

# **The Agent in the Deed: Towards an Expressivist Theory of Action**

**By Sara Jaramillo**

Respectfully submitted in partial fulfillment  
of the requirements for receiving  
the Degree of Bachelor of Arts  
in Philosophy at Haverford College

**Joel Yurdin, first reader**  
**Kathleen Wright, second reader**

April 24, 2015

## **Acknowledgements:**

I would like to thank my advisors for their help in shaping my project and their support throughout the process. This undertaking would not have been possible without Joel Yurdin's passion and encouragement from the beginning,

and his insightful commentary and suggestions throughout. Kathleen Wright was invaluable in pushing me to hold on to my original ideas and follow them to the end. Any errors in my paper are not due to their influence.

I would also like to thank Danielle Macbeth, advisor throughout four years and professor of my first philosophy class at Haverford. I so appreciate her constant support and enthusiasm for my intellectual journey. I also owe many thanks to Jerry Miller for challenging me to new ways of thinking and writing philosophically, and whose class provided the reading that inspired this paper.

Finally, I would like to thank my parents for providing love, intellectual curiosity, and a scientific perspective when discussing philosophy. Thank you for supporting me always.

**Abstract:**

My aim in this paper will be to examine the relation between an agent's intentions and her actions. I will look at two different ways of approaching the relation: the causal framework and the expressivist framework, and I will attempt to put forth my own critiques of each. The causal theorists I refer to in my paper approach the question from an analytic standpoint, discussing the mechanisms of the act and attempting to understand human action by breaking it down into its simplest and most basic form. I will claim that in doing so, they lose sight of the big picture and

the importance of understanding action phenomenologically. The expressivist theorists oppose the causalists from the beginning by assuming an essential difference between human action and mere events, thus seeking to characterize that difference rather than explain it. Expressivists provide what I consider to be a more phenomenologically accurate view of action, producing the beginnings of an artful perspective on human action. I will attempt to defend and enrich the expressivist understanding of human action, while also illuminating the places where the theory still needs to be filled out.

## **Introduction**

My essay falls under philosophy of action, a category that deals broadly with issues of agency, free will, and determinism. The question I will mainly be concerned with in this paper is that of how an agent's intentions are related to her deeds. Inspired by Robert Pippin's essay "Agent and Deed in Nietzsche's *Genealogy of Morals*" I will explore two opposing approaches to this question, the causalist and the expressivist. I will lay out a preliminary picture of each and comment on the ways they engage with each other. I will then show that although

the causalists and expressivists may seem to be in conversation with each other, directing themselves towards the same question, I find that the positioning of each framework points them in different directions from the outset. Because they disagree on whether there is an essential difference between action and nonaction, the expressivist and causalist frameworks, despite their respective merits, cannot be brought together to form a hybrid view. To that end, I will pursue the expressivist framework, attempting to show that it gets right what the causalist model gets wrong. My final aim in this paper will be to affirm and enrich the expressivist notion of human action and thereby arrive at a positive view of human agency. First, let me begin with a short introduction to some of the central themes and questions of the philosophy of action.

### **A brief overview of action theory**

Action theory sets out as one of its central questions the following: What distinguishes an action that an agent performs from something that merely happens to her? What is that factor that allows us to distinguish between my diving into the pool (an action) from my friend pushing me into the pool (an accident)? The question is part of a network of other related questions about the relation between the agent's mind and her material body, the way in which actions are defined socially, and the ways in which our inner mental states interact with the observable world.

Before crafting a theory that will help us distinguish between actions and events, it is important to identify at least some of the constitutive elements that

comprise our general understanding of what makes up an action. We might imagine desires, beliefs, knowledge, intentions, bodily motions, social consequences, and more. What adds some complexity to the issue is the difficulty of locating our exact meaning when we refer to "action." Some locate the action in the agent's initial impetus to act, and see the resultant bodily movement as a consequence of the action. Others claim that the action simply *is* the bodily movement. The difficulty in locating the moment of the *act* is that there is disagreement over which are the relevant factors leading to the action. This example hopefully illuminates the uncertainties surrounding the discussion of how to define action and distinguish it from mere occurrences.

#### **A preliminary look at both accounts**

The causalist view that I will present takes as its central question "how are we to know an action from a non-action?" or in other words, "what is an action?" This question arises from the fact that from a third-person observer's point of view, a person's action—here analogous with physical movement—may be indistinguishable from other cases of bodily movement in which the person has no or little control. In seeking to locate and characterize the distinction between these externally indistinguishable events, action theorists in the analytic tradition break action down into its constituent parts and attempt to determine how the act itself comes about. Thus the focus of the search turns to what *causes* an action, and how its cause distinguishes it from nonaction. Due to their denial of human action as inherently different from nonaction, the causalists already presuppose a difference that lies in the causes leading up to an action. So we can see how the framework

that results from the causalist project is limited in what it can reveal about human action.

In contrast to the causal theory, some thinkers take the opposite approach towards understanding action. Rather than zooming in on the inner mechanism of action to see how it is different from nonaction, they take for granted that action is inherently different from nonaction. As Charles Taylor puts it, "action [is] qualitatively different from nonaction, in that actions are what we might call intrinsically directed...inhabited by the purposes that direct them, so that action and purpose are ontologically inseparable" (Taylor 23). This "qualitative" view of action, what Robert Pippin refers to as the expressivist approach, looks at the relation between the agent and her actions from a broader perspective than the causalists. It assumes "an ontological difference between actions and nonactions which cannot be captured in terms of their causal origin" (Laitinen and Sandis 7). That ontological difference is recognized by virtue of the expressivists' phenomenological approach to understanding action. In acknowledging that it feels different to us when we purposively act, as opposed to cases in which we are merely acted on, the expressivists take that difference in our experience to constitute a distinction inherent to the nature of action.

One thing to note before I proceed: I am attempting to put the causalist and expressivist accounts in conversation with one another despite the fact that the expressivist theory lacks the cohesion of the causalist theory. The causal approach to action is well known, popular, and contains within it various distinct causal theories of action; in other words it has great breadth and depth of understanding. The expressivist approach to action, on the other hand, is neither as well-known nor widely-accepted, at least partly because, in my mind, it suffers from a dearth

of thinkers attempting to put forth their own understanding of expressivism. In essence, the expressivist approach has yet to be fully laid out, understood, and developed within the philosophy of action. Therefore my bringing the two approaches together in this essay might be awkward, and at times it will be necessary for me to speak on behalf of the expressivists where I do not have the authority. The point here is not to develop a more full or nuanced view of expressivism, but to really weigh the two theories against each other and ultimately put forth expressivism as the theory with the most potential for coming to understand human action.

### **The causal theory of action**

As its name suggests, the causalist approach posits a causal relation between an agent's intentions and actions. As Markus Schlosser writes, "a causal theory of the nature of actions...says, basically, that actions are events with a certain causal history (they are actions *in virtue of* this history)" (Schlosser 13). In other words, it identifies action according to what precedes it. If the conditions preceding an event are appropriate and the correct causal chain is in place, then that event counts as an action. Michael Costa uses the term *causal ancestry* to refer to the right kinds of antecedents: "The causal ancestry must be traced back to some kind of intentional state (or combination of intentional states) in the agent, such intentional state (or states) must have the right kind of content, and it must cause the bodily movement (or other effect) in the 'right' way" (Costa 831). This wordy definition boils down to the idea that actions are caused by the agent's intention. Although there are different opinions on the definition of intention, for time's

sake, in this essay I define intention for action as a desire (or some other positive attitude) plus a belief that the desire will be fulfilled by the specific action.

Donald Davidson uses the term "primary reason" to refer to this combination of an agent's pro-attitude towards a certain kind of action plus a belief that her action is of this kind (Davidson 4). This description of action makes intuitive sense. Let's take a relatively simple action: turning on the light. I have the desire for my bedroom to be more brightly lit than it currently is, and the related belief that by flipping the switch on the wall the light bulb will turn on and light up the room. The combination of my desire and belief comprise my intention to brighten the room by turning on the light. Therefore, my intention is the cause of my bodily action. However, this is the simplest and most direct case, and there are many counterexamples of deviant causal chains that complicate the account. For example, I may have the right desire and belief that lead to my intending to switch on the light, but instead of walking over to the light switch and flipping it on, at the last second I trip over the rug and fall into the wall, consequently flipping the light switch. Although I had an intention that was technically fulfilled by my bodily movements, it would be obviously incorrect to claim that my tripping into the light switch was intentional. This and other complicated cases of deviant causal chains make it necessary to retain some aspect of vagueness in the definition, formulated more or less as: if an event occurs *in the right situational context*, then it will count as an action. So the causal theory defines actions as events correctly situated within the causal chain.

### **Davidson's causal account**

I now want to turn to Donald Davidson to fill out the preliminary framework I've laid out for causal theory. In his essay "Actions, Reasons, and Causes" Davidson wants to address the relation between an action and the agent's reason for acting or her intention. He wants to defend what he calls the commonsense view that "rationalization is a species of causal explanation" (Davidson 3). Some philosophers deny this possibility on the grounds that causal explanations are necessarily lawlike, whereas reasons-explanations, which deal with psychological states, are not strictly lawlike. It must follow that a rationalization of an action, insofar as it is not lawlike, cannot be considered a kind of causal explanation (Mele 4). Davidson wants to resist this picture by arguing that to give a reason why the actor did what she did, i.e. to rationalize or justify the action, is to give a certain kind of causal explanation, although it doesn't explicitly lay bare the causal mechanism at work. He wants to convince us of the explanatory power of rationalization, which makes action meaningful by contextualizing it rather than by explicitly laying out the teleology of the cause and effect.

Davidson begins his essay by examining what we mean when we speak of an actor's *reason* for acting. He distinguishes two elements: what he calls a *pro-attitude* towards actions of a certain kind; and a corresponding *belief* that his action is of that certain kind (Davidson 4). Giving a reason for acting often consists in giving either the pro-attitude or the related belief; Davidson calls the pair the actor's *primary reason*. For him, understanding the formation of a primary reason is necessary and sufficient for understanding how a reason can explain an action (Davidson 4). For example, if I flip the switch for the space heater in my room, my primary reason for doing so consists in my desire to warm up, and my belief that the action of flipping the switch on the space heater will lead to me warming up. It

also becomes obvious from this example that in everyday communication about actions, one needn't give *both* the attitude *and* the belief in order to explain one's action. Indeed, to provide both would appear redundant and be met with confusion. If someone asked me why I switched the space heater on, it would suffice to reply, "because I'm cold and want to warm up" without including "and I believe that switching on the space heater will cause me to warm up." We are able to understand each other's reasons for acting without needing to have all steps in the causal chain explicitly spelled out because we live in a social world of familiar actions. If someone had never seen a space heater before, my justification of my action as "I'm cold" will not be sufficient for an explanation. This gets into the importance of the fact that human actions are enmeshed within an always already social world.

It is imperative for Davidson that he bring our attention to the way in which actions occur within the world of other people who are being shaped by the same norms, values, and realities—and who live under the same causal laws of the physical world. It is because of this shared situational context that we are able to intuitively make sense of a given action within the context of the agent's history. What's more, knowledge of how the agent has behaved in the past will provide even more context within which to understand their reason for acting. "In the light of a primary reason, an action is revealed as coherent with certain traits, long- or short-termed, characteristic or not, of the agent" (Davidson 8). I am much more likely to accurately identify and characterize an agent's action if I know their unique personal history, tendencies, personality, etc. Observing the action of a stranger, on the other hand, will prove a much more difficult task for me, lacking the context through which to judge their action.

This also highlights the importance of the agent's own conception of their action. My interpretation of my brother's action will be colored and enriched by learning his *own* understanding of his action. "To learn, through learning the reason, that the agent conceived his action as a lie, a repayment of a debt, an insult [etc.] is to grasp the point of the action in its setting of rules, practices, conventions, and expectations" (10). This is to understand the network of forces that were acting on him both in that moment and over time—social conventions, moral values, personal wishes, habituation, etc. "The most primitive explanation of an event gives its cause; more elaborate explanations may tell more of the story, or defend the singular causal claim by producing a relevant law or by giving reasons for believing such exists. But it is an error to think no explanation has been given until a law has been produced." (17) It seems that we are able to grasp an action's cause by learning the reason, without needing to have the causal mechanism explicitly spelled out. In other words, the reason *contains within it* the explanation.

This is the only way that reasons *can* take on an explanatory role for us in everyday life: because we as observers of others' actions are able to grasp and understand how the reason is part of a broader causal chain. And we needn't always be right about the explanation for others' actions. Indeed, just as the individual does not always have immediate and infallible access to all her beliefs and pro-attitudes, neither do outside observers have immediate and infallible access. Rather, the individual has the unique and privileged perspective of being inside her own mind, whereas the outside observer has the unique perspective that comes with distance from the agent's mind. An outside observer may have a clearer perspective of the agent's actions than the agent herself, due to the agent's

obliviousness to the beliefs and attitudes driving her actions. This brings up the related important point that a great deal of actions are social, and thus only count *as* actions insofar as they are recognized by *others* as actions. For example, I may mean to raise my hand in class to indicate that I wish to speak, but I think that it is enough to raise only one finger and leave my arm lying on the table. In this case, although I *think* I am performing the action of raising my hand, because the professor and my classmates do not recognize my action as the act of raising my hand, it doesn't count as such.

So, how does Davidson's discussion of reasons-explanation fit into the causalist-expressivist debate? Davidson wants to claim that an actor's *primary reason* for acting is analogous to her *intention* for acting. And since he is arguing that to give a primary reason for acting is to give a causal explanation, he is essentially claiming that an agent's intention causes his action. Therefore, although there are moments in which he agrees with some expressivist points, Davidson is ultimately a causalist. Thus far I have attempted to lay out a preliminary explication of the causalist viewpoint; I now want to give an account of the expressivist viewpoint to see where the two accounts agree and disagree.

### **The expressivist theory of action**

Where the causal theory takes there to be a causal relation between the agent's intention and his action, the expressivist theory posits an expressive relation. That is, rather than the intention *causing* his action, his intention is *expressed* in the action. It might be helpful at this point to remind ourselves of Hegel's claim that action is *qualitatively different* from nonaction. In causal theory, action is something

to be explained by referring to the agent's prior intention; in expressivist theory, action is a given, to be described as inhabited by the agent's intention.

In his essay "Hegel and the Philosophy of Action" Charles Taylor aims to investigate the basis for Hegel's argument that action is qualitatively different from nonaction, as well as exploring the nuances of that difference. From the outset, Taylor sets up his discussion of Hegel's expressivism as in opposition to causalism. Thus he begins his essay by characterizing the causal view as predicated on the Cartesian, dualistic model of mind separate from world, which presupposes "a clear ontological separation between outer event and inner background" (23). The cause of an action is wholly within the mind, and the action itself is wholly within the world, separate and cut off from its cause. For the causalist, then, the body is still a foreign object to the mind; the action, taking place in the body, is in some sense cut off from the agent. Hegel disagrees with this conception of the act as beyond the realm of the agent's control; to act, he says, is not to set into motion an event that is a foreign, untouchable object for the agent. Rather an action is something purposive, directed and inhabited by the agent's intentions. Someone's reason or intention or purpose in acting "can only exist in animating this action; or its only articulation as a purpose is in animating the action" (Taylor 24). After all, to have an intention and not act on it is to keep the intention in a merely potential rather than an actualized state. To make the intention manifest in the world is simply *to act*. Herein lies the key difference between the causalist and expressivist: the causalist sees the intention and action as ontologically separate, the expressivist as ontologically inseparable.

Pippin, also a Hegel scholar, finds the same argument poetically imagined in an aphorism from *Genealogy of Morals*. In his essay "Agent and Deed in

Nietzsche's *Genealogy of Morals*" Pippin analyzes a passage in which Nietzsche compares the splitting of the agent and her deed to the splitting of lightning and its flash. The central passage:

just as the popular mind separates the lightning from its flash and takes the latter for an action, for the operation of a subject called lightning, so popular morality also separates strength from expressions of strength, as if there were a neutral substratum behind the strong man, which was free to express strength or not to do so. But there is no such substratum; there is no "being" behind doing, effecting, becoming; "the doer" is merely a fiction added to the deed—the deed is everything. (Nietzsche qtd in Pippin 372)

Here Nietzsche employs a bold and sweeping metaphor that at first glance seems to suggest that there are no subjects at all, but only actions. Pippin wants to show that granted a longer look, the aphorism invites various interpretations. The first possibility that Pippin addresses is perhaps the most glaring one: that Nietzsche wants to deny the existence of a subject altogether. Pippin suggests that when understood in the context of what Nietzsche writes elsewhere, *GM I. 13* does not endorse such a radical denial of a subject. What he suggests instead is the following, slightly more moderate interpretation: "that Nietzsche is not denying that *there is* a subject of the deed; he is just asserting that it is not *separate*, distinct from the activity itself; it is 'in' the deed...[Nietzsche] appears to be relying on a notion of expression, rather than intentional causality, to understand how the doer is in the deed" (Pippin 379). This interpretation seems reasonable; more moderate than the claim that there is no "doer" at all, it puts Nietzsche alongside Hegel as an expressivist.

Thus far I have laid out the basic premises of the expressivist view: that action is qualitatively different from nonaction; and that the agent's intentions are expressed in and ontologically inseparable from her action. At this point I want to fill out the expressivist view and explore its points of intersection with causalism. I

find Pippin's characterization helpful here in terms of structuring the discussion: he identifies and characterizes two key theses within Hegel's thought, and briefly explains their ramifications. I will base my discussion on that structure, and through those ideas begin to illuminate some of the more subtle interactions between causalism and expressivism. The first key thesis that Pippin identifies is what he calls the *non-separability thesis*, which emphasizes the imperfect nature of agential self-knowledge. The second thesis is the *non-isolatability thesis*, and stresses the importance of contextualizing the action within the correct context of social and personal history. I will say more about each thesis below.

### **The non-separability thesis**

As you might imagine, the non-separability thesis states that intention and action are non-separable, and thus even the agent herself must to some extent refer not to her prior mental articulation of her intention but to the act itself in order to come to know what her intention was. Hegel argues that "the individual human being is what the deed is. [...] even if he deceives himself on the point, and, turning away from his action into himself, fancies that in this inner self he is something else than what he is in the deed" (Hegel qtd in Pippin 380). Here Hegel wants to call into question two assumptions: first, the authority of the agent who claims a privileged position of self-knowledge; and second, the existence of an "inner self" that is somehow truer or more accurate version of who the person truly is. In reference to the former point, Taylor wants to object that "[we] have indeed a different mode of access to what we are doing, but it is questionable whether we should tub this access 'privileged'. Neither immediacy nor incorrigibility are marks of agent's knowledge" (Taylor 26). Yet just as agential self-knowledge is not perfect, neither

is it always mistaken or incomplete; if expressivists tried to argue otherwise, their theory would cohere very little to our lived experience as subjects who direct ourselves towards certain actions. For Hegel, it is precisely this element of agential self-awareness that constitutes the distinguishing factor between action and nonaction: "there is a knowledge we are capable of, concerning our own action, that we can attain as the doers of this action; and this is different from the knowledge we may gain of objects we observe or scrutinize" (Taylor 25). In other words, an agent's understanding of her action is different from her awareness of an external object or an action done by somebody else. Thus we have a distinctly unique position regarding our own actions, but we shouldn't mistake the singularity of the position for authority.

If Hegel means to topple the authority of the agent, he also means to relocate the "true self" from the inner mind to the observable world of action. After all, if I cannot presume to have perfect knowledge of my intention in any given action, then I cannot claim to have privileged knowledge of who I truly am. In fact, the entire notion of a "true self" is predicated on the idea that our actions do not faithfully represent who we are. On the contrary, it is our actions which reveal us *as* we are. And although we have already pointed to the fact that we are sometimes wrong about the reasons why we are acting, we have not yet said much about why we as agents might be motivated to accept a mistaken understanding. I believe that it is in order to preserve and promote a generous image of ourselves that we occasionally choose to believe that the actions we take are faithful to our true selves.

On the one hand, we may fail to complete an action and blame our failure on arbitrary external factors. For example, if I believe myself to be a talented poet,

and yet have failed to write a great poem but instead have written a recognizably bad one, I may say to myself "Well, the idea of the poem I have in my head is much better than what came out; if I put a little more effort into the translation of the mental poem into the written one then I'm sure the outcome would be great." Thus I needn't relinquish my self-image as a great poet. On the other hand, I may succeed in completing an action and subsequently misrepresent my intentions in order to create a more pleasing personal narrative. Davidson says something to this effect: "you may err about your reasons, particularly when you have two reasons for an action, one of which pleases you and one which does not...You may be wrong about which motive made you do it" (18). In order to maintain a positive self-view, agents may lay claim to the supposed privileged authority of their position in order to craft a narrative of a "true self" as set apart from the actions they take. Accordingly, Hegel wants to expose the idea of the "true self" as a construction, while turning to actualized deeds in order to accurately reveal the agent.

Pippin adds that "I may start out engaged in a project, understanding my intention as S, and, over time, come to understand that this was not really what I intended; it must have been Y, or later perhaps Z" (Pippin 381). The idea is thus that we cannot know our intentions merely from our articulation of them alone. Rather, in much the same way as observers of my action must interpret and understand it without access to my inner mental state, so must I attempt to articulate an interpretation of my action not only before it occurs, but also afterwards.

I see at least two possibilities following from the non-separability thesis: first, that all Hegel is proposing is that we as subjects need to be more self-aware

of the way that we *actually* are in the world, so that our idea of how we act lines up with how we really *do* act. This interpretation leaves intact the causal theory of action and simply requires that we as agents regularly reflect on our actions and intentions. Such a shift might include periodically asking those who are close to us how they have interpreted our actions, incorporating their input into our own understanding of how well our intentions match up with what we actually do.

The second possibility, and the more radical one, argues that human action simply cannot be understood as causal. Rather, what Hegel suggests is a more complicated unfolding of agency in which the individual's intentions are constantly changing: "we should understand successful action as a continuous and temporally extended, everywhere mutable translation or expression of inner into outer, but not as an isolated and separated determinate inner struggling for expression in imperfect material" (Pippin 382). At the core of this reimagining is, I believe, a rethinking of how the human mind, which feels like such an inner, private experience, interacts with the outer, observable world through the body. In other words, it calls for a different understanding of humans as embodied beings.

### **The non-isolatability thesis**

The non-isolatability thesis claims that "attending only to a specific intention as both accounting for why the act occurred and what is actually undertaken distorts what is necessary for a full explanation of an action" (Pippin 381). Humans are inherently social beings, always already within the context of a world of others. Just as it is impossible to isolate a person from her environment, it is impossible to isolate one intention from the social forces shaping it. Since our values, beliefs,

desires, and habits are formed in the context of specific social norms and traditions, it makes sense to interpret actions with that context in mind.

Now, in pointing out the expressivists' insistence on the importance of contextualizing an action, I do not mean to suggest that causalists are ignorant of the role that social forces play in shaping our intentions. Indeed Davidson, in his definition of a pro-attitude towards a certain kind of action, lists not only desires, urges and promptings, but also "a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values" (Davidson 4). Clearly Davidson appreciates the importance of the social structure in constructing a solid understanding of the agent's action. The difference between Davidson and the expressivists is that his model of interaction between the individual and society is still predicated on a relatively strong distinction between the two. That is, society acts on and affects the individual, who then, within his mind, forms an intention which causes the external movement of his body. The expressivist rejects this rigid division between the "internal" mind of the individual and the "external" world of observable behavior.

With this in mind, it is easy to draw the connection between a person's location within a social context, the formation of their intentions, and the development of their character. The term "character" has not been used in my essay up to this point, but the essence of the concept has been present throughout. The expressivist's main claim that the agent's intention is expressed in her action might be reworded as "the agent's character, or the agent herself, is revealed in her action." To deepen our understanding of how the notion of character plays into the idea of human action, I now want to turn to Aristotle.

### **Aristotle's Account of Character**

Of course, Hegel was not the first person to suggest that an agent's intention is not *behind* or *before* the deed, but wholly within the deed. The idea of the unity of form and matter goes back to Aristotle; the roots of the expressivist tradition lie with him (Laitinen & Sandis, 7). With that in mind, I want to now add a bit of Aristotelian thought into the discussion. I will focus on his discussion of virtuous action in book II of *Nicomachean Ethics*, which takes as its subject matter virtue of character. Aristotle begins the section by stating what was discussed in the previous section: that there is virtue of thought and virtue of character (1103a10). Aristotle's aim in book II is to lay out a picture of what kind of a thing virtue of character is, as well as how it can be brought about. Throughout Aristotle's work, a meaningful discussion on the topic at hand will be one that sheds light on how one might live a good life, so the debate on how virtue of character arises within an individual is vitally important for his thought.

From the outset of book II, Aristotle states that it is from habituation that virtue of character arises, as opposed to virtue of thought which arises largely from teaching (NE II.1 1103a15). Based on this observation, Aristotle concludes that since character virtue requires a process of habituation, it must not reside naturally within the individual, or else it would arise naturally, right away, without requiring any additional process for its manifestation. However, he doesn't leave the acquisition of character virtue entirely up to chance; it is not the case, he claims, that humans could be habituated to acquire any possible state of character. Rather, we are able to acquire the virtues through habituation because we already have the potential to acquire those virtues. "Hence the virtues come

about in us neither by nature nor against nature, rather we are naturally receptive of them and are brought to completion through habit" (NE II.1 1103a24). In other words, our potential to take on the various virtues of character is partially based in our given nature, and partly based on the kind of upbringing we receive.

Therefore, for Aristotle, the role of habituation is vital in the production of a virtuous character.

We already have a commonsense understanding of how habituation works, but Aristotle's conception of the process of habituation is helpful for our discussion of action. According to his account, to become habituated in virtues of character is to learn by engaging in virtuous actions. In essence, you learn by doing what you don't yet know how to do:

The virtues...we acquire by first engaging in the activities, as is also true in the case of the various crafts. For the things we cannot produce without learning to do so are the very ones we learn to produce by producing them...Similarly, then, we become just people by doing just actions, temperate people by doing temperate actions, and courageous people by doing courageous ones" (NE II.1 1103a30).

Here we see Aristotle making a distinction between the individual *action* that a person performs and that person's *character* over a period of time: a person who does one just act is not thereby a just person. Rather, a just person is something you must *become*, and therefore is something that has to be worked on, practiced, acted out over time, until you have acquired a sort of second nature or what Aristotle calls a *state of character*. A just person is not only someone with the disposition to act justly, but someone who *actualizes* that disposition. And furthermore, once that state of character has been acquired, there is no guarantee that it will remain in place—although the individual's habituation will dispose her to take just actions, she must still at times consciously reflect on the nature of her

actions. A just character is meaningless unless it is manifested through acting justly. In this model, it is the repetition of virtuous *action* that leads to the formation of a virtuous *character*. It is in this distinction—between *person* and *character*—that is where I think Aristotle's account provides helpful insight as well as an interesting tension in relation to expressivist thought. On the one hand, it is a practical albeit tautological roadmap of how to acquire states of character, i.e. through doing the kinds of action that a person with that state of character would perform.

Let us then agree that a virtuous state of character is a kind of good habit of the soul that arises over time, as habits do, through the practicing of good actions. Aristotle believes that we learn the right kinds of actions to take by learning to enjoy the right kinds of things in our childhood. He thus is in agreement with Plato that "we must be brought up in a certain way straight from childhood...so as to enjoy and be pained by the things we should" (NE II.3 1104b10). In other words, the way in which you are taught to respond to pleasure and pain as a child will influence the way you act as an adult. Later, Aristotle adds that "pleasure has grown up with all of us from infancy. That is why it is difficult to rub out this feeling that is dyed into our lives" (NE II.3 1105a1). I find the metaphor of a dye that colors our lives to be illuminating to what he wants to say about states of character; just as a dye transforms what is already there and presents it in a new way to the world, habituation acts on and transforms an individual's given nature. That "dye" is what Aristotle is referring to when he talks about a person's state of character—it is something that is formed, maintains relative stability over time, and determines the way in which the person experiences the world and reacts to it.

This historical context provides useful insight on how to interpret the expressivist picture. If we return to Pippin's essay, we might notice that all discussion of the doer/deed relation refers merely to individual actions. A synthesis of Aristotle with the more modern expressivist theory might look something like this: an actor's intentions can be found within the action she takes, and is determined by the state of character she has developed. So the agent, through habituation, comes to develop her own state of character, which becomes the second-nature framework through which she sees the world and acts within it.

Whereas an individual action might be said to display or express a subject's momentary intention, how might a series of actions taken by the subject display or express that subject's character in the long term? Pippin too has some questions for expressivism; while he is clearly sympathetic to the expressivist view, he is not uncritical. After all, the expressivist view that he locates within Hegel's theory is not a finished theory at all. One key question that Pippin has regarding the expressivist account is how we are to address the unity of a person's experience if the stability or knowability of their character is called into question. "If whatever it is that is expressed in such deeds is not a stable core or substantial self, neither as an individual soul nor as a substantial type, what could form the basis of the temporal story that would link these manifestations and transformations?" (385). How can we reconcile the expressivist picture of a temporally fluid intention-formation and articulation with our experience of seeming to have relatively stable beliefs, desires, intentions, and plans for acting?

At this point in the discussion I want to turn to Harry Frankfurt's essay "The Problem of Action" in which he critiques the causalist approach and puts

forth the beginnings of a theory that I think can be usefully integrated into expressivism.

### **Frankfurt and intentional movement**

Frankfurt enters the debate by clearly articulating the central problem of action as he sees it: "to explicate the contrast between what an agent does and what merely happens to him" (Frankfurt 69). This is our question from above, and Frankfurt agrees that it is the causal theory that is the most popular way of answering that question. He defines the causal theory of action as one that distinguishes between actions and mere events by virtue of a causal chain, that is, a certain set of conditions that precede the action: "a bodily movement is an action if and only if it results from antecedents of a certain kind" (Frankfurt 69). However, for Frankfurt, this seems unlikely to be the characteristic that separates actions from events. He claims in his essay that contrary to what the causal theorists hold, human action is not to be defined by referring to a prior causal history.

Frankfurt argues that the reason the causal theorists turn to the events preceding the action in order to explain it is because they can find no difference between actions and mere happenings by looking at the action *at the time it is occurring*. As the event is occurring, there is nothing in that bodily motion to signal its action-ness to the observer (or, for that matter, to the actor herself). And yet as humans who often have the experience of acting with agency, our experience tells us that there *is* such a difference between actions and events. Frankfurt imagines the causal theorist's response to this puzzle: if the difference between actions and events is not located in the observable movement at the time of the act itself, the only possibility for locating the distinction is to go back in time, to the context of a

history of certain causal events. Here Frankfurt articulates his critique quite clearly: just because the motions of a human body do not reveal whether or not it is an action or a mere happening, “it does not follow...that the only way to discover whether or not a person is acting is by considering what was going on *before* his movements began—that is, by considering the causes from which they originated. In fact, the state of affairs *while* the movements are occurring is far more pertinent” (72). In other words, what the causal theorists fail to consider is that although observable bodily movements might be unable to reveal the nature of the event as action or not, it is possible that something other than the bodily motions (a mental event or state, perhaps) occurring *at the time of the event* could be the distinguishing factor. Thus there is no immediate or logical need to jump back in time in search of the distinction.

I agree with Frankfurt’s suggestion. The causal theorists, in looking back at the causal history preceding the action, are not returning to *observable* bodily motions but to *mental events*, such as intentions. In other words, the causal chain preceding an action is not made up of observable movements, but of mental occurrences. If this is the case, then I see no reason why we should look to prior mental events to explain action rather than mental events contemporary to the action itself.

Based on his critique of the causal theory, Frankfurt introduces the notion of guidance. He proposes that we think of actions as events that occur under the agent’s guidance, thus bringing the attention back to the importance of understanding what’s going on in the moment of action. The key part of his theory is best summarized here:

Given a bodily movement which occurs under a person's guidance, the person is performing an action regardless of what features of his prior causal history account for the fact that this is occurring. He is performing an action even if its occurrence is due to chance. And he is not performing an action if the movements are not under his guidance as they proceed, even if he himself provided the antecedent causes—in the form of beliefs, desires, intentions, decisions, volitions, or whatever—from which the movement has resulted. (73)

This account relies on a phenomenological understanding of action. Frankfurt is bringing our attention to the fact that we sometimes (or often) have the experience of guiding the course of our actions as they occur, with or without prior mental events. He does not mean to lay out a complete metaphysics for what that kind of guidance is or how it operates, rather he wants to show that there is something to our experience that is left out in the picture proposed by the causal theorists of action.

Frankfurt distinguishes two ways in which our movements could be said to be guided, or purposive: one kind of purposive movement is the unconscious way in which our breathing or the blinking of our eyes is guided by some biological mechanism. This kind of purposive movement, although it is guided by the person's *body*, could not truly be said to be guided by the person *himself*. The other kind of purposive movement is seen in the movement, for example, of my spoon to my mouth, or of raising my hand in class when I want to speak. These movements are guided by the agent herself, not by biological mechanisms of the agent's body. Frankfurt proposes to call these kind of purposive movements that are guided by the agent "intentional movements." These are not to be confused with what he calls "intentional actions," or actions "which the agent intends to perform" (73). Thus an *action* can be defined as intentional movement (i.e. action

under the guidance of the agent) and an intentional action designates those actions which are premeditated and intended.

### **Towards Expressivism**

I now want to give a few of the reasons why I believe the expressivist theory provides a better framework for understand the nature of human action. I will then put forth my own tentative claim for how I think the expressivist view can start to define human action.

First, a critique of causalism that I think expressivism manages to avoid: in restricting their examples of human action to relatively basic, discrete acts, the causalists needlessly and mistakenly restrict their analysis. The examples of action that I encountered in the causalist literature include the act of raising your hand, turning on the light, turning left at the fork in the road. Even the examples that I have utilized have followed this model. These examples might be described as basic actions, involving bodily movements that cannot meaningfully be broken down into more basic movements. Turning on the light, for example, occurs in one motion of my hand. Thus these examples are also discrete actions, in that the action is complete once the bodily movement has occurred. These are undeniably actions, and yet they are far from the only kind of action that humans perform. First of all, there are actions that occur over a longer period of time and involve more complex bodily movements, such as the performance of a piece of music. And second, there are actions that have not only to do with everyday decisions to turn on the light or dive into a pool, but are also inherently value-laden ethical decisions. The causalist, in limiting herself to basic, discrete actions and choosing to focus on the mechanism of the act rather than its ethical value, seems to focus in

on the part of an act that is perhaps the least distinctive of human action. After all, animals are also in control of their bodily movements to some extent, and the focus of our discussion is human action; the expressivist approach, as opposed to the causalist, seems to highlight examples of action that are more distinctively "human."

The second reason that I have for favoring the expressivist approach is related to Frankfurt's critique. As I showed earlier in the essay, Frankfurt posits that the causalists, failing to locate an observable difference between action and nonaction, presume that the distinction lies before the action. And yet in declaring the prior causal chain as the distinguishing factor, causalists make use of the agent's internal state of affairs to define action, despite not having referred to the internal state of affairs during the moment of the action's taking place! In other words, their criteria for distinguishing action from nonaction seems to change. The only explanation I can think of for this is that the phenomenological difference that the agent experiences between cases of action and nonaction is difficult to pin down and analyze, and is therefore dismissed as a merely experiential difference rather than a metaphysical one.

I think that one way of characterizing my problem with the causalist account is well explained in Jonathan Dancy's essay "Arguments from Illusion." In this essay Dancy attacks the conjunctivist standpoint in the philosophy of perception. One school of thought in the philosophy of perception, conjunctivism, holds that "there is a basic state in common between cases of illusion and cases of genuine perceptual awareness" or in other words, that it is impossible for me to introspectively determine whether I am having a veridical perception or an illusory one (Dancy 118). Therefore there must be a common element between the

two cases, and cases of genuine veridical perception can be explained by having the right kind of cause in the external world. Similarly, the causalists of action theory imagine an analogous indiscriminability between cases of action and nonaction. Their argument from illusion would look something like this: cases of action and nonaction are indistinguishable from a third person perspective, so there is a common element; yet there is a difference between cases of genuine intentional action and mere happenings; therefore cases of genuine action consist in mere happenings plus being caused in the right sort of way.

But in the case of human action the argument from illusion seems unnecessary, for although it might be the case that from a third person perspective action perfectly resembles nonaction, as first person agents we *do* experience a phenomenological difference in cases of acting as opposed to cases in which things merely happen to us. Perhaps we do not explicitly articulate that difference; perhaps we are not even conscious of it at times. Yet it would be a mistake to claim that to the agent herself, cases of acting and non-acting are indiscriminable.

Having given what I hope has been a convincing critique of causalism, I want to end by suggesting some potential ways for expressivism to be developed. Earlier in my essay, I referred to Taylor's explication of Hegel's conception of the nature of agential self-knowledge: "there is a knowledge we are capable of, concerning our own action, that we can attain as the doers of this action; and this is different from the knowledge we may gain of objects we observe or scrutinize" (Taylor 25). I want to suggest that the knowledge he is referring to is a kind of enactive knowledge, rooted in the present moment of the action. In other words it is a behavioral knowledge. Drawing support from Aristotle's account of state of character and Frankfurt's idea of intentional movement, I want to suggest that

human actors develop a kind of second nature which allows for an in-the-moment control of events.

Aristotle described a process of developing a certain kind of character by performing the kinds of actions that a person with that state of character would perform. Thus his idea of a state of character, far from being something essentially inner, is rather something that is essentially active—to be just, it is not enough to *try* to do just things, or to imagine that given the right circumstances you *would* do just things. You must *actually do* the just thing in various situations over a period of time. Thus your intentions do not matter unless they are actualized by animating the action. However, to gain a state of character is to gain a propensity for acting in a certain way. It will become an unthinking second nature—much like Frankfurt's notion of the agent guiding or directing her action without needing to consciously put effort into forming or articulating her intentions.

In addition, I think it will be helpful to widen our analysis of actions from the basic, discrete actions that populate the examples of the causalists, to the more distinctively human kinds of actions that are present in expressivist literature. In particular, I think that expanding the analysis to instances of performance will be extremely illuminating. A musical performance, or perhaps the act of painting a picture, as kinds of extended action, are capable of revealing much more about an agent's character than more basic actions such as turning on the light.

## **Conclusion**

My own interpretation of the expressivist picture, based on Pippin and Taylor's presentation of Hegelian expressivism, is that it does not deny the existence of a subject, nor does it deny that subject's agency. Rather it denies the existence of a

causal agent who exists in a kind of privileged inner sanctum, separate from and behind her actions. In other words, it doesn't mean to negate the existence of the inner mind that thinks, desires, wishes, believes, and imagines—to deny that would go against our experience. What it denies is the dualistic Cartesian model in which that inner mind is wholly apart from the outer world. Indeed, that is what I find most appealing about the expressivist theory of action: that it is not predicated on that model in which mind and body are sharply distinct and set apart from each other. It acknowledges the complexity of the mind-body relation.

In addition, it acknowledges that many of the everyday actions we engage in seem to be so inherently physical that it doesn't make sense to think of them as causal in the traditional sense. For example, when I wake up in the morning I unthinkingly get out of bed, put on my slippers, and go to the kitchen to make myself tea. That involves actions such as turning a doorknob, stepping into my slippers, turning on the stovetop, and reaching for a mug. These are all actions that I performed, and yet it is not the case that I was articulating my intention to do them prior to my actually doing them. Could I be said to have caused them with my intention? No—I merely *did* them, and in doing them my intention was not only displayed but made manifest.

Although I am positive that the expressivist view has the right way of thinking about human action and has the potential to radically transform our commonsense understanding of human agency, I also believe that there is still much work to be done. Furthermore, I believe that it will be difficult to do this work if a slightly more analytic lens is not taken up by modern expressivist thinkers. Hegel was attempting to develop a new way of thinking about human agency such that it does not continue to falsely assume that to act is for the inner

mind of a person to cause some outer event. Although I believe in the mission of this project, I think the fact that action is identified as a primitive given from the outset makes it more difficult to investigate the ontology of action. In a way, it would be nice to merge the vision of the expressivists with the analytic nature of the causalists.

Thus my project has been rooted in a long philosophical tradition of seeking to understand the relation of the inner mind of the individual to the outer world of observable behavior. By coming to a better understanding of that connection between the internal and the external, we will be able to construct a better understanding of the way that the internal and the external interact. In other words, we must first understand the mind and world as either sharply distinct or as somehow more whole before we can aspire to say anything about the way that human beings act in the world.

#### **Works Cited:**

- Aristotle. *Nicomachean Ethics*. Translated and edited by C.D.C. Reeve. Indianapolis: Hackett Publishing Company, 2014.
- Costa, Michael J. "Causal Theories of Action." *Canadian Journal of Philosophy* 17, no. 4 (1987) : 831-854.

- Dancy, Jonathan. "Arguments from Illusion." In *Disjunctivism: Contemporary Readings*, edited by Alex Byrne and Heather Logue, 117-136. Cambridge: The MIT Press, 2009.
- Davidson, Donald. "Actions, Reasons, and Causes." In *Essays on Actions and Events*, edited by Donald Davidson, 3-19. Oxford: Clarendon Press, 2001.
- Frankfurt, Harry G. "The Problem of Action." In *The Importance of What We Care About*, edited by Harry Frankfurt, 69-79. New York: Cambridge University Press, 1988.
- Laitinen, Arto and Sandis, Constantine. Introduction to *Hegel on Action*, edited by Arto Laitinen and Constantine Sandis. New York: Palgrave Macmillan, 2010.
- Mele, Alfred. Introduction to *The Philosophy of Action*, edited by Alfred Mele. New York: Oxford University Press, 1997.
- Pippin, Robert B. "Agent and Deed in Nietzsche's *Genealogy of Morals*." In *A Companion to Nietzsche*, edited by Keith Ansell Pearson, 371-386. Malden: Blackwell Publishing, 2008.
- Schlosser, Markus. "Agency, Ownership, and the Standard Theory." In *New Waves in Philosophy of Action*, edited by Jesús H. Aguilar, Andrei A. Buckareff, and Keith Frankish, 13-31. New York: Palgrave Macmillan, 2011.
- Taylor, Charles. "Hegel and the Philosophy of Action." In *Hegel on Action*, edited by Arto Laitinen and Constantine Sandis, 22-41. New York: Palgrave Macmillan, 2010.

#### Works Consulted:

- Haddock, Adrian. "Action". In *Oxford Bibliographies Online: Philosophy*. 21-Nov-2014.
- O'Connor, Timothy and Sandis, Constantine, eds. *A Companion to the Philosophy of Action*. Malden: Blackwell Publishing, 2010.
- Pippin, Robert B. "Hegel's Social Theory of Agency: The 'Inner-Outer' Problem." In *Hegel on Action*, edited by Arto Laitinen and Constantine Sandis, 59-78. New York: Palgrave Macmillan, 2010.
- Sneddon, Andrew. "Does Philosophy of Action Rest on a Mistake?" *Metaphilosophy* 32, no. 5 (2001) : 502-522.