
Kant and Wittgenstein on Reflective Judgment and Rule-Following



Hannah Ginsborg
UNIVERSITY OF CALIFORNIA, BERKELEY

De Gruyter Kant lecture delivered at the one hundred nineteenth Eastern Division meeting of the American Philosophical Association on January 5, 2023.

I. INTRODUCTION

Philosophers interested in Kant's mature account of knowledge typically focus primarily on his *Critique of Pure Reason*. There, as is well known, Kant identifies two fundamental faculties of the mind from which knowledge arises: the passive faculty of sensibility, through which objects are given to us, and the active or spontaneous faculty of understanding through which they are thought. Sensibility gives us intuitions, or singular representations of individual objects; understanding gives us concepts, or general representations of what different objects have in common. Each of these faculties has its own a priori representations: in the case of sensibility these are the pure intuitions of space and time; in the case of understanding, the pure concepts derived from the a priori forms of judgment. As Kant develops this picture of our cognitive faculties in the first Critique, it becomes more complex, with the introduction of two further faculties required for cognition. One is the faculty of imagination, whose role is to synthesize or combine the material given to us through sensibility in order to form a coherent experience. The other is the faculty of judgment, through which we apply concepts to objects that are given to us. Both imagination and judgment seem to play a mediating role between understanding and sensibility. But although Kant sometimes describes imagination as if it were an active faculty in its own right, in other passages he suggests that its workings are governed by understanding. And he is very clear that the faculty of judgment has no

rules or principles of its own, since its role is confined to subsuming objects under the rules or concepts given by understanding.

In the *Critique of Judgment*, however, Kant revisits the faculty of judgment, this time arguing that it is a faculty in its own right, with its own a priori principle. In contrast to how he treated the faculty of judgment in the first Critique, he now sees judgment not merely as subsuming objects under concepts that are already given, but as making concepts possible. He puts this by saying that there are two ways in which judgment can operate: as determining, in which case a concept is given and we subsume the particular under it, and as reflecting, in which case only the particular is given and the universal has to be found for it. This new distinction seems to imply that he now wants to allow for an exercise of judgment that, unlike the use of judgment in the first Critique, does not consist simply in the application of concepts which we already possess. Rather, Kant is now saying, there is a kind of judging—reflective judging—that does not depend on our already grasping concepts, but that enables us to acquire concepts, or makes concepts possible in the first place. On the face of it, this seems like an important and radical idea, deserving of being taken seriously both as a philosophical idea in its own right, and as a contribution to Kant's view of cognition. If reflective judgment is indeed a condition of the possibility of concept-acquisition, or of concepts themselves, then it would seem that it should play a crucial role both in Kant's account of cognition, and in our thinking about cognition generally.

But there are a number of factors that have led to the idea of reflective judgment's not getting much attention except from scholars who have a particular interest in the third Critique. One factor is that, although Kant says in a few passages that reflective judgment makes concepts possible, he tends to focus primarily on what seems to be a quite different role of reflective judgment, that of giving rise to judgments of beauty, which, he emphasizes, do not involve concepts and are not cognitive. Moreover, when Kant does describe reflective judgment's contribution to cognition, he often emphasizes its role, not in the conceptualizing that leads to everyday concepts like *dog* or *chair*, but rather in the kind of scientific theorizing that yields hierarchical systems of empirical concepts and laws. A second factor is that the *Critique of Pure Reason*, in investigating the "conditions of the possibility of experience," is often thought to provide an exhaustive account of empirical cognition. If this is so, then there is no need to look to the *Critique of Judgment* for an account of concept-acquisition, or of the possibility of concepts, since that issue will already have been addressed in the *Critique of Pure Reason*. These two factors suggest a division of labor between

the first and third Critiques on which the task of the first Critique is to explain the possibility of cognition, both empirical and a priori, whereas the task of the third Critique is to address issues that are more peripheral to cognition: aesthetic judgment, teleology, and our capacity for systematic scientific enquiry. It can seem difficult, then, to see the reflective judgment as making an important or interesting contribution to the understanding of cognition as such, whether for Kant or for us.

I believe that this relative lack of attention to reflective judgment is a mistake. In this paper, I am going to try to bring out what I take to be the philosophical importance of reflective judgment both for Kant, and for our own thinking about cognition, by drawing a parallel between Kant's idea of reflective judgment and Wittgenstein's rule-following considerations. I will argue that, with the introduction of reflective judgment in the third Critique, Kant is addressing broadly the same problem that engages Wittgenstein in his discussion of following a rule. Very roughly speaking, the problem is that of how acquaintance with a finite number of particular things can put us in a position to acquire a general rule—a rule which we can then go on to apply to an indefinite number of possible future instances. Without a solution to that problem, we cannot make sense of our capacity to think the individual things we encounter as having common properties and thus as exemplifying concepts. I take the idea of reflective judgment to be Kant's answer to this problem, and I take it to correspond, again broadly speaking, to Wittgenstein's approach. As we will see, the answer appeals to our natural tendencies, both to sort things together in some ways rather than others, and, in each instance of sorting, to regard our sorting in each instance as correct or appropriate. Kant's view that we have a faculty of reflective judgment with its own a priori principle is, in a nutshell, the view that we are entitled to do this. And although Wittgenstein doesn't use the language of "entitlement," let alone "faculty" or "principle," I think his answer amounts to the same thing. We can grasp general rules, according to Wittgenstein, because we can recognize the correctness of going on in one way rather than another from finite sequences of behavior, and we can do so without having already grasped rules which tell us that this rather than that way of going on is correct. In effect, this amounts to saying that we are capable of reflective judgment: of a kind of judging that makes rules or concepts possible rather than applying rules or concepts that we already grasp.

II. EMPIRICAL CONCEPTS IN THE *CRITIQUE OF PURE REASON*

I noted earlier that the *Critique of Pure Reason* is often viewed as offering a complete account of how cognition is possible. If that is so, then it should not be necessary to invoke reflective judgment to explain the possibility of empirical concepts. So we need to consider whether Kant does in fact account, in the first Critique, for our acquisition and grasp of empirical concepts.¹ This is difficult to determine because Kant says very little in the first Critique about empirical concepts specifically; the concepts that most interest him there are the pure concepts of the understanding. One of his primary aims in the first Critique is to show that the use of these concepts is required for experience, and he does so, in part, by arguing that they are built into our faculty of understanding, without whose involvement experience is not possible. Now, empirical concepts too are required for experience, and, without them, there can be no use of the pure concepts. I cannot perceive something as, say, a substance standing in causal relations without perceiving it as a substance of some particular kind standing in particular causal relations: for example, as a dog that might bite me if I approach it, or a stone which will get warm if the sun shines on it. But the concepts through which I represent these particular kinds of substance or causal interaction are not built into the faculty of understanding the way the pure concepts are. Nor can I derive them directly from sensibility, since sensibility gives us only intuitions, that is, singular representations. We don't immediately intuit things as having general features or as sorted into kinds; rather representing things as instances of kinds or as having common features is a kind of thinking, and so requires spontaneity. But on the face of it, all spontaneity belongs to the understanding, which, as we have seen, is not, at least on its own, the source of empirical concepts.

The obvious place to look for an answer to the question of empirical concept-acquisition is Kant's account of the imagination, which as I noted earlier, mediates between sensibility and understanding by synthesizing the intuitions given to us by sensibility into a coherent form. Kant is not very explicit about the workings of imagination; in the sketch I am about to give I am relying on a brief but illuminating interpretation offered by Strawson.² On this interpretation, imagination works in part by incorporating into my perception of an object, on any one occasion, elements drawn from previous perceptions, whether of the same object or of the same kind of object. For example, if I see an unfamiliar dog, look away, and then look at it again, imagination combines my sense-impressions in such a way that I see the dog as a continuously existing thing. Here, when I see the dog for the second time, I am in some sense recalling, or to use Kant's term, "reproducing" my previous perceptions

of the same dog as part of my present perception, so I do not perceive it as a new dog which has just come into existence. Or if the dog—let's call him Spot—is curled up in a basket so I can't see his feet, imagination allows me to represent him as having four paws, rather than as legless or as having hooves. Here I am "reproducing," in my present perception of Spot, elements drawn from my previous perceptions, say of Rover, Lassie, and Fido. This picture is something like Hume's account of the imagination, on which present impressions call to mind ideas "copied" from previous impressions according to psychological laws of association. But there are two differences. One is that, for Kant but not for Hume the imaginative activity is not additional to perception but rather, "a necessary ingredient in perception itself" (A 120n.).³ Second, synthesis for Kant is, on the face of it, a rule-governed activity rather than one which is determined by psychological laws. This second point is directly relevant to the question we're now considering, because Kant describes empirical concepts as either signifying, or as identical to, rules for the synthesis of imagination. He says in the Schematism that "the concept *dog* signifies a rule in accordance with which my imagination can delineate [*verzeichnen*] the shape of a four-footed animal in a general way" (A141/B180). In the first edition of Transcendental Deduction, he identifies empirical concepts with rules. "The concept of body . . . serves as a rule for our cognition of outer appearances' (A106). And he goes on to say something about how the concept *body* serves as a rule: "it represents in given appearances the necessary reproduction of their manifold . . . thus the concept of body in the perception of something outside us makes necessary the representation of extension, and with it, those of impenetrability, shape, etc."

The idea of concepts as signifying or as identical with rules for synthesis suggests a possible answer to the question of how we acquire the empirical concepts that figure in our explicit judgments. I come to grasp the general representations *dog* or *body*—and am thus in a position to make judgments like "some animals are dogs" or "bodies are heavy"—by reflecting on, and thus bringing to conscious attention, the rules that I have been following, more or less unconsciously, in my synthesis. Kant indeed suggests a picture like this when he describes the understanding as "bringing synthesis to concepts" (A78/B103). On this picture, we synthesize according to rules to which we do not attend; we then arrive at concepts, of the kind that figure in judgments, by reflecting on what we are doing and so bringing the rules to full consciousness. But this answer only gets us so far. For we now need to ask, what is the source of the rules themselves that govern the activity of synthesis? We have seen that these originate neither in sensibility alone nor in understanding alone. So it seems that they must somehow derive from imagination in

its role of mediating between sensibility and understanding. But what does this mean? One answer is that Kant's view of imagination is more like Hume's than I suggested earlier. Perhaps the rules governing my synthesis are nothing more than psychological laws describing how I, and other human beings, do in fact synthesize representations. On this understanding of synthesis, when I represent Spot as having paws rather than hooves, I am not following a normative rule requiring me to call to mind past dogs rather than past horses, but simply succumbing to a natural tendency or inclination to do so. I arrive at the concept *dog* by reflecting on this natural tendency, and bringing it to consciousness in the form of a general representation of what is common to dogs to the exclusion of other things.

However, it is unclear how we could ever come to grasp concepts if our only basis for doing so were the conformity of our imaginative activity to psychological laws. For these laws govern me, so to speak, only from the outside. I could come to know what they are only by observing myself and making generalizations about which representations I reproduce or, correspondingly, which objects I classify with which other objects. This would require me to have a whole repertoire of concepts in order to describe myself and my own imaginative activity, and these would almost certainly include the concepts, such as *dog*, whose origin we are trying to understand. Kant's suggestion that we can come to grasp empirical concepts by reflecting on the rules governing our synthesis requires that we can become aware of the rules, so to speak, from the inside. And this requires that they be already available to us prior to the synthesis on which we reflect: unlike psychological laws of association, or for that matter laws governing the subpersonal workings of my visual system, they must be potentially accessible to consciousness without the need for observation from a third-person perspective. So when I perceive Spot as having four paws, it must be that I am aware, or at least can become aware, that this is the right way to represent Spot, that it would be wrong or inappropriate to perceive him as having hooves or a fish tail. In effect, this amounts to saying that I must be aware, or capable of becoming aware, that it is right to classify Spot together with Fido, Lassie, and Rover and wrong to classify him with Bucephalus and Seabiscuit. For it is only this kind of awareness that could yield, on reflection, a concept of Spot as having the general property of being a dog. But this awareness is excluded on a view which takes this kind of classification solely as a matter of conformity to natural psychological laws. So we still lack an answer to the question how, according to the *Critique of Pure Reason*, we can arrive at empirical concepts.

III. EMPIRICAL CONCEPTS IN THE LOGIC

The considerations raised so far suggest that the *Critique of Pure Reason* does not give a full account of empirical cognition: it leaves us with a question about how we arrive at empirical concepts and are thus in a position to make empirical judgments in which those concepts figure. However, there is a passage in the *Logic* which is often seen as addressing that question, and for our purposes it will be useful to take a critical look at it. The core of the account offered in that passage is presented in the following example: "I see e.g., a spruce, a willow and a linden. In first comparing these objects among themselves, I notice that they are different from one another with respect to the trunk, the branches, the leaves and so forth; but now I go on to reflect only on what they have in common, the trunk, the branches, the leaves themselves; and I abstract from their size, shape and so forth; thus I receive [*bekommen*] a concept of a tree" (*Logic* §6, note I; 9:94–5). The passage describes a threefold activity of comparison, reflection, and abstraction: I compare the three trees with a view to identifying their differences (in the shape and size of their trunk, branches, and leaves), I "reflect" on the trees to identify what they have in common (their having a trunk, branches, and leaves), and I then disregard the differences and attend only to the common features. This supposedly yields the concept of a tree.

But there are two respects in which this account fails to give a satisfactory explanation of empirical concept-acquisition. First, it takes for granted that I am already capable of recognizing features like having branches or having leaves, and this implies that I already grasp some empirical concepts, if not the concept *tree*. So there is a threat of regress: in order to explain our grasp of the concept *tree* we need to explain how we arrive at the concepts *trunk*, *branch*, and *leaf*, and this explanation in turn will presumably depend on still further concepts. Second, and more importantly for our purposes, the account takes for granted that I already, in some sense, grasp the concept *tree* itself. For it assumes that I already classify the spruce, willow, and linden as trees, rather than as members of some other kind. We can see this by focusing on the phase of "reflection," and asking why it is that, in identifying what the three trees have in common, I come up with their having a trunk, branches, and leaves. For there are many other features they have in common, which reflection could also have identified. For example, the three trees all provide shelter from sun and rain, they all harbor insects, they are all composed largely of woody material, and none of them bear edible fruit. On the face of it, I could equally well have fastened on the features I have just mentioned, and disregarded the possession of trunk, branches, and leaves. I would thus have acquired a concept that included in its

extension non-fruit-bearing trees and wooden houses, and excluded, say, apple trees. The fact that we instead focus on the features we do—the possession of trunk, branches, and leaves—indicates that we are already conceiving the spruce, willow, and linden in such a way that they fall under the concept *tree* rather than, say, the concept *non-fruit-bearing tree* or *wooden house*. We do learn something from the exercise of comparison, reflection, and abstraction, namely, that a tree can be characterized as something with a trunk, branches, and leaves. But we don't, for example, acquire a capacity to identify things as trees, or to determine of some new object whether or not it is a tree. For we already needed that capacity in order to identify possession of trunk, branches, and leaves as distinctive features of being a tree.

As with our earlier discussion of the synthesis of imagination in the first Critique, it might here be suggested that our classification of the objects we have considered into kinds corresponding to concepts like *tree* or *house*, rather into kinds corresponding to concepts like *non-fruit-bearing tree* or *wooden house*, is nothing more than a reflection of our natural tendencies to classify objects in some determinate ways rather than others. It simply comes naturally to us to sort the apple tree with the spruce, the willow, and the linden, just as it comes naturally to us to sort Spot with Rover, Lassie, and Fido. But this suggestion leaves us with the same problem, namely, that it leaves mysterious how we could move from having those natural sorting tendencies to grasping the corresponding concepts. The knowledge that I am naturally inclined to sort trees with other trees and dogs with other dogs is available to me only third-personally, through observation of my psychological activity, and it presupposes grasp of the concepts *dog* and *tree*. It cannot be the basis of my coming to grasp those concepts.

A way to put the difficulty here is that, in order for me to acquire the concept *tree* by reflecting on the spruce, willow, and linden, it is not enough that I be presented with those three trees and identify features that they have in common. For there are an infinite number of such features, and only a few of them are distinctive of trees. What I need, rather, is to represent the spruce, willow, and linden in such a way that I recognize that it is correct to sort, say, the apple tree with them and incorrect to sort the house with them. My observation of the spruce, willow, and linden has to put me in a position to recognize, for each of an indefinite number of things I might encounter in the future, whether it should or shouldn't be classified with them. But what could entitle me, for any one of those future things, to take my classification to be correct? If I have already brought the spruce, willow, and linden under the concept *tree*—as opposed to the concept *non-fruit-bearing tree*

or wooden house—then I can reason that the apple tree should be classified with the spruce, willow, and linden because it is like them in having a trunk, branches, and leaves, and that the house should not be classified with them because it lacks those features. But if reflection on the spruce, willow, and linden is supposed to yield the concept *tree*, then I cannot assume that I have already brought them under that concept. So we seem to be left without any way of explaining how my acquaintance with the spruce, willow, and linden can make it possible for me to acquire the concept.

IV. WITTGENSTEIN'S PUZZLE ABOUT RULE-FOLLOWING

At this point, I want to bring in the puzzle about rule-following that Wittgenstein raises in *Philosophical Investigations*. The *locus classicus* of the puzzle, at §185, is a passage about a child who is learning the elements of arithmetic. The child is being taught to write out sequences of numbers in response to simple instructions like “add 1” or “add 2”: in the example, he is learning to respond to the order “add 2” by writing segments of the series 0,2,4,6,8,10... This child has shown himself successful for numbers up to 1000, but then, to our surprise, he continues the sequence by writing 1000, 1004, 1008. When we challenge him, he insists that he is going on correctly, that 1004 is what he should write after 1000. And he also insists that, in writing 1004 after 1000, he is doing the same as he did in the earlier part of the series. Wittgenstein comments: “we might perhaps say: this person finds it natural, once given our explanations, to understand our order as we would understand the order “Add 2 up to 1000, 4 up to 2000, 6 up to 3000, and so on” (*Philosophical Investigations* §185). In other words, we might think that the child, through some psychological quirk, has extrapolated a different rule from our examples than the one we intended. We typically take for granted that the rule or concept exemplified by the sequence of even numbers from 0 to 1000 is the rule *add two*, but that same sequence of numbers could equally well be viewed as instantiating the rule to add 2 up to 1000 and then to add 4 (which I shall abbreviate as the *schmadd-two* rule). And if the child has indeed understood the sequence as instantiating the *schmadd-two* rather than the *add-two* rule, then he is quite correct, it appears, to write 1004 after 1000.

The case of the aberrant child presents a problem, not because we are ever likely to encounter such a child, but because it raises a worry about our own case. We might initially be confident that, when we write 1002 after 1000, we are going on correctly, and that the child who says 1004 is going on wrong. But once we consider that the sequence of even

numbers from 0 to 1000 can be seen as instantiating not just the add-two rule, but any number of rules that did not previously occur to us, then it can seem that we have no ground for our confidence. If we cannot determine whether the sequence should be understood as instantiating the add-two or the schmadd-two rule, then it looks as though there is no answer to the question whether 1002 or 1004 is the right way to go on after 1000. And that suggests that our confidence in the correctness of what we write at each step, when we continue a series like 0,2,4,6,8,10... is illusory. Any one way of going on, it appears, is just as good as any other, with the upshot that there is no such thing as going on correctly, and no such thing as following the rule of adding two.

Now, it can look as though the worry is about numbers or arithmetic, but in fact it is far more general. For the source of the worry is that any finite set of particulars—whether particular instances of behavior or particular objects—can in principle be understood as instantiating an indefinite number of rules or concepts. Take the set of all dogs that have existed up until the present day. None of those dogs has ever been to the moon. So what determines that we should understand them as instances of the concept *dog* rather than the concept *dog that has never been to the moon*? We are confident that if, say, Spot were to travel to the moon, we would continue to sort him with the other dogs; for example, we would continue to use the word “dog” in referring to him. But the worry is that this confidence is unjustified: perhaps we should instead call him a “horse” on the grounds that, like all the horses we have so far encountered, he falls under the concept *horse which has never been to the moon or dog which has been there*. As in the case of the arithmetical series, this can be put as an issue of whether there can be right or wrong ways to “go on” from a sequence of behavior. Just as we can ask whether writing “1002” after “1000” is going on in the right way from what we or our teachers wrote earlier, we can ask whether calling Spot a “dog” is going on the right way from our earlier uses of “dog” in connection with dogs which hadn’t been to the moon. But the worry can also be put as an issue about the right way to sort things: whether we can think of it as correct to sort Spot with the dogs rather than the horses. And put in that way the worry is that it makes no sense to speak of right or wrong ways of sorting or classifying things: that any one way of classifying things is as good as any other. This worry, if justified, calls into question the very idea of rules and concepts. For if there are no right or wrong ways of going on in a sequence or of sorting one object with others, then it is hard to see how there can be any such thing as following a rule or thinking of an object as falling under a concept.

What is Wittgenstein's answer to this worry? As I understand him, his answer is that the problem is based on a misconception: namely, that in order for me to be correct in recognizing that 1002 is the right way to go on, or in recognizing that Spot should be classified with previously encountered dogs, I must first have extrapolated a rule or concept from the preceding instances. According to Wittgenstein, I do not first need to have identified those instances as exemplifying the add-two rule in order to know that 1002 comes after 1000 in the series. Rather, my capacity to know how to go on from examples presented to me, to recognize what goes with what, is more primitive than my capacity to identify general rules corresponding to features which the examples have in common. Wittgenstein puts this by saying that there "is a way of grasping a rule that is not an interpretation but which, from case to case of application, is exhibited in what we call 'following a rule' and 'going against it'" (*Philosophical Investigations* §201). My grasp of the rule *add two* in the first instance consists, not in my having interpreted the previous numbers in the sequence as exemplifying one of all the possible rules that it could exemplify—the rule of adding two rather than *schmadding* two, *say*—but rather in my being able to recognize, without having antecedently grasped a rule, that writing (say) 1002 after 1000 is correct and that writing (say) 1004 after 1000 is incorrect. I might describe what I recognize by saying that writing 1002 is a case of "following the rule" and that writing 1004 is a case of "going against it." But the idea that I am "following the rule" in writing 1002 does not imply that I previously grasped an item of intentional content—a general representation of what it is to add two—and am now being guided by it. Rather, it simply means that I recognize writing 1002 as the right thing to do after having written the previous sequence of numbers. This is not to say that I don't have a general representation of the add-two rule, nor that I cannot conceive of possible variants like the *schmadd-two* rule. On the contrary, I do have a general representation of the add-two rule, and I can use it to define other rules, such as the *schmadd-two* rule and all its variants. But I have that general representation only in virtue of my more fundamental capacity to know, at any point in the series, what to write next: my grasp of the concept consists in my knowledge of how to go on in the series. This knowledge involves my being able to recognize, at any point in the series, that what I write is correct. But it is a mistake to believe that this recognition—this exercise of judgment, we might say—must be justified in terms of an antecedent grasp of the rule.⁴

If I am right in thinking that this is Wittgenstein's view, then he is answering the worry by saying that we have a capacity not just for determining, but for reflective judgment. He is drawing our attention to a kind of judging that does not consist in the application of an already

grasped rule or concept, but that instead makes concepts possible. This is the kind of judging that we carry out when, in the context described by Wittgenstein, we take 1002 rather than 1004 to be the right way to continue the series, or when, in a context where no general concept or rule of sorting has been specified, we take Spot to belong with Rover, Fido, and Lassie rather than with Bucephalus and Seabiscuit. The fact that we judge the way we do in these contexts is a matter of empirical psychological fact: it comes naturally to us to go on, or to sort things, in these ways rather than others. But that does not stand in the way of our taking ourselves to be judging correctly. Normally, we do not question our knowledge that the right way to continue the series is to write 1002, or that Spot belongs with the dogs, not the horses. We would not hesitate to say that a pupil learning to continue the series was going wrong if he wrote 1004. Similarly, to take an example closer to reality, we do not hesitate to judge that a three-year-old in an intelligence test who sorts a picture of a dog with pictures of horses rather than other dogs is making a mistake. And we describe ourselves in these cases, not just as taking ourselves to be going on or sorting correctly, but as knowing that we are. Wittgenstein's answer vindicates this ordinary understanding of ourselves. If, after philosophical reflection, we come to think there is a problem about how we can know how to go on, it is only because we mistakenly assume that the only kind of judgment we can exercise is determining. We fail to see that the possession of the concepts we apply in determining judgment presupposes the exercise of reflective judgment. That is, it presupposes a capacity to recognize, without having to rely on a prior grasp of concepts, the right way to go on, or what should be sorted with what.

V. THE PRINCIPLE OF REFLECTIVE JUDGMENT

I want, now, to return to Kant. I argued in sections II and III that the *Critique of Pure Reason* and the *Logic* fail to account fully for empirical cognition because they fail to account for the acquisition of empirical concepts. To account for empirical concepts we need to make sense of our capacity to recognize that things belong together, or should be sorted together, in determinate ways. This requires understanding how I can recognize, for example, that Spot should be sorted with Fido, Rover, and Lassie, but without presupposing that I already recognize them as dogs, or as having any other feature in common. Kant is unable to address this problem in the first Critique, I now want to suggest, because he does not recognize the possibility of a kind of judging that does not depend on a prior grasp of concepts. But in the third Critique he broadens his conception of judgment to include the kind of judging

whereby we can recognize things as belonging together without already having subsumed them under a common concept. This allows him to explain how, by reflecting on our activity of sorting, we can abstract general representations for use in explicit subject-predicate judgments. Although I do not first need to bring Fido, Rover, and Lassie under the concept *dog* in order to recognize that Spot should be sorted with them, my coming to recognize that Spot should be sorted with Fido, Rover, and Lassie amounts to my coming to recognize them all as dogs. In this way, the exercise of reflective judging through which I sort Spot with the other dogs amounts to the acquisition of the concept *dog* and helps put me in a position to make judgments like “dogs bark” and “some animals are dogs.”

While I do not have space here for a detailed defense of this interpretation, I do want to relate it to Kant’s discussion of the a priori principle of reflective judgment. Kant claims, in the Preface to the third Critique, that judgment, “the correct use of which is so necessary and universally required” that it is just what we mean by sound understanding, must have a principle, since otherwise it would not be exposed to the “commonest critique” (5:169). (And here by “judgment” he is referring to judgment as reflective, since he makes clear elsewhere that determining judgment does not need a principle.) But he also says that there must be “great difficulties” involved in finding a principle for judgment. The main difficulty, he suggests, is that the principle has to be one “through which we do not actually cognize anything.” It has to serve as a rule for the faculty of judgment, “but not as an objective rule, to which it could adapt its judgment, since then we would need another faculty of judgment in order to decide whether or not the judgment is a case of that rule.” And he refers to the resulting difficulty as an “embarrassment” about a principle (5:169). The difficulty is apparently resolved, however, by identifying a subjective principle, whose application does not require that we determine whether or not we are successfully according with the principle.

But what is the principle? In the section of the First Introduction in which Kant introduces reflective judgment and also puts the most emphasis on its role in empirical conceptualization (section V), he describes it as the principle that “for all things in nature empirically determinate concepts can be found” (20:211). He then goes on to offer, both in the First Introduction and in the published introduction, a bewildering variety of alternative characterizations.⁵ A number of these invoke the idea that nature is what he calls “purposive” for our cognitive faculties or for our faculty of judgment, and several bring in the idea that concepts of natural things, and the corresponding natural laws, are organized in

the form of a system, so that they meet our need not just to bring natural things under empirical concepts, but to understand nature in terms of a systematic hierarchy of empirical concepts and laws. Commentators discussing the principle have often focused on the formulations in terms of the systematicity of nature, wondering in particular what licenses such an ambitious and apparently contentful principle. But I want to focus on the initial formulation, on which the principle is simply that we can find empirical concepts for objects, and on the related formulations in terms of the purposiveness or suitability of nature for reflective judgment itself. For these allow us to see most clearly why the principle is not an objective principle, and not one through which we cognize anything about the world. What the principle says, according to these formulations, is that there is a fit or conformity between things in the world ("nature," broadly construed) and our own activity of sorting or classifying those things, so that our classifying does in fact yield empirical concepts, that is, representations of which things should be sorted with which other things. To state the principle in an even more bare-bones way: it is that, in our sorting of things into kinds, we are sorting them as we ought to sort them. This is not to state any objective fact about nature, nor, more generally, about the objects presented to us, whether natural or otherwise. Rather, it is simply to endorse the correctness of something that we take to be correct anyway. When I sort Spot with Rover, Lassie, and Fido, rather than with the horses, or when I sort the apple tree rather than the house with the spruce, willow, and linden, I take myself to be sorting appropriately. And I do so even though I cannot justify it by saying that the things I am sorting together are all dogs or all trees. What the principle of reflective judgment says, on the minimal construal I am offering, is simply that I am entitled to do this, that the absence of conceptual justification does not make it illegitimate. I can take at face value my conviction, in exercising reflective judgment, that I am judging appropriately. In a sense, the principle of reflective judgment, at its most minimal, is nothing more than the principle that there is such a thing as reflective judgment. That is, it is possible, without already having grasped empirical concepts, to engage in a sorting activity through which we can arrive at empirical concepts.

The principle of reflective judgment, then, does not actually guide us in our sorting of objects—it does not tell us what we should sort with what. In this respect it is unlike, say, the principle that we should sort organisms together with other organisms in preference to artefacts, or that we should privilege shape and internal constitution over color and size, or that we should avoid sorting in ways which correspond to disjunctive concepts or to concepts which make reference to specific times and places. These might be useful heuristic principles, but they

ultimately depend on, rather than replacing, the principle of reflective judgment, since they are plausible only to the extent that they capture our more basic judgments about what should be sorted with what. The principle of reflective judgment is also unlike the principle that we should sort in ways corresponding to the natural sorting dispositions of members of the human species. This point deserves emphasis because we might otherwise be tempted to confuse the principle of reflective judgment with a principle of this kind, and that would be a mistake. When I take myself, on some given occasion, to be sorting appropriately or correctly, my entitlement to do so is not based on my recognizing the empirical fact that this is how human beings typically do sort. The principle that we should sort in accordance with the natural tendencies of human beings would be an objective principle, and adopting it would require us to carry out empirical investigations and form hypotheses about how human beings are in fact disposed to sort things. Such a principle, even if we wanted to adopt it, would largely be unusable as a way of coming up with concepts, since grasping it would require a grasp of the very concepts that reflection according to this principle was seeking to arrive at. The principle of reflective judgment, on the other hand, makes no reference to the capacities of human beings as such, but only to “our” cognitive capacities: it is essentially first-personal. It allows us to judge, in any case in which we take ourselves to be sorting one thing correctly with another, that we are in fact sorting correctly, but it provides no justification for the correctness of that judgment. Nor could it, since otherwise it would not avoid the “embarrassment” referred to in the Preface.

What then is the point of the principle, if it neither guides us nor justifies us in the exercise of our judgment? As I see it, the point is simply to disavow skepticism about our ordinary capacity for recognizing what should be sorted with what—a capacity which, as we saw, is so pervasive, so universally required, that Kant describes it as interchangeable with “sound understanding.” We do, as a matter of course, trust our ordinary judgments about how to go on in our use of words, which similarities among things are important and which can be disregarded, which comparisons do and do not make sense, and so on. The principle of reflective judgment says no more than that we should keep on trusting those judgments, that is, that we should refrain from taking the absence of justificatory concepts or principles as a reason for doubting our entitlement to make them. Kant’s point, in introducing the principle of reflective judgment, is not to prescribe new ways of thinking about the world, but rather to vindicate our ordinary ways of thinking, which include the thought that these ordinary ways are legitimate even if they do not allow for conceptual justification. Here we can see an affinity,

at a more general level, between Kant and Wittgenstein. For, at least in this domain, Kant's philosophy is fundamentally nonrevisionary. Our ordinary ways of thinking, Kant holds, are in order as they are: in this respect his philosophy, like that aspired to by the later Wittgenstein, "leaves everything as it is" (*Philosophical Investigations* §124).

NOTES

1. I address this question in more detail in my "Thinking the Particular as Contained Under the Universal" and "Lawfulness Without a Law: Kant on the Free Play of Imagination and Understanding," both reprinted in *The Normativity of Nature: Essays on Kant's Critique of Judgement* (Oxford: Oxford University Press, 2015).
2. P. F. Strawson, "Imagination and Perception," in *Freedom and Resentment* (Oxford and New York: Routledge, 1974), 50–72.
3. I follow the standard practice for citing Kant, using A and B page numberings for the first and second editions of the *Critique of Pure Reason*, and volume and page number of the *Akademie* edition (Kant 1902–) for other texts. Translations are my own.
4. I defend this reading in "Wittgenstein on Going On," *Canadian Journal of Philosophy* 50, no. 1 (2020): 1–17.
5. See, for example, First Introduction V, 20:215–16, and Introduction IV, 5:180.