

## Knowing What We Are Doing

Heidi L. Maibom

Suppose you are standing with a group of your friends outside a busy restaurant waiting for your table. A group of old people leaves and walks up the road. Suddenly one of them falls to the ground. There is a pause. No one moves. Then someone says: "He fell over, maybe we should go help?", to which another replies: "No, they'll take care of it". The conversation resumes briefly. The waiter comes out and lets you know that your table is ready. As you walk in, you note that the person is still on the ground while the other old people are attempting to help him to his feet. Are you and your friends blameworthy for your inactivity? After all, it is more than evident that every person in that group is very old and in a very poor position to help the fallen man. It is lucky that the old people do not harm themselves while helping the fallen person. All of you, on the other hand, are relatively young and spry.

Now pretend that you are the chair of an academic department in the US perusing Lakisha Johnson's cv. If you are like most, you will judge her to be less qualified than a male candidate with an English sounding name. The candidate is in the unfortunate position of sounding African-American, but also applying while being a woman. However, you certainly don't take yourself to be discriminating against her. You are a self-avowed feminist after all. You think she is a less good candidate. You are, we may say, the victim of stereotyping as much as the candidate is; only the consequences are more dire for her. Can we blame you for discriminating against her?

On the one hand, it seems that not helping those in need and discrimination are things people *do*, and if they are not cognitively or volitionally impaired, they can be held responsible for it. On the other, they not conceive of their actions as we do. The bystanders genuinely thought that the old person did not help and the chair really did believe that Lakeisha is less qualified that

Robert. It therefore seems unfair to hold them responsible for actions that they were unaware of performing. We typically demand that people know what they do to be responsible. On the third hand, as it were, it certainly *seems* as if the youngish, spry people and the chair *ought* to have known what they were doing, and if so, they can be held responsible for their actions.

The difficulty is encapsulated in some of the debate surrounding employment discrimination cases (e.g. Fiske & Borgida 2008, Williams 2003). Given the available evidence that much discrimination is due to stereotyping based in unconscious bias, how do we evaluate defendants' guilt? Stereotyping appears to be an automatic process, i.e. "unintentional, uncontrollable, efficient, autonomous, and outside conscious awareness" (Fiske & Borgida 2008, 127). The trouble is that there is a tendency to think of responsibility in terms of awareness, both in the law and in philosophy. People talk as though an agent's responsibility extended only as far as her awareness of what she is doing (Sher 2009, 4). This is not meant to exclude cases of neglect, but these instances are downplayed. This is nowhere more clear than when John Martin Fischer and Mark Ravizza maintain that "an agent is responsible only if she [...] knows (or can reasonably be expected to know) the particular facts surrounding her action" [etc.] (1993, 8). The parenthesis suggests that cases where the agent does not know what he or she is doing are the exception and should not mislead the reader when she thinks of the main cases, which almost invariably concern reasons and considerations that the agent is aware of culminating in action that the agent conceives of under the description or descriptions relevant to moral assessment. But even if they are exceptions—and more argument is needed to show that they really are—the justification of why agents can be held responsible for such actions must connect in some interesting way to their responsibility in ordinary cases.

In this paper, I consider whether or not we are responsible for acting in ways that we do not conceive of in this way, because our actions are largely determined by situational influences—of which the two examples above are examples—that we are unaware of at the time of action. Put somewhat bluntly, the question is whether we are sometimes responsible for doing what we do not know that we do. The two cases given at the outset of the paper are both

examples of situationist influences. I argue that we sometimes are responsible for acting under such influences though we do not know what we do at the time. Accounts that take character, identification or a deep self to be central to responsible agency cannot account for responsibility for situationist influences, as these are largely impersonal. Instead, I develop a version of the idea that an agent could and should have known what they were doing. I think of this in terms of ‘inferential routes’ to a relevant description of the action. The capacity to know what they do required for agents to be responsible is not necessarily or even primarily direct, private, first personal, or introspective. There is no need for the agent to be able to be directly aware of her action under a relevant description. Rather, it is the ability to conceive of our action from a perspective other than our own that is central to responsibility; to know what we do as from the outside. Or so I shall argue.

### **1. Not Knowing What We Do**

A person who leaves her dog in the hot car or a distracted driver who does not give himself enough time to merge with traffic are both typically held responsible for their actions (Sher 2009). Yet, they did not perform the action we hold them responsible for under that description. In fact, it is less what they do and more what they *failed* to do that we are concerned about. Their fault is one of *inaction* or *omission*. Omission is tricky. We cannot but omit to do an infinite number of things. *C’est la vie*, you might say. But though we cannot help but *not* do a great number of things, we do have some control over which particular omissions we make. Typically omissions are only thought to be relevant if the agent has knowledge relevant to the situation or could reasonably have acquired such information. The person who leaves her dog in her car on a hot day can reasonably be expected to know that she should leave the windows ajar or not leave the dog in the car at all, whereas her husband cannot be expected to come to the dog’s rescue even if he is only a block away if he does not know what is going on.

Neglectful actions are actions agents fail to undertake without awareness of such failure. The woman does not think “to hell with the dog”, for instance, thus leaving it to its fate. Neglectful inaction is typically ‘undertaken’ for no

reason at all. That is, there might be *a* reason that the agent fails to consider performing the action that we judge he should have performed, but more often than not *he* does not have a reason *not* to perform it. According to George Sher (2009), neglectful action is a prototypical example of agents acting without knowledge of what they are doing, yet they are responsible. It is a curiously opaque situation: the agent is unaware of omitting to do something.

We should agree with Sher, I think, that agents can be held responsible for neglectful actions even though it means that they are sometimes responsible for doing what they do not know that they do. The types of neglect Sher talks about are cases where the subject is distracted, preoccupied, overly focused on other tasks, etc. So it does seem quite reasonable to say that the agent *could* have known what he was doing and, if the action is morally significant, we can praise or blame him for it. But what should we say of cases where the agent is unaware of what she is doing in a more principled way? Typical neglect cases are cases where *if* the agent considered the situation—‘the dog is in the car and it’s hot and sunny’, or ‘the other drivers on the road won’t have time to avoid me if I speed up to merge’—*then* the agent would have acted differently (assuming that they are *only* cases of neglect). But the cases we are considering are ones in which the agent *does* consider the situation, but fails to see that his or her action falls under a description that makes it blameworthy. It is a description that the agent is typically disposed to deny is appropriate. So where the neglectful agent can be made to see what he neglected to do pretty easily, the person who fails to help someone in need or who discriminates against a candidate is apt to question our description of his action. Acting under unconscious psychological influences seems different from being neglectful.

Once we want to hold people responsible for actions that they do not conceive of under that description, we are faced with the problem of limiting the potentially infinite number of descriptions any action can fall under to just those an agent is responsible for. The most straightforward way to do this is to anchor such descriptions in what the agent has conscious *access* to. The agent either knows what she is doing, or she has access to that information. In the neglect cases, it is a matter of changing her attention to other features of the situation. She can consciously access the information she requires to act better, but she does

not. But there are many cases where we act in ways where the best or most salient explanation of what we do is not consciously accessible to us in this way.

Situationist influences, of which the bystander effect and stereotyping are examples, are instances where conscious access to a morally relevant description of what one does is affected (in ways soon to be specified). Some situationist influences are surprising, but hardly morally disquieting. Richard Nisbett and Timothy Wilson (1977) found that people tend to show a right side bias when asked to choose the best quality item in a row (position effect) and that they tend to value a mug that they own more than an identical one (endowment effect, see also Kahneman, Knetsch, and Thaler 1990, Kahneman 2011). Subjects were not only unaware of these effects, but would also categorically deny that they had any influence on their actions. One might fairly claim that they did not know what they were doing. However, determining the relevant description of their action—how to weigh their account of what they were doing against psychologists' descriptions of their actions—may seem relatively unimportant. Other types of situational influences are more troubling.

Bib Latané and John Darley discovered that people are much less likely to help a person in need if other people are present (Latané & Darley 1970). People are also less likely to help if they are in a hurry (Batson & Darley 1973), and more likely to help if in a good mood (Carlson, Charlin, and Miller 1988). Yet, these are not aspects of the situation that people reference when explaining their reasons. In bystander scenarios, people typically say that they did not help the person in need because they were not quite sure what happened, whether the person was in need of help, or whether the situation was serious enough to require intervention (Latané & Rodin 1969). People's tendency to cooperate is also easily influenced by the environment they are in. Describing a prisoner's dilemma as *the Wall Street Game* as opposed to *the Community Game* has large effects on cooperation. Under the former description, only about a third of subjects chose to cooperate compared to two thirds when the game was presented under the latter description (Ross & Ward 1996).

Stereotyping is now a well-known phenomenon. It affects all aspects of group belonging: ingroup vs. outgroup, men vs. women, white vs. black, white vs. Asian, high socioeconomic status vs. low socioeconomic status, old vs. young,

and so on (for overviews see Kunda 1998 and Fiske & Taylor 2013). For instance, the prevailing stereotypes associated with women is that they are submissive, intuitive more than rational, understanding, empathetic, bad at math (and philosophy<sup>1</sup>), emotional (sad, not angry), and so on. Women and African Americans are assumed to be less qualified than men (Superson 1999) even if the evaluator has the same information about both, e.g. identical cv's (Steinpreis, Anders, and Ritzke 1999, Bertrand & Mullainathan 2004, Moss-Racusin et al. 2012). This is thought to be due to stereotyping, (hopefully) mostly of an unconscious kind.

In the above cases there is reason moral concern about situational effects. People see certain features of the situation as salient or ignore other, morally important, features, without having conscious access, even in principle, to the fact that this is occurring. The types of explanations that psychologists typically give of such situational effects do not have a counterpart in what the subject can become aware of by introspection. This is why someone like Timothy Wilson (2002) says that much of the time we do not know *what* we are doing, or *why* we are doing it. There is some truth to this. Stated like this, though, it suggests a vision of human kind staggering about like zombies operated by situational and other factors as if by a puppet master. But just because people don't conceive of their action under a certain description doesn't mean that they have no idea what they are doing. They may conceive of their actions *under a different description*. In stereotyping, for instance, the subject takes himself to be doing whatever he is doing without undue influence of extraneous factors. In his own mind, the chair judges our female African-American candidate to be less qualified than her white male peer because she *is* less qualified.

The bystander effect distinguishes itself from stereotyping by causing, or contributing to, *inaction*. In John Darley's and Dan Batson's (1973) cases, where subjects rushed past a person slumped in a doorway seeming to be in dire straits, many maintained that they hardly even noticed the person in need, so busy were they getting to where they were going. In Latané and Rodin's (1969) study where the experimenter pretended to have an epileptic seizure behind a curtain

---

<sup>1</sup> See, e.g., Haslinger 2008.

separating her from the experimental subjects, who were waiting for her, people claimed not know whether she really needed help. In bystander cases, people typically consider the other person's situation and decide—incredibly—that the person is not in need of help, or something along those lines.

A bystander who does not help need not quibble about the description of his action as 'you did not help'. Instead, she might maintain that the person was not in need of help, that it was not clear that he needed help, or that she did not know that he needed help. After all, we do not help others most of the time. Often others would not want us to help them. Sometimes it would be inappropriate to do so. It may therefore be better to describe the agent's failure to help as failing to help a person who is in need of help. This would help distinguish the agent's description of her action and that of others (other non-bystanders, as it were). The debate, then, seems to be about whether or not the person who was not helped was in need of help. But this cannot be quite right either. We do not blame people for not rushing to the help of someone who, say, has just unwittingly consumed a deadly poison, but is not yet showing symptoms. It must therefore be relatively obvious that the agent is in need of help.

Prototypical bystander cases seem fairly straightforward. Think back at the experimenter trashing around on the floor in the next room. Surely, she is in need of help! Just last week, I saw a man tumble out of his wheelchair onto a busy road. As I was on my balcony on the 13<sup>th</sup> floor, I assumed people at the ground level would be faster to come to his aid than I would. But as no one stopped to help, I descended 13 floors and crossed half a block to be the first person there to help. He had been flailing around on the street for 5-10 minutes in the rush hour! If we assume that the pedestrians and the drivers would all say that they were unsure whether the person needed help—in line with subject reports from other bystander effect studies—they are surely self-deceived. Presumably, the most obvious assumption when seeing a man flailing in the street next to his wheelchair is that he is in need of help. Yet, there is no particular reason to think that bystanders are lying when they avow their ignorance of the neediness of the person. Why is this so hard for people to see when they are in the company of others?

The situationist literature suggests that we are far more sensitive to others' actions and judgments than we think. Solomon Asch (1951, 1956) found that people are likely to override their own, correct, judgments of the comparative length of lines if the majority of the people they are with judges the lines to be of equal length. One interpretation of the bystander effect, therefore, is that since we look to others as guides to what to think and do, and they do nothing, we don't do anything either. Perhaps we are responding to what we assume is a norm of non-helping or we assume that others are not helping because help is not required or desirable. When we say that we did not help because we were unsure about whether the subject needed help or not, it may be interpreted as a reasonable shorthand for this process.

It strikes me as quite plausible that bystanders typically do not help because they are unsure of whether the subject is really in need of help or whether it is appropriate to help. This uncertainty arises from observing that others do not help, but they probably do not help because they are unsure of what to do given others' inactivity.<sup>2</sup> Most likely, they *would* help were they aware that they were subject to the effect. I know knowledge of the effect had that effect on me. This fact makes it clear that there is an essential lack of access here for most people. And this is where it gets tricky. For if we say that it is obvious to the agent that the patient is in need of help, the agent would disagree. That is the very point about the bystander effect; *people question the need of the other person*. In order to hold bystanders responsible, it cannot simply be that it was relatively obvious that the other was in need of help, because one might always ask: 'obvious to whom?'. There must be some sort of connection between the mind of the agent and the obviousness of the need of the other.

If we give up on the idea of conscious awareness, there are a couple of candidates for the responsibility fixing relation between the agent and the action:

- i. the agent *would* agree with performing the action if she conceived of it under the relevant description (or even weaker: the agent would not *refrain* from performing the action if she thus conceived of it),
- ii. the act is *in character*, or
- iii.

---

<sup>2</sup> Through my exposure to social psychology, I have long been familiar with the bystander effect and have become relatively good at ascertaining when it operates. I, therefore, tend to help. And as soon as I do, I am usually thronged with previously passive onlookers so eager to help that I often do not get to do much at all.

The agent *should and could* have known that her action is an action of the relevant kind, and if blameworthy, should have refrained from performing it.

Should we hold people who neglect to help others in need, people who discriminate against others unknowingly, etc. responsible? It seems to me that we should, or at least that we sometimes should. Some cases are clearer than others. Think of the toddler Wang Yue in Foshan, China, who wandered from her house onto a busy laneway where she was run over by a car. After the motorist leaves the scene without helping, a truck runs over her legs as she lies immobilized on the street. She is eventually picked up and taken to a hospital by an old lady. In the meantime, however, she lies crying and bleeding on the road for a full 7 minutes while 18 people skirt around her. No one helps. She subsequently died from her injuries. In this case, there seem to be little doubt that the 18 are responsible and blameworthy for not helping.

I have discussed the bystander effect at length, but similar reflections apply to the case of discrimination. Evidently, Lakisha is unfairly evaluated. Yet, the chair is unaware of the effects that stereotyping has on his evaluation and regards it as fair. This type of covert, unconscious, discrimination is the most serious obstacle facing women and minorities today (at least in the West). As in the bystander effect, it seems compelling to hold people responsible for discriminating against others even when they do not know that they are doing so. How can we make this intuition fit with a decent account of responsibility? That is the topic of the next section.

## **2. Responsibility for Situational Influences**

As previously mentioned, there are 3 good candidates for responsibility attribution that circumvents the need for conscious access: i. the agent *would* agree with performing the action if she conceived of it under the relevant description (or: the agent would not *refrain* from performing the action if she thus conceived of it), ii. the act is *in character* or expresses the agent's deep self, and iii. the agent *should and could* have known that her action is an action of the relevant kind, and if blameworthy, should have refrained from performing it. Here I rule

out i. and ii. as decent contenders for holding people responsible in situations where they are under situationist influences. I then develop the remaining candidate in the following section.

The first option is easy to dismiss as a candidate for holding people responsible who are subject to situational effects. Bystanders are not likely to agree that they should refrain from helping someone in need. And if they knew what they were doing, they would stop, i.e. they would aid the person in need (by and large). Though blatant discrimination exists, most of the discrimination women and members of racial minorities face in the West these days is of an implicit sort. Discriminators, too, are likely to disagree that it is ok to discriminate against others. Many, or most, of those who engage in implicit discrimination would cease to do so were they to conceive of their actions as discriminatory. The problem is, evidently, that we are demanding some form of conscious or explicit agreement, where susceptibility to situationist effects is really mostly implicit. But perhaps some sort of *identification* will succeed where explicit agreement fails (e.g. Frankfurt 1971).

So-called 'Deep Self' views (Wolf 2003) base responsibility in identification with the reasons that bring about the action, or perhaps with what the action represents (e.g. Frankfurt 1971). This is quite a popular approach. Crucial to responsibility, on such a picture, are the deeper structures of the self that have to do with self-evaluation, self-transformation, and longer-term goals and projects. Here responsibility is inextricably linked to our sense of who we are, who we want to be, and what we embrace or reject in ourselves. To make such views accommodate cases of neglect, *conscious* avowal of one's deeper self cannot be necessary to such a picture. People that are inconsiderate, for instance, typically do not identify with this tendency. Nevertheless, they rarely stop up and think about how others think or feel in response to their actions. And it is this fact more than any act of identification, actual or counterfactual, that makes us critical of them. Perhaps one can think of it as passive acceptance. So it is not enough to object, if blamed for something, that the action flowed from motives that are not part of who we want to be. We want to avoid holding people responsible for truly *alien* ideas and motives. In the case of being inconsiderate,

the person is arguably responsible for it to the extent that she consistently fails to take the relatively easy and obvious steps to avoid being inconsiderate. This lack of consideration can, therefore, be thought to form part of her deep self in a suitably expanded framework. But how, you might ask, is such a framework to make sense of the sorts of unconscious situational motivations and influences that we have seen examples of above?

Although John Doris (2002) worries that the impersonal nature of situationist influences sits uncomfortably with the rather rich notion of a deep self, he argues that minor adjustments to this picture will accommodate responsibility for acting on such influences. Instead of attempting to fit situational effects into a narrative structure of who we take ourselves to be or who we want to be, Doris introduces the notion of plans and long-term commitments. Through our plans and projects we plot a route through situationist space. By making it a habit to fly busily from one commitment to another, to hang out with a bad crowd, and so on, we betray a commitment or lack of commitment to certain ways of leading our lives. Someone who hurries past a person in need is blameworthy because his being hurried reflects a way of life that he endorses and though he does not endorse not helping others because he is in a hurry, he *does* endorse the type of lifestyle that predictably leads to such effects. So, situationist influences can be understood in a personal way after all, Doris maintains. And where they cannot, it might indeed be inappropriate to assign praise or blame to a subject for influences of the situationist kind.

It seems to me that Doris strikes an uneasy balance between situationism and Deep Self views. The situationist literature contrasts the person and the situation so as to stress the *impersonal* nature of situationist influences. An action performed under situationist pressures is not an action that we should understand in terms of the person's *character*, but in terms of the situation that she is in. If we consider the bystander effect again, the literature suggests that *had* there not been other people present, *then* the person would likely have helped. This is not to say that character can play no role in such situations, but where it does, it is more likely to do so in cases of *helping*. Susceptibility to situationist influences has little to do with character, life plans, etc. It is a more universal

feature of people (though culture does, of course, have an effect). It is therefore difficult to reconcile *situationism* with an identificationist account, which tend to focus on the *personal*. We may be better off thinking about ways to flesh out the idea that the subject *should* have known or been aware of the relevant reasons.

Sher (2009) fleshes out the 'should have known/been aware of' in terms of the causal aspects of a subject's reasons. He argues that if an action springs from "some combination of his own constitutive attitudes, dispositions, and traits" (p. 87) and it is a failure to live up to some applicable standard, we can hold the agent responsible for it. For something to be constitutive of a person, it must be "among the elements of the system whose causal interactions determine the contents of the conscious thoughts and deliberate activities in whose absence he would not count as responsible at all." (p. 121) Are bystanders who do not help responsible on this account? It is hard to say. On the one hand, people's sensitivity to the judgments of others is causally related to the conscious capacity for deliberation that marks what it is to be a person. So it seems that they can be held responsible. On the other hand, the suggestion that the action must spring from "some combination of *his own* constitutive attitudes" etc. suggests that a personal element remains in Sher that conflicts as much with the situationist spirit as does Doris's suggestion does.

If I am right that we can hold someone responsible though we think she was subject to the bystander effect (or engaged in implicit stereotyping), it looks like we are giving up linking responsibility to the long-term plans and policies, a person's character, or her deep self. What a person is responsible for need not be something that she is aware of, or that pertains to her character; it need not be (causally) related to her life-plan, her policies, or her values. There is nothing personal at all about being subject to the situational effects outlined by the social psychology literature. In fact, in many cases, there is something deeply impersonal about it. These are tendencies that we share with many or, in some cases, most people. They are part of who we are, of course, but not in a way that individuates us as particular persons. It may be part of our inner, deeper core, but it is a core that we have in common with many others.

### 3. Responsible Self-Reflection

Let us stop up for a moment and remind ourselves why having conscious access to relevant aspects of one's actions (or omissions) is so important. *Ceteris paribus* an agent who has conscious access to such aspects is linked to it directly by her understanding. She knows what she does; she *owns* her action. Epistemic access is also important since if she does not know what she does or is proposing to do, she lacks an important reason to change her course of action. Thus, conscious access gives you rational control and authenticity or ownership, all else being equal, which are ingredients in responsibility. But conscious access cannot account for responsibility for aspects of actions or outcomes that the agent is not aware of. It excludes simple cases of negligence. So if we are to account for why people are responsible for such actions, we must widen the scope of the epistemic condition to include the *possibility* of awareness, suitably defined. As we saw, Fischer and Ravizza demand that the agent "could reasonably be expected to know" (1993, 8) the relevant facts pertaining to her action. This seems to be a fair demand in instances of negligence where, in effect, we are requiring the agent to attend to relevant consciously accessible information ("I am leaving the dog in the car, it's hot, who knows how long I'm going to be", etc.). This line of thinking is the most promising of the possibilities that we have considered so far. The way I'm going to develop it, however, takes it in a potentially different direction by stressing an agent's capacity for reasonable self-reflection and scrutiny of motives over conscious access.

When we hold people responsible for neglect, we reflect a demand for cognitive effort on the part of the subject that goes beyond what she might be naturally disposed to expend. So neglect points in the direction that I want to take us. Consider moral education. Children have relatively little authority over how their actions are interpreted. They are blamed for actions that they did not conceive of under the description that their parents think of them. A child does not usually get off the hook by denying that he was hurting his sister and only trying to get his toy back. A delicate balance is struck whereby children are blamed for actions that they did not conceive of under that description, but not before they are able to conceive of that action as falling under such a description.

Doing otherwise is pointless, not to mention unfair. Being blamed for doing something that one was not entirely aware of doing, and accepting such blame, is part of learning how to be a moral agent. It is an ongoing process. We never quite finish learning. Life as a social being is an ongoing negotiation among individuals about how to conceptualize situations, events, and actions.

How to describe an action is not simply a function of how the agent conceives of it, or how those affected by it, or those observing it, conceive of it. (And these conceptions are sometimes incommensurable.) It is also relative to a culturally sanctioned way of conceiving of certain types of actions: their most socially salient aspects, and so on. A canonical description of an action relevant to the assignment of praise or blame is typically an interaction between these potentially different perspectives on the very same trajectory through cognitive or behavioral space. Sometimes the socio-cultural perspective is paramount. Stealing, lying, cheating, hurting, and killing are all dominant descriptions of actions, which are morally and socially sanctioned, and we are expected to consider whether or not our actions fall under any of these. We cannot excuse ourselves by pointing to our own perspective on the action in such cases. At other times the individual's perspective plays a prominent role. I cannot be blamed for giving my friend strychnine if unbeknownst to me someone has laced my bottled water with it, but I *can* be blamed for neglecting to consider that leaving a dog in a car on a hot day might kill it.

Returning to the bystander effect again, in many cases the person *should* have been aware that there was something suspect about their easy acceptance of the idea that the subject is not in need or that it is not appropriate to help. People flailing around in rooms or on streets are surely in need of help. A further complicating factor in accepting the bystander's defense of her not helping is that most people help if they are the only ones to witness the event. And they do so because they *think* that the person is in need of help. So, it is hard to make out that bystanders are *incapable* of seeing the subject as being in need or seeing that help is appropriate. In fact, it very much seems to be the case that they *should* have seen the situation in this way. The conclusion seems inescapable. In many instances of the bystander effect—where the subject is in somewhat serious need

of help and there is some obligation to help—the inactive bystander is responsible and, therefore, the proper subject of blame. We may be more understanding towards him because we realize that many or most of us would have acted similarly, but understanding why he did not help is not absolving him of responsibility.

Considering one's actions under a variety of different descriptions requires a degree of self-reflection. We expect others to critically examine themselves, their actions and patterns of action, and to question their own interpretation of them. Deluding oneself or failing to be properly reflective about one's actions is something we can hold people responsible for, assuming that they possess the relevant capacities. As long as there is an *inferential route* from a person's values and beliefs, or from values and belief relatively easily accessible to her, to an important conceptualization of her action, we may hold her responsible (Maibom 2013). For instance, the male academic who professes to be a feminist, yet excludes female academics, ignores their opinions, and fails to take their concerns seriously, is deluded, but not necessarily in a way that is impenetrable to him. For surely he can think about his actions in a more objective fashion, or from a perspective other than his own, and consider what motives may underlie them. He does not. But this is not because he is *unable* to conceptualize his action in such a way as to see it as being wrong.

Why should it matter that the conceptualization of the action is important and how should we understand 'important' anyway? Because of the potentially infinite number of descriptions an action falls under, one must by necessity restrict the range that someone is responsible for. As we have seen, being aware of performing the action under the relevant description falls within that range, but does not exhaust it. For this excludes neglect. So the relevant description of the action may not be one that the agent is aware of, but it must be within her grasp. She can conceive of this description without undue epistemic hardship. But surely there are all sorts of descriptions that an action falls under that an agent *could* conceive of without straining herself cognitively! That is the agent can come to consider each one individually, but we cannot expect her to think of

very many of such descriptions, let alone all of them. This is where 'important' comes in.

There is nothing terribly precise we can say about what is an important description and what is not. Descriptions that are closely linked to moral prohibitions and prescriptions are obvious candidates. Descriptions that people affected by our actions or that reasonable objective observers would give are also candidates for descriptions that an agent might plausibly be expected to consider when acting. And the more severe the consequences of our actions, e.g. death or injury, the higher the demand for self-reflection. On the other hand, if there is a small possibility that we might hurt someone's feeling, it may be overly harsh to blame us or even hold us responsible. Given all the other things we need to do in our lives, it may not be reasonable to expect us to consider our actions under such descriptions.

How does all this apply to unconscious states and processes? We have the capacity to reflect on our actions from the outside, as it were. Ordinary human interaction forces other perspectives on us from early on. We can imagine how someone else would describe the action we consider performing. When people get angry with us, it is often because they see our actions in ways that we do not. They may be wrong, they may be right. In either case, we are in a position to consider whether such descriptions are reasonable. The situation is symmetrical, of course. We have all experienced feeling fobbed off with an unlikely story about somebody's motives, which is supposed to alter our conception of her action. We suspect her of being insincere. This complex interaction between our own and others' perspective on thoughts, motives, and actions sets the scene for the ability to reflect critically on what we think, say, and do. It should not, therefore, be ruled out that a person is responsible even if she does not conceive of her action as being of the relevant kind. For instance, the inconsiderate person or the bystander *is* responsible if they fail to be properly self-reflective. We have to be careful, of course, that our conditions on responsible agency are not too strict. We cannot demand rigorous self-reflection of a sort that would put anyone but Socrates or Buddha to shame. It is unlikely that people possess the

intellectual energy for such exercise, and if they do, it is not at all clear that it is best expended on such an enterprise.

The demand for self-reflection is a function of the severity of the consequences of not engaging in it. This is no different, however, from what we require when the moral stakes are high generally: we demand greater scrutiny, more care, etc. from the people faced with difficult moral choices. If a person makes a hash of his life by wandering blindly through it, never stopping up to examine his actions or reasons, to question himself, he is responsible for failing to exercise his capacity for self-reflection. Someone who tends to regard female candidates as less qualified than male candidates ought to question his judgment. Determining whether, or the degree to which, someone is responsible for unconscious mental activity may often not be straightforward. Nevertheless, the fact that it seems reasonable to hold people responsible in many such cases suggests that we should not translate the demand for knowing what we are doing into a demand for conscious access to what we are doing under the relevant description.

My suggestion takes the ability for self-reflection and inferential routes to socio-morally important descriptions of our actions to be basic to responsibility and moral agency. Conscious awareness is important, yes, because the inferential route culminates in the agent thinking about her action under relevant descriptions. However, we should not think of the central cases of such a process as involving conscious *access* to information as from a first person perspective. We are not simply “looking in” to see whether we have considered all the relevant information. Often, we must consider different ways of thinking about our actions, thereby engaging our ability to take up different perspectives on our own actions. As we develop as moral agents, we achieve facility with classifying our actions as belonging to relevant moral kinds. At the outset of our journey, however, it is often by taking another perspective on ourselves and our actions that we come to see actions as belonging to certain (moral) kinds. Much of the time, situations and our actions within them are nuanced, complex, and do not present themselves to us with ready-made descriptions or in categorical terms. We need to do a bit of work conceptualizing them. This work is often done by

perspective taking. We consider what we do from the perspective of others. We do not arrive at the relevant description(s) of our action by introspection, but by seeing our action as from the outside. What matters is that we are able to see our action as others would see it, i.e. *from their perspective(s)*. Thinking about our own actions in a morally and socially productive way often, even typically, involves thinking about them as we would think of others' actions, from a standpoint outside that of the actor, without the curious privacy that typically characterizes our decision process. This brings out the impersonal nature of the suggestion, which fits nicely with situationism. The capacity for self-reflection, for perspective taking, and so on are entirely impersonal and apply to moral agents across the board. There is, of course, a personal touch insofar as the inferential route is not indifferent to the agent's capacities, her knowledge, the situation she is in, and so on. Where one agent can be held responsible, another might not be depending on the inferential route available to her. Nevertheless, the capacity for self-reflection is impersonal and universal. It has nothing to do with the kind of person someone is.

#### **4. Conclusion**

What I have been concerned with here is sometimes called the epistemic condition on responsibility. I have argued that the best way of extending this condition to apply to *prima facie* responsible actions, such as those that result from people being subject to various situational effects, is in terms of the capacity for self-reflection. Conscious access to the (socio-morally) relevant perspective(s) on an action is relevant in a limited way. It comes from the experience of interacting with others, taking their perspectives, reflecting on one's actions more broadly, and so on, and allows for fluid and adept action. But the required experience for responsible action comes from examining one's actions as from the outside. We need flexibility in how to conceive of our actions. We must be able to conceive of it from someone else's perspective and according to a publicly sanctioned description, and we need the ability to scrutinize motives or reasons for action. We require the ability to take a perspective on what we are doing that is *not* purely our own. Identification or the possession of a deep self is not central

to this view of responsibility. We are responsible for *more than* what flows in any direct way from our characters, who we want to be, our long-term goals and projects. Scrutiny of motives and capacity for flexible conceptualization of our actions and their consequences is entirely impersonal. It has little to do with the kinds of persons that we are. It is a universal human capacity to re-conceptualize our proposed actions in categories or terms that are publicly available and important or that capture how *others* will see our actions.

## References:

- Asch, S. E. 1951. Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.) *Groups, Leadership, and Men*. Pittsburgh, PA: Carnegie Press.
- Asch, S. E. 1956. Studies of independence and conformity. *Psychological Monographs*, 70 (9).
- Bertrand, M. & Mullainathan, S. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94, 991-1013.
- Carlson, M., Charlin, V., and Miller, N. 1988. Positive mood and helping behavior: A test of six hypotheses. *Journal of Personality and Social Psychology*, 55, 211-29.
- Darley, J. & Batson, C. 1973. From Jericho to Jerusalem: A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, 27, 100-8.
- Doris, J. 2002. *Lack of Character: Personality & Moral Behavior*. New York: Cambridge University Press.
- Fischer, J. & Ravizza, M. 1993. (Eds.) *Perspectives on Moral Responsibility*. Ithaca, NY: Cornell University Press.
- Fiske, S. & Borgida, E. 2008. Providing expert knowledge in an adversarial context: Social cognitive science in employment discrimination cases. *Annual Review of Law and Social Science*, 4, 123-48.
- Fiske, S. & Taylor, S. 2013. *Social Cognition: From Brains to Culture*. London, UK: SAGE Publishers.
- Frankfurt, H. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy*, 68, 829-39.
- Haslinger, S. 2008. Changing the Ideology and Culture of Philosophy: Not by Reason (Alone). *Hypatia*, 23, 210-23.
- Kahneman, D. 2011. *Thinking Fast and Slow*. Toronto: Doubleday Canada.
- Kahneman, D., Knetsch, J. L., and Thaler, R. H. 2005. Experimental Tests of the Endowment Effect and the Coase Theorem. In D. Kahneman, J. L. Knetsch, and R. H. Thaler (Eds.) *Negotiation, Decision Making, and Conflict Management*. Northampton, MA: Edward Elgar Publishing, 92-115.
- Kunda, Z. 1999. *Social Cognition: Making Sense of People*. Cambridge, MA: MIT Press.
- Latané, B. & Darley, J. 1970. *The Unresponsive Bystander: Why Doesn't He Help?* New York: Appleton-Century Crofts.
- Latané, B. & Rodin, J. 1969. A lady in distress: Inhibiting effects of friends and strangers of bystander intervention. *Journal of Experimental Social Psychology*, 5, 189-202.

- Maibom, H. 2013. Values, sanity, and responsibility. In D. Shoemaker (Ed.): *Oxford Studies in Responsibility*. New York: Oxford University Press.
- Milgram, S. 1963. Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67, 371-78.
- Moss-Racusin, C., Dovidio, J., Brescoll, V., Graham, M, and Handelsman, J. 2012. Science faculty subtle gender bias favors male students. *PNAS*, 109, 16474-16479.
- Nisbett, R. & Wilson, T. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-59.
- Ross, L. & Ward, A. 1996. Naïve Realism in Everyday Life: Implications for Social Conflict and Misunderstanding. In T. Brown, E. Reed, and E. Turiel (Eds.): *Values and Knowledge*. Hillsdale, NJ: Erlbaum, 103-35.
- Sher, G. 2009. *Who Knew? Responsibility without Awareness*. New York: Oxford University Press.
- Superson, A. 1999. Sexism in the classroom: The role of gender stereotypes in the evaluation of female faculty. *APA Newsletter on Feminism and Philosophy*, 99, 46-51.
- Steinpreis, R., Anders, K., and Ritzke, D. 1999. The impact of gender on the evaluation of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex Roles*, 41, 509-28.
- Williams, J. 2003. The social psychology of stereotyping: Using social science to litigate gender discrimination cases and defang the 'cluelessness' defense. *Employee Rights & Employment Policy Journal*, 7, 401.
- Wilson, T. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Wolf, S. 2003. Sanity and the metaphysics of responsibility. In G. Watson (Ed.) *Free Will*, 2<sup>nd</sup> Ed. New York: Oxford University Press, 372-87.