the concepts of obligation and entitlement. These concepts are obviously normative and seem already to presuppose the reality of norms and rules. Is it therefore necessary to explain what it means to satisfy these norms by reference to further norms?

However, the new model presented here focusses not on correct performances but on legitimate sanctions. And the legitimacy of a sanction is subject to
conditions different from those employed in determining the correctness of a performance. While a correct performance is defined by conformance to a rule which might only be capable of being explained by recourse to further rules of interpretation, the legitimacy of a sanction may have requirements which are much easier to fulfill: In the normative communities described, there is the default case of legitimacy of criticism because of the special relation it bears to recognition. That means that each recursion to further norms will happen only in the exceptional case of strong normative disagreement. While the infinite regress is still a theoretical problem it has ceased to be a practical one. Social norms can be fully understood even if there is only a finite set of norms, given the normative background of a community of mutual recognition and common standards of discourse.

Therefore, such a normative community is not a community of conformists. Since a person whose performance is sanctioned is able to point to a meta-rule which is believed to show that the sanction is wrong, this individual, in turn, is entitled to enjoy the same authority of interpretation. In this case, a normative conflict is unavoidable. It can be solved either if other individuals step in as undisputed authorities or if there is a higher-level norm which is interpreted in the same way by both individuals and which settles the conflict. If neither of these conditions can be fulfilled, the community between both individuals is bound to dissolve. This mutual attack on the normative authority of the other is the classical figure of a “struggle for recognition”.

Now, the notion of a collective attitude can be spelled out: a group of people displays a collective attitude if and only if they are ready to be bound by a common norm, which in turn is based on the mutual acceptance of a special authority of criticism. This model tries to avoid the metaphysical connotations of traditional accounts of collective attitudes or collective intentions by basing them on practical commitments of the individuals involved; this strategy entails, however, that the collective attitude cannot be understood as a non-normative matter. It depends on the normative statuses of the members of the relevant group.

A Moderate Collectivist Proposal

On the basis of this discussion, an approximate account can be given if both the theory of recognition sketched here and Haugeland’s model are combined. Such an account must start from a community of simple, censorious conformists, but require that they do not censor in regard to conformity with their own behaviour. Rather, they censor in regard to conformity with the community’s attitude, and this, in turn, is determined by the normatively justified disposition to accept criticism from fellow agents. Whenever these agents censor each other’s reactions they do so in the name of a shared way of life. This picture is compatible with the fact that the censoring individuals themselves vary in their interpretation as to what this way of life demands in each specific case, since
their relation to this way of life is accepted to be fallible but nevertheless having a standard authority. And, if we add the unproblematic empirical assumption that for most people under favourable conditions one way of interpreting a rule will be salient (cf. Pettit 1990), the actual probability for disagreement will be low. And since disagreement can be resolved by citing a higher-level rule (about how interpretations are justified, who is an expert for the given rule, etc.), the probability of justified claims to an exception from the standard authority of interpretation shrinks at every level.

There would, however, still be a problem with such a holistic normative account. It seems to be subject to the same problem as the traditional social theories: Does it not again presuppose that the community is infallible, only this time by referring, not to an actual, but rather to an imaginary community? Does it provide the necessary objectivity which rule-following seems to need?

There must still be a difference between what the rule demands and what an imaginary community of all mutually recognizing persons is normatively committed to. There is, however, no need to limit this imaginary community to a given set of persons at a certain time. We, for example, feel that our interpretations of epistemic obligations should also be accepted by people that lived five hundred years ago and by people who will live five hundred years from today. And, furthermore, we assume that even those persons to whom we are not in a mutual relation of recognition as authoritative interpreters of the rules of science (maybe because they were socialised in a culture which has either a vastly different or no conception of what scientific rules are) could theoretically enter into such an relationship, and we expect then our present interpretation to be acceptable for them. In other words, the obligation is to a universal community of all competent interpreters. It now seems reasonable to conclude that the only interpretation of what a rule demands which could prove acceptable to anyone who understands the rule, is what the rule really demands. But contrary to the classical interpretation of rule-following, this objective content of the rule is only an idealized notion and conceptually linked with the equally idealized notion of such an abstract community. The objective content is not something given before any interpretation, but is constituted by assuming that there is exactly one notion which would prove acceptable to all relevant persons and that we can refer to this content even if this community is never empirically constituted.

Therefore, the difference between what a person or a community takes a rule to demand and what it really demands, is at least on all practical levels, not reducible. It follows that no actual claim to the correct interpretation can have a justified claim to an exception from (justified) higher-level criticism.

While this would divert the classical skeptical attack on rule-following, there is a lot more to say about the objectivity of rules. This outline of an account is therefore only meant to demonstrate that one can avoid some of the attacks against ordinary collectivist theories.
All solutions which can be built upon this discussion are skeptical solutions, in the sense of Kripke, insofar as they do not presuppose that the content of a rule is given, even before any specific application of the rule. Before an agent applies the rule to a case the first time, its application is not somehow anticipated as an abstract Platonic form. What is the correct way of continuing the rule is decided in practice. They are, however, not skeptical solutions in the sense that what is objectively correct is fully identical with what any set of agents take to be correct. Their application of the rule as well as their criticism and censorship of other agents can only be justified if they refer to the rule “itself”.

What the rule “itself” should be is, however, a question which cannot be simply decided at any level. What can be decided, however, is the question of whether further higher-level criticism is justified in any given case.

Therefore, rule-following is only possible if there are specific forms of institutions in a community which regulate at the different levels how normative disagreement should be resolved. Since there is room for conflict at all of these levels, these institutions have to specify procedures rather than material norms. They can, however, refer to a certain way of life, which is the pre-discursive condition of membership in the relevant community.

Such types of normative institutions are more than can be established on the basis of Wittgenstein’s references to “ways of life”. They require a community to be able to bring their implicit presuppositions of normative criticism out into the open and at the same time allow the development of the normative structure of the community by means of individual disagreement.

This account proposes a collectivist solution to the rule-following problem insofar as the entity towards which the behaviour conforming to the rule should be oriented is that of communal approval. It is a moderate collectivist solution, however, insofar as no actual instance of criticism or appraisal it taken to be the “last word” on the correctness of the specific performance. Wherein the rule consists cannot be read off of anything; it can only be discerned by fallible guesses. The normative ideal on which these guesses must be oriented is the approval of an idealised community of all possible past, present and future community members after taking account of all possible criticism on any level.

Conclusion

I have argued that there is a need for a social solution to the rule-following problems, but that all existing social accounts lack sufficient arguments against the objection of non-normativity. As a model for possible solutions, the notion of a community which is constituted by intersubjective recognition as default-authorized evaluators is introduced. I have argued that such a community, which is intrinsically based on practically acknowledged norms, could provide sufficient conditions for genuine rule-following.
In retrospect, the capability for rule-following, interpreted in this way, is not a capability that creatures acquire all at once. Rather, it comes in different grades. It is dependent on certain social conditions, namely certain forms of interaction in communities, mental capabilities to refer to abstract objects like norms, propositions and evaluations and, finally, the existence of certain institutions.

It seems reasonable to distinguish between communities where all these conditions are given and communities where only a certain minimum of them is fulfilled but which are nevertheless capable of exhibiting rule-following on an “as if” basis; that is, instances in which a practice that claims adherence to rules can be sustained even if most of the conditions are not satisfied. Real communities, of course, are somewhere in between these two idealized poles. They are, however, committed to the institutional preconditions of full social normativity by their own claims to the existence of rules.

The idea of a mutual dependence between social institutions and the capability for rule-following hints to a wider importance of the social ontological issues connected with normativity to classical political and social philosophy.

References