Actions, Mistakes, and Rule Following: An Account of Mistaken Actions

By
Andrew Lampus

Submitted to Central European University

Department of Philosophy

In partial fulfilment of the requirements for the degree of Master of Philosophy

Supervisor: Professor Ferenc Huoranszki

Vienna, Austria

2022

Abstract

This thesis gives an account of the nature of mistaken action. I analyze the nature of mistakes through their relation to intentional action on the one hand and rule-following on the other. I focus on the question of how to account for incorrect and correct applications of a rule and how this in turn gives us an explanation of mistakes. In Part one I begin by giving an account of the relationship between basic action and the nature of agency. I defend the claim that the core of agency is basic actions that are not necessarily intentional, but rather 'pre-intentional'. Next, I introduced levels of generality in descriptions of intentional actions and argue that every description of intentional and unintentional action involves a nested hierarchy of act types. I argue that an action is a mistake if and only if an unintended occurrence is realized in the external action that preserves the intended type of action. Finally, I bring these results to bear on the nature of rule-following and I argue for an understanding of Wittgenstein's remarks on blindly following a rule as a kind of basic action.

Table of Contents

1.Introduction

The central thought of this thesis is that action, on a fundamental level, is mistake-free. This is not the claim that it is impossible for an action to be a mistake, but rather the idea that the notion of a mistaken action cannot be directly derived, or 'read-off' of the action itself. Its source must be sought on a level 'above' the action, above the bare movements that constitute the basic level of action. But what is the basic level of action? Is such a concept coherent? And furthermore, could it possibly be mistake-free?

Any bodily movements performed by an agent are presumably constituted by the elements of the body (molecules, atoms, elementary particles) and the same types of elements constitute the movements of non-agential bodies. So, it cannot be at this level that mistakes come into the picture, but what then allows the concept to get off the ground? There is a corresponding question related to action. What separates the 'mere happenings' of my body from the movements of my body that we could characterize as actions? We can interpret Wittgenstein as raising a similar question when he says "Let us not forget this: when 'I raise my arm', my arm goes up. And the problem arises: What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?" (Wittgenstein, 621 PI). The question then is what separates a mistake from a mere unwanted occurrence? One plausible answer to this is that the concept of a mistake becomes applicable if and only if the concept of action is applicable. In this case, we need to be able to attribute agency to the object and along with this some notion of an intention and an intentional action. If we introduce the concept of intention, then any action that is classified as a mistake ought to be seen in relation to this concept, where some sort of intention is needed as a *precondition* for a mistake to occur.

1.1 The Logical Basis for the Possibility of Mistakes

My purpose in defending the notion of basic action is to provide a basis from which an analysis of mistakes can occur. For one, in order for a mistake to be possible, so I shall argue, there must be something an agent does which is *intentional*. By this I mean to indicate that only

against the background of an action that is not a mistake, can the result of that action be considered a mistake. In other words, the concept of a mistake *simpliciter* is incoherent. This claim might at first glance seem to be either absurd or too strong to be plausible. Common sense and the way we use language seem to suggest that individual actions are themselves mistakes. In the case of actions that involve mistakes we often seem to attribute the mistake to the action itself. For instance, "John mistakenly hit the gas pedal at the red light" or "John typed 's' instead of 'a' when spelling 'action' by mistake". In both these cases, the act itself seems to be a mistake, yet it should be noted that in each case we would naturally consider John to be the *agent* of the action. But if John is the agent of the action, then it would also seem to imply that he did *something* that was not a mistake. For instance, he intentionally moved his foot, but it was just that he moved it in the wrong direction. Or he intentionally engaged in the act of typing, but his finger missed the 'a' on the keyboard.

In these examples I am assuming that the agent knew where the gas pedal was located, and he knew where 'a' is located on the keyboard. In these cases, he didn't make a mistake because of some false belief, but rather he made a mistake in the *performance* of the action. To make sense of the conflicting sense in which the agent both performed an action that we would call a mistake and the idea that they were nonetheless the agent of the action, we should first ask whether we can separate these two aspects in the action itself. For instance, was there simply one action that, in itself was a mistake, or were there two or more actions, one which was a mistake and another that was not? But before we answer this question (which I will answer by saying that there is *one action with multiple descriptions and multiple act types that the descriptions can take*) we must ask what the prior conditions for a mistake are in the sense in which the action is an instance of agency.

The way an action is an instance of agency can in turn be brought out through the relation between basic actions and agency. I will argue that a basic action is one that exhibits agency, but is not necessarily *intentional*. This will provide an explanatory basis for an account of what it is to make a mistake by securing a teleological starting point for intentional action to arise. So, in order for us to say an action was a mistake, we first need a kind of action that is *logically* more fundamental and where the concept of a mistake is not applicable. From this basis we then need a

way of introducing the possibility that an action is intentional or unintentional, only after this notion is possible, will it then be possible to derive conditions for an action to be a mistake. My proposal is that basic actions are coextensive with actions that, logically, can't be mistakes. Though they are mistakes free we can still make sense of them as actions because they are acts of agency.

2. The Concepts of Basic Action and Agency

Danto introduced the term 'Basic Action' by which he means an action that is not the result of another action, or, in other words, an action that is not caused by another action(s). The main condition that Danto sets down for basic actions is

> If a is an action performed by M, then either a is a basic action of M, or else it is the effect of a chain of causes the originating member of which is a basic action. (Danto, p.142)

From this condition, it seems clear that the reason for introducing basic actions is to stop a type of infinite regress that would result if we assumed that every action is caused by another action. Hence, given that some actions should be understood as parts in the context of larger actions, there must be some primitive level, or set of basic actions, that compose the larger action. There are two types in which an action can be said to be basic (Hornsby 1980). One says that an action is basic if it is *causally basic*, in the sense that there must be some action that initiates a causal chain of other actions, that itself was not caused by further action. This is not to say the initial action was uncaused, but just that it was not caused by another *action*. The second sense is *teleological basic* actions. This says that an action is basic if it is something we do "just like that". This can be explicated from two directions, either by saying that it is an action that doesn't require another action as a *means* to accomplish the action or by saying that it is an action that is the *end* of no action (Lavin, p.275). Another way of putting this is that teleological basicness is teleologically simple or 'unstructured'. In giving an account of what it is to make a mistake I will begin by focusing on how an action can be teleologically simple since what we want in an account of a mistake is an account of the conditions by which an intention is not fulfilled. In other words we need an account of what it means for an action to not achieve the end an agent

sets out to achieve. So, my question will be what condition can be given, if any, (in a way akin to Danto's causal condition) for teleological basic actions that allows us to understand how every mistake requires actions that are teleologically basic?

Consider a case in which I am walking from my house to a nearby park. Here is a case where I am attempting to fulfill the end of arriving at the park, by the means of walking. Now, what kinds of mistakes are possible here? For one I can make a wrong turn, or I can set off in the wrong direction. But can I make a mistake in the walking itself? The question itself seems odd. To be sure I could trip while walking, or twist my ankle, but even in these cases I must move my legs, I must move them intentionally and my movement of them must be characterized by the fact that I am the agent of their movements. There seems to be an essential reference to an agent in order to make sense of any action that involves the verb 'walking'.

What I will argue, however, is that in order for the action of walking to the park to fail, or for me to make a mistake in this (or any) action there must be some part of it that is not a mistake and this part is what characterizes the basic core of my agency of the action. But say I set off in the wrong direction or trip once I step out of the door, what agency have I exercised if the intended action itself seems to fail to even get off the ground? Well, the answer is that I have exercised my agency by performing an action *type* say of 'walking', or 'stepping out the door' even if the bodily movements that constitute this action can be characterized as a mistaken application or failed performance of this action type. Furthermore, the agency involved in these actions should be seen as an instance of agency *only if* it can be described as a type of action, and the particular manifestation of that action instantiates the type in some way. But if I intend an action type, this can only be possible if I know a description or some concept that characterizes the action I intend to perform. And this description or concept must be distinguishable from the manifestation of the action in the physical world. There is a distinction at work between the abstract and the concrete.

However, what we want is a notion of a basic action that draws a limit to the practical knowledge that the agent exercises in performing the action. As Hornsby has put it we want a

notion akin to the notion of know-how that provides us with a set of actions that the agent is able to do 'just like that' in an immediate sense (Hornsby 1980). She argues that

> Some things—at the end of these 'by'-chains, as it were—must be done without possession of knowledge of procedures. These are things an agent does 'directly'. They are basic things, in one sense of that action-theoretic notion. They are things which we are inclined to say the agent is able to simply do (Hornsby 2005, p.114)

One gets to the park *by* walking and one walks *by* moving one's leg but the ability to move one's leg in the way required for walking is itself a basic ability the agent has. Lavin has criticized this view on the grounds that it leaves a gap between knowing how to do something in *general* and knowing how to do something at a particular moment. He formulates the point thus:

> There is, then, a gap between Hornsby's thesis that there must be basic actions (concepts)—teleologically unstructured conceptions of types of things to do—and any answer to our question whether there must be basic actions (deeds)—teleologically unstructured doings of things (Lavin, p. 286)

The idea goes that the ability to move one's legs, or the ability to walk, when exercised in a particular context is no longer an act type but an act token. Part of Lavin's worry is that if we admit the existence of basic actions then the agent becomes *alienated* from his actions (Lavin, p.288). The agent becomes a mere observer of the actions that occur and that the causal work is not that of the agent but that of the causal nexus of nature which the agent simply observes when their actions come to completion. In carrying out an action that is basic, the agent seems to disappear, and what is left is a mysterious act that the agent does that initiates the causal order that follows in nature, without any subsequent intervention of the agent. There are two problems with this.

The first is that when an event occurs that corresponds with an action, at the very least, the action type does not thereby become a token action of that type. When the agent knows what they do, the agent does not have *exact* knowledge and cannot have exact knowledge. If they did then it would itself destroy the agency of the action, for any knowledge of the exact manifestation of the event that constitutes the action would require knowledge of the exact microstructure of the event. If this is not the case, then it is hard to see that the event is not

interpreted, even when it occurs, under the guise of an action type. In other words, the type doesn't disappear when the event of the action occurs, but on the contrary, the agent can't evaluate the event if they don't employ action types (or concepts).

Secondly, we couldn't make sense of a given action as a mistake unless we employed action types in the evaluation of the event that constitutes the action. The act token, if it exists, stands in a logically internal relation with the event that the action type describes, but this raises a problem. It should be pointed out that it seems correct to say that act types are obviously not identical with events. Nonetheless, there must be some notion that links act types with events and this notion must itself imply agency. A plausible candidate would be that act tokens are events and are in addition identical with events caused in 'the right way'. But this won't due either. An act token makes sense only as a token of a particular type. There is obviously something that is an event that can be described as 'walking to the park'. However, for any given event, there seems to be *multiple* act tokens that can be true of a single event. For one, if 'John is walking to the park' is a possible true description of the event, then this would also seem to imply that 'John is walking' is also true and if this is true then necessarily 'John is moving his legs' is true. If we take the action verb 'moving' in a transitive sense, then the fact that John's legs are moving implies that this is an action John is performing. The converse of this implication, though, need not necessarily hold. Even if it does hold, if we are thinking about the relation between an act type and its token, if there are two disguisable act types, there must be two distinguishable act tokens, regardless of the context in which these occur.

But then we can ask is the event that grounds the walking to the park itself identical with all the act tokens of the different act types that the event admits? For instance, is the event an act token of 'moving my legs' and 'walking to the park' and 'walking at a leisurely pace' and so on? Authors like Alvin Goldman would say 'yes' to this question and this is based on a *fine-grained* view of action individuation (Goldman 1971). If we take this line though and zero in on a particular event, we should then be able to identify different events that correspond to different act tokens. An act token of walking, and one of moving my legs, though doesn't seem to be able to be separated in the event in question. Walking is just a higher-order activity of moving my legs. This suggests that the act type 'walking', in this case, contains 'moving my legs' and that

there is no part of the event where the former occurs but the latter doesn't. If there is such a distinction to be made, it is one of *aspect*. What is happening is that we are picking out an aspect of a single event and relating the act tokens to the different aspects. But we need not do this and thereby unnecessarily multiply events. Instead, we can view the situation where one act type doesn't necessarily have an act token that can then be identified with a particular event, but rather we can have one act type that has another act type as its *occurrence*[1]. At some level though we must touch basis with the event. This is the *basic action*. We thus have the following picture: A single event can admit of different act types that are all co-extensional, in that they are all true of the same event. Higher-order act types, such as walking to the park, can take lower-order ones such as 'walking' as their type occurrences. But if the action is an event this type hierarchy must eventually make contact with the event. So, we can say, 'walking to the park' is satisfied if 'walking' is satisfied, with both being types. 'Walking' is satisfied if 'moving my legs' under conditions C is satisfied, and finally 'moving my legs under condition C' is made true directly by the event. Thus 'moving my legs under condition C' is a basic action. So, any act type is made true of an event if and only if three conditions are met

1) There is an ascending relation among the act types
2) There exists a basic action B that grounds the hierarchy
3) There is some event e that is identical to B

A full defense of this response will be the task of part three of this thesis, where I will make use of the 'levels of generality' in the description of an action to bring out the idea of nested action types in the evaluation of actions as mistakes. For now, what we need is simply an argument for the *existence* of basic actions that will provide grounds for assuring that every mistake requires an action that cannot be a mistake. The basis of such an argument lies in the idea that mistakes require agency while not necessarily being intentional. The key is to say that although the basic actions need not be intentional, if they are not intentional this does not imply that they are unintentional. Instead, we may say that an important feature of many basic actions is that they are 'pre-intentional' or 'sub-intentional' and form a conceptual basis from which we

[1] By using the word 'occurrence' instead of 'token' I mean to indicate that a when one type takes another as its occurrence, there is still a residue of conceptualization in the occurrence. An occurrence is thereby not necessarily concrete particular and so cannot be 'read off' directly from the event. We must *interpret* the occurrence as an appropriate conceptualization of the event

may say that an action is intentional. The description 'under conditions C' should be understood to pick out the pre-intentional states of the action.

The point is that by describing the action we can come to classify it as a mistake. There is nothing in the movements or 'pre-intentional' states of the agent that allows the concept of a 'mistake' to be applicable but what makes an event agential is that they are guided by reasons that generate intentional actions. The pre-intentional states are states that lie below the level of intentional action in the sense that they involve aspects of the action which we do not directly form intentions about, and for which we do not aim at, per se, in carrying out our actions. Though this is true, any event characterized as involving an agent is *co-extensional* with the events of my body that are characterized as intentional. The events of my body, say of my brain and nervous system, are themselves expressions of agency and would not have occurred as they did without my being the agent of them. However, this is not to say I as agent stand in a causal relation to the acts of my body, in the sense that my agency is a distinct object (or substance) that causes the objects or substance of my body to undergo the changes they in fact do. There is of course a question of in what sense the changes my body undergoes are instances of my agency. I will not attempt to answer this question fully in this paper, but rather to show the constraints that an account of mistakes has for any theory of agency.

To begin with we should resist as much as possible any talk of the changes the body itself undergoes as mistakes. If this were so, we could then ask what the changes in the body consisted of. Surely we would be given an answer that made reference to various physiological, chemical, and general physical facts and laws. And even if this doesn't provide us with a full explanation of *why* the body changed in the way it did, it seems likely that it explains *how* it changes. At this level, however, it is very implausible to say that we could find any room for the notion of a mistake or error. Furthermore, every physical change, when viewed as a physical event, cannot be an error unless we impute error or mistakes to the laws of physics themselves. Now consider the case of moving my legs while walking. This is a physical event. Because it is an event it is also a particular, dated, entity. The particularity of the event also implies that the full physical constitution of the objects (or substances) that are in the event, are necessary conditions for the event to be exactly the event it is. The particular event that is my walking would not exist unless

my muscles contract, unless the molecules in my muscles undergo the changes they do in fact undergo, and so on down to whatever is the fundamental physical level. This implies that the event itself cannot be a mistake. Yet, my legs would not have moved the way they did if I was not the agent of that action. The agency should therefore be viewed as a *higher-order property* of this event. Since the agency is a property of this event but no mistake can arise from this event itself, the agency alone cannot be the cause of any mistake. Rather, it is a necessary pre-condition for the possibility of a mistake. My agency simply is the changes my body undergoes, insofar as they are guided by higher-order faculties that allow us to evaluate if any particular change my body undergoes is in line with what I intend my body to do.

Walking is an expression of my agency, the movements of my limbs are something I do without really thinking about it and are things I 'just do' in the sense that they are a basic power I possess and use (causally). We can then say that the movements of my limbs are a means for the action I seek to accomplish, though they are not the ends of any other action. However, considered in itself as a series of physical movements, the movements of my legs are basic in the sense that qua event the relation of my agency to the mean/end structure of the walking is simple. That is to say, the means and the end of the physical act of walking is manifest at every instance and in every part of the event of 'walking'. The higher-level properties that are the expression of my agency extend *vertically* from the teleologically simple structure, but not *horizontally* along the complex teleological chain of means to end relations. In that sense the means and the end of each 'pre-intentional' act of walking are identical and hence it is teleologically basic.

To see why agency is preserved at the pre-intentional level, and why this is relevant, take the case where I suddenly stop walking because I trip and hurt myself to the point where I can't continue. Every movement of my legs up to this point would still be an expression of my agency, even the tripping itself would not have arisen if I have not been the agent of my actions. There is no disappearance of agency in any descriptions of this action. However, the intentional action I was performing (walking to the park) and the end I sought to achieve (arriving at the park) would be thwarted and must transform into their negations. There is thus a logical difference between the expressions of agency and the structure of intentional action, which if what I am saying is right, is reflected in the former being teleologically basic and the latter being teleologically

complex. One way of putting the idea is to say that the *will* stands in an internal relation to the event. On the other hand, the teleological structure of intentional action results from our thought and reasoning about our actions and results from abstracting complex combinations of action types from the events that ground the actions.

If every mistake is to be made sense of only where there is *something* the agent does which is intentional, and any action that admits a means-ends structure is to open to the possibility of failing to fulfill the end one intends, and failing to fulfill the end one intends is an instance of a mistake, then no action can have every one of its parts be mistakes. This would then imply that basic actions cannot be mistakes since they have no parts where one part would be intentional and the other unintentional. In other words, basic actions only have a single element, and that element cannot be a mistake since if it was then there would be no description of the action that the agent did that was intentional and hence the concept of a mistake would not be applicable. For basic actions no such gap exists, when I move my arm if I have the power of movement, I cannot fail to move it though I may fail to accomplish the goal I intend by moving my arm. The basic action of the movement is something I cannot fail to accomplish.

2.1 Sense of Agency

Michael Thompson and Douglas Lavin both point out a difficulty in identifying basic actions. Lavin offers a case where I am driving with a companion and we are both looking at a map, one of us then proceeds to show the other the route to our destination (Lavin 2010). In doing so I move my finger along the route we should take, say I am showing him how to get from Vienna to Budapest and in doing so my finger passes over the city of Győr. I might not even notice that my finger passed over Győr, and my intention was merely to move my finger from Vienna to Budapest. The movement from Vienna to Budapest can, however, be divided up into parts, one of them being the movement of my finger from Vienna to Győr. We can then iterate this process for every sub-movement of my finger between Vienna and Győr. At this point, it seems that if we carry this process far enough, we must end somewhere, yet at the same time, it is not clear where. Furthermore, none of the sub-processes need be intentional, in fact, at some point, it seems that the movement becomes so fine-grained that none of the individual

movements were themselves intentional. The problem for the basic action theorist is to find a set of minimum movements that are intentional and whose union yields the 'larger' intentional action of moving my finger from Vienna to Budapest. But it seems any minimum set of actions can only be arbitrarily chosen. As Lavin points out this is not a decisive argument against the existence of basic action, but rather it is meant to point toward the intuitive difficulty that the concept involves (Lavin, p.276). The proper response to this though is to grant the intuitive difficulty in finding any set of minimum movements that are intentional and to instead focus on how pre-intentional movements can become intentional.

Lavin suspects that the concept of basic action is meant to do fundamental conceptual work for both standard and non-standard theories of action (Lavin, p.278). He says both standard and non-standard theories "share an allegiance to the very general framework of the causal theory of action: that X did A intentionally is the arithmetic sum of what merely happens and something else" (Lavin, p.278). The work to be done by the 'something else' can be seen in the light of a distinction between 'mere happenings' and whatever counts as the conditions for action. The idea is that the mere happenings, which include my bodily movements, are not themselves intentional, and so for us to say that some event is not a mere happening but an action, there must be a basic action that brings agency into the picture. This is indeed the general picture I was working with in the previous section. The main problem Lavin sees with this is related to his claim we saw in the previous section that if this were true then we would become alienated from our actions (Lavin, p.288). The reason is that in order for an end to be accomplished, the subject of a basic action would have to delegate to nature the coming to fruition of the goal of his action. As I already indicated, part of Lavin's argument for this view resides in the idea that the subject (or agent) that carries out basic actions, their practical knowledge of how to carry out the action involves a form of action or an act-type.

Basic actions are, in addition, meant to halt a regress in procedural knowledge. For example, if someone knows how to build a house we can ask them 'How?' and they can proceed to give us instructions, 'first you lay the foundations, then you erect the frame, then… etc.'. If we then ask 'how do you lay the foundations?' they would respond, 'first you dig a hole…etc.' but if their knowledge is to be *productive,* then they must do something directly without having an

answer as to how they do it, for example, they must be able to operate a shovel and perform the bodily movements required to dig a hole, or the whole enterprise never gets off the ground (Lavin, p.285). Furthermore, every step in the process of building a house requires the builder to perform bodily movements that cannot be explained or justified with further knowledge but must be presupposed in order for them to carry out the various steps in the process.

Let us now focus on the claim that if the agent is the mere initiator of their basic actions, then they are alienated from their actions because the success or failure is 'up to nature'. I think we should view this claim in the context of the question 'How deep does the possibility of making a mistake reach into the structure of an action?'. There is a sense that accompanies the movements of our limbs, of our willing and intentional action in general. We feel and know what we are doing without having to check. We may call this our *sense of agency*. But the fact that we have a sense of agency that accompanies our intentional actions does not mean that we have a 'sense' of intentional action. The difference between intentional and unintentional action shows this. For an action to be intentional, the world must cooperate, and generally be as we suppose it to be when we act. The 'mere happenings' must not transform our intentional action into an unintentional one.

To make a geometric analogy, our sense of agency is point-like in that it is fully present at every point in time when I am performing a basic action. Thompson has argued, we need not view specific intentional actions as 'point-like' or simple structures in order to provide an explanation of action (Thompson, p.6). It can hardly be denied that many actions take time, for instance building a house, making an omelet, or even raising my arm. But there is a temptation to think that each sub-action of these larger actions must resolve itself into unanalyzable units that are the 'minimum' actions that are intentional. Thompson argues that instead, we should view intentional action thus:

> The nature of intentional action, or of the kind of being-subject-of-an-event that characterizes a rational agent and a person, resides in the peculiar "synthesis" that unites the various parts and phases of something like housebuilding, e.g., mixing mortar, laying bricks, hammering nails, etc (Thompson, p.6)

We are now in a position to see how pre-intentional basic actions can be synthesized and transformed into a structure that characterizes intentional (or unintentional) action. Our sense of agency forms a pre-conceptual link that synthesized the events that radiate from our basic actions *horizontally* and allows us to recognize them as my doings. The mere happenings in the world are thereby transformed and seen under the guise of the action verbs that indicate intentional and unintentional action. Thus, the trajectory of the ball becomes a 'kicking' the disappearance of the nail into the wood becomes a 'hammering' the rotation of the mortar becomes a 'mixing'. In a similar fashion, the laying of a foundation becomes a step or a means in the construction of a house.

Without a sense of agency that we all have in a primitive sense, no teleological structures would emerge in the world. We interpret our agency at every point in the conceptual chain that accompanies the events that admit a teleological structure. If we had no sense of agency, we would have no basis from which we could recognize an event as something I made happen, everything would be mere happenings. At the same time though, no sensation of ours can be a mistake, and nor can any mere happenings be a mistake. Only failures in the teleologically interpreted unfolding of events can give rise to mistakes. Thus, only if the world shows us something different from what we intended to occur can the notion of a mistake arise. Without basic unities that are teleologically unstructured and mistake-free though, we would need to look into nature itself to find the mistakes. But the mistakes don't go this deep, they arise not from any sensation, or physical event, but from our interpretation of the events in the context of what our goals and intentions are. This does not imply that we are alienated from our actions in any implausible sense, it merely shows that outside a context of an environment where events happen that we didn't intend or want to happen, there is no possibility of a mistake. This does not destroy our sense of agency and alienate us from our actions, rather it merely shows the limits of our powers

2.2 Basic Action and Knowledge

The reduction of the core of actions to basic actions, bears a resemblance to the reduction of all propositions to combinations of elementary propositions, or 'simple ideas' in the theory of knowledge. Danto's remarks that basic action can be compared to the 'givens' or non-inferential

propositions and that an analysis of action is only possible if we assume some ground (Danto, p.147). In a more positivist outlook, these basic elements from which all knowledge is derived would be associated with sense-datum, or sensory qualities or something which is phenomenally basic. Now, even if we are skeptical of the existence of any type of epistemological foundation, the interest of this analogy in relation to mistakes can be brought out by considering the argument from illusion. Thompson introduces this analogy in relation to a type of reductionism he attributes to 'sophisticated theories of action'. He presents the 'Argument From Illusion' with the hope of countering a common argument that propositions of the form 'I am doing A because I am doing B' or 'I want to do A because I am doing B' find their real content only after the 'doing/do' is reduced to some primitive psychological state such as 'wanting' 'intending' or 'trying' (Thompson, p.23-24).

Briefly, the argument from illusion says that in forming a perceptual judgment, say that this is a cup in front of me, there is room for error in that we think there is a cup but in fact there is not, but rather only an illusion of one. But, so the argument goes, we cannot be mistaken about the fact that we have a sense-datum of a cup. Likewise, in the case of action, we cannot fail to know what we 'want' or 'intend' since in either the case where the agent is successful or unsuccessful in their action, what is the same in both cases is not the 'doing of A' but the 'wanting' or 'intending of A'. The impossibility of a failure for us to know our wanting or intending (as opposed to our doing, where the world may 'be uncooperative') is a cognition requirement on the rationalization of our actions. Thompson responds to this by saying "If it shows this, then it also shows that all that ever really gets explained is itself wanting. For the same possibility of error and failure also afflicts my would-be doing of A" (Thompson, p.24). If it was the case that only the wanting gets explained, then it is doubtful that we can ever explain action itself. But if action can't be explained then the concept of wanting or intending has no real meaning, since only if there is the possibility of action is there the possibility of wanting or intending (Thompson, p.25).

I am sympathetic to this response, and what we should take from it is that no psychological reduction is necessary to explain action. But what this does not show is that there is no reduction to normative rules in evaluating actions. Furthermore, we can ask how could we explain actions

if we were not first able to evaluate them? In fact, Thompson seems to suggest as much in the symmetry of his table 'between possible explanantia and possible explananda' (Thompson, p.24). If we can establish a separation in the notion of an agent following a rule from the notion that the agent cognitively grasps the rule by which action is rationalized and evaluated, then we can make room for an aspect of action that is immune to the possibility of error. This will in turn require an idea of how we can follow a rule 'blindly' without recourse to a reduction of rule-following to a relation of the meaning or content of a psychological state.

Nonetheless, though a reduction to a psychological state is not possible, (or helpful) there is still higher-order faculties of the mind that are necessary for an evaluation of an action. There is then a question of what these higher-order faculties consist in and whether they are sufficient to evaluate whether an action was a 'correct' or 'incorrect' one. This links up with general questions on the norms that allow us to evaluate an action as either correct or incorrect in the context of following a rule. There is a tendency to ignore the question of how these norms are so much as possible and to instead simply suppose that the action itself is either incorrect or correct. This is like saying that the action demands the particular evaluation that we in fact give it. Maybe this is so, but then there is a question of what is the force behind this demand, why are we forced into one interpretation of the event we witness, rather than another?

For one, our talk of action types makes it evident that a linguistic capacity is essential for evaluation of what aspect of any event we witness is a mistake. There is of course a further question about the relation between language and thought, if one is more fundamental than the other or if there is any meaningful sense in which they can be separated. I will not attempt to answer this difficult question here. What does need to be assumed, however, to be able to give an account of how we come to recognize some action as a mistake is an ability to form judgments about which act-types characterize any phenomena we take to be the result of a mistake. In addition, we must be able to infer that one act type implies another if we are to recognize that something an agent does is both a mistake and a consequence of their agency. This inferential ability must itself be able to take the contents of our perceptions and linguistic judgments as elements that instantiate the inferential forms that govern our inferences. It should also be noted that individual acts of inferences are governed by rules of inference. But an inferential form, like

modus ponens cannot itself be governed by another rule. If this is the case a vicious regress would result, since we could ask what is the rule governing the rule of the rule? At some point in order for an act of inference to be possible, we must blindly accept and *use* some rule, in the sense that we are not further justified by another rule. This, as I will argue in chapter 4, is a kind of basic act, and though its consequences in a particular circumstance can be subject to criticism, error and mistakes, the form of the basic act itself cannot be since it has no need or possibility of justification.

3. Intentional Action and Descriptions

The motivation for introducing levels of generality is to provide a framework where an action can be described as both intentional and unintentional depending on how general the description is. For instance, consider a case where an agent makes a mistake by intending to press a button W, but instead mistakenly presses a button Z. If we describe the case of pressing the button Z as 'A pressed a button' where A represents the agent, we can then further add 'A intentionally pressed a button'. If we make a less general description of this action and instead say 'A pressed button Z' then we cannot say 'A intentionally pressed button Z' but rather that it was unintentional.

The idea is to extend this idea to all actions and to argue that

1) every mistake is an action
2) every action involves some description under which it is intentional
3) The level(s) of generality under which the action is rendered intentional are more basic or fundamental than those descriptions under which the action is rendered unintentional

The relation between mistaken actions and intentions can be seen by how intentions for the future explain actions. There need not be a causal relation in the sense that there is some causal link between a mental state of intending and the event that constitutes the action. What there does need to be, to make sense of mistakes, is a transfer of the intention into the action. What I shall argue is that this transfer can be explained logically as a form of inference that allows us to infer 'A intends to X' from 'A intentionally X's' and 'A intends to not X' from 'A unintentionally X's'. The inference where we infer 'intends' from 'intentionally' is

straightforward enough, though even it hides what I will call a hierarchy of action types. The case of inferring ‘not X’ from ‘unintentionally X’ is where the difficulty lies, and this brings out the necessity of viewing the situation through a nested series of action types. Furthermore, since it is implausible to say that ‘A intends to not X’ explicitly for every mistake the agent commits, this should count as an argument against the view that intentions are mental states, yet nonetheless the structure of the mental states that relate mirror the rules of evaluation we apply to actions.

What is the relation between intention and action? For one, how can I make a mistake in my calculation if I don’t intend to get it right? Or how can I be mistaken about who is president if I don’t intend my statement to relate to the facts? For instance, a joke that I am president is not a mistake, precisely because there is no intention to state a fact, though of course a joke could be used to implicate a fact, what it is not doing is stating any fact that corresponds with any proposition about me actually being president. What is distinctive about the type of intentions connected with these questions is that the failure to conform to the intention is always the consequence of the action of an agent. Donald Davidson makes this point thus “making a mistake must be doing something with the intention of achieving a result that is not forthcoming” (Davidson, p.46). On this account, for any description of an action that involves a mistake, there must be some description of the action that fails to fulfill the intention, yet that description must preserve the agency of the person performing the action.

If a mistake requires that an agent does something intentionally and that there is a description of the event that results from the agent’s intention, that is not intentional, is it then necessary that any description of an intentional action that leads to a mistake, be mutually exclusive with any description of, the same event, which describes the result of the action that is unintentional? One way to disentangle this issue is Anscombe’s distinction between

(1) ‘A intends to X’

where A is an agent and X is a description of the action an agent intends (typically represented through an action verb, like ‘run’ ‘shout’ ‘speak’ etc.) and its adverbial form

(2) ‘A intentionally X’s’

With (1) we have the schema of a description of a future action, and with (2) we have a schema of a description of an action that has already occurred (or is in the process of occurring). Furthermore, (1) tells us something about the agent, and some have construed it as a description of some 'inner act' of the agent, whereas (2) tells us something about the event that is caused by the action. We gain information about the latter through observation, whereas it seems the former can only be known by interrogating the agent. But what happens when (1) is an accurate description of an agent but (2) fails as a description of the resulting event?

To answer this question, consider the following example used by Davidson in his essay "Agency". The example is of someone who unintentionally pours out a cup of coffee when they intend to pour out a cup of tea. The person thinks his cup is full of tea and intends to pour out the tea (let us even suppose he wants to then fill his cup with coffee), but really the cup is full of coffee and so he ends up pouring out the coffee that he wants. So, (1) can be represented as 'A intends to pour out tea'. If we represent (2) in this case as 'A intentionally poured out the liquid in the cup' then the action under that description seems to be correctly describing the intentional part of the action, insofar as at least part of his original intention was to pour out what was in the cup, which he believes is tea. Let us further suppose that our agent believes tea is a type of liquid, then part of his intention was to pour out at least some type of liquid (other than coffee). However, if we describe the action as (2*) 'A intentionally poured out the coffee in the cup' then the corresponding description does not hold and is a description of a mistake. In other words (2) fails as a description of the event and so we must instead say

(2*) 'A unintentionally poured out the coffee in the cup'

to make the description true.

Yet (2*) is simply a more determinate description of the same event that (2) describes. The question then becomes, how, or why, can giving a more specific description of a single event eliminate the intentional aspect of that event? Must we suppose that there are really two different events one being the agent's intention to perform the action, viewed as a kind of 'inner act' and the other being the external action that results? For if the agent's intention involves an abstract type as an element, then it seems that the actual action cannot include this abstract type, or in other words since the event is not, by definition, abstract, if the intention has an abstract

element then they must be intrinsically different. What we have then is two descriptions of the same event, one being intentional which only references an abstract type of action, the other being unintentional and as such references an occurrence of the abstract type of action. This leads to the idea that intentions have a type/occurrence structure. When everything goes according to plan, the type/occurrence structure that the agent intends to carry out is reflected in a type/occurrence structure of the external action.

A mistake occurs, however, with an unintended realization of a type. In our example, the intended type of action is 'pouring out the liquid in the cup'. The intended realization was that the type of liquid (to be poured out) be 'tea'. The description that eliminates the intentional aspect of the event does so by instantiating an unintended occurrence of the intended type. In this case 'coffee' is a different unintended occurrence from that of 'tea', but both are of the type 'liquid'. It should be noted that if some intended type of action is not reflected in the external action, then the corresponding external action cannot be a mistake in relation to the original intention if that type of action can be inferred to be part of the intention. For instance, if no action type of 'pouring out the liquid in the cup' occurs, and instead, say, the person simply picks up the cup and puts it down again, the action is not a mistake because it has no relevant relation to the original intention. To return to the example of pouring out the coffee, since the external action, viewed as a physical event admits two incompatible descriptions namely

(3) 'A intentionally poured out the liquid in the cup'

and

(3*) 'A unintentionally poured out the coffee in the cup'

We might be tempted to ask 'Well which is it! Was the action intentional or unintentional? For there was only one external action and it can't be both.' It is certainly correct that the action itself cannot be both, but the problem really lies in supposing that mistakes are a class of actions. As the two incompatible descriptions suggest, mistakes are not a class of actions as such, but to attribute a mistake to something is a consequence of the logical form of the description of events that constitute intentional actions. The relevant logical form is the type/occurrence structure we use to describe the event that constitutes the action. My claim is that when we judge an action as

a mistake we must necessarily view the action in relation to an intention we infer the agent has, and when we represent this relation in our judgment we implicitly make use of a type/occurrence structure in comparing the original intention with the intentional action that the original intention causes. Therefore, we judge the action to be a mistake if and only if an unintended occurrence is realized in the external action that preserves the intended type of action.

Once we see that our descriptions of actions invariably make use of types of actions, we can then rank the types of actions by how general their descriptions are relative to the events that constitute the action. For instance, the type of action referred to in (3) is a more general description than that referred to in (3*). By ascending to higher levels of generality the actions acquire a description that renders them intentional, the question then is does this allow us to claim that higher-level descriptions are 'more basic' or more closely aligned with basic actions? The answer to this lies in our results about basic actions in section two. A basic action is invariably one that is simple and lacks the necessary complex structure for a mistake to be possible. Under the current formulation, there needs to be a structural distinction between a type and an occurrence of that type. If there is no difference between the two then there is no room for a mistake. Further, it should be noted that given a hierarchy of action types, we can evaluate whether a given type has an occurrence by taking the occurrence of the type as itself a type of a lower-level occurrence and asking if this type itself has an occurrence that is true of the event. This itself is not a problem as long as we hold that there is an internal relation between action and thought, and that every action is itself an instance of our conceptual capacities in operation (McDowell, pp.147). A full defense of this will require us to understand how the notion of following a rule (conceptually) is internal to the notion of action and its evaluation.

But to see how higher-level descriptions are more basic consider the following. When I reach for the cup to pour it out, we can describe this by saying 'A moved his body towards the cup' and that was presumably intentional. Even if I miss the cup when I go to reach for it, the act of moving my arm was still intentional. The precise path my arm takes towards the cup might have been misguided, yet when we normally reach or move our bodies, we scarcely have the power to intend one specific path rather than another. At best what we normally do is intend to move in some more or less general direction. The means-end structure of a goal-directed intentional action is not unique when it comes to the paths traced by bodies that ground the

action. Consider now what happens when we initiate our movement, since our power of movement is not precise enough to direct our movements perfectly along a unique path, we should rather understand the criterion of success for the action being a *family of paths* that all accomplish the goal (given certain constraints). The idea then is that there is a direct relation with the family of paths that constitute the success of the action and action types that make a description of an action intentional or unintentional. By reducing the level of generality of the type of action in the description, we render the family of paths the action can take (for it to be either successful or unsuccessful) narrower and narrower. Conversely, if we expand the scope of the description so that it becomes more general, the family of paths becomes wider and wider and we eventually reach a level where success is guaranteed if the action type perseveres. But since the possibility of a mistake requires that the action type perseveres in the relation between the intention and the action, if the action type is guaranteed and the occurrence becomes identical to the type (by widening the scope) then any description of the action must be intentional and therefore not a mistake. An action is a teleologically basic action if there is no gap between the means and the end of the action. When the type is general enough so its family of paths is identical with the initiation of the action then the occurrence equals the type and the means equals the end. In this case, the action is teleologically basic.

Returning to the claim I made at the beginning of this section that we can infer 'A intends to not X' from 'A unintentionally X's', this is made explicit if we take the agent to be able to recognize a particular action as unintentional only if they already possess a concept of the act type that truly describes the unintentional result of the action. If something totally unexpected occurs, and the agent (or others who witness the action) have no idea what to make of the events that unfold, then they don't recognize the action as a mistake. Suppose I pour out the coffee thinking it was tea. I think to myself that I made a mistake. But unbeknownst to me, there is a camera on the ceiling that is recording the cup. I later look at the videotape of this camera and realize that there was tea in the cup but that it transformed into coffee right before I picked it up. This would then not be a mistake. The main point is that to say 'A unintentionally X's' we must have an understanding of what it is to X and what types characterize X. In bizarre cases we should reformulate our descriptions and adjust the types of generalizations we would make. In short, we must presuppose a kind of regularity in the rules governing our concepts if our

descriptions of actions are to be able to ascribe mistakes to the agent. If such a regularity exists then we can infer 'A intends to not X' from 'A unintentionally Xs' without recourse to any explicit intention that A entertains.

4. Rule-Following and Normativity: Kripke's Account

In the previous section I argued for a criterion of what it is to be a mistake by saying that a mistake occurs if and only if an unintended occurrence is realized in the external action that preserves the intended type of action. This is a rule for evaluating whether an action is a mistake. There is then the question of what justifies this rule and how it can be said that agents follow this rule when they evaluate their own and others' actions. So, far it might seem that the notion of rule-following has little to do with that of making a mistake, and in some sense this is true. The notion of a mistake can be given in terms of action (either mental or physical) and the relation action has with an intention, specifically with whether or not an event that is associated with an action conforms to (or not) an intention-for-the-future. But this naturally raises a question as to how we can know if an intention has been satisfied.

If an action is to be a mistake, there must be a prior intention that can be attributed to the agent which itself links up with the world in such a way that an event in the world can be said to satisfy the content of the intention. The key issue regarding mistakes that the notion of rule-following impinges on is that of objectivity and factualism. Whether there is a fact to the matter whether a given action is a mistake is itself dependent on whether the is a fact about whether my action can be said to be an application of a rule. To bring this out consider the case introduced by Kripke, in his work on Wittgenstein's notion of following a rule. Kripke presents a meaning skeptic who challenges that there is any fact to what we mean by our use of a concept. He gives the example where a skeptic challenges us to say what we mean when we use the symbol '+' with the intention to mean the addition function. The main idea that the meaning skeptic challenges us to account for is that it is impossible to distinguish between two different functions, 'plus' and 'quus' merely from our past usage or any mental states (or occurrent thought) we had or have when we think we are using the symbol '+'. Plus is the ordinary

addition function that takes any two numbers and produces their sum. ‘Quus’ is, however, a function that produces the sums of plus up until a given number, for argument’s sake suppose this number can be arbitrarily large, for the sake of argument Kripke uses ‘125’. We can formalize the function thus for any two numbers x and y thus:

x +* y = x + y, if x and y < 57
x +* y = 128 otherwise

Where ‘+*’ is the quus function and ‘+’ is the normal addition function. Suppose we are then asked what x+y is where x and y are numbers we have never added in the past and are greater than any numbers we have added in the past.

The skeptics’ challenge is to ask how we can be sure that we didn’t mean ‘quus in the past instead of plus. Suppose someone hands be a piece of paper with ‘What is 58 + 68?’ printed on it and I write ‘128’. Suppose further I am such a skeptic that Kripke presents, and even though I intended my answer to accord with the normal addition function, the person who handed me the paper could say ‘Your answer is mistaken, the answer is ‘126’. Being the clever skeptic I am, I could respond by saying ‘There was no mistake here, even though I thought I was using the addition function, what I really was doing was using the quus function’. Furthermore, one could insist that there was no fact about me, or about my intention that determines which function I meant to use, and if there is no fact then there is nothing that would determine that I, in fact, made a mistake. But the skeptic does not stop with this example. The idea behind the skepticism can be applied globally to any word or concept that I use. In effect, the skeptic’s line of reasoning will open up the possibility that they could deny that mistakes ever occur. Every action that I perform, every time I make use of a word or concept to say or mean something, whatever ends up happening, I can always invent some abnormal rule that accounts for the action I perform and brings it in line with this rule so that under this new interpretation of my action I have not really make a mistake at all.

However, Wright and Boghossian have argued global skepticism about facts of meaning is unstable (Wright 1984 & Boghossian 1989). A non-factualist about meaning seems to be committed to a sort of projectivism that denies that any sentence in our language is truth-

conditional and instead views the situation as one where we project truth onto the utterances of our language. But as Wright has pointed out, this would itself seem to be a discovery that we make about our language, and it is hard to see how a statement equivalent to the conclusion of the skeptical argument is not itself a sort of projectivism (Wright 1984, p.770). But if this in the case then the skeptic himself has found sentences that have meaning, namely the conclusions of his argument, hence his global skepticism is unsound. Boghossian provides another reason along the lines that if global projectivism about meaning is to be coherent it must presuppose a theory of truth richer than a deflationary conception of truth. But if we don't have a deflationary conception of truth it is hard to see how there could fail to be a fact to the matter about judgments about a sentence's truth value (Boghossian 1989, p.526). There is a corresponding simple argument about mistakes. If there is no fact to the matter of whether a given action (or judgment) is a mistake then humans who made such claims in the past have been mistaken about the existence of mistakes. But it is a fact that humans have claimed that certain actions and judgments are mistakes. Therefore, this cannot be true since if we are mistaken about the existence of mistakes, this itself is an example of a mistake that must exist that we have just now discovered, hence the claim that mistakes don't exist is false.

We are now left with a question of how to respond to the skeptic. The main view that is attacked by Kripke's argument which the possibility of the skeptic is meant to cast doubt on is the dispositional view of meaning. A concept cannot be given a dispositional account, where the meaning of a certain concept is explained by what we are disposed to answer in a given circumstance. He also interprets Wittgenstein as accepting the skeptic's claim that there is no fact about the person (no mental state) that determines that they meant plus instead of quus (Krike, p.66). The problem is then how can we avoid the conclusion that meaning, and hence all language is impossible? There are several arguments he gives in support of this, one of them involves the idea that the dispositional view cannot account for mistakes, since we might have an innate disposition to make mistakes. He says, "But a disposition to make a mistake is simply a disposition to give an answer other than the one that accords with the function I meant." (Kripke, p.30). We need some other criteria to evaluate our dispositions, that is itself not another disposition since if it is then the skeptic's challenge can be used against these dispositions, creating an infinite regress.

Instead Kripke argued that the relation between meaning and intention on the one hand, and future action on the other, is normative. The normativity of following a rule is a consequence of the normativity of meaning and content in our application of a word or a concept. This is contrasted with any descriptive relation between a subject and their use of a concept that would determine whether the use is a correct or incorrect one. If the dispositional view was true, then we could give a descriptive account of how I would respond when presented with an addition problem. This in turn would explain what I do in fact mean. Kripke puts the point thus:

> 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'. (Kripke, p.37)

Given the problems that the dispositionalist has in accounting for mistakes, even if a theory of meaning could be rescued for their view, I think that at the very least, they must grant some notion of normativity must come into any explanation of what it is to make a mistake. For any descriptive account that made reference only to a set of dispositions couldn't establish anything more than that some of our dispositions were abnormal the very notion of a mistake itself presuppose some type of normative assessment of the events that realize our dispositions. As we have seen, physical events and processes themselves can't be mistakes. So even if what we mean when we use a word is ultimately a set of dispositions grounded in the physical processes that make up our bodies, mistakes only arise when we interpret our dispositions under the form of act-types that provide the basis from which evaluation is possible. We thus must reject the idea that the relation between meaning and intention to future action can be given in purely descriptive terms and that it lacks any normativity.

4.1 Escaping The Community View and Platonism: The Modus Ponens Model

The source of Kripke's Skeptic in Wittgenstein comes from Kripke's interpretation of the famous passage on the so-called 'rule-following paradox'. Wittgenstein writes:

> 201. This was our paradox: no course of action could be determined by a rule, because every course of action can be brought into accord with the rule. The answer was: if every course of action can be brought into accord with the rule, then it can also be brought into conflict with it.

And so there would be neither accord nor conflict here. That there is a misunderstanding here is shown by the mere fact that in this chain of reasoning we place one interpretation behind another, as if each one contented us at least for a moment, until we thought of yet another lying behind it. For what we thereby show is that there is a way of grasping a rule which is not an interpretation, but which, from case to case of application, is exhibited in what we call "following the rule" and "going against it". That's why there is an inclination to say: every action according to a rule is an interpretation. But one should speak of interpretation only when one expression of a rule is substituted for another. (Wittgenstein, PI)

This is immediately followed with the claim that it is impossible to follow a rule 'privately' (Wittgenstein, 202 PI). Following a rule, Wittgenstein claims is a practice, it is a custom that has a sense only in the context of rule-following institutions (Wittgenstein, 198 PI). If this were not the case, then as Wittgenstein notes, any case where I took myself to be following a rule, would be sufficient for me to be actually following it (Wittgenstein, 202 PI). What are we to make of this? Wittgenstein does clearly say that there is a paradox about an action being in accord (or not) with a rule, but in the next sentence he says that 'there is a misunderstanding here'. The problem is that we suppose that every rule must be interpreted in some way in order for us to follow it. So, this then raises the question of what kind of resolution we are to make of this paradox.

Kripke interprets Wittgenstein as offering a skeptical solution to the skeptical problem that following a rule seems to require (Kripke, p.68). The skeptical problem is that every action, can be made under a suitable interpretation to accord with the rule and if this is so, then it would be impossible to not follow a rule. But the notion of a rule essentially involves a possibility that it is not followed, hence making the notion of following a rule impossible. This in turn makes it impossible to not follow a rule so any mistaken application of a rule is also deemed impossible. The skeptical solution, as Kripke sees it, is to deny that there really is any fact to the matter to what we mean but to instead replace the truth conditions of our statements with the notion of assertability conditions (Kripke, p.74). Assertability conditions are themselves given warrant by the community the speaker (or agent) finds themselves in. There is nothing over and above the community that gives a particular word or concept its meaning.

There are two issues with this. One has to do with whether this is the right reading of Wittgenstein. I will not be that concerned with that in what follows and will only briefly sketch

some of these issues with this. I suspect, however, that Wittgenstein's position is more nuanced than Kripke makes it out to be. In short, there is evidence that Wittgenstein has both a deep *interest* in the skeptical problem, and that he does not offer a skeptical solution to it. In other words, it is highly contentious whether he fights skepticism with skepticism. As Baker and Hacker have pointed out there is a distinction between the claim that *empirically* language and other rule-based systems (such as games like chess) always involve a community to establish standards of correct and incorrect usage and the claim that *logically* the possibility of following a rule involves a community (Baker, G. P. & Hacker, P., p.151). Regardless of what Wittgenstein's settled view on this is (maybe he oscillated between these or even rejected the logical-empirical distinction in this case) Kripke's Community based solution has interesting consequences in its own right.

This brings me to the second issue with Kripke's idea, namely, if something like the community view is right, then what consequences does this lead to in regards to our standards of objectivity and truth? If the meaning of our words and concepts is nothing more than what is accepted by the community, then the community itself cannot be wrong. There would be no external authority, no criterion of objectivity and truth that the community could appeal to, since the meaning of the very concept of objectivity and truth, or any time we used these words as they are accepted by modern users of English, would be whatever the community took them to be. This surely stands in the face of what we as a community currently take these concepts to mean! Modern Society is so permeated with the scientific outlook, and the notions of truth and objectivity that the sciences seems to imply, it is a hard pill to swallow to say that at bottom all it all merely amounts to what we take them to be. There is an air of paradox that resembles that which we saw about the skeptic who denies the existence of mistakes. For if we are wrong to think that truth is something beyond community assent, then it is true that truth means whatever the community thinks it means. But the community doesn't think it means whatever the community takes it to mean. Therefore, we have found a truth that is not governed by the rules of the concept 'truth' that the community assents to.

Boghossian has pointed out a related problem with any pure communitarian view. If I as an individual have a systematic disposition to make mistakes, it is equally likely, given that I am

not abnormal, that the community itself would also have a systematic disposition to make mistakes (Boghossian 1989, p.536). This raises a difficulty since the conditions that the individual is subject to can also be applied to the community itself and so any problems with the dispositional view would be inherited by the communitarian view. Consider the case where, in poor lighting conditions I have the systematic tendency to mistake an object in my visual field for a horse, which is in fact a cow. The problem is then that the community cannot regard any of its systematic dispositions as mistakes. Boghossian writes

"He must insist, then, firm conviction to the contrary notwithstanding, that 'horse' means not *horse* but, rather, *horse or cow*. The problem, of course, is general." (Boghossian 1989, p.536)

So, it seems that by appealing to the community to ground the possibility of mistakes, we are left in the same position as we were in when the individual was confronted by the skeptic. We thus need to find a way to rejoin to the skeptic without recourse to a pure communitarian view.

What should we turn to? A first reaction might be some form of Platonism about rules. Rules are abstract objects that exist independently of our practices and usage of them, and the conditions for correct or incorrect applications of a rule are determined by objective conditions that exist outside of time and space. The reason why it is wrong to answer '128' to the question 'What is 58 + 68' is that there is a fact about the structure of reality that makes this answer a mistake. Maybe this is true. But as I have indicated above, we cannot locate mistakes in physical events or objects themselves. Where are we then to locate the source of the possibility of making a mistake? In relation our actions have to abstract propositions and their truth conditions? The problems that beset any kind of Platonism would beset us here as well, for we can then ask how are actions that are situated in the nature to reach out to stand in a meaningful relation of the kind Platonism requires?

Avoiding both a community view and Platonism about rules then seems desirable. Wright offers such a middle ground within the context of with what he calls the 'Modus Ponens Model' of rule-following (Wright 1989). To be precise it is the case where the Modus Ponens model breaks down that the middle ground between Platonism and the Community view opens up (Wright 2007, p.498). The Modus Ponens Model gives us a general model for what Wright calls

non-basic rule-following. It is non-basic because it involves explicit interpretation of the rule in a particular circumstance. In other words, we must recognize that the circumstances are such that the rule is *appropriate and applicable*. Wright gives the following example of the Modus Ponens Model:

Rule: If neither King nor one of its Rooks has moved in the course of the game so far, and if the squares between them are unoccupied, and if neither the King nor any of those squares is in check to an opposing piece, then one may Castle

Premise: In this game neither my King nor this Rook have yet been moved, the squares between them are unoccupied, and . . .

Conclusion: I may castle now. (Wright 2007, p.490)

Here what we have is a general rule about castling in chess, the agent then needs to recognize that the circumstances that he is currently in allow the rule to be applied. This is the premise in the model. He then acts on this recognition. But the problem is that this recognition is always an interpretation of the rule. We then would seem to need another rule that would govern whether this interpretation is correct. And then we are off into a vicious regress. At some point, we need to simply apply the rule. This is the basic case. We can compare this with Wittgenstein's idea that we must at some point obey the rule 'blindly'. He says,

217. "All the steps are really already taken" means: I no longer have any choice. The rule, once stamped with a particular meaning, traces the lines along which it is to be followed through the whole of space. – But if something of this sort really were the case, how would it help? No; my description only made sense if it was to be understood symbolically. – I should have said: This is how it strikes me.
When I obey a rule, I do not choose.
I obey the rule blindly
(Wittgenstein PI)

If we take this in combination with remark 201 in PI, and apply our results about action types we can say the following. By replacing one form of the action types in a description of intentional action with another form, we are providing a new interpretation of the action. In

doing this we change the criterion by which the action can be evaluated as a mistake. This process must, however, come to an end. The end of this process yields descriptions of actions that cannot be mistakes and ground the description that gives conditions for which aspects of the action were mistakes. In an analogous way, the ability to follow a rule on a particular occasion must be grounded in a general capacity to apply the rule. This general capacity *is* my possession of the rule governing the concept and is not rationally grounded in anything but how it strikes me to apply it. When I apply it in this way I apply the rule blindly. Though I might be mistaken in a particular application of applying the rule, the general capacity to apply the rule as it strikes me is a general enough description to be identified with a basic ability I have, and so is a kind of basic action. Right above this level of generality, where I describe the circumstances that I use the rule (think of a description of the castling case) the possibility of a mistake arises. This is equally where normative assessment arises.

We avoid the pitfalls of the community view since the possibility of a mistake is not something that is governed by what others take to be the description of the situation, but rather emerges out of a determination of the general capacity for an individual to apply a rule *and* to *interpret* that determination within the context of a system of practices and customs. We avoid Platonism because we don't need to give an account of how an agent gains access to some abstract rule, rather it emerges out of the blind application of the basic cognitive powers they inherently possess as rational agents. Finally, because the blind application grounds itself, there becomes a fact of whether an agent has applied a rule or not. Furthermore, his fact itself provides grounds for the normative assessment and so we, therefore, have at least a minimum requirement of objectivity.

4.2 Finding Bedrock: Basic action and Rule-Following

Normative assessment operates at the level right above basic action, which is mistake-free. There is a level at which justification for the action ends, this can't be regarded as open to a possible mistake since justification is internally linked with criticism of the action (as a mistake). Compare this with Wittgenstein's Bedrock metaphor of following a rule.

217. “How am I able to follow a rule?” If this is not a question about causes, then it is about the justification for my acting in this way in complying with the rule. Once I have exhausted the justifications, I have reached bedrock, and my spade is turned. Then I am inclined to say: “This is simply what I do.” (Remember that we sometimes demand explanations for the sake not of their content, but of their form. Our requirement is an architectural one; the explanation a kind of sham corbel that supports nothing.) (Wittgenstein, PI)

So if this is ‘simply what I do’ does that mean that I cannot evaluate what I do? That at some point in order for me follow a rule, I must also simply act in a way that I cannot justify any further. Though Wittgenstein might balk at the suggestion, there is an analogy here between our account of basic actions and the idea that eventually actions cannot be justified. Basic actions are simply things we ‘do just like that’ and though they may lead to mistakes, there must be some point at which I act as an agent that cannot be further justified by answering any question ‘Why?’. If no such answer is possible, I cannot have unjustified grounds for acting, since in an important sense I have no grounds for acting as I do. This is just simply the way I must act if I am to do what I intend to do at all.

And here the key point arises, if I am to make a mistake in a non-basic application of a rule, along the lines the modus ponens model indicates, I must first have a set of basic competencies, a set of basic concepts whose criterion of competency simply *is* the fact the I possess the concept. Take the example of someone seeing a red object and being asked what color it is, *if* I have the concept of red I cannot fail to recognize that this object in front of me is red. I cannot make a mistake, having the concept of red simply is the basic power to recognize red objects. If I am then told this is not a red object, well I must question much more than the particular situation at hand. I must either think the other person doesn’t have the concept, or that I simply don’t have it. Maybe I use the wrong word, but this is akin to me characterizing an action under an inappropriate action type, it is not akin to me making a mistake by exercising a basic action. In fact, as I have argued this is not possible. And if I am to take the doubt that what I see is red seriously, or at all, I must then doubt in such a way that in doubting I use concepts that I do have competency with. I cannot doubt everything and anything, since then I could doubt that I even have the power of doubting, which is a contradiction. I must eventually arrive at bedrock.

To conclude, an instance of basic rule-following is an instance of acting at the level of bedrock. The ability to follow a rule blindly is akin to a creative ability of the mind to acquire new concepts and to act with these concepts simply by the fact that it has acquired them. We act 'as it strikes us' and critical evaluation emerges out of this primitive creative act. The capacity to follow a rule blindly is like the basic ability I have to move my arm. The possibility of making a mistake arises only in the context of non-basic applications of a rule, in the same way that the possibility of making a mistake is only possible in the context of non-basic actions where external circumstances can thwart the intended occurrences of the types of actions we seek to accomplish. In short, in order to make a mistake, there must always exist an aspect of our action that is not a mistake. This is the case for both physical actions and mental actions.

Bibliography

Anscombe, G.E.M. 1963. Intention. 2nd ed. Basil Blackwell.

Baker, G. P. & Hacker, P. 2005. *Wittgenstein: Rules, Grammar and Necessity*, Volume 2 of an analytical commentary on the *Philosophical Investigations* Oxford: Blackwell Publishing

Boghossian, Paul, 1989. "The Rule-Following Considerations," *Mind*, Vol.98, 507–549

Danto, Arthur C. 1965. "Basic Actions." *American Philosophical Quarterly*, (2): 141–48.

Davidson, Donald. 1971. "Agency." Reprinted in his *Essays on Actions and Events*, 43–61. Oxford University Press, 1980.

Goldman, Alvin. 1971. "The Individuation of Action" The Journal of Philosophy , Vol. 68, No. 21, Sixty-Eighth Annual Meeting of the American Philosophical Association Eastern Division, 761-774

Hornsby, Jennifer. 1980. *Actions*. Routledge and Kegan Paul.

———. 2005. "Semantic Knowledge and Practical Knowledge." Proceedings of the Aristotelian Society, Supplementary Volumes 79: 107–45

Kripke, Saul, 1982. *Wittgenstein on Rules and Private Language*, Cambridge MA: Harvard University Press

Lavin, Douglas. 2013. "Must There Be Basic Action?" *Noûs* 47 (2): 273–301

McDowell, John. 2015. "Acting as One Intends" printed in *Philosophy of Action: An Anthology*. Edited by Jonathan Dancy and Constantine Sandis

Thompson, Michael. 2008. "Naïve Action Theory," chapter 2 of *Life and Action: Elementary Structures of Practice and Practical Thought* . Harvard University Press.

Wittgenstein, Ludwig, 1953, *Philosophical Investigations*, G. E. M. Anscombe (trans.), Oxford: Blackwell. Revised fourth edition with P.M.S. Hacker and Joachim Schulte (eds.), Chichester: Wiley-Blackwell, 2009.

Wright, Crispin, 1984. "Kripke's account of the Argument against Private Language", *Journal of Philosophy* LXXI, 12, 759-778

———.1989. "Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics" in *Reflections on Chomsky*, ed. Alexander George (Blackwell), 233-64.

———. 2007. "Rule-following without Reasons: Wittgenstein's Quietism and the Constitutive Question" in *Wittgenstein and Reason*, ed. John Preston; *Ratio* vol. XX no. 4

Abbreviations

PI *Philosophical Investigations.*